HAL
open science

Deciphering the regulatory network controlling blood cell specification and reprogramming

Samuel Collombet

To cite this version:

Samuel Collombet. Deciphering the regulatory network controlling blood cell specification and reprogramming. Genomics [q-bio.GN]. Université Paris sciences et lettres, 2017. English. NNT : 2017PSLEE032 . tel-01798690

HAL Id: tel-01798690

https://theses.hal.science/tel-01798690

Submitted on 23 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Ecole Normale Supérieure

Deciphering the regulatory network controlling blood cell specification and reprogramming

Investigation du réseau de régulation contrôlant la spécification et la reprogrammation des cellules du sang.

Ecole doctorale n°515

Complexité du vivant

Spécialité Génomique

COMPOSITION DU JURY :

M LE CROM Stéphane
Université Pierre et Marie Curie, Président du jury

Mme GOLUB Rachel
Université Paris Diderot, Rapporteur

M XENARIOS Ioannis
Université de Lausanne, Rapporteur

Mme ROTHENBERG Ellen
California Institute of Technology, Membre du jury

Mme CHAOUIYA Claudine
Instituto Gulbenkian, Membre du jury

M SOLER Eric
Institut de génétique moléculaire de Montpellier, Membre du jury

Mme THOMAS-CHOLLIER Morgane
Ecole Normale Supérieure, Membre invitée du jury

M GRAF Thomas
Centre for Genomic Regulation, Membre invité du jury

Soutenue par Samuel COLLOMBET
le 30 octobre 2017

Dirigée par **Denis Thieffry**

ENS ÉCOLE NORMALE SUPÉRIEURE | PSL RESEARCH UNIVERSITY PARIS

Acknowledgments

First and foremost, I thank my teachers and supervisors: Denis Thieffry, who guided me during all these years and offered me so many amazing opportunities; Thomas Graf, who welcomed me in Barcelona like a member of his family and always trusted me; and Morgane Thomas-Chollier, who supervised all my work and was always present when I needed her.

Without them, I would not be here today.

षडङ्गादिवेदो मुखे शास्त्रविद्या कवित्वादि गद्यं सुपद्यं करोति ।
मनश्चेन्न लग्नं गुरोरंघ्रिपद्मे ततः किं ततः किं ततः किं ततः किम् ॥३॥

One may have mastered the vedas, have the knowledge of all sciences
And write prose and verse

But if his mind does not rest at the feet of his teacher,
What use? What use? What use?

(Sri Adi Shankaracharya, guruashtakam)

I thank the members of my jury, who graciously gave of their time to evaluate my work and come to my defense. I thank specially Rachel Golub and Ioannis Xenarios, for reviewing my thesis and for their much-appreciated comments.

I thank all the people I worked with and who made this work possible, in particular all past and present members of the Thieffry lab, the Graf lab and the IT facility at IBENS. I am particularly thankful to Bruno Di Stefano and Jose Luis Sardina Ortega for their trust, their help and their friendship.

I thank my parents, my brother, my sister, my wife, and all my friends for their support during these years.

Finally, my gratitude goes to my wife Yana Komarnitskaya and my yoga teachers, Elena Figarola and Eddie Stern, for introducing me to a practice that has changed my life and gave me another perspective on my work.

Without them, I would not be here today.

एकं सद्विप्रा बहुधा वदन्ति

The truth is one. Sages talk about it in various ways.

(Ṛgveda)

Table of content:

Abbreviations **2**

Section 1: Introduction **1**

1.1 Development and cell fate: historical perspective **1**

1.1.1 Epigenesis, the making of different cells 1

1.1.2 Stem cells 3

1.1.3 Cell fate reprogramming 3

1.2 Molecular mechanisms controlling cell fate **5**

1.2.1 Mechanisms of gene regulation 5

1.2.2 Non cell-autonomous effects 7

1.3 Methods to decipher the molecular regulation of cell fate **8**

1.3.1 Identifying factors involved in cell fate regulation 8

1.3.2 The advent of pan-genomic approaches 9

1.3.3 Bioinformatic analysis of omics data 11

1.3.4 Data integration 13

1.4 Molecular regulatory networks **15**

1.4.1 Molecular regulatory networks: mathematical approaches 15

1.4.2 Circuits functionality 16

1.4.3 *In silico* analysis of gene regulatory networks 17

1.4.4 Iterative modelling: from the computer to the bench, back and forth 18

Goals of my PhD **21**

Section 2: Deciphering the molecular regulation of B cell reprogramming **23**

2.1 Transcription factor interplay during B cell trans-differentiation **23**

2.1.1 Haematopoiesis 23

2.1.2 C/EBPa induces B cell reprogramming into macrophages 24

2.1.2 Publication: *C/EBPa Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis (Stem Cell Reports, 2015)* 26

2.2 Chromatin associated factors regulating B cell reprogramming into induced pluripotent stem cells (iPSCs) **57**

2.2.1 Regulation of pluripotency and early differentiation 57

2.2.2 C/EBPa transient expression in B cell allows a very fast induction of pluripotency 58

2.2.2 Publication: *C/EBPα creates elite cells for iPSC reprogramming by up-regulating Klf4 and increasing the levels of Lsd1 and Brd4 (Nature Cell Biology, 2016)* 60

2.3 Discussion: on the importance of dynamical analyses **85**

Section 3: Dynamical modelling of gene regulatory networks **87**

3.1 Gene network controlling B cell and macrophage specification **87**

3.1.1 How to integrate high-throughput data into gene regulatory network model 87

3.1.2 The logical formalism 88

3.1.3 Model analysis and dynamical simulations 88

3.2 Publication: Logical modelling of lymphoid and myeloid cell specification and trans-differentiation (PNAS, 2017) **90**

3.3 Discussion **105**

Section 4: Conclusion and perspectives **107**

4.1 Mechanisms of cell reprogramming and cell fate specification **107**

4.2 Workflows for the integration of epigenetic data **108**

4.3 Contribution of omic assays to the deciphering of molecular regulation in cell fate .. **110**

4.4 Integrating high-throughput data into gene networks **111**

4.5 Inferring gene regulation and factor combinations **111**

4.7 Extending our model to other cell types **112**

4.6 Integrating pathways, TFs and chromatin into gene networks **112**

References **115**

Section 5: Annexes **122**

5.1 Publication: C/EBPa poises B cells for rapid reprogramming into induced pluripotent stem cells (Nature, 2014). **122**

5.2 Publication: Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells (Scientific Data, 2016) **140**

5.3 Publication: A Transcription Factor Pulse Can Prime Chromatin for Heritable Transcriptional Memory (Molecular and Cellular Biology, 2016) **150**

5.4 Publication: Krox20 hindbrain regulation incorporates multiple modes of cooperation between cis-acting elements (PLoS genetic, 2017) **180**

Abbreviations

Cell types:

ESC: embryonic stem cell

iPSC: induced pluripotent stem cell

HSC: hematopoietic stem cell

LMPP: lympho-myeloid primed progenitor (also called lymphoid primed progenitor)

CLP: common lymphoid progenitor

CMP: common myeloid progenitor

GMP: granulocyte-macrophage progenitor

MEP: megakaryocyte-erythrocyte progenitor

Mac: macrophages

Gran: granulocytes

Techniques:

RNA-seq: RNA deep sequencing

ChIP: chromatin immuno-precipitation

ChIP-seq: chromatin immuno-precipitation followed by deep sequencing

ATAC-seq: assay for transposase-accessible chromatin and sequencing

MNAse-seq: microccal nuclease sensitivity assay

3C: chromosome conformation capture assay

4C: chromosome conformation capture combined with deep sequencing

CRE: cis-regulatory element

Section 1: Introduction

1.1 Development and cell fate: historical perspective

1.1.1 Epigenesis, the making of different cells

Multi-cellular organisms are made of different types of cells, which can exhibit very different morphologies (in shape as well as in size) and fulfill different functions. All these cells originate from a single cell, the fertilized egg (or zygote), which divides and gives rise to all the cells in the body, starting with the three embryonic tissues (or germ layers): the ectoderm, which will form the skin and most of the nervous system; the endoderm, which forms the lung, the pancreas, the gastrointestinal tract and other organs; and the mesoderm, which forms the blood and muscles.

This process of cellular differentiation happens **hierarchically**, involving first the specification of multipotent progenitors, which evolve into more and more specified cells up to terminally differentiated ones. Once a cell goes into one lineage, it becomes **committed** and cannot switch to another cell type in normal conditions.

This process is also recapitulated in the **homeostasis of adult tissue**. Cells have to be renewed on a daily basis to replace older ones or in response to aggressions (e.g. in case of infection, lymphocytes expand and proliferate), and have to adapt to a change of environmental conditions (e.g. in case of a decrease of oxygen, more red blood cells are produced).

Both processes have to be **tightly regulated**: errors during development can lead to anomalies like the absence of a cell type or body part (Grieshammer et al., 1996). Mutations affecting key regulators can also affect adult tissues, and developmental defects are often involved in cancer (Hutchinson, 2008). Therefore understanding the mechanisms controlling cell specification is of crucial importance.

In 1893, 30 years before the discovery of DNA, August Weismann proposed a theory linking **epigenesis** (how different cell types are made) and **genetics** (how hereditary characters are transmitted) (Weismann, 1893). He proposed that the substratum of the hereditary information (which he called the *germ-plasm*) was formed of functional entities (*biophors*), which were the "*bearers of the characters or qualities of cells*", and that cell fate commitment was achieved by the gradual loss of genetic information during development.

In the beginning of the 20th century, Thomas Hunt Morgan showed that genes were carried on chromosomes, and further studies suggested that the genetic information was not lost during cell specification. This led Conrad Waddington to introduce the notion of the **epigenetic landscape** (Waddington, 1957), where cell differentiation was represented by a ball rolling down from a hill into valleys, each one ultimately representing a different cell type (Figure 1). This landscaped is shaped by "*a complicated network of guy ropes which are attached to pegs in the ground*" (the pegs

representing genes and the ropes their "*molecular forces*"). This analogy introduced the ideas that the complexity of phenotypes is an emergent property of the interactions between genes. Modelling cell specification as a **dynamical system,** where differentiated cells would be stable states, Waddington further considered that cells could be "*pushed partway up the hillside*" into another stable state, i.e. that a cell could be reprogrammed towards another cell fate.

The discovery of **gene regulation** mechanisms in bacteria by Jacob and Monod (Jacob and Monod, 1961) offered a mechanistic explanation for the regulation of cell fate, something they had already foreseen at the time of their discovery (Monod and Jacob, 1961). Later work identified the regulators to be DNA binding proteins, called **transcription factors** (**TFs**), able to activate or repress the expression of other genes (Ptashne, 1967; Ptashne and Gann, 1997). In the 1970s-80s, Edward Lewis, and later Eric Wieschaus and Christiane Nüsslein-Volhard described the roles of different transcription factors as master regulators in the early development of the fly (*Drosophila melanogaster*) (Lewis, 1978; Nüsslein-Volhard and Wieschaus, 1980). The function of some master regulators is even conserved across distant species, like in the case of Pax6, which controls eye development in species ranging from drosophila to human (Gehring and Ikeo, 1999).

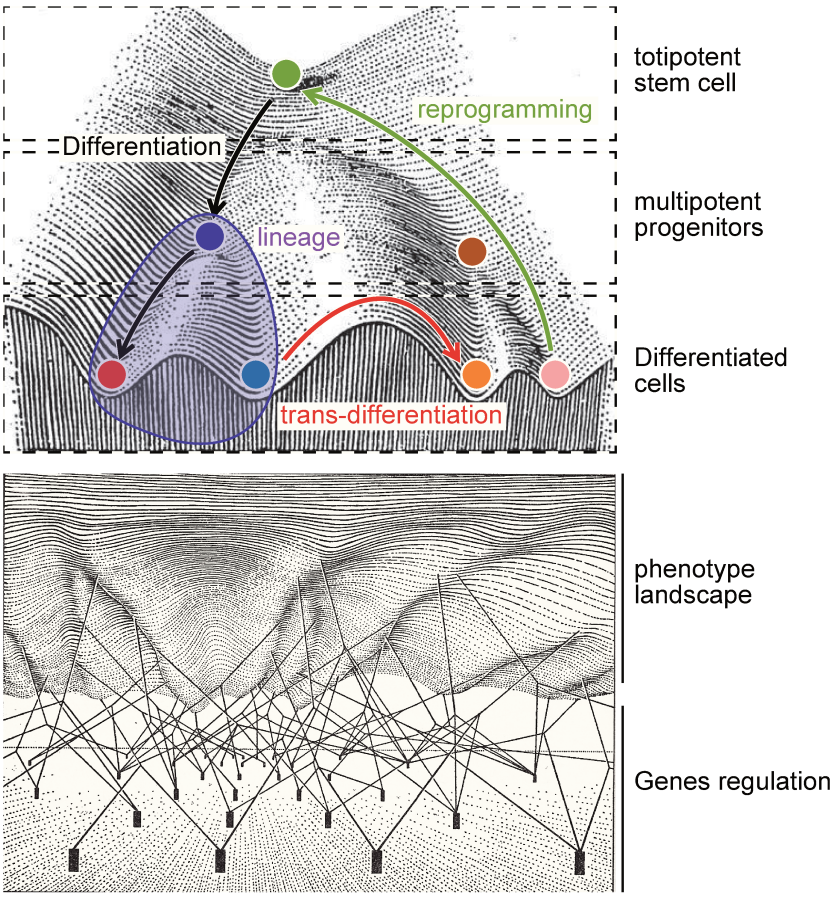


Figure 1: Waddington's epigenetic landscape

(A) The epigenetic landscape represents the potential of cell differentiation, from totipotent stem cells to progenitors of different lineages, towards terminally differentiated cells; the "hills" represents the barriers between different cell fates. (B) This landscape of cellular phenotypes is shaped by the underlying forces of genes regulations. Adapted from (Waddington, 1957).

1.1.2 Stem cells

The process of epigenesis is often depicted by a tree, going from a totipotent egg to less and less multipotent progenitors, and leading to fully differentiated cells. In 1877, Ernst Haeckel introduced the term "**stem cell**" to describe the fertilized egg, comparing cellular differentiation to the emergence of new species during evolution (Haeckel, 1877).

In 1964, Kleinsmith and Pierce isolated a cell line from an embryonic carcinoma, which was capable of differentiating into the 3 main embryonic layers (ectoderm, mesoderm and endoderm), and was used as a model to study early development (Kleinsmith and Pierce, 1964).

In 1981, the first **embryonic stem cells** (**ESCs**) were isolated from mouse epiblast (Evans and Kaufman, 1981; Martin, 1981). Protocols have then been developed to trigger ESCs differentiation, first in aggregates containing cells of the different lineages (called embryoid bodies), and then into specific cell types. These systems became major tools for the study of molecular mechanisms controlling pluripotency and cell differentiation.

Stem cells in adult tissues have been particularly studied in the hematopoietic system, mainly due to the accessibility of the blood and because irradiation of mice causes a complete depletion of their immune cells. In the early 20th century, several scientists hypothesized the existence of adult **hematopoietic stem cells** (**HSCs**) (Maximow, 2009; Ramalho-Santos and Willenbring, 2007), a small population of cells giving rise to all blood cells such as lymphocytes (B and T cells), myelocytes (macrophages, granulocytes, etc.) and erythrocytes (red cells). This was proven true in the early 1960s by Till and McCulloch, who demonstrated the existence of HSCs able to restore the complete bone marrow of irradiated individuals (Becker et al., 1963). Later on, adult stem cells have been identified in virtually every tissues, including muscle (Mauro, 1961), intestine (Barker et al., 2012), and the central nervous system (Temple, 1989).

1.1.3 Cell fate reprogramming

Another approach to understand cell fate was to study its plasticity. In the 1950s, Briggs and Kings (Briggs and King, 1957) hypothesized that, if no genetic information was lost during development, the molecules contained in the cytoplasm of an enucleated totipotent cell should be able to reprogram the nucleus of a somatic cell, a technique called **nuclear transfer** or **cloning** (Figure 2A). Although Briggs and Kings were not able to obtain viable individuals and concluded (wrongly) that some genetic information is irreversibly lost during development, John Gurdon later managed to perform successful cloning using another frog model (*Xenopus laevis* instead of *Rana Pipiens*) and obtained viable, fertile adults (Gurdon, 1962), thereby proving that fully differentiated cell could be reprogrammed towards totipotency.

In 1988, the lab of Harold Weintraub discovered that fibroblasts, when subjected to the ectopic expression of a single gene coding for the transcription factor **MyoD**, could be reprogrammed into muscle cell (Tapscott et al., 1988). This experiment highlighted the importance of transcription factors, one single protein being capable of reprogramming

cell fate (as opposed to nuclear cloning were the whole cytoplasm of an egg was used). It also confirms Waddington's view that cells could be taken back up the epigenetic landscape (*de-differentiation*) or pushed over the hills separating cell types (*trans-differentiation*).

Many other **trans-differentiation** cases have been reported (Graf and Enver, 2009). One that will be presented in section 2.1 is the reprogramming of B cells into macrophages (Figure 2B), discovered in the laboratory of Thomas Graf in 2004 (Xie et al., 2004). In this system, the ectopic expression of the transcription factor CEBP/a in B lymphocytes can convert them into macrophages, in approximately one week. As B cells undergo a re-arrangement of their DNA at the immunoglobulin locus (V-D-J gene recombination), which can be verified by DNA sequencing, their reprogramming offers an elegant proof that the resulting cells are indeed reprogrammed B cells and not the result of a contamination.

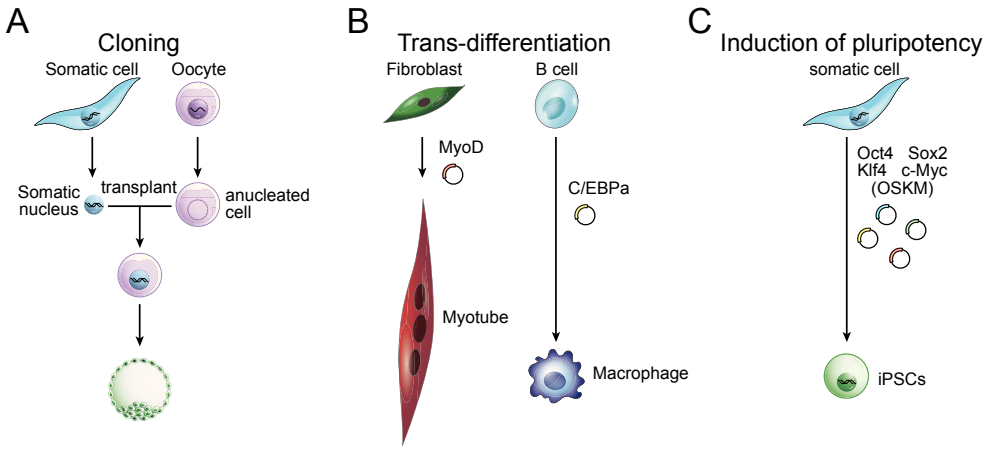


Figure 2: Cell fate reprogramming.

(A) Cloning: The nucleus of a somatic cell is extracted and transplanted into an enucleated oocyte. The resulting cell is pluripotent and can develop into a blastocyst. (B) Trans-differentiation: ectopic expression of a transcription factor can reprogram differentiated cells into another cell type. (C) Induction of pluripotency: ectopic expression of the "Yamanaka factors" (OSKM) can induce reprogramming of somatic cells into induced pluripotent stem cells (iPSCs). Adapted from (Graf and Enver, 2009; Yamanaka and Blau, 2010).

In 2006, Shinya Yamanaka hypothesized that the ectopic expression of some master regulator(s) of pluripotency could be sufficient for **reprogramming somatic cells toward pluripotency**, as in the case of MyoD for muscle (Fig.2 C). He and his team identified a combination of four transcription factors (Oct4, Sox2, Klf4 and cMyc, or OSKM, the "Yamanaka cocktail"), capable of reprogramming fibroblasts into induced pluripotent stem cells (iPSCs) in mouse (Takahashi and Yamanaka, 2006), and human (Takahashi et al., 2007). Although the extremely low efficiency of reprogramming casted some doubt about potential contamination, later work inspired from the trans-differentiation of B cell allowed the group of Rudolf Jaenish to reprogram V-D-J rearranged B cells into iPSCs (Hanna et al., 2008), thereby confirming their somatic origin.

1.2 Molecular mechanisms controlling cell fate

1.2.1 Mechanisms of gene regulation

Following the major discoveries of molecular biology in the 20th century, we know that genes are transcribed by polymerases into messenger RNA (mRNA), which are then exported to the cytoplasm to be translated into proteins by the ribosomes. Transcription is one of the major checkpoints for the regulation of gene expression. In mammals, the polymerase requires the helps of **transcription factors** (TFs) to initiates and/or sustains transcription (Fig. 3). TFs recognise and bind to specific sequences of DNA (usually around 10 base pairs) in regions of the genome called regulatory elements. These regions are almost exclusively on the same molecule of DNA as the gene they target, and therefore are called *cis*-regulatory elements (CREs). TFs can recruit other **factors that activate the polymerase** and start the transcription (or increase its rate), such as the mediator or the P-TEFb complexes.

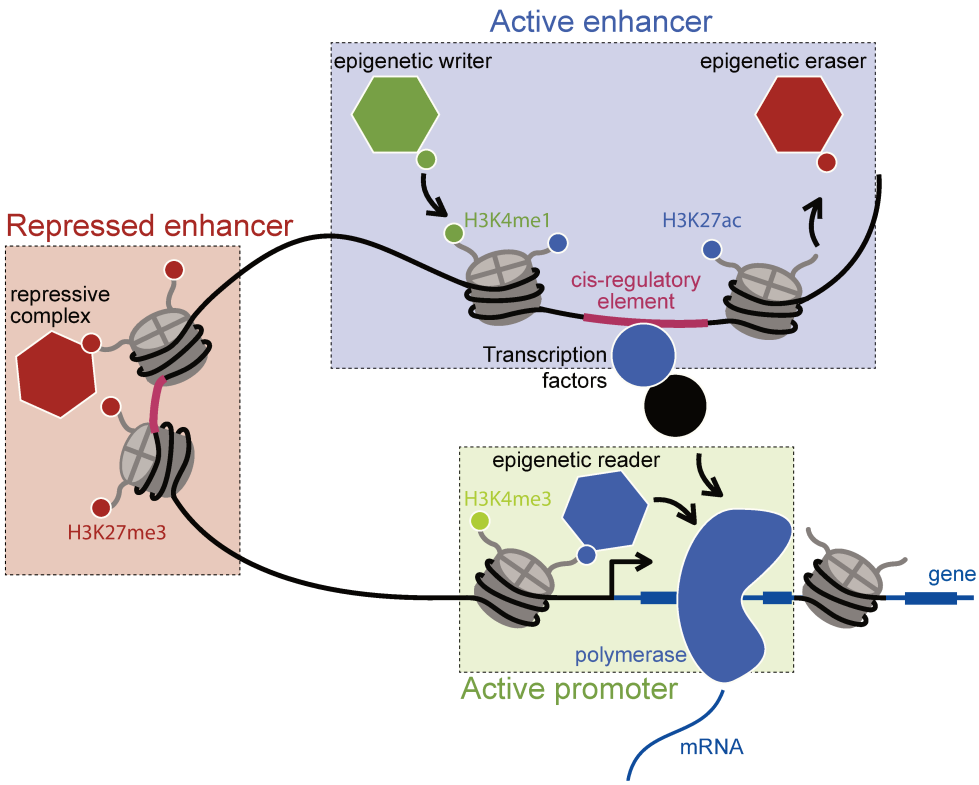


Figure 3: Regulation of gene expression.

Schematic representation of the molecular mechanisms controlling the regulation of gene transcription in mammals.

Regulatory elements are classified into two main categories. The region located just upstream of the transcription start site (TSS) of a gene is called the **promoter**. In bacteria, many TFs regulate genes by binding at their promoters, whereas in mammals genes are often regulated through distant elements, called **enhancers**, which interact with promoters through the formation of DNA loops. Other types of CREs have been described, such as

the insulators, which preclude the formation of loops between regions separated by them (Gaszner and Felsenfeld, 2006).

Recently, a subset of enhancers has been described as **super-enhancers**, because of their stronger effect on transcription activation than "classical" enhancers (Whyte et al., 2013). These regions exhibit a much higher density of the general TF Med1, as well as the histone mark H3K27ac and cell type specific TFs. Although it remain controversial whether super-enhancers are entities clearly distinct from classical ones (Pott and Lieb, 2014), they are clearly stronger regulatory elements, often associated with genes coding for developmental regulators or factors involved in cancer (Hnisz et al., 2013; Meng et al., 2014; Qian et al., 2014), and are therefore of interest for the study of cell fate.

In the nucleus, DNA is strongly compacted and coiled around octamers of proteins called **histones**, which includes different variants (mainly the histones H2A, H2B, H3 and H4). The structures formed by DNA around histones are called nucleosomes. Histones hinder DNA accessibility to transcription factors, and their positioning/removal at regulatory elements constitutes a mechanism of regulation of transcription. In this respect, complexes involved in nucleosome displacement, such as the SWI/SNF complex, are important for the regulation of cell fate and in cancer (Kadoch et al., 2017; Mathur et al., 2017).

Each histone has a tail protruding outside of the nucleosome, whose amino-acids can be subjected to various **post-translational modifications**, such as methylation, acetylation and others. These modified residues can serve has docking platforms for some regulatory complexes. For example, the acetylation of lysine 27 of histone 3 (H3K27ac) is recognized by BRD proteins, which are required for transcription; H3K27me3 is recognized by proteins of the polycomb complex (PRC), which have a negative effect on gene expression (Aranda et al., 2015).

Finally, chemical modifications can also be added to the DNA itself. The most studied DNA modification is the **methylation of cytosines** (5mC). 5mC has a general repressive role on transcription initiation when occurring at promoters but its role at enhancers his still under debate (Jones, 2012). 5mC is added by DNA methyl-transferases (DNMT), and can be removed either passively by not being maintained after cell division, or actively through the enzymes of the **TET family** (Wu and Zhang, 2017).

Although transcriptional regulation is extremely important in the regulation of cell fate, there are **other important levels of regulation**. **RNA splicing** can be a major regulation checkpoint for protein expression, and it has been shown to regulate the transition from neural progenitor to mature neuron (Zhang et al., 2016a). Micro-RNAs (**miRNA**), targeting mRNA and causing their degradation, have been shown to be involved in hematopoietic cell development (Mehta and Baltimore, 2016), while the reprogramming of somatic cells into iPSCs can be achieved using only miRNA (Anokye-Danso et al., 2011).

1.2.2 Non cell-autonomous effects

Epigenesis and homeostasis do not only pertain to the question of how a cell can make different cell types, but also how to make the right cell in the right place, at the right moment. This requires the integration of **information from the environment**, which is done through **signalling pathways**. Signalling molecules, such as morphogens, cytokines or growth factors are secreted in the inter-cellular environment, and bind to and activate receptors on the cell membrane. This results in cascades of enzymatic modifications, ultimately leading to the activation of transcription factors and hence the induction/repression of many target genes.

In some cases, the **signalling molecules are capable of passing through the cell membrane**, bind to transcription factors directly in the cytoplasm, activate them and induce their translocation to the nucleus. This is the case of the estrogen receptor (ER), which has been widely used to generate fusion proteins made of a transcription factor of interest and the ER, allowing to induce the activation of the TF by simply adding some estradiol to the culture medium, and stop induction by washing the cells. Such tools enable a precise control of transcription activity.

Signalling pathways do not only control the expression of cell specific gene programs and differentiation, but also control **cell survival** and cell division. Indeed, most cells cannot survive in the absence of specific signals, which are normally present in their environment. This level of regulation complicates the study of cell fate *in vivo*: the readout of a mutant or knock-out experiment *in vivo* is often measured at the population level (increase/decrease in the number of some specific cells), and this can be due to a block of cell differentiation, or to an effect on cell proliferation and survival.

1.3 Methods to decipher the molecular regulation of cell fate

1.3.1 Identifying factors involved in cell fate regulation

The most classical method to identify factors involved in lineage commitment has been the study of **mutants exhibiting particular phenotypes**, like the lack of a certain type of cells or anatomical structure. For example, a loss-of-function of Pax6 in drosophila results in the absence of eye formation (Gehring and Ikeo, 1999); in hematopoiesis, the loss of Cebpa blocks the development of granulocytes (Zhang et al., 2004). Many master regulators have been discovered based on similar analyses.

This was taken to a larger scale by the development of saturating **genetic screens** in drosophila, where mutations were induced in many individuals, leading to observable phenotypes, and responsible mutations were then mapped. In mammals, such screens for loss-of-function became possible with the development of knockdown and knockout technologies, such as shRNA (Cooper and Brockdorff, 2013) or more recently CRISPR-CAS9 (Parnas et al., 2015). Screens have also been performed to identify factors allowing cellular reprogramming or forcing stem cell differentiation (Eguchi et al., 2016).

Gain- or loss-of-function can be due to a mutation inside a coding gene, but also to a **mutation in a regulatory element**. For example, a mutation in an enhancer of PTF1A, a master regulator of pancreas development, has been shown to result in a complete block of pancreas formation (Weedon et al., 2014). In this case, the characterisation of the mutated region involved the confirmation of its cis-regulatory effect, through a reporter assay.

TFs binding sites can be predicted by **sequence analysis**. The DNA sequence recognized by a TF can be defined by a consensus sequence (string of nucleotides), or by a motif accounting for varying affinity for similar sequences (modelled by a position-weight matrix, PWM). Databases collecting consensus and PWMs have been built, such as JASPAR (Mathelier et al., 2016) and TRANSFAC (Matys et al., 2006). These collections of motifs can be used to scan regions of interest to predict TF binding sites (Turatsinze et al., 2008), which can then be confirmed experimentally using chromatin immuno-precipitation (**ChIP**), followed by PCR or deep sequencing.

More recently, techniques to **capture the conformation of chromosomes** (3C techniques) have allowed for the measurement of physical interaction between enhancers and promoters (Dekker et al., 2002). Although the original 3C approach was combined with PCR amplification targeting regions of interest, this approach has been scaled up to the genomic level (4C and Hi-C) (Lieberman-Aiden et al., 2009; Simonis et al., 2006).

1.3.2 The advent of pan-genomic approaches.

Since the development of pan-genomic experimental approaches, and in particular high-throughput sequencing, many techniques have been developed to study gene regulation at the genomic level (Table 1).

Table 1: Genomic approaches to study gene regulation

Feature studied	Technique	Description
TF binding	ChIP-seq	- Binding sites genome wide (resolution: few 100bps) - Relative quantification of binding
Open chromatin	DNAse-seq	- Open chromatin (resolution: few 100bps)
	MNAse-seq	- Nucleosome positioning - Nucleosome free regions (equivalent to open chromatin)
	ATAC-seq	- Open chromatin (resolution: few 100bps) - Nucleosomes positioning (requires high sequencing depth)
Histone modification	ChIP-seq	- Regions of enrichment genome wide (peaks or domain) - Relative quantification of the enrichment
Chromatin conformation	4C	- "One-to-all": capture interactions between a region of interest (view point) and any other regions - Capture interaction ~2Mb around the view point - PCR bias due to the use of restriction enzymes
	HiC	- "All-to-all": capture interactions between all regions of the genome - Poor sensitivity - Resolution: of 5 to 20kb
Transcriptome	Microarray	- Levels of expression of a comprehensive set of targeted genes (several thousands)
	RNA-seq	- Levels of expression for all genes (sensitivity depending on the sequencing depth) - Discovery of novel gene/transcripts
Proteome	SILAC	- Protein quantification relative to a control sample
	Label free quantification (LFQ)	- Protein quantification, independently for each sample

The measure of the transcription level of all, or at least several thousands of genes (**transcriptomic**), has allowed the determination of all the genes expressed specifically in one cell type, or differentially expressed during differentiation. This was first performed using DNA probe hybridation chips, and later with RNA sequencing (RNA-seq) (Wang et al., 2009).

Expression can also be studied at the level of proteins using **quantitative proteomic**, with some sensitivity limitations, in particular for nuclear regulatory factors, as they are usually lowly abundant compare to cytoplasmic proteins.

High-throughput sequencing has further opened the way to study TF binding genome-wide through chromatin immuno-precipitation followed by deep sequencing (**ChIP-seq**) (Johnson et al., 2007). This approach can be

used to predict targets genes (with some limitations), but also to study factor cooperation at regulatory elements, by comparing different ChIP-seq or by sequence analysis.

De novo **motif discovery** applied to TF binding regions can reveal the motif(s) of the ChIP-ed protein (Thomas-Chollier et al., 2012), as well as potential co-factor binding in its vicinity. However, one should keep in mind that such motifs represent an average picture of a factor binding specificity, which might neither reflect the strongest binding, nor the more relevant ones from a functional point of view.

ChIP-seq can also be performed using antibodies targeting specific **histone modifications**. These modifications have been extensively mapped in different cell types, and characteristic marks have been described to be associated with different genomic features (Table 2). Different enzymes can add or remove specific marks (chromatin remodeling factors), or bind them and affect transcription (chromatin readers) (Table 3). Therefore, changes in histone marks during differentiation/reprogramming can be used to identify enhancers becoming active (or inactive) from one condition to another one, and DNA motif analysis (*de novo* discovery or scanning) can then be used to predict TF binding (Ziller et al., 2015). One such application is described in the publication presented in section 2.2.3).

Table 2: Histone marks and their combinations associated with regulatory elements.

Regions	Marks	Annotation	Description
Promoter	H3K4me3 only	Poised promoter	- Mostly enriched at promoters - Marks active and "to be active" promoters
Enhancer	H3K4me1 only	Enhancer	- Marks active and poised enhancers.
Domains	H3K27me3 only	Repressed domain	- Enriched in large domains of several kilobases - Mark of the polycomb repressive complex
Promoter	H3K4me3 + H3K27ac	Active promoter	- H3K27ac marks the promoter transcribed genes
Promoter	H3K4me3 + H3K27me3	Bivalent promoter	- Promoter having both active and repressive mark - Repressed but ready to be active - Specific of ESCs
Enhancer	H3K4me1 + H3K27ac	Active enhancer	- H3K27ac mark active regulatory elements
Enhancer	H3K4me1 + H3K27me3	Poised enhancer	- Similar as bivalent promoter

(Adapted from (Rivera and Ren, 2013))

Table 3: Chromatin factors associated with histone marks.

Marks	Depositor	Eraser	Reader
H3K4me1/2/3	SETD1/MLL complex	Lysine-Specific histone Demethylase (e.g. LSD1)	Double chromodomain proteins (e.g. CHD1)
H3K27ac	Histone Acetyle Transferases (e.g. p300-CBP)	Histone DeACetylases (e.g. HDAC1)	BRomo Domain proteins (e.g. BRD4)
H3K27me3	Polycomb Repressive Complex 2 (PRC2)	Lysine-Specific histone Demethylase (e.g. LSD6A)	Polycomb Repressive Complex (PRC)

(Adapted from (Calo and Wysocka, 2013; Musselman et al., 2012))

Other techniques allow mapping regions of **open chromatin** accessible to transcription factors, using a deoxyribonuclease (DNase) (Boyle et al., 2008) that cannot access nucleosomal DNA. These have been supplanted in the last years by the assay for transposase-accessible chromatin (**ATAC-seq**), which requires much less cells (Buenrostro et al., 2013). Smaller and more centred on the binding sites than the regions defined by histone marks, the resulting "open chromatin regions" are more amenable to TF motif analysis, as the noise generated by surrounding sequences is reduced.

Finally, the combination of chromosome conformation capture with deep sequencing has allowed scaling up the "one-to-one" interaction analysis to a "one-to-all" view (**4C**), where all interactions between one locus of interest (the view point) and any other loci can be assessed. Some "many-to-all" methods (4C-seq, (Zhao et al., 2006), ChIA-PET (Fullwood et al., 2010)) and "all-to-all" methods (HiC, (Lieberman-Aiden et al., 2009)) have been developed, but currently lack the sensitivity and resolution necessary to systematically associate enhancers with genes.

1.3.3 Bioinformatic analysis of omics data

The processing and analysis of high-throughput data requires specific **statistical and computational methods**, in order to account for technical biases (Robinson and Oshlack, 2010). Many methods and tools have been established in the context of large consortia such as Encode and Roadmap (Bernstein et al., 2010; The ENCODE Project Consortium, 2012), which systematically produced genomic datasets (RNA-seq, ChIP-seq, etc.) in many cell types, allowing for systematic analysis and fostering the establishment of guidelines (Bailey et al., 2013). The main steps of pan-genomic data analysis, their pitfalls and some of the major software are listed in Figure 4.

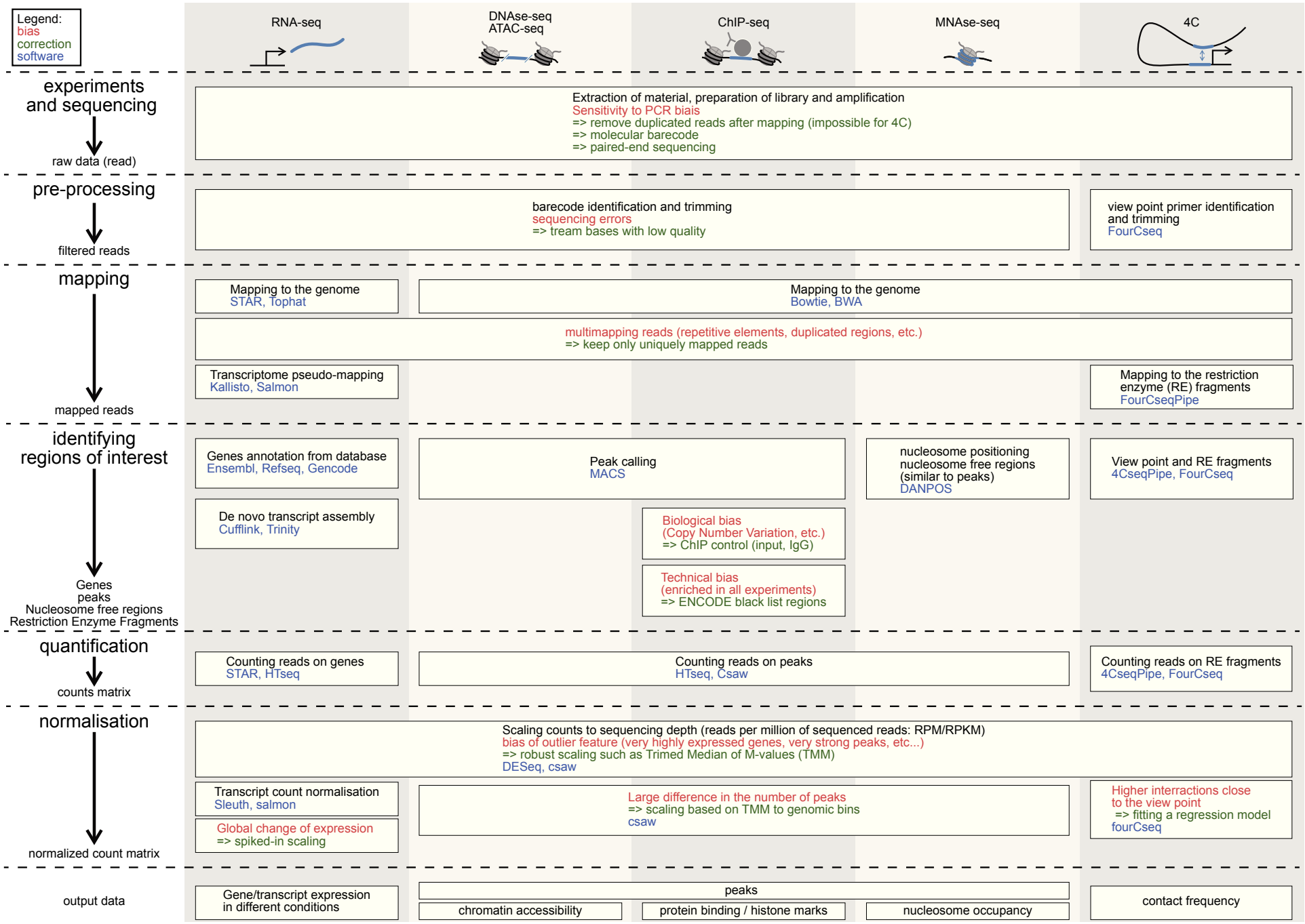


Figure 4: workflow for the analysis of omics data and possible biases

One major step of the analysis of ChIP-seq data is the peak calling, i.e. the identification of regions with enrichment of signal compared to the surrounding background.

For **peak-shaped signals** such as TF ChIP-seq (or DNAse-seq, ATAC-seq, etc.), this has been well characterised and many software tools have been proposed, the most standard one nowadays being MACS (Zhang et al., 2008). Specific variants have been developed to identify precisely the **summit** of sharp peaks, or to delineate **broad peaks** (in particular for histone marks).

For histone marks forming **larger domains with fuzzier border**, the current methods usually perform poorly with default parameters, and often require some parameters tweaking or even some *in house* software development. It is in fact possible that these marks form domains with very different characteristics (shape, size, level of enrichment, etc.), which cannot all be identified using one single method.

One crucial step for the comparison of any sequencing data is the correction of differences in quantity of sequenced material, or **scaling** (often abusively called *normalisation*). Such bias can be the result of differences in initial quantities of material, sequencing depth, etc.

The most common method is the scaling of signal by the number of millions of reads sequenced per sample (reads per million, **RPM**). More refined methods have been developed to account for outlier features with extremely high counts (e.g. very highly expressed genes), such as trimmed median of M-values (**TMM**) (Robinson and Oshlack, 2010).

In ChIP-seq (or similar assays), large variations in the number of enriched regions (peaks) can affect the results of such scaling, and it is preferable to estimate the scaling factors from the signal in large genomic bins rather than in all merged peaks (Lun and Smyth, 2015). Other technical biases have been described for ChIP-seq, and they are discussed at length in (Lun and Smyth, 2015).

1.3.4 Data integration

The analysis of cell fates often involves the **comparison of many cell types** or the analysis of **time course** during differentiation/reprogramming experiments. A first approach is to use methods for the **reduction of dimensionality**, such as principal component analysis (PCA), independent component analysis (ICA), or canonical covariance analysis (CCA). These methods are often used to analyse the data in a qualitative way, looking at the proximity between cell types, or the trajectories along time courses in the reduced space.

A second approach is the identification of features showing **significant changes between conditions/cells** (gene expression, ChIP-seq peak intensity, etc.). Although such comparisons can be done in a pairwise fashion, it is often preferable to take advantage of statistical methods such as **generalized linear models,** and statistical tests such as likelihood ratio test (Love et al., 2014). Such methods allow identifying features (genes, peaks, etc.) changing in any condition, which can further be used for **unsupervised**

clustering (hierarchical clustering, K-means, etc.) in order to identify patterns.

With the democratisation of genomic assays, it has become easier to assess the dynamics of multiple epigenetic features, such as different histone modifications or chromatin accessibility. The **integration of different epigenetic features** is complicated by the different sizes of enriched regions (peaks, domains, etc.). This can be addressed by using **genome-partitioning** algorithms, such as chromHMM (Ernst and Kellis, 2012). Other programs have been developed to take into account the information of time series or different cell types (Zhang et al., 2016b)

The interpretation of epigenetic changes at enhancers relies on their association **to target genes**. A simple approach is to link peaks with their closest gene, but this fails to capture distant enhancers. More refined methods based on signal correlation have been proposed (Corradin et al., 2014), and this can be further improved using **chromatin conformation capture** experiments (3C) to identify physical contacts between enhancers and promoters.

1.4 Molecular regulatory networks

1.4.1 Molecular regulatory networks: mathematical approaches

TFs, epigenetic factors and signalling pathways are not isolated but interact with and regulate each other, forming **complex regulatory networks**. Feedback- and feedforward-circuits (positive or negative) can produce non-trivial behaviour and therefore requires the use of mathematical models, which can be represented as a graph.

Molecular entities (gene, proteins, signalling molecules, etc.) are represented by nodes and mathematical variables are associated with them, which can take continuous or discrete values depending on the formalism. Molecular regulations are represented by oriented edges (arrows) between the nodes, and mathematical functions define the behaviour of a variable depending on the values of its regulators. The type of functions used depends on the precision of modeling required and the available data (Table 4).

Table 4: Mathematical formalism for modeling gene regulation

Formalism	Representation of regulations	Advantages	Limitations
Ordinary Differential Equations (ODE) model	Deterministic, continuous differential equations	- Quantitative - Precise analysis of specific phenomena (bifurcation, etc.)	- Requires precise, quantitative information (or strong assumptions) - Simulations only for large and complex networks - Deterministic
Boolean/logical model	Logical rules	- Requires only qualitative data - Global view of the dynamics (Attractors, reachability, etc.)	- Assumes the existence of non-linearities - Over-approximation of the system's dynamic
Stochastic model	Probabilistic functions	- Discriminate between high and low probability events	- Requires precise data on probabilities (or strong assumptions)

Ordinary differential equations (ODE) have been widely used to model biochemical systems such as enzymatic reactions, using quantitative data. ODE models have been successfully used to model cell fate regulation (Laslo et al., 2006). However, measuring kinetic parameters in mammals is hampered by the complexity of gene regulation (combinations of enhancers, different layers of regulation, etc.), and by experimental limitations (impossibility to measure all kinetic parameters *in vivo*, such as TFs binding, TFs-polymerase interactions, etc.).

On the other hand, **logical formalisms** require only qualitative data and are therefore well adapted to the modelling of gene regulation (Bérenguier et al., 2013; Kauffman, 1969). In this formalism genes can be represented by simple Boolean variables (ON or OFF), or by multi-level ones to

represent several levels of activity. This formalism allows the modelling of very large and complex networks, at a moderate computational cost.

Stochastic approaches can further be used to model cell dynamic at the population level. These methods generally rely on continuous Markov chain and the use of Gillespie algorithm (Stoll et al., 2012). Logical models can be expanded into stochastic ones (see application in section 3.2).

1.4.2 Circuits functionality

As it became clear that single factors cannot, on their own, explain complex phenotypes such as the existence of multiple stable cell types, it was hypothesised that regulatory circuits could be responsible for these behaviours. In 1981, René Thomas made two conjectures about the **roles of regulatory circuits** in dynamical system (Thomas, 1981):

1. Coexistence of several stable states requires a positive circuit in the regulatory graph.
2. Sustained oscillations require a negative circuit.

These conjectures were furthered demonstrated to be true in different formalisms (Remy et al., 2008; Soulé, 2004).

A **positive circuit** is an oriented, circular path in the regulatory graph, where the product of the signs of the regulations is positive (i.e. it contains no or an even number of negative regulations). For example, **Positive feedback loop** allow for the stabilization of gene expression after a transient input signal (Figure 5A) (Becskei et al., 2001). **Cross-activation** are also frequent between TFs which regulate a common gene program, hence locking cell fate (Figure 5B) (Lin et al., 2010).

Another case of positive circuits is the **cross-inhibitory circuit**, where two factors repress each other (Fig. 5C). One example in development is the GATA1 – PU.1 cross-inhibitory circuit. The transcription factor GATA1 controls the induction of the erythroid gene expression program and represses PU.1, the later fostering the myeloid program and repressing GATA1. Such circuits offer a simple explanation for mutually exclusive gene programs, and have been proposed to control other branching points in development, such as Gfi1 - Egr2 for granulocyte - macrophage specification (Laslo et al., 2006), or Oct4 – Cdx2 for ICM – trophoblast specification (Niwa et al., 2005).

More **complex circuits** have also been involved in cell fate specification. In fact, in the case of GATA1 – PU.1, it has been shown that their cross-inhibition was not sufficient to completely account for cell behaviour, and that a third factor was necessary (Figure 5D and (Chickarmane et al., 2009)). Other circuits have been shown to support more complex behaviours, such as bi-stability and excitability (Figure 5E and (Panovska-griffiths et al., 2013)).

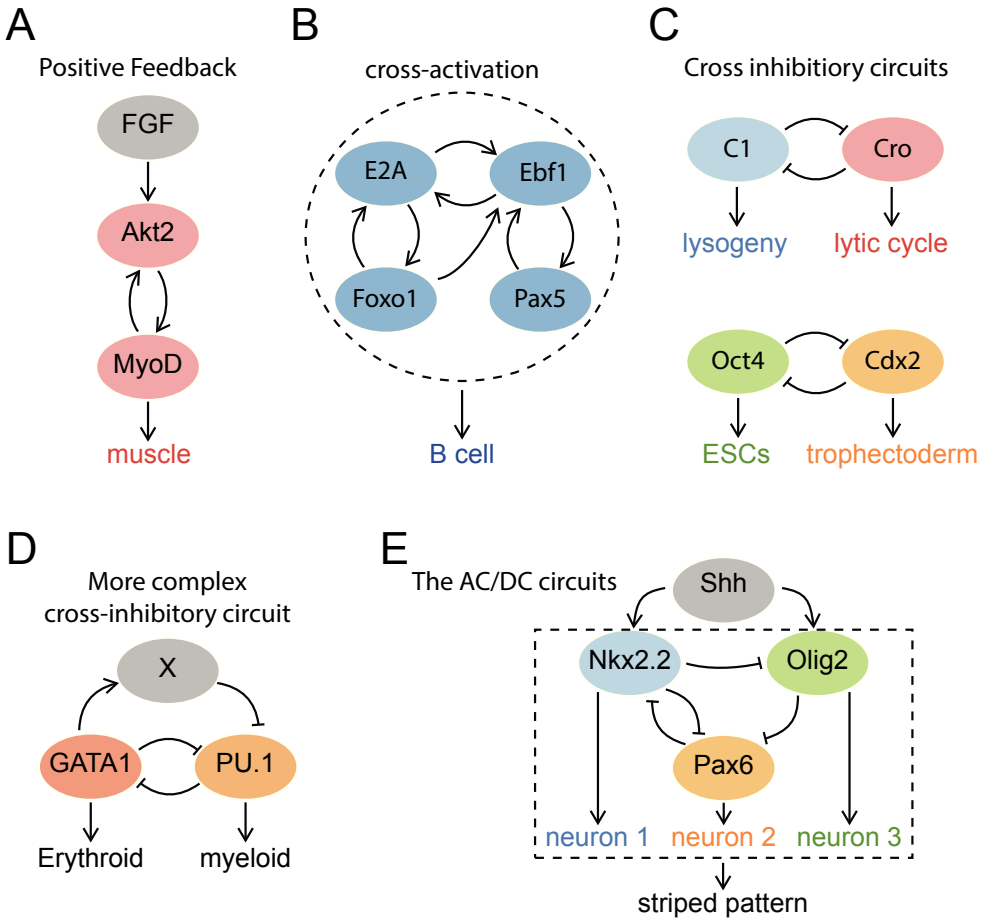


Figure 5: circuits in gene networks

(A) Example of a positive feedback, allowing for the stabilisation of muscle differentiation after at transient stimulation (Kaneko et al., 2002). (B) Example of cross-activation between TFs controlling B cell specification. (C) Examples of cross-inhibitions where two factors control alternative cell fates and repress each other, excluding each other program. (D) More complex circuit controlling erythropoiesis and myelopoiesis, based on Gata1 - PU.1 cross-inhibition (Chickarmane et al., 2009). (E) The AC/DC circuits controlling patterning in the neural tube (Panovska-griffiths et al., 2013).

1.4.3 *In silico* analysis of gene regulatory networks

Different kind of analysis can be performed on dynamical model of a gene network. A first analysis is the identification of the **attractors** of the model, in particular **stable states**, which usually correspond to cell types.

Simulations are used to assess the behaviour of the system starting from particular states, e.g. to simulate **differentiation** from a progenitor state. One can also study the effect of a **perturbation** on a specific state (e.g. corresponding to a cell type). They can be used to reproduce *in silico* the effects of gene knock-downs or ectopic expressions, or yet treatments with drugs (Flobak et al., 2015).

Finally, **reachability questions** can be addressed, i.e. asking whether a certain state can be reached, from an unstable state or from a perturbed state (knock-down, treatment, etc.). This can be done using simulations, but other methods, (e.g. model checking approaches) enable more systematic analysis (Abou-Jaoudé et al., 2015).

1.4.4 Iterative modelling: from the computer to the bench, back and forth

Although gene network modelling usually comes after experiments, it is not a finality in itself, and its main added value is to make **predictions that can be tested experimentally**.

The most common ones are **predictions of cellular phenotypes**, like the effect of molecular perturbations on cell fate (knock-out affecting differentiation), reprogramming protocols or prediction of novel cell types (e.g. cells expressing mixed gene programs (Naldi et al., 2010)). However, such predictions on limitations of the model (absence of some regulators in the model, etc.) and should therefore be taken with a grain of salt.

Models can also be used to make predictions regarding **missing regulations** or regulators. For example, computational analysis can be used to test if a hypothetical regulation is consistent with known phenotypes and/or improves the consistency of the model with available gene expression data. This can be used to prioritize hypotheses before experimental validation. One case of such application is described in the paper in the part 2 of the results section.

Finally, the combination of modelling and subsequent validation should be seen as an **iterative process**: once predictions are tested and validated, they are integrated in the model and further used for another round of predictions.

Goals of my PhD

Following previous work of the laboratory of Thomas Graf, we aimed at **deciphering the regulatory network controlling B cells reprogramming**. In close collaboration with experimentalists, who performed functional genomic experiments and produced many datasets (RNA-seq, ChIP-seq, ATAC-seq, Proteomics, 4C), I have combined various computational methods to analyse the data and model *in silico* the regulatory network controlling cell specification. My work has been focused on the following points:

- **Identify the most adapted methods** for the analysis of the different types of data.
- **Develop workflows for the integration** of these datasets and computational analyses.
- **Predict key regulators** (TF, chromatin factors) that could be tested experimentally.
- Integrate these data with previous information into a **dynamical model of the underlying regulatory network**.

Use this model to **predict novel regulations or cellular behaviours**, and test them experimentally.

In section 2, I present two studies where we performed various pan-genomic analyses to decipher the mechanisms of B cell reprogramming. We first focused on the interplay between different transcription factors and chromatin remodellers during **B cell trans-differentiation** (section 2.1), by integrating data from transcriptomic, ChIP-seq targeting TFs, histone marks and MNAse-seq.

We then studied the mechanism of **B cell reprogramming into iPSCs** by combining time course data from RNA-seq, proteomic, ChIP-seq for histone marks and ATAC-seq, to identify key factors and regulatory elements (section 2.2).

In section 3, we then integrated this information in a **dynamical model of the underlying gene network**. We used this model to predict unknown molecular regulations, which we confirmed experimentally, and simulated cell reprogramming experiments (section 3.2).

We also **compared the molecular dynamics of reprogramming with physiological processes**, like normal differentiation (section 2.1 and 3.2), as well as inflammatory response (annex 5.3), to assess whether the mechanisms at play in reprogramming reflect more general physiological processes.

Section 2: Deciphering the molecular regulation of B cell reprogramming

2.1 Transcription factor interplay during B cell trans-differentiation

2.1.1 Haematopoiesis

The different populations of blood cell have been extensively characterised by the study of **surface markers** using flow cytometry, allowing for the discrimination of different progenitors and differentiated cell types (Figure 6). In the classical view, hematopoietic stem cells (HSCs) form a small population on-going slow, asymmetric divisions, sustaining themselves and producing multipotent progenitors (**MPPs**), which divide symmetrically. MPP can further differentiate into common lymphoid progenitors (**CLPs**), granulocyte-macrophage progenitors (**GMPs**), or megakaryocytes-erythrocytes progenitors (**MEPs**) (Orkin and Zon, 2008). They further differentiate and ultimately give rise to terminally differentiated cells, such as macrophages, granulocytes, B and T lymphocytes.

The specification of these three main types of progenitors has been the subject of controversies, as it was first thought that MPP can differentiate into CLP or into a common myeloid progenitor (CMP, which could further differentiates into MEP or GMP); however a population of lympho-myeloid-primed progenitors (**LMPP**) has been described to have lymphoid and myeloid potential but unable to give rise to erythroid cells (Adolfsson et al., 2005). It has also been questioned whether the population described as CLP, which has a T cell potential *in vitro*, is really giving rise to T cells *in vivo*, or if the cells migrating to the thymus are in fact earlier progenitors. Recent single cell experiments have led to the identification of novel sub-populations of progenitors (Nestorowa et al., 2016; Paul et al., 2015), and these progenitor definitions should be considered as useful models rather than an absolute truth.

Transcription factors regulating blood cell specification have been identified based on the effect of their loss-of-function on cell populations. **PU.1** is a general factor in haematopoiesis, and is required for both lymphoid and myeloid development (McKercher et al., 1996). **Cebpa** is required for the development of GMPs, its loss-of-function causing a developmental arrest at the CMP stage, leading to acute myeloid leukaemia (Zhang et al., 2004). The transcription factors **Ebf1**, **E2a**, **Foxo1** and **Pax5** are cooperating for the development of B cells.

Different **signalling pathways** have been reported to be involved in blood cell specification. The **CSF1** or M-CSF pathway induces macrophage specification through PU.1 up-regulation (Mossadegh-keller et al., 2013); the **IL7** pathway is required for B cell specification, while **Notch** signalling leads to T cell specification. Other pathways, such as Flt3 or the GM-CSF have also been described to play key roles in hematopoiesis, but it is less

clear whether their function is really to induce cell type specific programs or to allow cell survival.

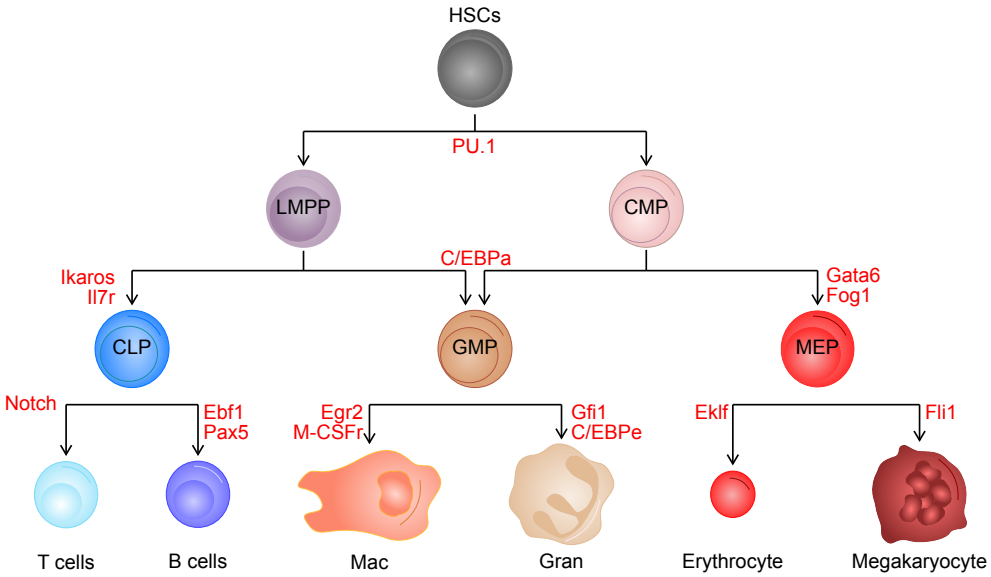


Figure 6: Hematopoiesis

Schematic representation of blood cell specification, from hematopoietic stem cells (HSCs) into different progenitors toward terminally differentiated cells. The genes indicated in red` are required for downstream cell specification. LMPP: lympho-myeloid primed progenitor; CMP: common myeloid progenitor; CLP: common lymphoid progenitor; GMP: granulocyte-macrophage progenitor; MEP: megakaryocyte-erythrocyte progenitor; Mac: macrophage; Gran: granulocyte.

2.1.2 C/EBPa induces B cell reprogramming into macrophages

In 2004, the lab of Thomas Graf discovered that the ectopic expression of the transcription factor **C/EBPa in B cells reprogrammed them into functional macrophages** (Xie et al., 2004). They later showed that this was a direct trans-differentiation, and not a de-differentiation toward an intermediate, multipotent states (Di Tullio et al., 2011). This system is extremely **efficient** (more than 95% of the cell get reprogrammed), relatively homogeneous (Figure 7A) and very fast (cells are reprogrammed into functional macrophages in approximately a week, Figure 7B); furthermore, it involves **barely no cell divisions** nor major cell death (Di Tullio and Graf, 2012). It is therefore a system of choice to study the molecular regulation of cell fate. A reprogrammable B cell line was later developed (Bussmann et al., 2009), allowing performing many genomic assays in time courses.

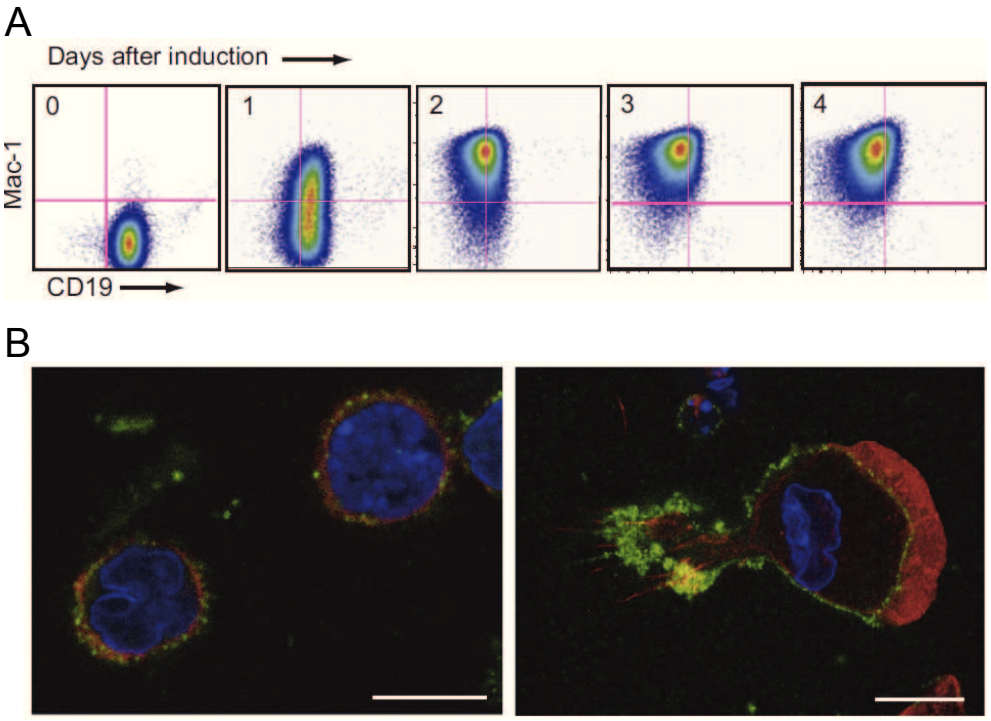


Figure 7: B cell trans-differentiation into macrophage through ectopic expression of C/EBPa

(A) Flow Cytometry analysis during B cell trans-differentiation into macrophages. The X axis corresponds to the expression for the B cell marker CD19, and the Y axis for the macrophage marker Mac-1. Vertical and horizontal lines mark the thresholds between positive and negative cells. (B) Fluorescence microscopy picture of B cell before (left) and 7 days after induction of C/EBPa (right). B cells are small, round and floating in culture; C/EBPa induced macrophage are larger, form lamelipodes, attach to the plastic culture plate and phagocyte bacteria. Adapted from (Bussmann et al., 2009).

What are the **molecular mechanism** controlling this process? Combining C/EBPa expression with other gain- or loss-of-function, the lab of Thomas Graf has shown that the B cell factors **PAX5 and EBF1 counteract** C/EBPa action and that **PU.1 is required** for the activation of the macrophages genes (Xie et al., 2004). The role of PU.1 was further supported by the fact that its expression together with that of C/EBPa in fibroblasts and T cells (as both cell types do not express PU.1, contrary to B cell), allowed their reprogramming into macrophages (Laiosa et al., 2006; Xie et al., 2004). As PU.1 has been described to be a **pioneer factor** (Zaret and Carroll, 2011) and is expressed in B cells, it could prime the chromatin for C/EBPa. C/EBPb might play a role in the trans-differentiation, as it is also induced by C/EBPa and can replace the later to reprogram B cell (Xie et al., 2004).

In order to decipher the role of each factor and their interaction, we studied the **dynamic of binding of each TF** to regulatory elements, genome-wide.

2.1.2 Publication: *C/EBPa Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis (Stem Cell Reports, 2015)*

Contribution:

I participated in the analysis of the ChIP-seq data, in particular in motifs analysis, and the comparison of TFs binding and histone marks in different sets of regions, using heatmaps and average plots.

Stem Cell Reports
Article
ISSCR
OPEN ACCESS

C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis

Chris van Oevelen,[1,*] Samuel Collombet,[2] Guillermo Vicent,[1] Maarten Hoogenkamp,[3] Cyrille Lepoivre,[4] Aimee Badeaux,[5] Lars Bussmann,[1] Jose Luis Sardina,[1] Denis Thieffry,[2] Miguel Beato,[1] Yang Shi,[5] Constanze Bonifer,[3] and Thomas Graf[1,*]
[1]Center for Genomic Regulation and Pompeu Fabra University, 08003 Barcelona, Spain
[2]Ecole Normale Supérieure, Institut de Biologie de l'ENS, INSERM, U1024, Centre National de la Recherche Scientifique (CNRS) 8197, 75005 Paris, France
[3]School of Cancer Sciences, Institute of Biomedical Research, University of Birmingham, Birmingham B15 2TT, UK
[4]CNRS, Aix-Marseille Université, IGS UMR7256, 13288 Marseille, France
[5]Harvard Medical School, Children's Hospital, Boston, MA 02115, USA
*Correspondence: chris.vanoevelen@crg.eu (C.v.O.), thomas.graf@crg.eu (T.G.)
http://dx.doi.org/10.1016/j.stemcr.2015.06.007
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

SUMMARY

Transcription-factor-induced somatic cell conversions are highly relevant for both basic and clinical research yet their mechanism is not fully understood and it is unclear whether they reflect normal differentiation processes. Here we show that during pre-B-cell-to-macrophage transdifferentiation, C/EBPα binds to two types of myeloid enhancers in B cells: pre-existing enhancers that are bound by PU.1, providing a platform for incoming C/EBPα; and de novo enhancers that are targeted by C/EBPα, acting as a pioneer factor for subsequent binding by PU.1. The order of factor binding dictates the upregulation kinetics of nearby genes. Pre-existing enhancers are broadly active throughout the hematopoietic lineage tree, including B cells. In contrast, de novo enhancers are silent in most cell types except in myeloid cells where they become activated by C/EBP factors. Our data suggest that C/EBPα recapitulates physiological developmental processes by short-circuiting two macrophage enhancer pathways in pre-B cells.

INTRODUCTION

The discovery that transcription factors (TFs) can convert somatic cells into both specialized and induced pluripotent stem cells (iPSCs) has revolutionized stem cell research and promises to have major clinical applications (Graf and Enver, 2009; Yamanaka and Blau, 2010). Lineage-instructive TFs activate and repress tissue-specific genes by recognizing sequence-specific DNA consensus motifs contained within enhancers and promoters (Ptashne, 2007). They establish gene regulatory networks (GRNs) of the novel gene expression program while dismantling those of the old program, involving the formation of feedforward, cross-inhibitory, and auto-regulatory loops (Bertrand and Hobert, 2010; Davidson, 2010; Graf and Enver, 2009; Holmberg and Perlmann, 2012). However, how these processes are coordinated and whether they recapitulate normal development remain unclear (Vierbuchen and Wernig, 2011), especially as neither TF-induced lineage conversions nor iPSC reprogramming appear to retrace normal developmental pathways (Apostolou and Hochedlinger, 2013; Di Tullio et al., 2011; Ladewig et al., 2013; Vierbuchen and Wernig, 2011).

Lineage-instructive TFs act through synergistic and cross-antagonistic interactions, are typically able to access closed chromatin (Zaret and Carroll, 2011), preferentially target sites with specific histone mark combinations, and bind to either nucleosome-depleted or nucleosome-dense regions (Soufi et al., 2012; Taberlay et al., 2011; Wapinski et al., 2013). However, what establishes these chromatin configurations in the first place and what proportion of the incoming reprogramming factors interacts with pre-existing TF complexes are largely unknown. A major reason for these gaps in our knowledge is that cell conversion frequencies in most cell systems are low, complicating efforts to study early events in a time-resolved fashion.

An exception is the transdifferentiation of pre-B/B cells into macrophages induced by the leucine zipper-type TF C/EBPα, which is arguably the most efficient and rapid system described so far (Bussmann et al., 2009; Di Tullio and Graf, 2012; Xie et al., 2004). C/EBPβ, like C/EBPα, can likewise induce B cell transdifferentiation into macrophages (Bussmann et al., 2009; Xie et al., 2004), but the two factors also have non-redundant functions. Mice ablated for C/EBPα die shortly after birth because they lack granulocyte/macrophage progenitors (GMPs, precursors of neutrophil granulocytes and macrophages, two closely related myeloid cell types) as well as granulocytes, while C/EBPβ-knockout animals are fully viable but contain macrophages and B cells with functional defects (Chen et al., 1997; Tanaka et al., 1995). C/EBPα cooperates with PU.1 (Spi1) to regulate myeloid gene expression (Friedman, 2007), the two factors interact physically (Reddy et al., 2002), and a combination of C/EBPα and PU.1 converts fibroblast into macrophage-like cells (Feng et al., 2008). The *Pu.1* gene encodes an Ets family TF specifically expressed in the early

stages of hematopoiesis and its knockout generates mice that lack both myeloid and lymphoid cells (Scott et al., 1994). Low-level expression of PU.1 in hematopoietic precursors induces B cell differentiation, whereas high levels favor myeloid differentiation (DeKoter and Singh, 2000).

Here we have analyzed, in a time-resolved manner, how C/EBPα establishes a myeloid expression program in pre-B cells, and we found that it binds to both pre-existing enhancers occupied by PU.1 and de novo enhancers where it acts as a pioneer factor. Strikingly, the combined activation of these enhancer types, regulating the expression of nearby macrophage genes, recapitulates the activation of myeloid enhancers and associated genes during normal hematopoiesis.

RESULTS

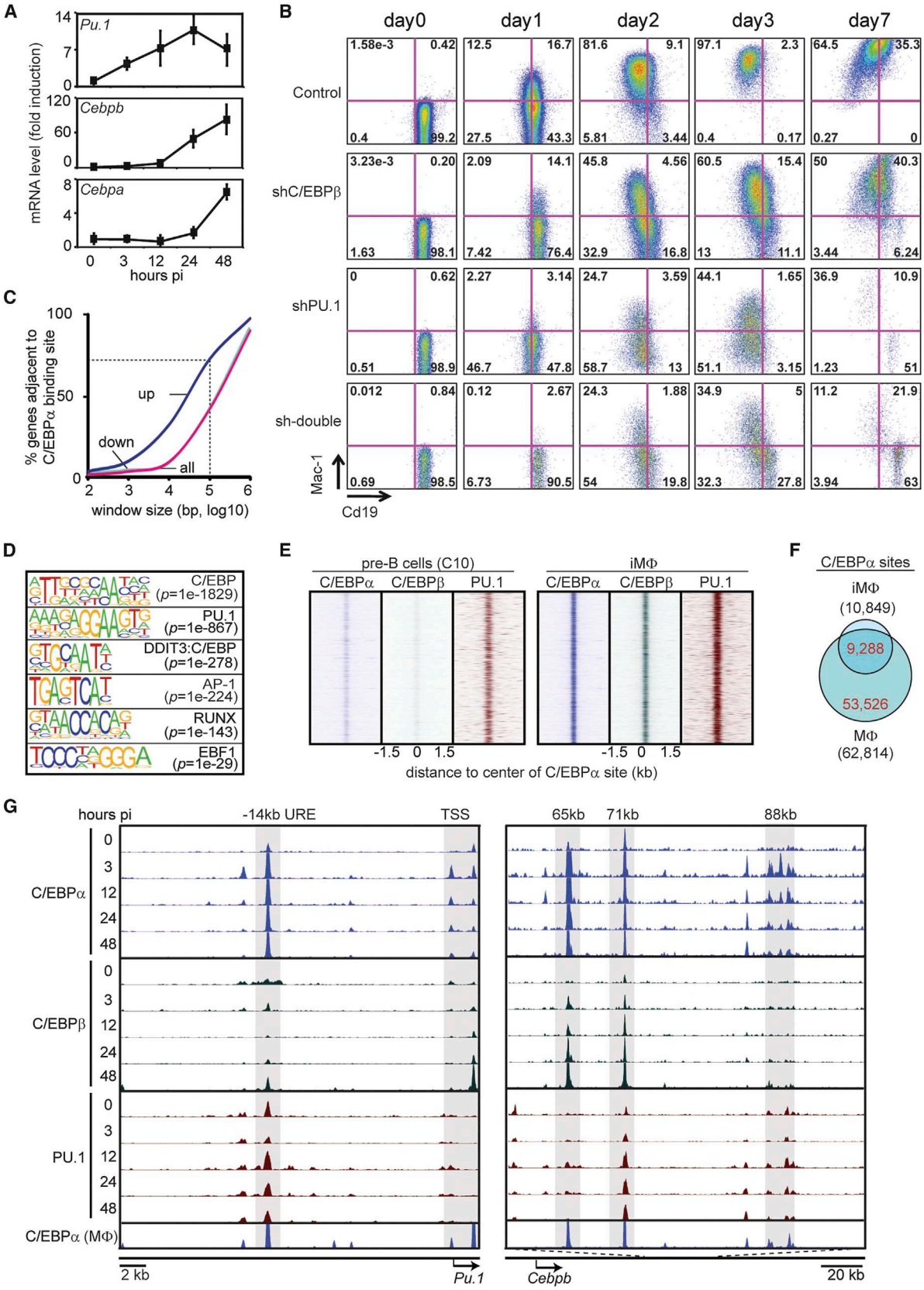
C/EBPα Induces High-Level Expression of *Pu.1* and *Cebpb*

To study how C/EBPα induces transdifferentiation, we used two pre-B cell lines that express an inducible C/EBPαER fusion protein tagged with either human CD4 (hCD4; C11 cells) or GFP (C10 cells). In both lines, treatment with 17 beta-estradiol (β-Est) shuttles C/EBPαER into the nucleus and induces the formation of macrophage-like cells within 2 to 3 days (Bussmann et al., 2009). Importantly, C/EBPα mRNA levels in C10 cells at 0 hr post-induction (hpi) or 24 hpi did not exceed C/EBPα levels observed in primary macrophages (MΦ) (Figure S1A). To monitor two important myeloid regulators known to cooperate with C/EBPα, we tested the expression levels of *Cebpb* and *Pu.1*. These genes were expressed at low to intermediate levels in pre-B cells (Figure S1B) and became upregulated within 3–12 hpi (Figure 1A). As C10 cells become transgene independent 24 hpi (Bussmann et al., 2009), i.e., before the expression of endogenous C/EBPα (Figure 1A), we determined whether the rapid activation of C/EBPβ and PU.1 is necessary for transdifferentiation. We generated C11 cells stably expressing a short hairpin RNA (shRNA) against C/EBPβ, PU.1, or both. Cells were induced with β-Est and analyzed by fluorescence-activated cell sorting (FACS) for the presence of Cd19 and Mac-1 (CD11b) at different days thereafter. At 3 days post-induction (dpi), the knockdown of C/EBPβ and of PU.1 resulted in a 35% and 50% reduction in the formation of Mac-1^{+}Cd19^{-} cells, respectively, while deleting both factors further enhanced the effect. At 7 dpi, Mac-1 expression in shC/EBPβ cells caught up with wild-type levels, whereas cells expressing shPU.1 exhibited extensive cell death (Figures 1B and S1C). These data show that C/EBPα rapidly upregulates *Pu.1* and *Cebpb,* that PU.1 is necessary to establish the myeloid GRN, and that C/EBPβ plays a more minor role.

A Limited Set of Sites Stably Bound by C/EBPα Correlates with the Upregulation of Macrophage Genes

To explore the mechanism by which C/EBPα turns on the myeloid program in pre-B cells, we treated C10 cells for different times with β-Est and performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) experiments, using antibodies to C/EBPα, C/EBPβ, and PU.1 (Table S1 gives a summary of ChIP-seq results and peak calling). A total of 54,198 non-redundant C/EBPα-enriched regions could be detected during the time course of which 10,849 sites were stably bound (i.e., up to 48 hpi, Table S2), whereas the remaining sites were transiently bound. Genes nearest stable binding sites, but not transient sites, were enriched for upregulated genes (Figure S1D). In addition, using a sliding-window approach, we observed that 70% of upregulated genes were localized within 100 kb of a stable C/EBPα-binding site, whereas no such enrichment was seen for downregulated genes (Figure 1C). Motif analysis of the stable sites in 48-hpi cells (hereafter referred to as induced macrophages or iMΦ) showed strong enrichment for consensus motifs of C/EBP and PU.1. The same sites also were enriched for AP-1 (Jun and Fos) and RUNX motifs, as previously reported (Figure 1D; Heinz et al., 2010) and more weakly enriched for EBF1 (Figure 1D; also see Figure 3). The majority of stable C/EBPα sites were co-occupied by C/EBPβ and PU.1 in iMΦ, and ~40% of these were pre-bound by PU.1 in pre-B cells, however, showing lower intensity signals (Figure 1E). Low-intensity signals in pre-B cells also were detectable for C/EBPβ, reflecting its low-level expression, as well as for C/EBPα (Figure S1E), suggesting some leakiness of the transgene.

A total of 10,849 C/EBPα sites were detected in iMΦ and 62,814 in bone-marrow-derived macrophages (MΦ) (Zhang et al., 2013), showing 9,288 common sites (Figure 1F). The larger number of sites in MΦ compared to iMΦ cannot be explained by a higher sequencing depth (Table S1). However, these differences became smaller when the numbers of associated genes were compared as follows: C/EBPα sites combined with 5,849 and 14,078 genes in iMΦ and MΦ, respectively, and shared 5,252 genes (Figure S1F). The shared gene set was enriched for genes that became upregulated during transdifferentiation of primary B cells into macrophages (Di Tullio et al., 2011), whereas the gene set unique for MΦ (8,826) was actually depleted (Figure S1G). In addition, shared upregulated genes were enriched for gene ontology (GO) terms associated with myeloid function, while upregulated genes unique for MΦ were not (Figure S1H). The induced rapid and efficient conversion of pre-B cells into highly motile, aggregating, and phagocytic macrophages within 51 hr (Movie S1; Bussmann et al., 2009) further supports the interpretation that C/EBPα



(legend on next page)

binds to a core set of enhancers in iMΦ required for myeloid cell specification.

The binding of C/EBPα, C/EBPβ, and PU.1 in induced C10 cells and of C/EBPα in primary MΦ is illustrated for the promoter and the −14-kb URE enhancer of the *Pu.1* gene (Yeamans et al., 2007), for three putative enhancers of *Cebpb* (Figure 1G), as well as for putative enhancers of *Fos* and *Il1b* (Figure S1I). (Genomic coordinates of these and other regions are described in Table S3.) Together, our data suggest that C/EBPα combined with C/EBPβ and PU.1 activates a core set of enhancers, shared between the cell line and primary macrophages, required to induce macrophage specification.

Prospective Myeloid Enhancers in Pre-B Cells Fall into Two Broad Classes

To characterize the epigenetic status of prospective myeloid regulatory regions in pre-B cells, we performed ChIP-seq experiments for histone modifications characteristic of poised (H3K4Me1), active (H3K27Ac, P300), and repressed enhancers (H3K27Me3) (Creyghton et al., 2010; Rada-Iglesias et al., 2011), and we analyzed levels of these marks at C/EBPα sites away from the transcription start site (TSS), representing putative enhancers (Figure S2A). We observed two broad classes of prospective myeloid enhancers in pre-B cells as follows: (1) pre-existing enhancers that were decorated with H3K4Me1, H3K27Ac, and P300 and depleted for H3K27Me3 (Figure 2A); and (2) de novo enhancers that lacked any of the active enhancer marks but were instead often decorated with H3K27Me3 (Figure 2A). Similar results were obtained by performing ChromHMM analysis (Figure S2C; Ernst and Kellis, 2012) as an independent analytical approach demonstrating that pre-existing enhancers are enriched for activation marks, whereas de novo enhancers are depleted for activation marks and enriched for H3K27Me3 (Figure S2D). Approximately two-thirds of the pre-existing enhancers were bound by PU.1 and exhibited high levels of activation marks compared to sites not bound by PU.1 (Figure S2B). Importantly, we confirmed the presence of pre-existing- and de-novo-type enhancers in primary B cells by using recently published ChIP-seq datasets (Figure 2B; Heinz et al., 2010; Lara-Astiaso et al., 2014). Furthermore, we confirmed binding of C/EBPα to selected pre-existing enhancers of the *Pu.1*, *Cebpb*, *Il1b*, *Ehd1*, and *Ifngr2* genes and to de novo enhancers of the *Mmp12*, *Cd14*, *Gbe1*, *Fos*, and *Fgd4* genes in primary pre-B cells induced to transdifferentiate (Figures 2C and 2D; Di Stefano et al., 2014).

To determine whether the enhancer activation state in pre-B cells correlates with gene expression, we first analyzed the promoter configuration of adjacent genes and found the following: 73% of the pre-existing enhancers paired with active promoters (as defined by the sole presence of H3K4Me3), 7% were decorated with inactive promoters (H3K27Me3 or no marks), and 20% were decorated with promoters containing a bivalent domain (H3K4Me3 and H3K27Me3) (Figure 2E). In contrast, only 36% of de novo enhancers paired with active promoters, 44% with inactive promoters, and 20% with bivalent promoters (Figures 2E and S2E). Based on these results, we redefined pre-existing enhancers as those that combine with active promoters and de novo enhancers as those that pair with inactive promoters. As expected, genes associated with pre-existing enhancers already were expressed at significant levels in pre-B cells and primary pre-B cells (Bussmann et al., 2009; Di Tullio et al., 2011), whereas genes associated with de novo enhancers only showed background expression levels (Figure 2F).

The finding that pre-B cells express genes associated with pre-existing enhancers predicts that PU.1 expression in cells devoid of PU.1 will selectively activate these genes. To test this, we expressed PU.1 in 3T3 fibroblasts (Figure 2G) and measured mRNA levels of a number of genes associated with either pre-existing or de novo enhancers. Supporting the hypothesis that PU.1 preferentially binds to pre-existing enhancers and activates associated genes, we observed that 6 of 11 pre-existing enhancer-associated genes tested were upregulated as compared to 2 of 10 de novo genes (Figure 2G; primer sequences are in Table S4).

Figure 1. Upregulation of *Cebpb* and *Pu.1* Genes by C/EBPα and Effects of Their Knockdown on Transdifferentiation

(A) Expression of endogenous *Pu.1*, *Cebpb*, and *Cebpa* RNA after β-Est induction of C10 cells as measured by qRT-PCR. Data are represented as mean ± SEM (independent triplicates) expressed as the fold induction relative to uninduced pre-B cells.

(B) FACS plots of C11 pre-B cell carrying either a scrambled short hairpin knockdown construct (control) or constructs against C/EBPβ, PU.1, or both, and induced by β-est treatment. See also Figure S1C.

(C) Percentage of upregulated or downregulated genes (>2-fold) within defined windows around C/EBPα sites. Dotted lines indicate that 70% of all upregulated genes are within 100 kb of a C/EBPα-binding site.

(D) Significantly enriched sequence motifs at C/EBPα-binding sites as determined by HOMER.

(E) Heatmaps visualizing C/EBPα, C/EBPβ, and PU.1 binding in pre-B cells and iMΦ. Window, 3 kb; bin, 10 bp. See also Figure S1E.

(F) Venn diagram showing the intersection of C/EBPα sites in iMΦ (n = 10,849) and primary MΦ (n = 62,814).

(G) Screenshots of C/EBPα, C/EBPβ, and PU.1 binding at selected enhancers in C10 cells and of C/EBPα in primary MΦ. Arrows indicate TSS, length of ORF, and direction of transcription. See also Figure S1I.

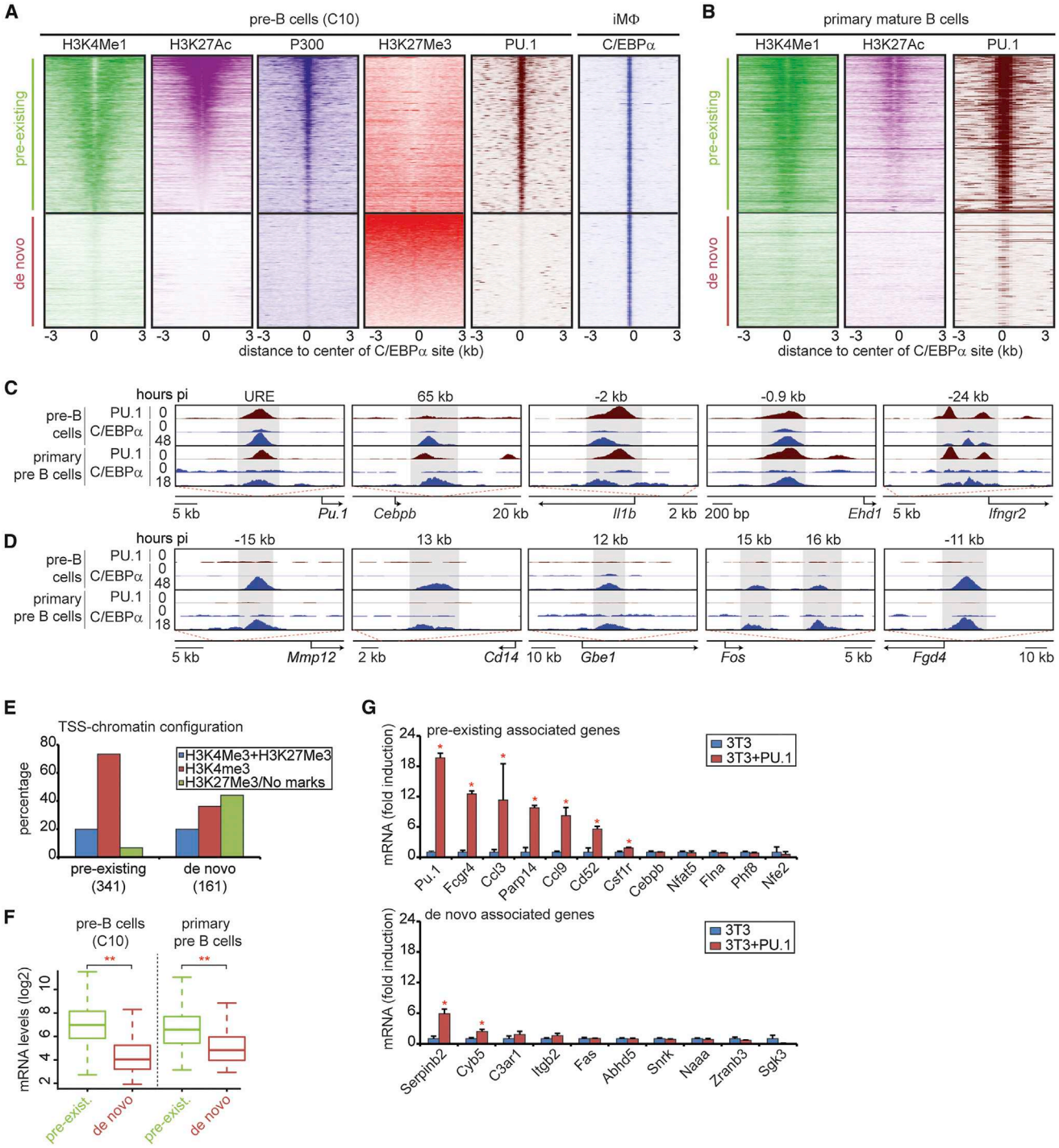


Figure 2. Prospective Myeloid Enhancers with Either Pre-existing or De Novo Configurations in Pre-B Cells

(A) Heatmaps visualizing H3K4Me1, H3K27Ac, P300, H3K27Me3, and PU.1 in pre-B cells and C/EBPα binding in iMΦ. Center of C/EBPα binding, 0; window, 6 kb; bin, 100. See also Figures S2B–S2D.

(B) As in (A), visualizing H3K4Me1, H3K27Ac, and PU.1 in primary mature B cells.

(C and D) Screenshots of C/EBPα and PU.1 binding in C10 cells (0 and 48 hpi) and primary pre-B cells (0 and 18 hpi) at pre-existing (C) and de novo enhancers (D).

(E) Distribution of genes nearest pre-existing or de novo enhancers marked with bivalent (H3K4Me3, H3K27Me3), active (H3K4Me3), or repressed (H3K27Me3 or no marks) promoters. See also Figure S2E.

(legend continued on next page)

Together, our findings suggest that C/EBPα is capable of activating two broad classes of prospective myeloid enhancers in pre-B cells as follows: (1) pre-existing enhancers with active enhancer marks that are predominantly associated with expressed genes, and (2) de novo enhancers lacking such marks that are predominantly associated with silenced genes.

A Subset of Pre-existing Myeloid Enhancers Is Bound by the B Cell TF Ebf1 in Pre-B Cells

Our finding that motifs associated with the B cell TF Ebf1 are enriched in myeloid enhancers (see Figure 1D) prompted us to study their relevance in transdifferentiation. Analysis of the Ebf1 motif distribution shows that it is specifically enriched in pre-existing enhancers (Figure 3A). To test actual binding of Ebf1, we performed ChIP-seq experiments in pre-B cells yielding 6,627 Ebf1 peaks that were predominantly located in intergenic regions (Figure 3B) and enriched for EBF1, ETS (PU.1), and E2A motifs (Figure 3C). In line with the motif analysis, intersection of Ebf1-bound sites with the two types of myeloid enhancers showed 725 that were associated with pre-existing enhancers, but virtually none with de novo enhancers (Figure 3D).

To determine the functional state of the Ebf1-bound enhancers targeted by C/EBPα, we determined the kinetics of Ebf1 and C/EBPα binding as well as H3K27Ac decoration after induction of transdifferentiation. The heatmaps in Figure 3E show that, while C/EBPα binding already was observed after 3 hpi, the loss of Ebf1 binding was not detected until 48 hpi. However, H3K27Ac enrichment at these enhancers was maintained throughout the time course (Figure 3E), suggesting that the relevant enhancers remain active even after the loss of Ebf1. Examples of enhancers bound by Ebf1, C/EBPα, and PU.1 are shown in Figure 3F. This includes the 88-kb putative enhancer of *Cebpb* (see Figure 1G), as well as *Nfe2* and *Cd40* enhancers. In addition to enhancers bound by Ebf1, C/EBPα, and PU.1, 27% lack PU.1 binding, as exemplified by the −150-kb *Tgfbr2* putative enhancer (Figure 3F; additional examples are shown in Figure S3A). Ebf1 binding to these regions was confirmed using an independent Ebf1 ChIP-seq dataset (Treiber et al., 2010; Figures S3B and S3C). Importantly, genes associated with putative Ebf1-C/EBPα-bound enhancers were upregulated during transdifferentiation (Figure S3D).

In conclusion, a significant proportion of pre-existing myeloid enhancers targeted by C/EBPα in pre-B cells are bound by the B cell TF Ebf1. This finding raises the possibility that pre-existing myeloid enhancers act as bona fide B cell enhancers and that C/EBPα converts them into enhancers active in myeloid cells.

C/EBPα Acts Both as a Pioneer and as a Secondary Factor at Prospective Myeloid Enhancers

To study how the two enhancer types become activated, we determined the binding kinetics of PU.1, C/EBPα, and C/EBPβ. As expected, at pre-existing enhancers PU.1 was bound throughout the time course, whereas it was initially absent at de novo enhancers, gradually increasing after induction (Figures 4A and S4A). In contrast, C/EBPα binding showed a steeper increase at pre-existing than at de novo enhancers, with both converging at 48 hpi and the rate of C/EBPα binding kinetics correlating with the starting levels of H3K27Ac or H3K27Me3, respectively (Figure S4B). Finally, C/EBPβ occupancy increased steadily at the two enhancer types (Figures 4A and S4A).

The binding profiles observed predict that at de novo enhancers C/EBPα binds before PU.1. Indeed, ChIP-seq experiments with induced C10 cells at early time points (10, 30, and 60 min post-induction) showed that C/EBPα binds to 74% of de novo sites *before* PU.1 (Figure 4B). An example of a putative pre-existing enhancer bound by PU.1 first is shown for the *Tyrobp* gene; examples of de novo enhancers are the 24-kb site of *Tlr4*, the 65-kb enhancer of *Cebpb*, and the −16-kb site of *Ctsd* (Figure 4C).

To further study the interplay between PU.1 and C/EBPα, we knocked down PU.1 in pre-B cells (Figure S4C), induced transdifferentiation for 3 and 24 hr, and analyzed C/EBPα binding at five pre-existing and five de novo enhancers. In control cells we observed higher binding of C/EBPα at 3 hpi for the pre-existing relative to the de novo enhancers. In addition, knockdown of PU.1 caused an initial decrease of C/EBPα binding at 3 hpi for both enhancer types on all loci tested (Figures 4D and 4E; primer sequences are in Table S4). However, at 24 hpi, C/EBPα binding recovered to control levels or even above in 9 of 10 enhancers tested (Figures 4D and 4E). This suggests that C/EBPα binding at pre-existing enhancers does not strictly require PU.1, raising the possibility that C/EBPα can access closed chromatin (see also Figures 2A and S4B).

To test this more directly, we performed micrococcal nuclease (Mnase) digestion experiments with chromatin isolated from pre-B cells and iMΦ cells and deep-sequenced nuclease-protected DNA. Average nucleosome profiles

(F) Distribution of mRNA levels of upregulated genes nearest to either pre-existing (n = 318) or de novo (n = 103) enhancers in pre-B cells (C10) and primary pre-B cells. Statistical analysis by Wilcoxon rank-sum test, $^{**}p < 0.001$.

(G) Expression of genes nearest pre-existing or de novo enhancers in 3T3 cells or 3T3 overexpressing PU.1 by qRT-PCR. Data are represented as mean ± SEM (independent triplicates) and expressed as the fold induction relative to 3T3 cells. Statistical analysis by Student's t test, $^{*}p < 0.05$.

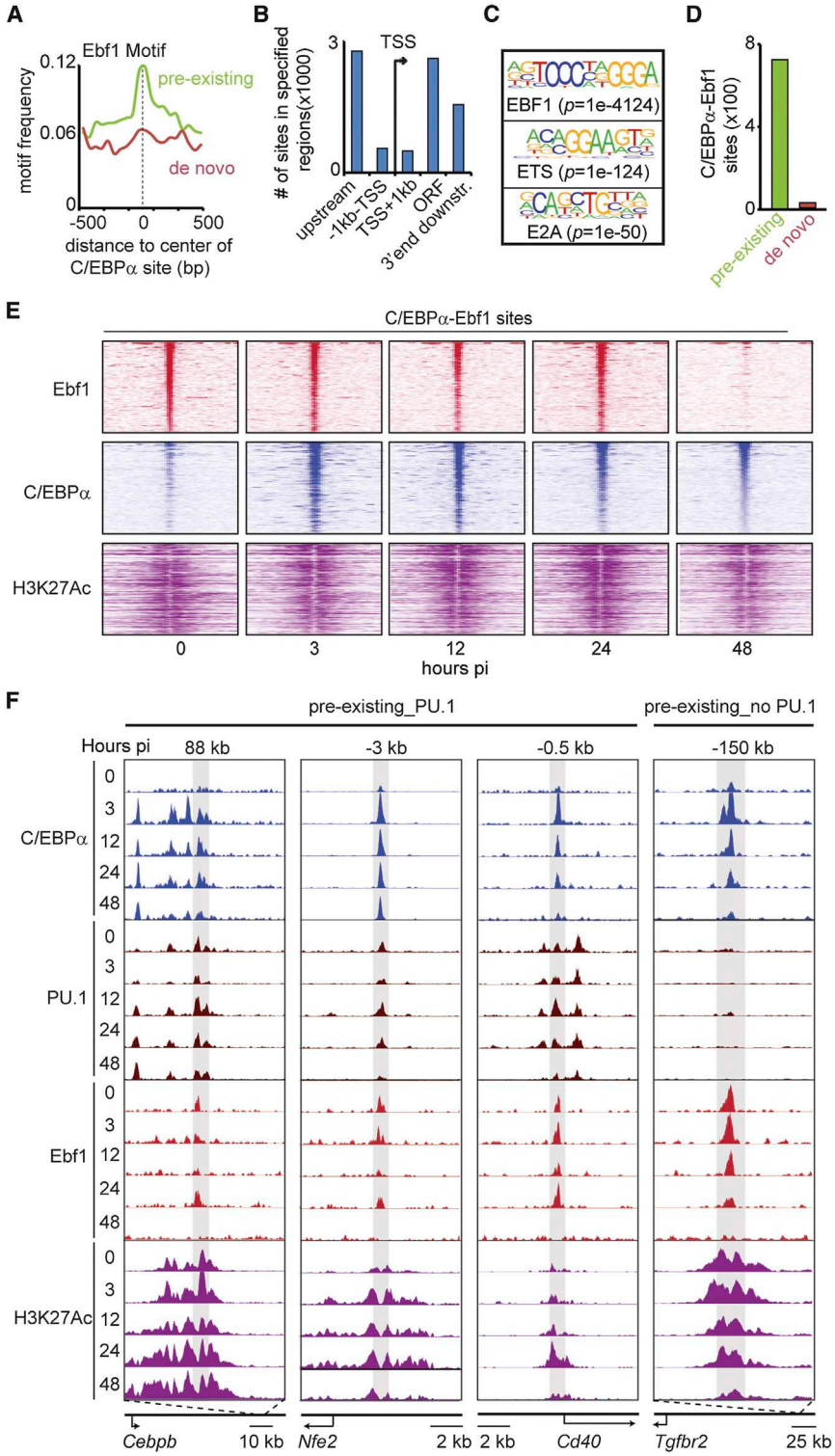


Figure 3. Binding of the B Cell TF Ebf1 to Pre-existing Enhancers

(A) Frequency of Ebf1 motif within pre-existing and de novo C/EBPα-binding sites by HOMER.

(B) Genomic distribution of Ebf1-binding events (n = 6,627) relative to the TSS in C10 cells. ORF, open reading frame.

(C) Significantly enriched sequence motifs at Ebf1-binding sites, as determined by HOMER.

(D) Number of C/EBPα sites bound by Ebf1 for each enhancer type.

(E) Heatmaps, centered on C/EBPα binding in iMΦ, visualizing Ebf1, C/EBPα, and H3K27Ac after the induction of transdifferentiation. Center of binding, 0; window, 6 kb; bin, 100.

(F) Screenshots of C/EBPα, PU.1, Ebf1, and H3K27Ac ChIP-seq profiles at selected enhancer regions in C10 cells. See also Figures S3A–S3D.

calibrated with sites uniquely bound by PU.1 revealed a nucleosome-depleted region (valley) flanked by two positioned nucleosomes (Figure S4D) that confirmed an earlier report (Heinz et al., 2010). Pre-existing enhancers bound by PU.1 in pre-B cells showed a small valley that became more pronounced in iMΦ (Figure 4F). In contrast, de novo enhancers targeted by C/EBPα in pre-B cells were contained in a nucleosome-dense region that changed into a profile

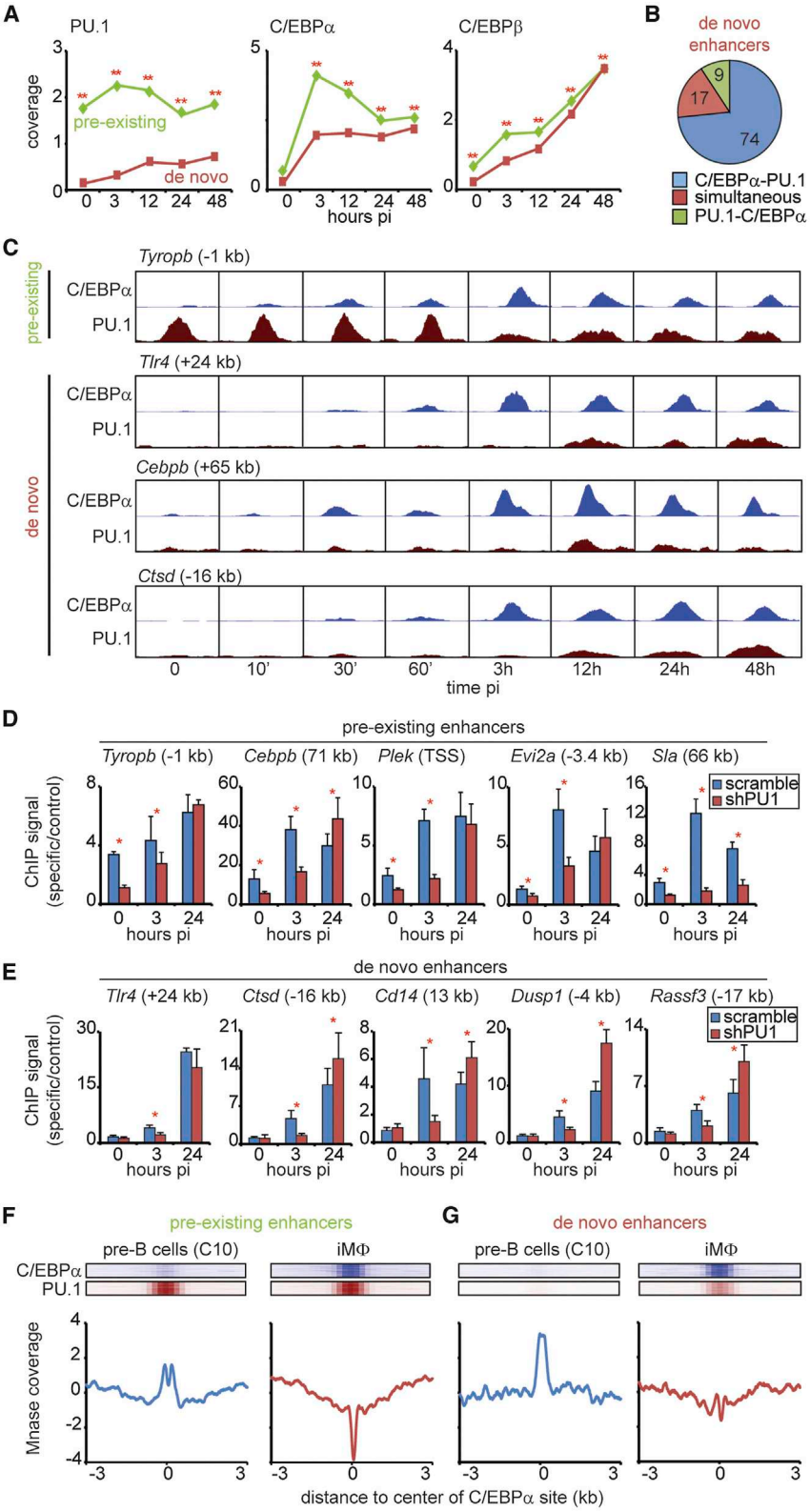


Figure 4. C/EBPα Binds to Open and Closed Chromatin

(A) Kinetics of PU.1, C/EBPα, and C/EBPβ binding at center position of pre-existing (n = 4,711) and de novo (n = 3,424) enhancers at indicated hpi. Statistical analysis by Wilcoxon rank-sum test, **p < 0.001. See also Figures S4A and S4B.

(B) C/EBPα and PU.1 binding order on de novo sites (C/EBPα first [C/EBPα-PU.1], simultaneously, or PU.1 first [PU.1-C/EBPα].

(C) Screenshots of C/EBPα and PU.1 binding at selected pre-existing or de novo enhancers (10, 30, and 60 min).

(D and E) C/EBPα binding at pre-existing (D) or de novo (E) enhancers in induced C10 cells knocked down for PU.1. See also Figure S4C. Data are represented as mean ± SEM (independent triplicates). Statistical analysis by Student's t test, *p < 0.05. Primer sequences are given in Table S4.

(F and G) Average MNase profiles at pre-existing (F) and de novo (G) enhancers bound by C/EBPα and PU.1 (shown in blue and brown). Profiles were centered on PU.1 binding in iMΦ and normalized by median subtraction. Window, 6 kb; bin, 1 bp. See also Figures S4D and S4E.

similar to that observed for pre-existing enhancers, although less pronounced, in iMΦ (Figure 4G). No ordered nucleosome patterns were obtained with profiles centered on random genomic positions (Figure S4E).

Our data show that C/EBPα binds to a nucleosome-depleted region in pre-existing enhancers and to a nucleosome-dense region in de novo enhancers. These findings support the notion that C/EBPα can act as a pioneer factor.

The C/EBPα and PU.1 Binding Order Determines the Activation Kinetics of Adjacent Genes

To determine whether the epigenetic status of the two enhancer types in pre-B cells influences their subsequent activation kinetics, we analyzed enrichment levels of H3K4Me1, H3K27Ac, P300, and H3K27Me3 during transdifferentiation. As judged by P300 and H3K27Ac, pre-existing enhancers became hyper-activated, albeit mostly in a transient manner. In contrast, de novo enhancers became gradually activated, starting from background levels (Figures 5A, 5B, and S5A). Both enhancer types followed a similar sequence of enhancer mark acquisition, consisting in P300 binding followed by H3K4Me1 and H3K27Ac decoration (Figure 5B). In contrast, the repressive H3K27Me3 decreased, predominantly on de novo enhancers (Figures 5C and S5B). These findings are illustrated for the pre-existing 3-kb FIRE enhancer of the *Csf1r* gene and the de novo −16-kb enhancer of *Ctsd* (Figure 5D). Additional examples are shown in Figure S5C.

To determine how the two types of prospective myeloid enhancers modulate the upregulation kinetics of adjacent genes, we interrogated gene expression data from C10 cells and primary pre-B cells induced to transdifferentiate (Bussmann et al., 2009; Di Tullio et al., 2011). Pre-existing enhancer-associated genes started from low expression levels and became gradually upregulated ∼4-fold, while de novo enhancer-associated genes started from background levels and were upregulated ∼9-fold (Figures 5E, 5F, S5D, and S5E).

In sum, the C/EBPα and PU.1 binding order determines the activation kinetics of targeted enhancers, with pre-existing enhancers becoming activated gradually from detectable base levels and de novo enhancers becoming activated more steeply and with a delay. These differences also are reflected in the activation kinetics of adjacent genes.

Pre-existing and De Novo Enhancers Are in an Active State in Distinct Hematopoietic Cell Types

Are the pre-existing and de novo myeloid enhancers identified during transdifferentiation relevant for normal hematopoietic differentiation? To study this we determined their activation state in various types of immature and mature hematopoietic cells and interrogated expression data of associated genes during hematopoiesis (Lara-Astiaso et al., 2014; see Figure 6A for the hematopoietic lineage tree and nomenclature used). Surprisingly, ∼58% of pre-existing enhancers already were active (i.e., marked by H3K27Ac) in long-term hematopoietic stem cells (LT-HSCs), and their proportion further increased in common myeloid and lymphoid progenitors (CMPs and CLPs, respectively), reaching ∼66% and ∼74% in terminally differentiated granulocytes (Gns) and MΦ, respectively (Figures 6B and S6A). Moreover, a substantial fraction of pre-existing enhancers remained active in B cells (60%) but decreased in megakaryocyte-erythroid progenitors (MEPs), erythroid cells (Erys), and T cells (∼30%) (Figures 6B and S6A). In contrast, activated de novo enhancers were essentially restricted to the myeloid compartment with 25%–28% being decorated with H3K27Ac in CMPs and GMPs and ∼40% in Gns and MΦs, while HSCs and multipotent progenitors (MPPs) showed lower percentages (7% and 14%) and MEPs, Erys, and B and T cells were essentially negative (Figures 6C and S6A). Similar trends were observed for pre-existing and de novo enhancers marked with H3K4Me1 (Figure S6B). Heatmaps of the two enhancer types during the transition from short-term hematopoietic stem cells (ST-HSCs) to macrophages (Figure 6D) were remarkably similar to those of pre-B cells transdifferentiating into macrophages (see Figure 5A). In contrast, the two enhancer types were not activated in T cells (Figure 6E).

These findings are illustrated for the pre-existing −14-kb URE enhancer of *Pu.1* and the 88-kb putative enhancer of *Cebpb* (Figure 6F), which are bound by PU.1 (Figure 1G). A de novo enhancer is exemplified by the 65-kb enhancer of *Cebpb* (Figure 6F). Strikingly, mRNA levels of genes associated with pre-existing and de novo enhancers reflected enhancer activity during hematopoiesis using two independently derived datasets analyzed by either RNA-seq or expression arrays (Figures 6G, S6C, and S6D; Lara-Astiaso et al., 2014; Di Tullio et al., 2011). The arrays also showed that in normal macrophages the expression levels of genes associated with the two enhancer types nearly converged (Figure S6D).

In conclusion, our data show that the majority of pre-existing enhancers targeted by C/EBPα during transdifferentiation are broadly active in hematopoietic stem cells, progenitors, and B cells, whereas de novo enhancers are largely restricted to the myeloid compartment.

The Activity of the Two Enhancer Types Reflect *Pu.1*, *Cebpa*, and *Cebpb* Expression during Hematopoiesis

How are the two enhancer types observed during C/EBPα-induced transdifferentiation controlled during normal hematopoiesis? To study this we analyzed the expression of *Pu.1*, *Cebpa*, and *Cebpb* during hematopoietic differentiation. *Pu.1* expression was found to closely correlate with that of pre-existing enhancers, *Pu.1* being broadly

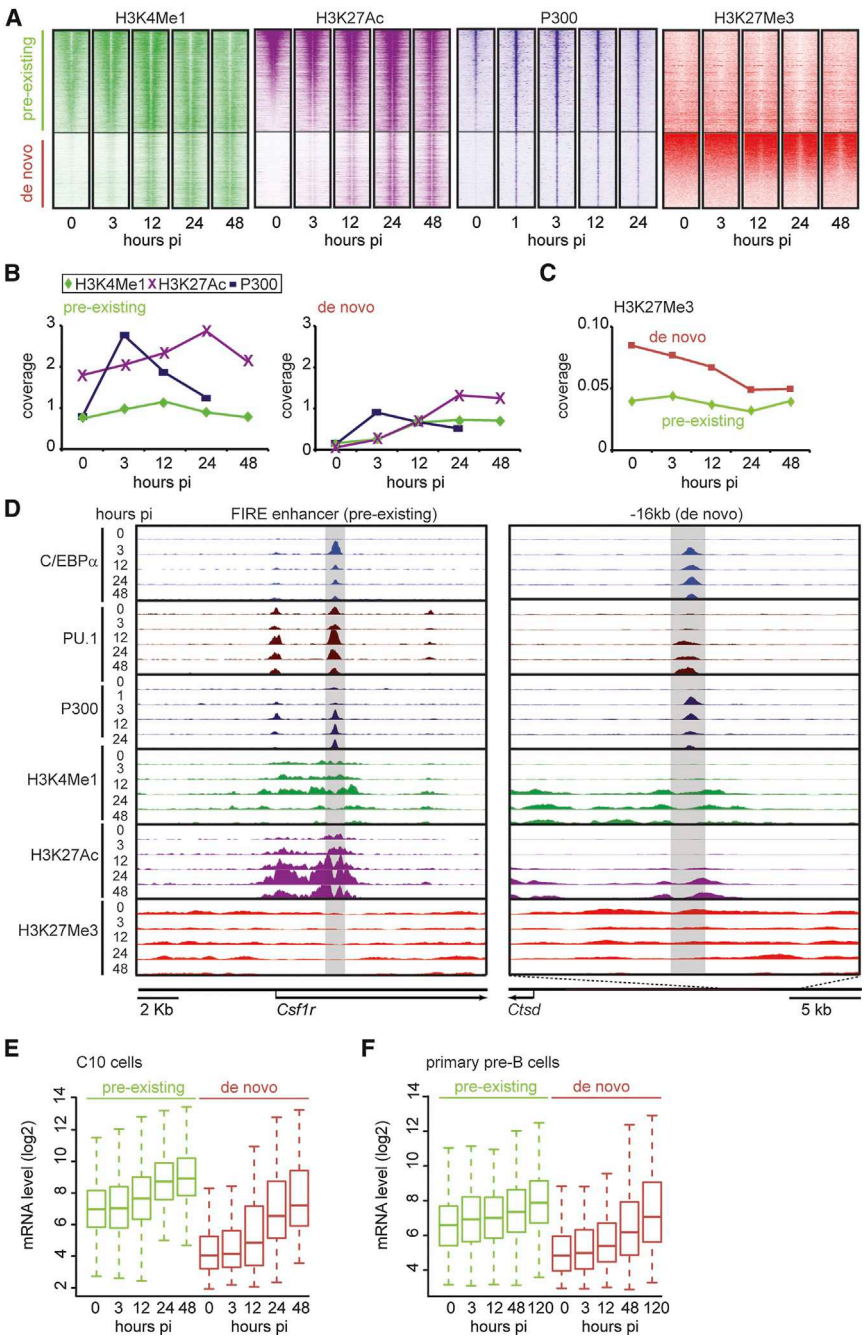


Figure 5. Kinetics of Histone Marks, P300 Binding, and Gene Expression Associated with Myeloid Enhancers

(A) Heatmaps visualizing H3K4Me1, H3K27Ac, P300 binding, and H3K27Me3 at pre-existing and de novo enhancers at different hpi of induced C10 cells. Window, 6,000 bp; bin, 100. See also Figure S5A.

(B) Quantification of H3K4Me1, H3K27Ac, and P300, as in (A). Bins with the highest coverage are shown.

(C) Quantification of H3K27Me3, as in (A), except values at the center position are shown. See also Figure S5B.

(D) Screenshots of selected enhancers showing C/EBPα, PU.1, P300, H3K4Me1, H3K27Ac, and H3K27Me3 profiles in C10 cells. See also Figure S5C.

(E and F) Distribution of mRNA levels of upregulated genes nearest to either pre-existing (n = 318) or de novo (n = 103) enhancers during transdifferentiation of C10 cells and primary pre-B cells is shown. See also Figures S5D and S5E.

expressed in stem and progenitor cells and weakly in T cells, MEPs, and Erys (Figure 7A). Similar expression patterns were observed at the protein level with PU.1 reporter mice (Back et al., 2005). In turn, *Cebpa* was expressed mostly in the myeloid compartment where its levels were highest in GMPs (Figure 7A; Wölfler et al., 2010), in agreement with the fact that mice lacking C/EBPα do not develop GMPs (Zhang et al., 2004). In contrast, *Cebpb* expression reached highest levels in macrophages and Gns (Figure 7A), suggesting that C/EBPβ takes over the role of C/EBPα in terminally differentiated myeloid cells. This interpretation agrees with the fact that macrophages from C/EBPβ-knockout mice have functional defects (Chen et al., 1997; Tanaka et al., 1995).

To test whether the *Cebpa* expression pattern reflects its binding specificity in the hematopoietic system, we analyzed the C/EBPα-binding sites identified in pre-existing and de novo enhancers in stem and progenitors cells

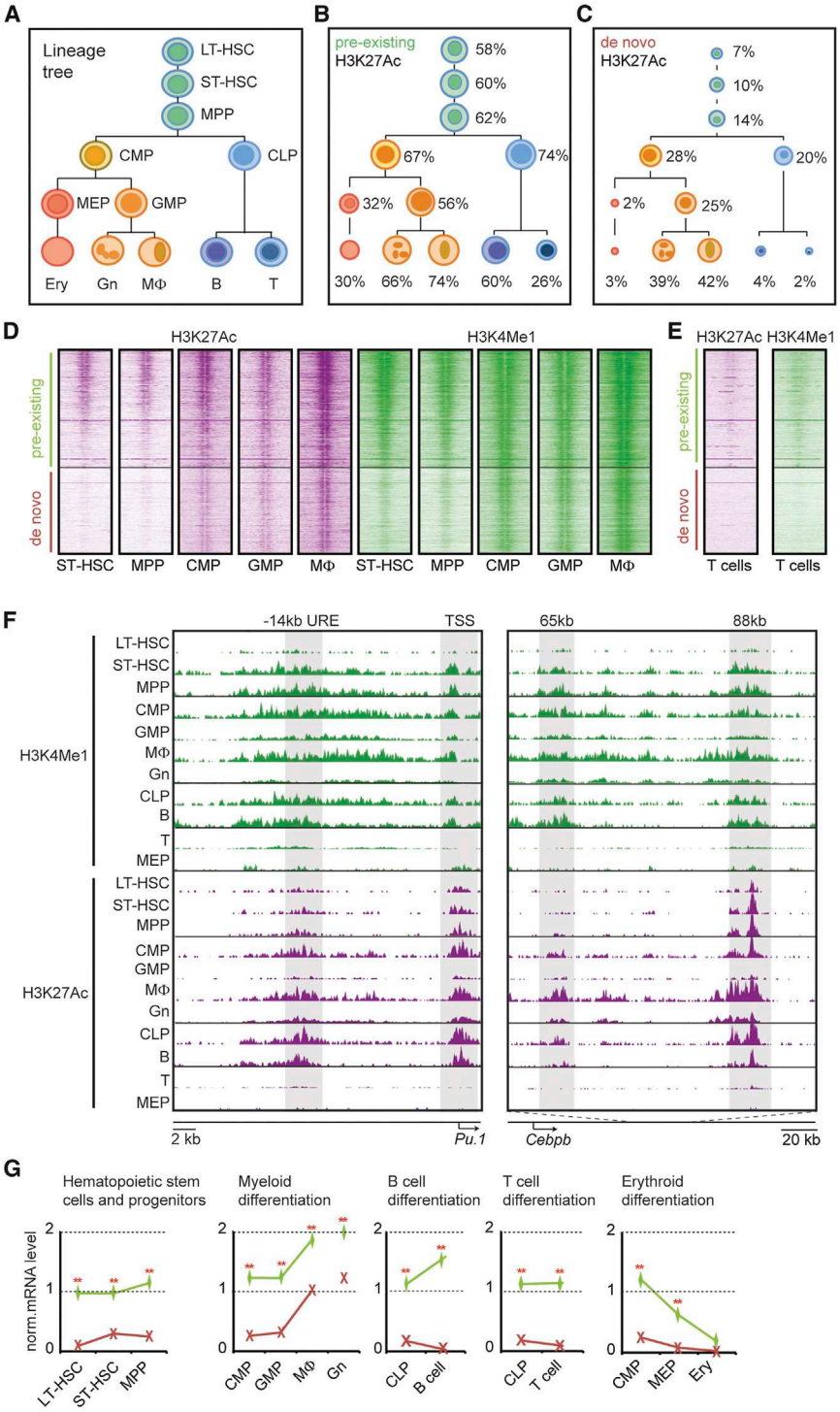


Figure 6. Distribution of Active Myeloid Enhancers during Hematopoietic Differentiation

(A) Cartoon depicting blood cell lineage specification.

(B and C) Percentage of pre-existing and de novo macrophage enhancers intersecting with enhancers decorated with H3K27Ac in different hematopoietic progenitors and differentiated cell types. The size of the circles relative to circles in (A) indicates the percentage of representation. See also Figures S6A and S6B.

(D) Heatmaps visualizing H3K27Ac and H3K4Me1 decoration at pre-existing and de novo enhancers during myeloid differentiation. Window, 6,000 bp; bin, 100.

(E) As in (D), but for T cells.

(F) Screenshots of H3K4Me1 and H3K27Ac profiles at selected C/EBPα-bound enhancers of *Pu.1* and *Cebpb* in the indicated hematopoietic cell types.

(G) Median mRNA levels of genes nearest either pre-existing (green lines; n = 318) or de novo (red lines; n = 103) enhancers in different hematopoietic stem/progenitors and differentiated cells. Statistical analysis by Wilcoxon rank-sum test, $^{**}p < 0.001$. See also Figures S6C and S6D.

and GMPs, as previously reported (Hasemann et al., 2014). Strikingly, <10% of prospective myeloid enhancers were bound by C/EBPα in the progenitors, while ~80% of the sites were bound in GMPs and in primary macrophages (Figure 7B). C/EBPα binding in progenitor cells and primary MΦ is illustrated for the *PU.1* and *Cebpb* genes (Figure 7C) as well as for the *Tlr4* and *Ctsd* genes.

Together, our observations indicate that, within the hematopoietic system, the combination of *Pu.1*, *Cebpa*, and *Cebpb* determines the activity of the two types of

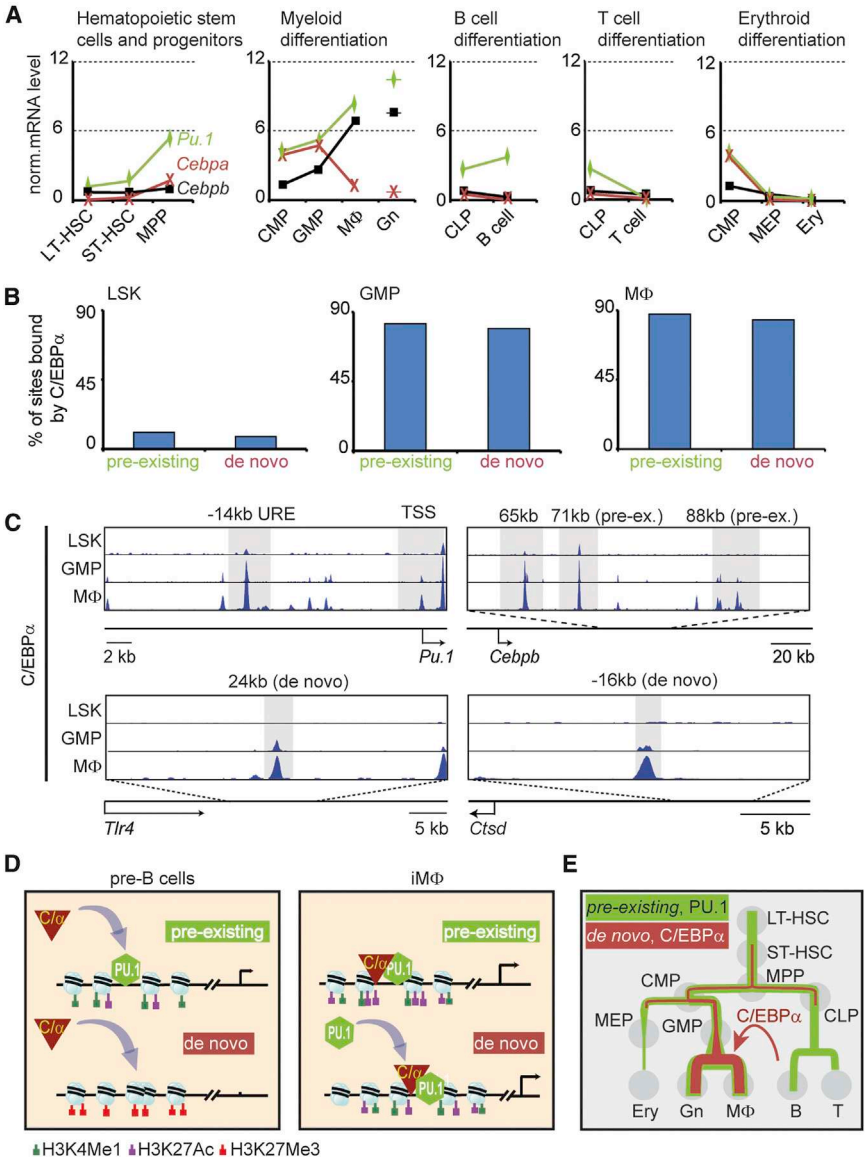


Figure 7. Expression Kinetics of Genes Associated with Pre-existing and De Novo Enhancers during Hematopoiesis

(A) mRNA levels of *Pu.1*, *Cebpa*, and *Cebpb* in different hematopoietic stem/progenitor cell types, based on RNA-seq data (Lara-Astiaso et al., 2014).

(B) Percentages of pre-existing or de novo enhancers bound by C/EBPα in early hematopoietic progenitors (Lin- Sca-1+ Kit+, LSK cells), GMPs, or primary MΦ.

(C) Screenshots of C/EBPα bound to selected enhancers of *Pu.1*, *Cebpb*, *Tlr4*, and *Ctsd* in the indicated hematopoietic cell types. Pre-ex., pre-existing.

(D) Pre-existing and de novo myeloid enhancers in pre-B cells and iMΦ, showing PU.1 occupancy, binding sites targeted by incoming C/EBPα (curved arrows), enhancer states, and gene expression. Nucleosomes are indicated by light blue balls.

(E) Artist's rendering of the trajectory of activated pre-existing enhancers within the hematopoietic lineage tree (in green) and de novo enhancers (in red). The arrow depicts how C/EBPα short circuits the two trajectories when expressed in pre-B cells.

prospective macrophage enhancers and, hence, the expression of adjacent genes in a manner that recapitulates C/EBPα-induced transdifferentiation.

DISCUSSION

Our study of C/EBPα-induced pre-B-cell-to-macrophage transdifferentiation has revealed two types of prospective myeloid enhancers that are activated by C/EBPα. Pre-existing enhancers in pre-B cells are decorated with active enhancer marks and bound in their majority by PU.1, while de novo enhancers are free of enhancer activation marks and free of PU.1; C/EBPα simultaneously hyper-activates pre-existing enhancers and newly activates de novo enhancers (summarized in Figure 7D). These enhancers drive a substantial part of the gene repertoire required for the formation of functional macrophages. Strikingly, we also observed a similar synergy between pre-existing and de novo enhancers during myeloid lineage specification during normal hematopoiesis (Figure 7E).

The finding that pre-existing-type myeloid enhancers drive low-level expression of adjacent myeloid-restricted genes in early hematopoietic progenitors provides a mechanistic explanation for the phenomenon dubbed "lineage priming" (Hu et al., 1997). The observed expression of the myeloid markers lysozyme and CSF-1 receptor in hematopoietic stem cells (Mossadegh-Keller et al., 2013; Ye et al., 2003) supports this interpretation. The following observations indicate that PU.1 is a key component in the

generation of pre-existing myeloid enhancers: (1) most pre-existing enhancers are bound by the factor; (2) PU.1 is expressed in stem and progenitor cells, but downregulated in T cells and Erys (Back et al., 2005; Lara-Astiaso et al., 2014; Nutt et al., 2005), and this strongly correlates with the distribution of pre-existing myeloid enhancers and the expression of nearby genes; and (3) overexpression of PU.1 in fibroblasts partially activates myeloid genes associated with pre-existing enhancers (Figure 2G) and C/EBPα further enhances their expression, while C/EBPα alone has no effect (Feng et al., 2008).

However, it is likely that, in addition to PU.1, other TFs participate in the initiation of the establishment of pre-existing enhancers and the activation of de novo enhancers. Thus, C/EBPα sites also were enriched for the RUNX motif, in line with the finding that during myelopoiesis Runx1 binds transiently to the URE element of the *Pu.1* gene to establish open chromatin, permitting the binding of PU.1 (Hoogenkamp et al., 2009). In addition, it is possible that the *Fos* gene acts as a downstream effector, as it is directly regulated by C/EBPα and we observed enrichment of AP-1 motifs in C/EBP-bound sites where it might co-operate with C/EBPs.

The weakly active pre-existing myeloid enhancers in hematopoietic progenitors appear to be in a stand-by state that can be fully activated by changes in the bone marrow microenvironment either during development or in adult life, such as after infections with pathogens. These signals may in turn increase the levels of PU.1, C/EBPα, and C/EBPβ expression. Thus, for example, bacteria or inflammatory stimuli can upregulate *Pu.1* expression in hematopoietic stem cells through the activation of M-CSF, a cytokine that in turn activates the CSF-1 receptor (Mossadegh-Keller et al., 2013). In addition, the yeast *Candida albicans* can induce emergency granulopoiesis in hematopoietic progenitors through upregulation of C/EBPβ (Hirai et al., 2006). Therefore, the ectopic expression of C/EBPα/β to induce transdifferentiation of pre-B cells might mimic processes that are normally triggered in hematopoietic progenitors by developmental cues or pathogens.

Surprisingly, a subset of pre-existing enhancers appears to be bi-functional. In B cells this subset is bound by the B cell TF Ebf1, typically in combination with PU.1, resulting in low-level expression. Binding of C/EBPα further activates these genes, raising the possibility that their products are themselves bi-functional. The *Cebpb* gene illustrates this scenario as its putative 88-kb upstream enhancer is bound by Ebf1, which is eventually replaced by C/EBPα during the conversion into myeloid cells. In addition the factor is required for the function of both B cells and macrophages (Chen et al., 1997; Tanaka et al., 1995). However, whether the 88-kb site is the physiologically most relevant *Cebpb* enhancer is unknown.

Previous work on TF combinations that induce cell fate conversions have postulated two alternative models as follows: (1) a symmetric collaboration between various TFs acting as pioneer factors, exemplified by Oct4, Sox2, and Klf4 that act during iPSC reprogramming (Soufi et al., 2012); and (2) a hierarchical model, exemplified by Ascl1 acting as a pioneer for the subsequent binding of Brn2 and Myt1l during induced neuronal transdifferentiation (Wapinski et al., 2013). Here we propose a mixed model, where the key lineage-instructive factors exert dual roles as both pioneer and secondary factors. The conclusion that C/EBPα can act as a pioneer factor is based on the observation that it binds to chromatin regions free of activating histone marks and to a nucleosome-dense region within de novo enhancers, agreeing with the reported pioneer activity of C/EBPβ (Siersbæk et al., 2011). It is possible that PU.1 also can act as a pioneer factor, as it is one of the earliest lineage-instructive factors expressed in the hematopoietic system (Dzierzak and Speck, 2008), and on its own can induce the expression of myeloid genes in non-hematopoietic cells.

In conclusion, our work revealed that the collaboration between an exogenous and an endogenous lineage-instructive TF (C/EBPα and PU.1) leads to the activation of pre-existing and de novo myeloid enhancers during transdifferentiation, resulting in macrophage differentiation. Interestingly, this mechanism recapitulates the way endogenous C/EBP factors and PU.1 collaborate to induce myeloid differentiation during normal hematopoiesis. It will be interesting to determine whether conversions of other cell types driven by TFs likewise recapitulate developmental processes that result from the superimposition of complementary enhancer types.

EXPERIMENTAL PROCEDURES

Cell Culture, Retroviruses, and shRNA Constructs

The origin of the HAFTL pre-B cell line, its derivatives C10 (C/EBPαER-GFP) and C11 (C/EBPαER-hCD4), and induction of transdifferentiation (treatment with 100 uM β-est and grown in the presence of 10 nM Il-3 and 10 nM CSF-1) have been described previously (Bussmann et al., 2009; Xie et al., 2004). The shC/EBPβ-KD07 directed to the ORF of *Cebpb* was purchased from Sigma (Mission shRNA System) in a pLKO.1-puro lentiviral backbone. An shRNA against PU.1 cloned into LMP-GFP virus (Open Biosystems) was a gift from Dr. M. Sieweke (Sarrazin et al., 2009). The 3T3 cell culture conditions and the PU.1-GFP construct have been described previously (Feng et al., 2008). Phagocytosis of yeast was performed as described previously by Rapino et al. (2013). To test for statistical differences of C/EBPα binding after knockdown of PU.1, we applied the Student's t test, one-tailed, alpha level (0.05).

FACS

FACS experiments were performed as described previously (Bussmann et al., 2009) using conjugated antibodies against Cd19 (550992) and Cd11b (552850) and combined with blocking antibody (553142) from BD Pharmingen. Unstained cells or an isotype control antibody (553932, BD Pharmingen) were used as a negative control.

ChIP, ChIP-Seq, and MNase-Seq Experiments

ChIP experiments were performed as described previously (van Oevelen et al., 2008). DNA libraries were prepared using Illumina's reagents and instructions. Nucleosome positioning was determined by MNase digestion using a modification of a published method (Cappabianca et al., 1999). All libraries were sequenced on the Illumina GA IIx or Hiseq2000 sequencer.

Processing of ChIP-Seq and MNase-Seq Data

High-throughput Illumina sequencing data were base-called using the Illumina pipeline, and sequencing reads were aligned to the mouse genome (mm9) using either the Illumina Eland alignment tool or Bowtie (Langmead et al., 2009) without mismatches. Aligned sequences were filtered to remove identical sequence tags and sequence tags not aligning uniquely to the mouse genome. To detect enriched regions, we used HOMER (http://homer.salk.edu/homer/ngs/index.html) (Heinz et al., 2010; and see Table S1). See the Supplemental Experimental Procedures for further details. To test for statistical differences in the level or reduction of coverage between sets of regions, we applied the Wilcoxon rank-sum test, two-tailed, alpha level (0.05).

Binding Site Annotation, Motif Analysis, and Gene Expression

Position of non-redundant regions relative to TSS of nearest gene (RefSeq mm9) was based on center position and calculated by in-house Perl scripts. For a subset of genes, the median expression level was calculated, and, to test for statistical differences in gene expression levels between sets of genes, we applied the Wilcoxon rank-sum test, two-tailed, alpha level (0.05). Gene expression values in hematopoietic cells (Lara-Astiaso et al., 2014) were normalized by dividing each presented mRNA value by the average mRNA of all listed genes per cell type. To annotate genes for enrichment of GO terms, we employed David with standard settings (Huang da et al., 2009). Motif discovery within selected regions was performed using HOMER (Heinz et al., 2010).

Gene Expression Analyses by qRT-PCR

To analyze mRNA levels of selected genes in either C10 cells induced with β-est or 3T3 cells overexpressing PU.1, we extracted RNA using trizol and reverse transcribed it with GeneAmp RNA PCR (Applied Biosystems). SybrGreen PCR Master Mix (Applied Biosystems) was used for amplification and detection of cDNAs, and PCR reactions were carried out with the AB7900HT detection system (Applied Biosystems). To test for statistical differences in mRNA levels, we applied the Student's t test, one-tailed, alpha level (0.05).

ACCESSION NUMBERS

The accession numbers for the ChIP-seq data reported in this paper are GEO: GSE53173, GSE53362, and GSE53460 for chromatin marks, factor binding, and MNase, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, four tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.06.007.

ACKNOWLEDGMENTS

We thank C. Berenguer, A. Ferrer, and S. Nacht for assistance with ChIP-seq and MNase experiments; the core facilities (T. Zimmermann) for time-lapse microscopy; and the T.G. lab for discussions. Dr. H. Xie provided the retroviral vectors. We also thank Dr. N. Dean for mCherry-labeled *C. albicans*. This work was supported by the Ministerio de Educacion y Ciencia, SAF.2012-37167, AGAUR 2009 SGR768, EU-FP7 project BLUEPRINT (282510), and Fundacio la Marato TV3. C.B. was funded by a program grant (LLR 7001) from Leukemia Lymphoma Research UK.

Received: February 19, 2015
Revised: June 24, 2015
Accepted: June 29, 2015
Published: July 30, 2015

REFERENCES

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

Back, J., Allman, D., Chan, S., and Kastner, P. (2005). Visualizing PU.1 activity during hematopoiesis. Exp. Hematol. *33*, 395–402.

Bertrand, V., and Hobert, O. (2010). Lineage programming: navigating through transient regulatory states via binary decisions. Curr. Opin. Genet. Dev. *20*, 362–368.

Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E., and Graf, T. (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell *5*, 554–566.

Cappabianca, L., Thomassin, H., Pictet, R., and Grange, T. (1999). Genomic footprinting using nucleases. Methods Mol. Biol. *119*, 427–442.

Chen, X., Liu, W., Ambrosino, C., Ruocco, M.R., Poli, V., Romani, L., Quinto, I., Barbieri, S., Holmes, K.L., Venuta, S., and Scala, G. (1997). Impaired generation of bone marrow B lymphocytes in mice deficient in C/EBPbeta. Blood *90*, 156–164.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911–920.

DeKoter, R.P., and Singh, H. (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science *288*, 1439–1441.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

Di Tullio, A., and Graf, T. (2012). C/EBPα bypasses cell cycle-dependency during immune cell transdifferentiation. Cell Cycle *11*, 2739–2746.

Di Tullio, A., Vu Manh, T.P., Schubert, A., Castellano, G., Månsson, R., and Graf, T. (2011). CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc. Natl. Acad. Sci. USA *108*, 17016–17021.

Dzierzak, E., and Speck, N.A. (2008). Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nat. Immunol. *9*, 129–136.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc. Natl. Acad. Sci. USA *105*, 6057–6062.

Friedman, A.D. (2007). Transcriptional control of granulocyte and monocyte development. Oncogene *26*, 6816–6828.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Hasemann, M.S., Lauridsen, F.K., Waage, J., Jakobsen, J.S., Frank, A.K., Schuster, M.B., Rapin, N., Bagger, F.O., Hoppe, P.S., Schroeder, T., and Porse, B.T. (2014). C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. PLoS Genet. *10*, e1004079.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S., Imanishi, J., Akashi, K., and Tenen, D.G. (2006). C/EBPbeta is required for 'emergency' granulopoiesis. Nat. Immunol. *7*, 732–739.

Holmberg, J., and Perlmann, T. (2012). Maintaining differentiated cellular identity. Nat. Rev. Genet. *13*, 429–439.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., et al. (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. Blood *114*, 299–309.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. *11*, 774–785.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Ladewig, J., Koch, P., and Brüstle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nat. Rev. Mol. Cell Biol. *14*, 225–236.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Mossadegh-Keller, N., Sarrazin, S., Kandalla, P.K., Espinosa, L., Stanley, E.R., Nutt, S.L., Moore, J., and Sieweke, M.H. (2013). M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. Nature *497*, 239–243.

Nutt, S.L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. J. Exp. Med. *201*, 221–231.

Ptashne, M. (2007). On the use of the word 'epigenetic'. Curr. Biol. *17*, R233–R236.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rapino, F., Robles, E.F., Richter-Larrea, J.A., Kallin, E.M., Martinez-Climent, J.A., and Graf, T. (2013). C/EBPα induces highly efficient macrophage transdifferentiation of B lymphoma and leukemia cell lines and impairs their tumorigenicity. Cell Rep. *3*, 1153–1163.

Reddy, V.A., Iwama, A., Iotzova, G., Schulz, M., Elsasser, A., Vangala, R.K., Tenen, D.G., Hiddemann, W., and Behre, G. (2002). Granulocyte inducer C/EBPalpha inactivates the myeloid master regulator PU.1: possible role in lineage commitment decisions. Blood *100*, 483–490.

Sarrazin, S., Mossadegh-Keller, N., Fukao, T., Aziz, A., Mourcin, F., Vanhille, L., Kelly Modis, L., Kastner, P., Chan, S., Duprez, E., et al. (2009). MafB restricts M-CSF-dependent myeloid commitment divisions of hematopoietic stem cells. Cell *138*, 300–313.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573–1577.

Siersbæk, R., Nielsen, R., John, S., Sung, M.H., Baek, S., Loft, A., Hager, G.L., and Mandrup, S. (2011). Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. EMBO J. *30*, 1459–1472.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Taberlay, P.C., Kelly, T.K., Liu, C.C., You, J.S., De Carvalho, D.D., Miranda, T.B., Zhou, X.J., Liang, G., and Jones, P.A. (2011). Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell *147*, 1283–1294.

Tanaka, T., Akira, S., Yoshida, K., Umemoto, M., Yoneda, Y., Shirafuji, N., Fujiwara, H., Suematsu, S., Yoshida, N., and Kishimoto, T. (1995). Targeted disruption of the NF-IL6 gene discloses its

essential role in bacteria killing and tumor cytotoxicity by macrophages. Cell *80*, 353–361.

Treiber, T., Mandel, E.M., Pott, S., Györy, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714–725.

van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Mol. Cell *32*, 359–370.

Vierbuchen, T., and Wernig, M. (2011). Direct lineage conversions: unnatural but useful? Nat. Biotechnol. *29*, 892–907.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Wölfler, A., Danen-van Oorschot, A.A., Haanstra, J.R., Valkhof, M., Bodner, C., Vroegindeweij, E., van Strien, P., Novak, A., Cupedo, T., and Touw, I.P. (2010). Lineage-instructive function of C/EBPα in multipotent hematopoietic cells and early thymic progenitors. Blood *116*, 4116–4125.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676.

Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature *465*, 704–712.

Ye, M., Iwasaki, H., Laiosa, C.V., Stadtfeld, M., Xie, H., Heck, S., Clausen, B., Akashi, K., and Graf, T. (2003). Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. Immunity *19*, 689–699.

Yeamans, C., Wang, D., Paz-Priel, I., Torbett, B.E., Tenen, D.G., and Friedman, A.D. (2007). C/EBPalpha binds and activates the PU.1 distal enhancer to induce monocyte lineage commitment. Blood *110*, 3136–3142.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., et al. (2004). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity *21*, 853–863.

Zhang, H., Alberich-Jorda, M., Amabile, G., Yang, H., Staber, P.B., Di Ruscio, A., Welner, R.S., Ebralidze, A., Zhang, J., Levantini, E., et al. (2013). Sox4 is a key oncogenic target in C/EBPα mutant acute myeloid leukemia. Cancer Cell *24*, 575–588.

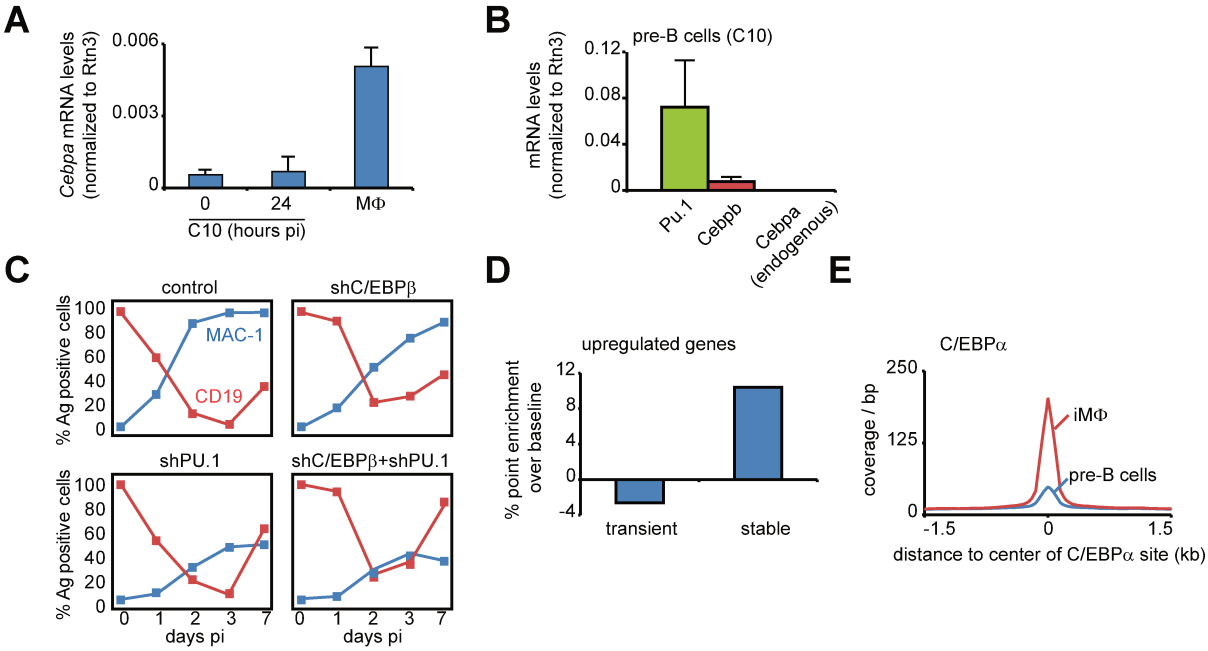
Stem Cell Reports

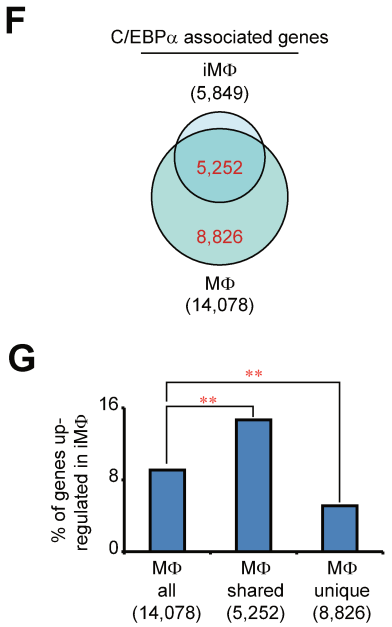
Supplemental Information

C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis

Chris van Oevelen, Samuel Collombet, Guillermo Vicent, Maarten Hoogenkamp, Cyrille Lepoivre, Aimee Badeaux, Lars Bussmann, Jose Luis Sardina, Denis Thieffry, Miguel Beato, Yang Shi, Constanze Bonifer, and Thomas Graf

Figure S1





H

genes upregulated in iMΦ

C/EBPα-shared (770)

Enrichment Score: 14.33	P_Value
response to wounding	9.4E-19
immune response	7.9E-15
defense response	1.1E-14
inflammatory response	1.4E-14
inflammatory response	2.1E-12
Enrichment Score: 12.24	**P_Value**
lysosome	6.9E-14
lysosome	1.1E-13
lytic vacuole	1.3E-13
vacuole	1.3E-12
Lysosome	5.2E-11
Enrichment Score: 8.46	**P_Value**
response to molecule of bacterial origin	1.0E-10
response to bacterium	1.3E-10
response to lipopolysaccharide	3.2E-6

C/EBPα-unique (451)

Enrichment Score: 6.87	P_Value
vacuole	1.1E-8
lysosome	1.7E-7
lytic vacuole	1.9E-7
lysosome	2.3E-7
Lysosome	5.3E-7
Enrichment Score: 3.5	**P_Value**
endosome	8.5E-5
late endosome	5.1E-4
endosome	7.3E-4
Enrichment Score: 3.46	**P_Value**
vacuolar membrane	2.7E-6
vacuolar part	1.3E-5
lysosomal membrane	4.4E-5
late endosome	5.1E-4
endosomal part	4.7E-2
endosome membrane	4.7E-2

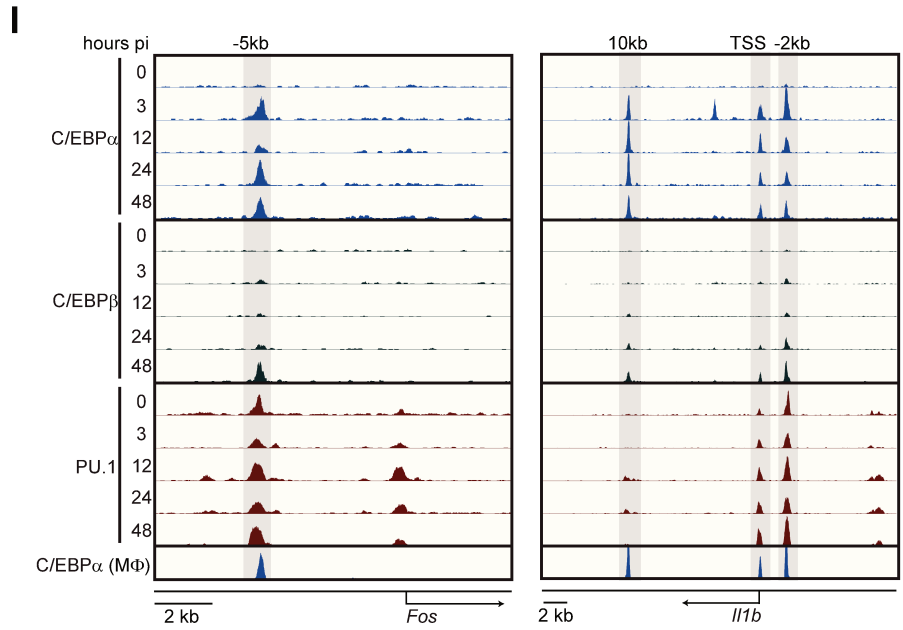


Figure S2

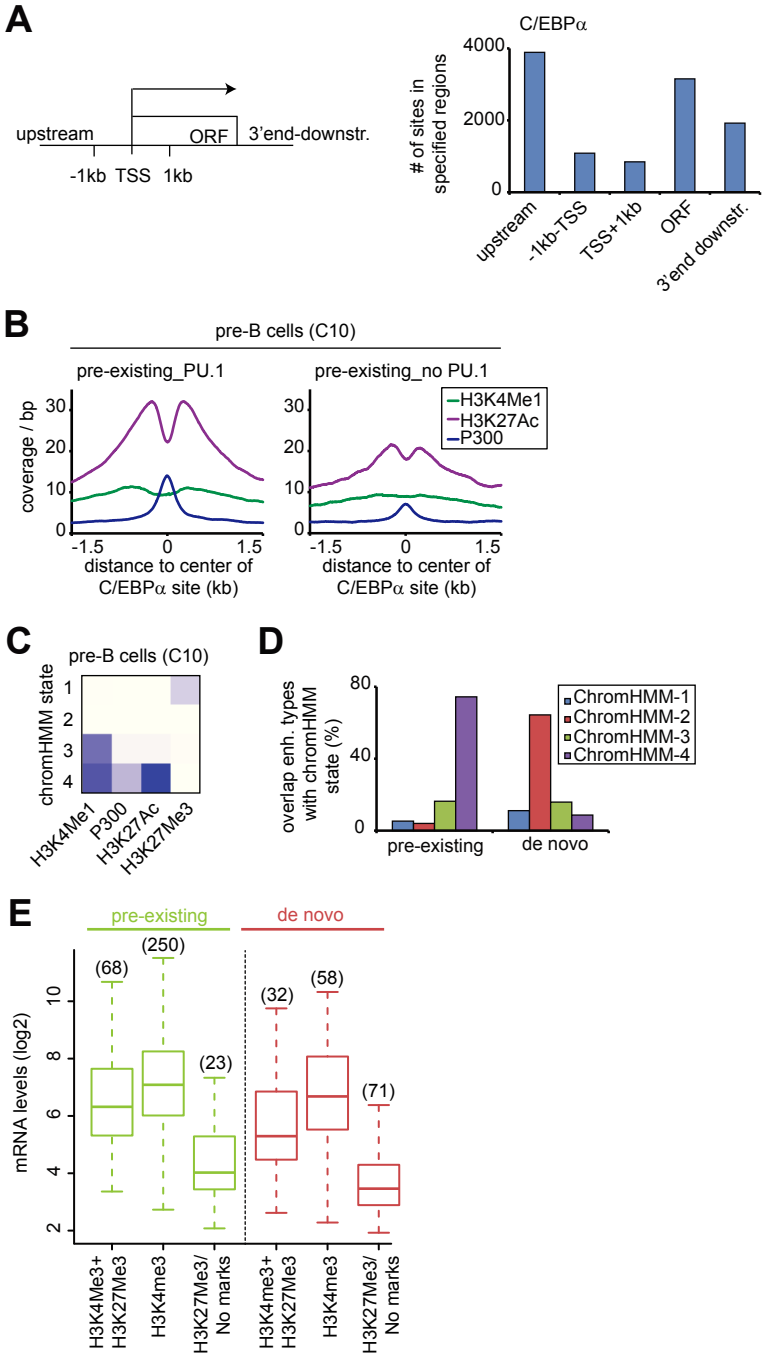

A
upstream
ORF
3'end-downstr.
-1kb TSS 1kb
C/EBPα
# of sites in specified regions
4000
2000
0
upstream
-1kb-TSS
TSS+1kb
ORF
3'end downstr.
B
pre-B cells (C10)
pre-existing_PU.1
pre-existing_no PU.1
coverage / bp
H3K4Me1
H3K27Ac
P300
distance to center of C/EBPα site (kb)
C
pre-B cells (C10)
chromHMM state
H3K4Me1
P300
H3K27Ac
H3K27Me3
D
overlap enh. types with chromHMM state (%)
ChromHMM-1
ChromHMM-2
ChromHMM-3
ChromHMM-4
pre-existing
de novo
E
pre-existing
de novo
mRNA levels (log2)
(68)
(250)
(23)
(32)
(58)
(71)
H3K4Me3+ H3K27Me3
H3K4me3
H3K27Me3/ No marks

Figure S3

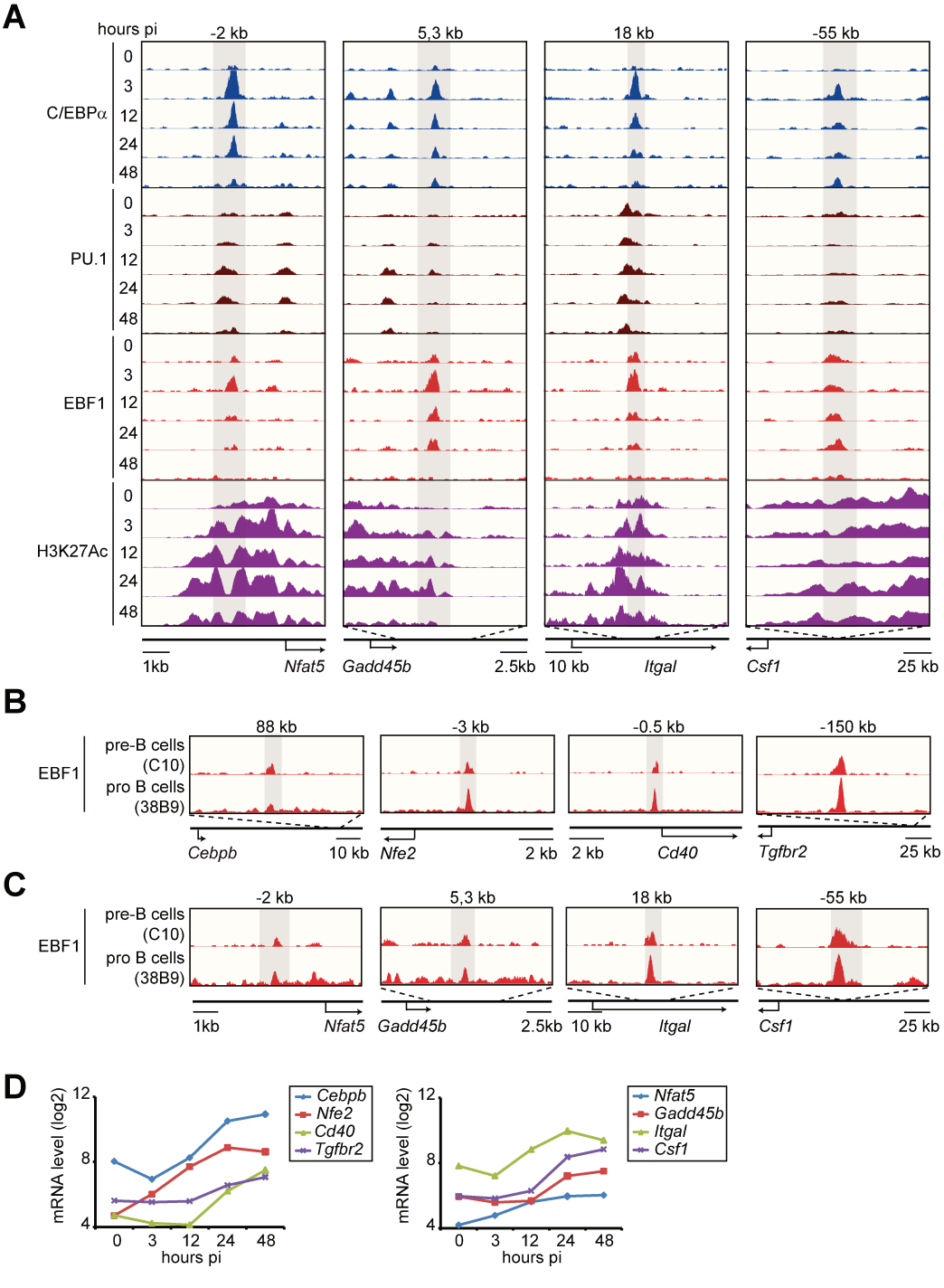

A
hours pi
-2 kb
5,3 kb
18 kb
-55 kb
C/EBPα
PU.1
EBF1
H3K27Ac
0
3
12
24
48
1kb
Nfat5
Gadd45b
2.5kb
10 kb
Itgal
Csf1
25 kb
B
88 kb
-3 kb
-0.5 kb
-150 kb
EBF1
pre-B cells (C10)
pro B cells (38B9)
Cebpb
10 kb
Nfe2
2 kb
2 kb
Cd40
Tgfbr2
25 kb
C
-2 kb
5,3 kb
18 kb
-55 kb
EBF1
pre-B cells (C10)
pro B cells (38B9)
1kb
Nfat5
Gadd45b
2.5kb
10 kb
Itgal
Csf1
25 kb
D
mRNA level (log2)
12
8
4
0
3
12
24
48
hours pi
Cebpb
Nfe2
Cd40
Tgfbr2
Nfat5
Gadd45b
Itgal
Csf1

Figure S4

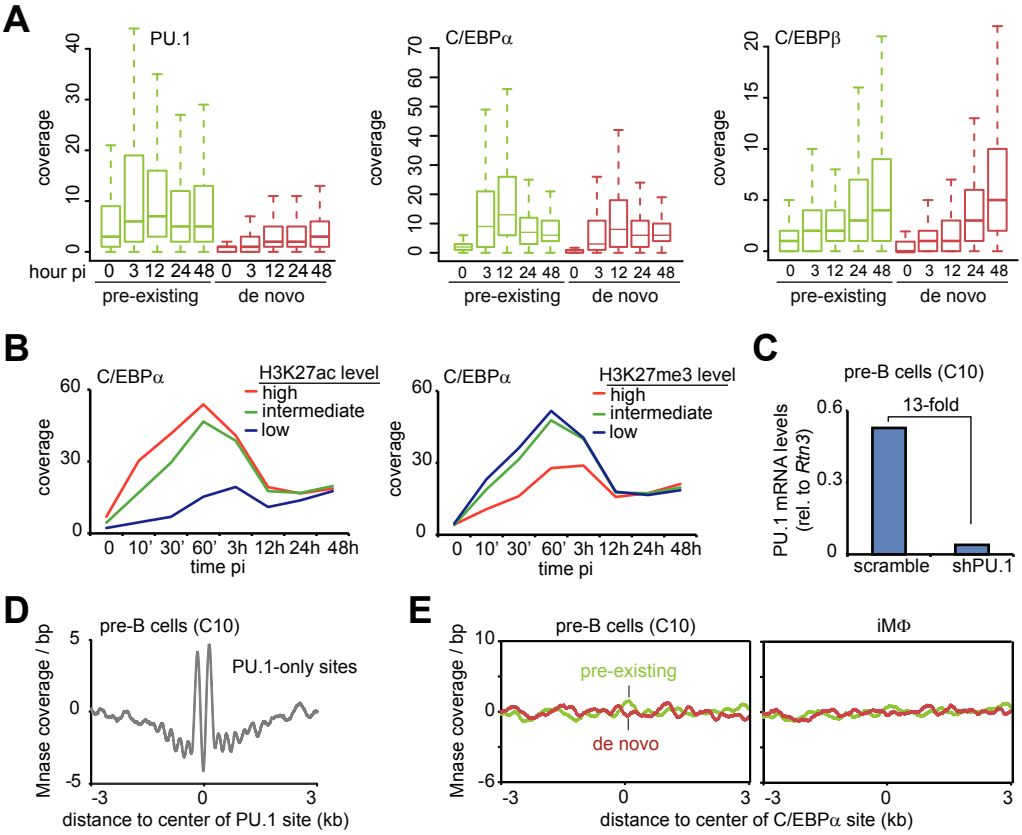

A
PU.1
C/EBPα
C/EBPβ
coverage
hour pi
0 3 12 24 48 0 3 12 24 48
pre-existing
de novo
B
C/EBPα
H3K27ac level
high
intermediate
low
H3K27me3 level
time pi
0 10' 30' 60' 3h 12h 24h 48h
C
pre-B cells (C10)
13-fold
PU.1 mRNA levels (rel. to Rtn3)
scramble
shPU.1
D
pre-B cells (C10)
PU.1-only sites
Mnase coverage / bp
distance to center of PU.1 site (kb)
E
pre-B cells (C10)
iMΦ
pre-existing
de novo
distance to center of C/EBPα site (kb)

Figure S5

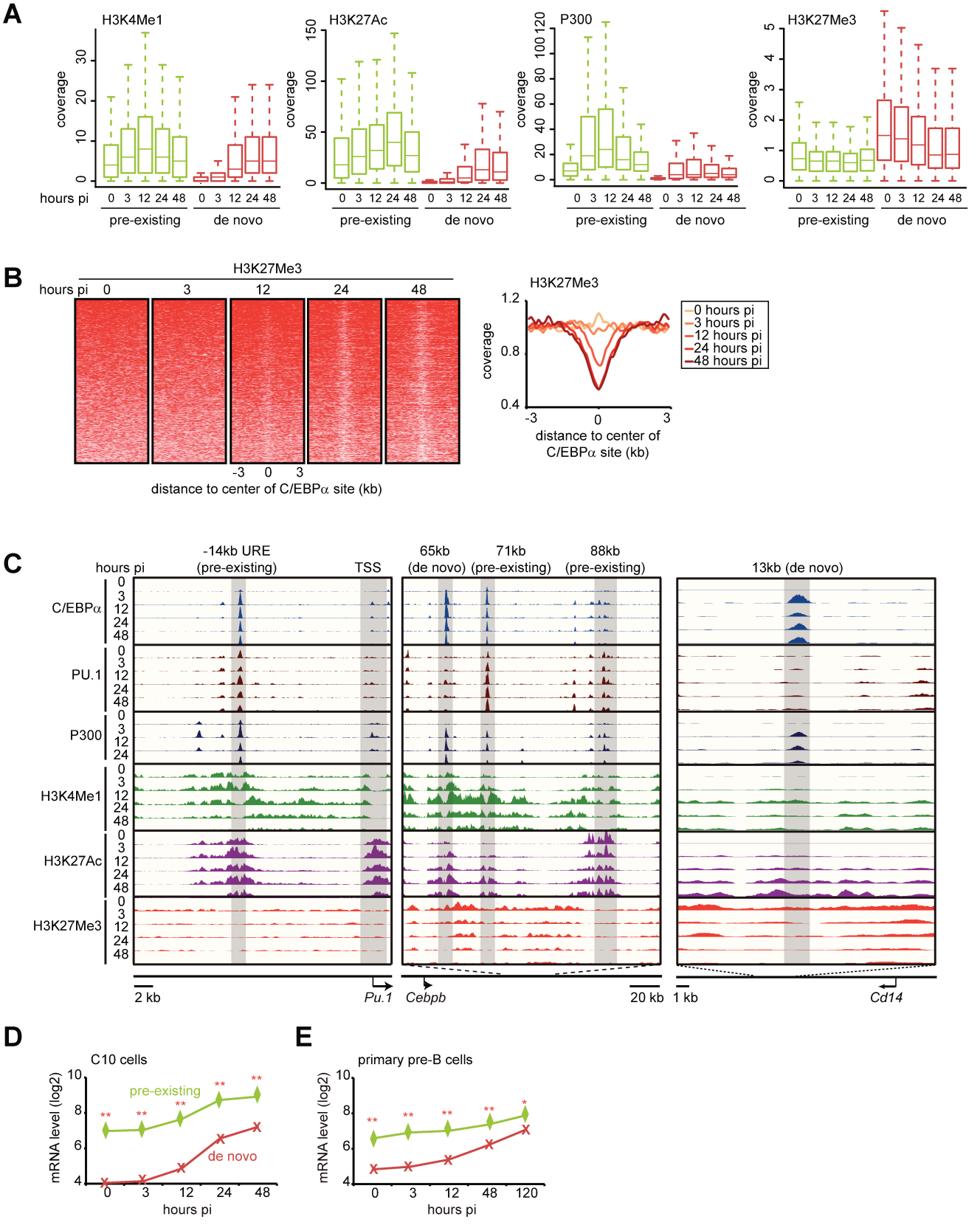
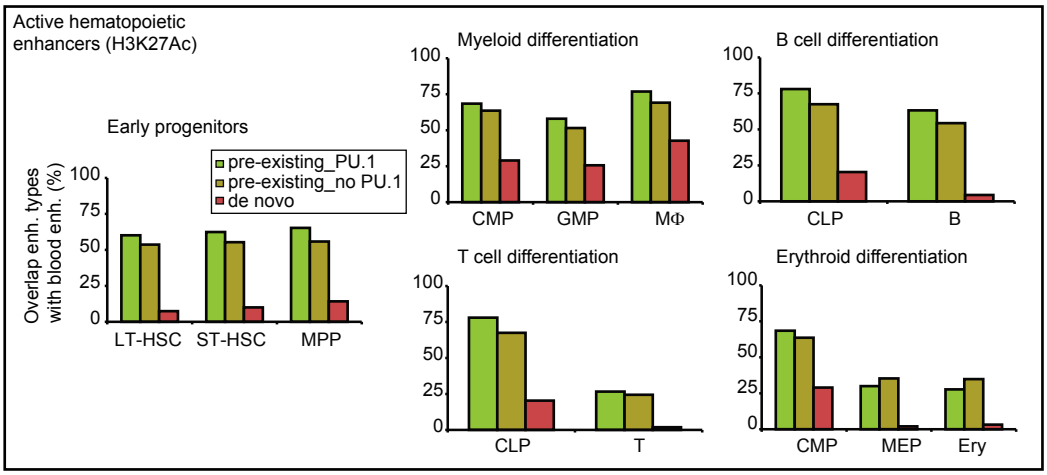
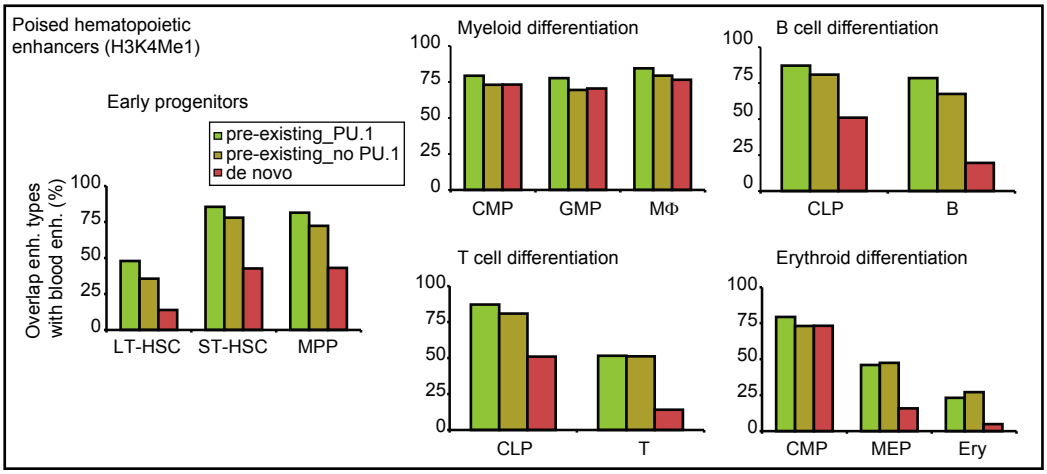

A
H3K4Me1
H3K27Ac
P300
H3K27Me3
coverage
hours pi
0 3 12 24 48
pre-existing
de novo
B
H3K27Me3
hours pi 0 3 12 24 48
-3 0 3
distance to center of C/EBPα site (kb)
0 hours pi
3 hours pi
12 hours pi
24 hours pi
48 hours pi
C
hours pi
-14kb URE (pre-existing)
TSS
65kb (de novo)
71kb (pre-existing)
88kb (pre-existing)
13kb (de novo)
C/EBPα
PU.1
P300
H3K4Me1
H3K27Ac
H3K27Me3
2 kb
Pu.1
Cebpb
20 kb
1 kb
Cd14
D
C10 cells
pre-existing
de novo
mRNA level (log2)
hours pi
E
primary pre-B cells
mRNA level (log2)
hours pi

Figure S6

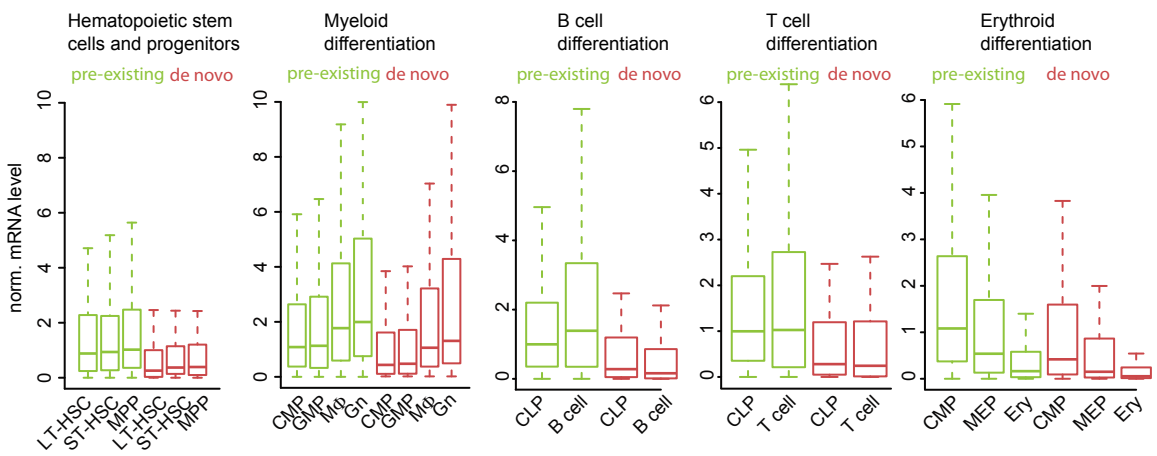
A



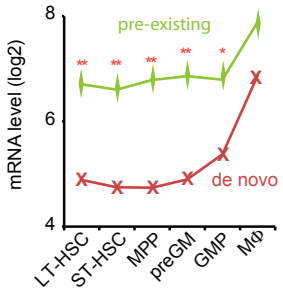
B



C



D



Supplemental Figures

Supplemental Figure S1. PU.1 and C/EBPβ are required for B cell to macrophage transdifferentiation and characterization of C/EBPα binding sites.

(**A+B**) Expression of C/EBPα (exogenous and endogenous) in C10 cells (0 and 24 hours pi) and primary MΦ (A) and *Pu.1*, *Cebpb* and *Cebpa* (endogenous) in uninduced pre B cells (B) as measured by qRT-PCR (technical triplicates) normalized to the mRNA levels of *Rtn3*. Data are represented as mean ± SEM. (**C**) Percentage of CD19 or MAC-1 positive cells (antigen: Ag) was determined by FACS after ablation of C/EBPβ, PU.1 or both. (**D**) C/EBPα sites (n=54,198) were grouped based on either transient or stable C/EBPα binding (stable binding: up to 48 hours pi). For each group the number of sites nearest upregulated genes was calculated, compared to all sites (baseline) and expressed as percent point difference. **(E)** C/EBPα binding in uninduced B cells and iMΦ (48 hours pi). Center of binding=0, Window= 3kb, bin=1 (**F**) Venn diagram showing the intersection of C/EBPα associated genes in iMΦ (n=5,849) and primary macrophages (MΦ, n=14,078) by nearest gene approach (**G**) Percentage of all, shared or unique C/EBPα associated genes that become upregulated during transdifferentiation of primary B cells. Statistical analysis by hypogeometric distribution; two asterixes $p<0.001$. **(H)** Functional annotation analysis of shared (n=770, left panel) or unique (n=451, right panel) C/EBPα associated genes which become upregulated during transdifferentiation by DAVID. First three enrichment clusters of each group is shown (**I**) Screenshots of C/EBPα, C/EBPβ and PU.1 binding at enhancer regions associated with the *Fos* and *Il1b* genes in C10 cells induced for different times and C/EBPα binding in primary MΦ. Arrows indicate TSS and direction of transcription.

Supplemental Figure S2. Chromatin state maps in uninduced B cells

(**A**) Genomic distribution of stable C/EBPα binding events (n=10,849) relative to the transcription start site (TSS) in induced MΦ. ORF: open reading frame (**B**) Coverage of H3K4Me1, H3K27Ac and P300 around *pre-existing* sites bound or not bound by PU.1. Window= 3 kb, bin: 1 bp (**C**) Chromatin

state map in uninduced B cells by ChromHMM. Number of states was pre-set to 4. (**D**) Comparison of *pre-existing* or *de novo* enhancers with ChromHMM state maps. (**E**) Distribution of mRNA levels of genes nearest *pre-existing* or *de novo* enhancers in uninduced C10 cells. Genes were subdivided based on promoter activity. Number of genes in brackets.

Supplemental Figure S3. Binding of the B cell transcription factor EBF1 to myeloid enhancers.

(**A**) Screenshots of C/EBPα, PU.1, EBF1 and H3K27Ac ChIPseq profiles at selected enhancer regions in C10 cells. (**B+C**) Screenshots of EBF1 binding at enhancer regions as shown in Figure 3F and panel A in uninduced C10 cells and a pro-B cell line (38B9). (**D**) Gene expression kinetics of genes associated with prospective myeloid enhancers bound by EBF1 (see Fig. 3F and panel A) by microarray.

Supplemental Figure S4. C/EBPα and PU.1 binding kinetics and nuclease accessibility at macrophage enhancers during transdifferentiation.

(A) Distribution of PU.1, C/EBPα and C/EBPβ coverage values of *pre-existing* or *de novo* enhancers at indicated times after induction. Values of the center position are shown. (**B**) Kinetics of C/EBPα (at center position) of *pre-existing* enhancers grouped by H3K27Ac levels or *de novo* enhancers grouped by H3K27Me3 levels at indicated times post induction. 10, 30 and 60 minutes, h: hours (**C**) *Pu.1* mRNA levels after ablation of PU.1 by shRNA in C10 cells as measured by qRT-PCR (technical triplicates) normalized to the mRNA levels of the *Rtn3* gene. (**D**) Average profile of MNase resistant DNA at PU.1 sites devoid of C/EBPα and C/EBPβ binding (n=8,765) in pre-B cells. Profiles were generated as in Figs. 4F,G. (**E**) Average profile of MNase resistant DNA at random positions in the genome. Profiles were generated as in Figs. 4F,G.

Supplemental Figure S5. Chromatin kinetics at macrophage enhancers during transdifferentiation

(**A**) Distribution of H3K4Me1, H3K27Ac, P300, and H3K27Me3 coverage values at center position of *pre-existing* or *de novo* enhancers at indicated times after induction. (**B**) Heatmaps visualizing H3K27Me3 decoration at *de novo* enhancers at different hours pi. of C10 cells. Window size: 6000 bp, bin size: 100. Right panel: coverage of H3K27Me3 around *de novo* enhancers. (**C**) Screenshots of selected enhancers, showing C/EBPα, PU.1, P300, H3K4Me1, H3K27Ac and H3K27Me3 profiles in C10 cells. (**D+E**) Median mRNA levels of genes nearest *pre-existing* or *de novo* enhancers during transdifferentiation of C10 cells (D) and primary pre-B cells (E). Statistical analysis by Wilcoxon test; one asterix, $p<0.01$, two asterixes $p<0.001$.

Supplemental Figure S6. Enhancer activity of macrophage enhancers during hematopoiesis.

(A+B) Bar graph showing the percentage of indicated enhancer type (see text and Figure 2) intersecting with enhancers either decorated with H3K4Me1 or H3K27Ac in various hematopoietic progenitors and differentiated cells. **(C)** Distribution of mRNA values of genes associated with *pre-existing* or *de novo* type enhancers in various hematopoietic progenitors and differentiated cells. (**D**) Similar as panel C but using cell samples prepared in slightly different ways and analyzed for gene expression by Affymetrix arrays (Di Tullio et al. 2011).

Supplemental Movie S1.

Transdifferentiation in action. The video shows a culture of C10 cells labeled with green fluorescent protein (GFP) surrounded by red fluorescent yeast (*Candida albicans*). As the transcription factor C/EBPα is activated within the B cells these aggregate and turn into macrophages that ingest the yeast, so that 51 hours after activation all pathogens became eaten. The microscope used for the acquisition was a Zeiss Cell Observer HS.

Supplemental Tables

Table S1: Summary of peak calling statistics of transcription factors and chromatin marks by HOMER.

Table S2: Peaks characteristics, annotation and gene expression of nearest genes of C/EBPα enriched regions in induced macrophages (iMΦ).

Table S3: Genomic coordinates of USCS screenshot examples and tested loci for ChIP (Refseq mm9/10).

Table S4: Primer sequences

Supplemental Experimental Procedures

Antibodies for ChIP-seq

Antibodies against C/EBPα (sc-61), C/EBPβ (sc-150), PU.1 (sc-352) and P300 (sc-585) were purchased from Santa Cruz Biotechnologies and against EBF1 (abnova1879) from Abnova. An antibody against H3K4Me1 (037-050) was purchased from Diagenode; antibody to H3K27Ac (ab4729) was purchased from Abcam (Cambridge UK) and to H3K27Me3 (17-622) from Millipore.

MNase-seq

To generate mononucleosomal DNA fragments, C10 cells were washed and resuspended in Buffer A (15 mM Tris-HCl pH: 7.5, 60 mM KCl, 15 mM NaCl, 2 mM CaCl2, 150 mM Sucrose, 0.15 mM spermin and 0,5 mM spermidin, lysolecithin at 0,5 mg/ml) for 1 min at 37 ° C. Then 30, 90, 200 and 300 units/ml of MNase were added and incubated for 2 min at 37° C. After stopping the reaction with 40 mM EDTA cells were lysed in Buffer B (50 mM Tris-HCl, pH: 8, 20 mM EDTA, 1% SDS), treated with RNase A and Proteinase K, DNA extracted with phenol-chlorophorm and precipitated. The DNA was run on an agarose gel (1% in TAE 1x) and the mononucleosomal band was excised, purified and used to prepare libraries.

Processing of ChIPseq and MNase data

To detect enriched regions we used HOMER (http://homer.salk.edu/homer/ngs/index.html) (Heinz et al. 2010) with the following settings: For factor binding (C/EBPα, C/EBPβ, PU.1, P300 and EBF1)

the peak localization/shape option was set to center and the style option to factor. For H3K4Me1, H3K4Me3 and H3K27Ac chromatin marks the style option was set to histone, the size option was set to 1000 and the minDist option was set to 1000. For H3K27Me3, the style option was set to histone, the size option was set to 5000 and the minDist option was set to 5000.

An overlap between C/EBPα enriched regions was defined by requiring the start or end position of an enriched region to lie between the start and end position of another enriched region (i.e. a minimal overlap of 1 bp). Coordinates of the shared region were merged and defined by the lowest and highest boundary. In this way, non-redundant binding sites for C/EBPα were identified by combining enriched regions of indicated time points post induction. Intersection of these non-redundant C/EBPα regions with histone marks required an overlap of 1 bp. Because coordinates of enriched C/EBPα regions were identified with the center option (see HOMER: http://homer.salk.edu/homer/ngs/index.html) for details and Supplemental Table 1, Supplemental Table 2: length of enriched regions), which defines a minimal centered region of the peak, intersection of these regions with other enriched regions of selected factors was defined as being in the proximity of maximally 1000 bp relative to start or end coordinate.

Heatmaps and average binding profiles

The position of aligned reads was extended by adding 150 bp to 5-3' aligned reads or subtraction of 150 bp from the position of 3-5' aligned reads. Data files in bed format were converted to Wig format using 'in house' awk scripts and subsequently converted to BigWig files using the 'wigToBigWig' utility from USCS and visualized by the USCS browser.

To generate heatmaps of factor binding or chromatin marks selected regions were centered on C/EBPα binding 48 hours post induction (hours pi) and extended 3000 bp up- and down-stream. For each 6000 bp region the coverage of aligned tags was calculated at 1 bp resolution and summed per 100 bp bin relative to the central position using Bedtools (Quinlan and Hall 2010) and "in house" awk /perl scripts. Matrix files were visualized using Treeview (Saldanha 2004).

To generate average profiles of factor binding or chromatin modifications selected regions were centered on the median position of the non-redundant C/EBPα binding sites and extended 3000 bp up and downstream. The average coverage was calculated per 1 bp bin relative to the central position. The maximum coverage value within the extended region (6000 bp) of each profile was plotted in either column or line graphs. For H3K27Me3 the coverage value of the center position was plotted.

To generate average profiles of MNase treated chromatin selected regions were centered on PU.1 at 48 hours pi. For each MNase profile we calculated the median value within the 6000 bp window and applied a median background subtraction. Control MNase profiles were generated by selecting random sites within 6000 bp used to generate PU.1 centered average MNase profiles. Graphs were based on the entire 6000 bp region.

GEO data tracks

GEO data tracks GSM1223648 (C/EBPα in primary macrophages) (Zhang et al. 2013) and GSM499030 (EBF1 in preB cells (Treiber et al. 2011), were re-aligned (mm9) using Bowtie without mismatches.

Series GSE60103 (H3K4Me1 and H3K27Ac in blood cells) (Lara-Astiaso et al. 2014), GSM1054815 (C/EBPα in Lsk) and GSM1187163 C/EBPα in GMP) (Hasemann et al. 2014) were re-aligned (mm10) using Bowtie2 standard settings. HOMER was used to detect enriched regions with the peak localization/shape option set to center and the style option set to factor for PU.1 peaks. For H3K4Me1 and H3K27Ac chromatin marks the style option was set to histone, the size option was set to 1000 and the minDist option was set to 1000. To convert mm9 to mm10 coordinates liftover utility of USCS was used with standard settings. GEO data GSE52373 (C/EBPα in primary B cells) (Di Stefano et al. 2013) and GEO data track mentioned above were converted to bigwig files using SRA toolkit from USCS and uploaded to the USCS browser.

Gene expression array analyses

To correlate factor binding with gene expression we used the gene expression databases described previously (Bussmann et al. 2009; Di Tullio et al. 2011; Lara-Astiaso et al. 2014). Each factor binding sites was associated with the mRNA level of the nearest gene defined in RefSeq (mm9). For subsets of sites, duplicate genes within each subset were removed and the median mRNA level of unique genes for each subset calculated. To correlate the spatial association of sustained C/EBPα binding sites with up-, down-regulated genes and all genes, genes were counted that are in the neighborhood of C/EBPα binding sites with windows of increasing size centered on these sites and expressed as a proportion of all genes for each group. Up-regulated genes were defined if the ratio of mRNA levels of 48 and 0 hours pi was ≥ 2, whereas down-regulated genes were defined as ≤ 0.5. To test for statistical differences in mRNA levels between subsets of genes we applied the Wilcoxon test, two-tailed, alpha level: 0.05.

Supplemental References

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

2.2 Chromatin associated factors regulating B cell reprogramming into induced pluripotent stem cells (iPSCs)

2.2.1 Regulation of pluripotency and early differentiation

Embryonic stem cell pluripotency is controlled by a **core network** comprising the transcription factors **Oct4, Sox2** and **Nanog** (Fig. 8) (Hackett and Azim Surani, 2014). These three factors cooperate and bind at regulatory elements, together with other factors such as **Klf4, Esrrb** and **Tcfcp2l1**, in order to activate the pluripotency gene program. These factors block differentiation (to a certain extent). Oct4 and Sox2 are important for the repression of the trophectoderm specification, whereas Nanog represses the primitive mesendodermal program (Teo et al., 2011).

Signalling pathways are particularly important for ESCs maintenance. Activation of the **LIF** signalling pathway is required to prevent ESCs differentiation, and LIF is required in ESCs culture medium. This pathway leads to the activation of the TF **STAT3**, which regulates many pluripotency genes, and represses differentiation genes such as those involved in endoderm specification (Ying et al., 2003). One main target of STAT3 is **Tcfcp2l1,** whose constitutive expression can (partially) rescue the loss of LIF, although STAT3 has other specific target such as **Klf4** (Hall et al., 2009).

Two other important pathways are the **BMP4** and **FGF** pathways, which have antagonistic effects on cell differentiation. BMP4 represses differentiation through the Id proteins (Ying et al., 2003) while FGF favours it through the MEK-ERK signalling cascade (Kunath et al., 2007). The **WNT** pathway is also important for pluripotency, one of its main effects being the activation of the transcription factor **Esrrb** through beta catenin (Martello et al., 2012). WNT is also involved in the regulation of the multipotency of adult stem cells (Ring et al., 2014).

Although these factors altogether maintain the pluripotency of ESCs, they individually are **regulators of differentiation programs**. For example, over-expression of Oct4 triggers mesoderm differentiation (Zeineddine et al., 2006), whereas Sox2 favours the activation of the neurectoderm program (Kopp et al., 2008). Hence ESCs factors do not only cooperate to maintain the expression of pluripotency genes, but also counteract each other (as well as in collaboration and against external signals) in an **unstable equilibrium** (Loh and Lim, 2011).

It has been shown that ESCs cultured in serum-containing medium (with LIF) are heterogenous, in particular regarding Nanog expression. These cells exhibit a transcriptomic profile closer to the epiblast than to the ICM, and were therefore designated as ***primed ESCs.*** Upon removal of serum and addition of two inhibitors targeting MEK and GSK3 (**2i conditions**) (Hall et al., 2009), ESCs display a more **ground** or **naïve state** of pluripotency, with homogeneous expression of Nanog and a global transcriptome closer to that of the ICM (Boroviak et al., 2014).

However, 2i conditions have recently been shown to trigger **irreversible epigenetic defects**, in particular de-methylation of imprinted loci (Yagi et al., 2017), and start to exhibit chromosomic instability after a long maintenance in culture (Choi et al., 2017). This confirmed the fact that the ground state of pluripotency is not a physiologically stable state, as these cells exist only transiently during development, and are intrinsically programmed for differentiation.

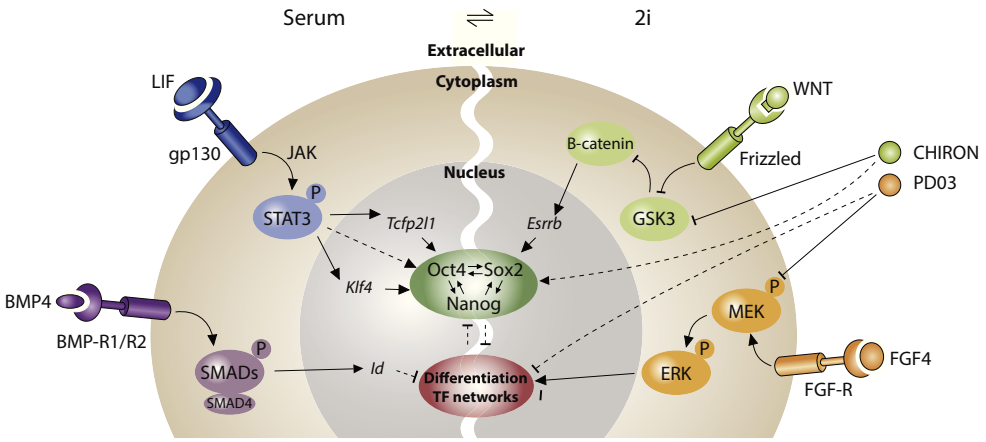


Figure 8: Molecular regulation of pluripotency in ESCs.

Schematic representation of the main pathways and TFs regulating the maintenance of ESCs pluripotency. Adapted from (Hackett and Azim Surani, 2014).

2.2.2 C/EBPa transient expression in B cell allows a very fast induction of pluripotency

After the discovery of iPSCs, the extremely low efficiency of MEF reprogramming by the Yamanaka factors raised the question whether a small **sub-population of more plastic cells** were at the origin of the iPSCs, or if completely differentiated cells could really be reprogrammed to pluripotency.

The best way to address this question was to reprogram B cells, which had undergone V-D-J recombination, and could therefore be traced genetically by sequencing of the immunoglobulin locus. However B cell are very refractory to reprogramming. Inspired by the work of Thomas Graf, the laboratory of Rudolf Jaenish showed in 2008 that ectopic expression of C/EBPa along with OSKM allowed the **reprogramming of terminally differentiated B cells into iPSCs** (Hanna et al., 2008), though with a low efficiency.

During trans-differentiation, C/EBPa induces first the repression of the B cell program, and then the activation of the macrophage program. Therefore, a hypothesis to explain the effect of C/EBPa on B cell to IPSCs reprogramming was that cells could go through a “less differentiated state”, in which OSKM could be more potent, thereby creating **elite cells for reprogramming**.

Following this idea, a PhD student in the laboratory of Thomas Graf found that a **pulse induction of C/EBPa for 18h followed by induction of OSKM** led to a striking increase in reprogramming efficiency (Di Stefano et al., 2014), allowing for more than 95% of the cells to be reprogrammed after 4 days of induction (one of the most efficient iPSCs reprogramming system). These results are presented in two publications in annex 5.1 and 5.2.

How can a short induction of one transcription factor results in such increased potential for reprogramming?

A first transcriptome analysis suggested that, in addition to allowing a more rapid induction of the pluripotency program, C/EBPa was allowing a rapid mesenchymal to epithelial transition (**MET**) (Figure 9 and annex 5.1), a step which has been shown to be mandatory for reprogramming (Li et al., 2010). As C/EBPa is normally a myeloid factor and does not directly activate the pluripotency program, we hypothesized that its effect would be mediated through other factors in particular **epigenetic and chromatin remodelling factors**. Indeed, further analysis showed that the **TET2** enzyme, responsible for active de-methylation of cytosines, was required for the reprogramming. Tet2 is upregulated after C/EBPa pulse and foster the demethylation of the enhancers of key pluripotency genes.

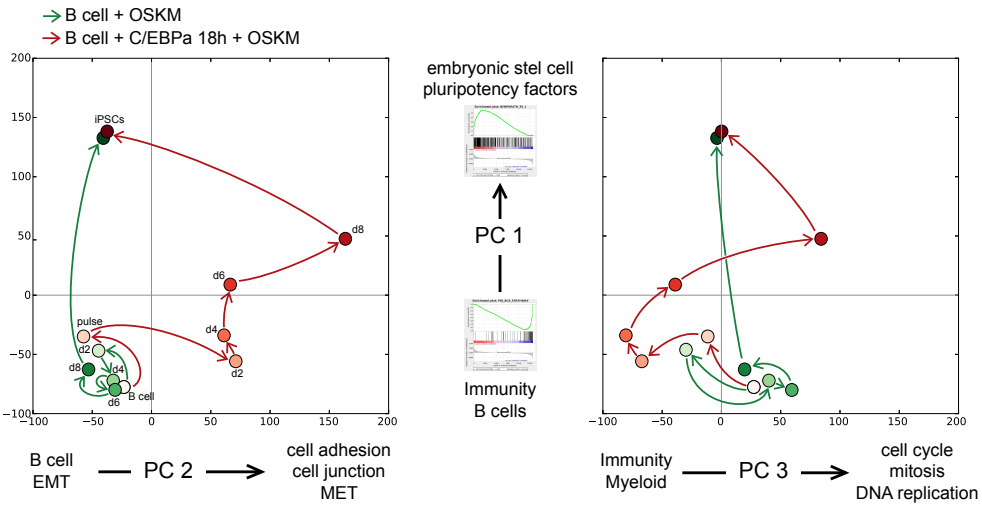


Figure 9: Preliminary transcriptomic analysis of B cell reprogramming into iPSCs

PCA analysis of gene expression microarray data from B cell reprogramming into iPSCs, after 18h expression of C/EBPa ("pulse") and induction of OSKM (red line and point), or just OSKM induction in B cells (green). Left: first component (Y axis) and second component (X axis); right: first component (Y axis) and third component (X axis). A Gene Set Enrichment Analysis (GSEA) was performed on the list of genes ordered by their PCA weight in each axis, and the enriched terms are shown next to the axis.

In order to identify other factors involved in B cell reprogramming into iPSCs, and to decipher their interaction at regulatory elements, **we combined time series of ChIP-seq for histone marks, RNA-seq and proteomic assays**.

2.2.2 Publication: *C/EBPα creates elite cells for iPSC reprogramming by up-regulating Klf4 and increasing the levels of Lsd1 and Brd4 (Nature Cell Biology, 2016)*

Contribution:

I performed all the bioinformatic analyses (with the exception of the processing and normalisation of proteomic data) and data integration. I participated to the identification of potential targets (Lsd1, Hdac1, Brd4 and Klf4) and to the design of subsequent experiments to confirm their role.

ARTICLES

nature cell biology

C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4

Bruno Di Stefano[1,2,8,9,10], Samuel Collombet[3,8], Janus Schou Jakobsen[4,5,6,8], Michael Wierer[7], Jose Luis Sardina[1,2], Andreas Lackner[1,2,9], Ralph Stadhouders[1,2], Carolina Segura-Morales[1,2], Mirko Francesconi[1,2], Francesco Limone[1,2], Matthias Mann[7], Bo Porse[4,5,6], Denis Thieffry[3] and Thomas Graf[1,2,10]

Reprogramming somatic cells into induced pluripotent stem cells (iPSCs) is typically inefficient and has been explained by elite-cell and stochastic models. We recently reported that B cells exposed to a pulse of C/EBPα (Bα′ cells) behave as elite cells, in that they can be rapidly and efficiently reprogrammed into iPSCs by the Yamanaka factors OSKM. Here we show that C/EBPα post-transcriptionally increases the abundance of several hundred proteins, including Lsd1, Hdac1, Brd4, Med1 and Cdk9, components of chromatin-modifying complexes present at super-enhancers. Lsd1 was found to be required for B cell gene silencing and Brd4 for the activation of the pluripotency program. C/EBPα also promotes chromatin accessibility in pluripotent cells and upregulates *Klf4* by binding to two haematopoietic enhancers. Bα′ cells share many properties with granulocyte/macrophage progenitors, naturally occurring elite cells that are obligate targets for leukaemic transformation, whose formation strictly requires C/EBPα.

The ability to reprogram somatic into pluripotent cells has revolutionized stem cell research with major implications for almost all fields of modern biology. Reprogramming to a pluripotent state can be achieved by overexpressing the transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM; refs 1,2). The resulting induced pluripotent stem cells (iPSCs) resemble embryonic stem cells (ESCs), being capable of contributing to chimaeric animals, including the germline[3,4].

iPS reprogramming of mouse embryo fibroblasts (MEFs) is accompanied by transcriptional and epigenetic remodelling[5,6], initiated by the downregulation of the somatic transcriptional program[7,8], followed by a mesenchymal-to-epithelial transition[9,10] (MET). After these changes, pluripotency genes, including *Oct4*, *Nanog* and *Sox2*, become expressed in a small number of cells in the course of about two weeks[7,8]. In pluripotent cells, ESC-specific super-enhancers[11] (ESC-SEs) bound by Brd4, Med1 and Cdk9, maintain the identity of the cells[12,13]. Increasing evidence indicates that reprogramming to pluripotency is a complex process, where multiple players synergistically establish new transcriptional networks and remove epigenetic barriers[14]. Among the factors that have been shown to affect the efficiency and kinetics of reprogramming are cell cycle regulators[15,16], chromatin remodellers[17–19] and facilitators of the MET transition[9,10,20,21].

However, a comprehensive understanding of the molecular mechanism of iPS reprogramming is still lacking, in large part because of the low reprogramming efficiency of most somatic cell types[14]. Yamanaka proposed two alternative explanations for this situation: according to the elite-cell model, reprogramming would take place only in a few predisposed cells within a population; and according to the stochastic model most or all cells are competent for reprogramming at low probabilities[22]. We have recently developed an approach that generates the equivalent of 'elite'-type cells, by transiently expressing in pre-B cells (henceforth called B cells) the transcription factor C/EBPα (ref. 23). Following OSKM activation,

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain. [3]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR8197, INSERM U1024, Ecole Normale Supérieure, PSL Research University, F-75005 Paris, France. [4]The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. [5]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. [6]Danish Stem Cell Center (DanStem), Faculty of Health Sciences, University of Copenhagen, 3B Blegdamsvej, Copenhagen 2200, Denmark. [7]Max Planck Institute of Biochemistry, 82152 Munich, Germany. [8]These authors contributed equally to this work. [9]Present addresses: Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA (B.D.S.); Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria (A.L.).
[10]Correspondence should be addressed to B.D.S. or T.G. (e-mail: distefano@molbio.mgh.harvard.edu or thomas.graf@crg.eu)

Received 29 July 2015; accepted 8 February 2016; published online 14 March 2016; DOI: 10.1038/ncb3326

© 2016 Macmillan Publishers Limited. All rights reserved

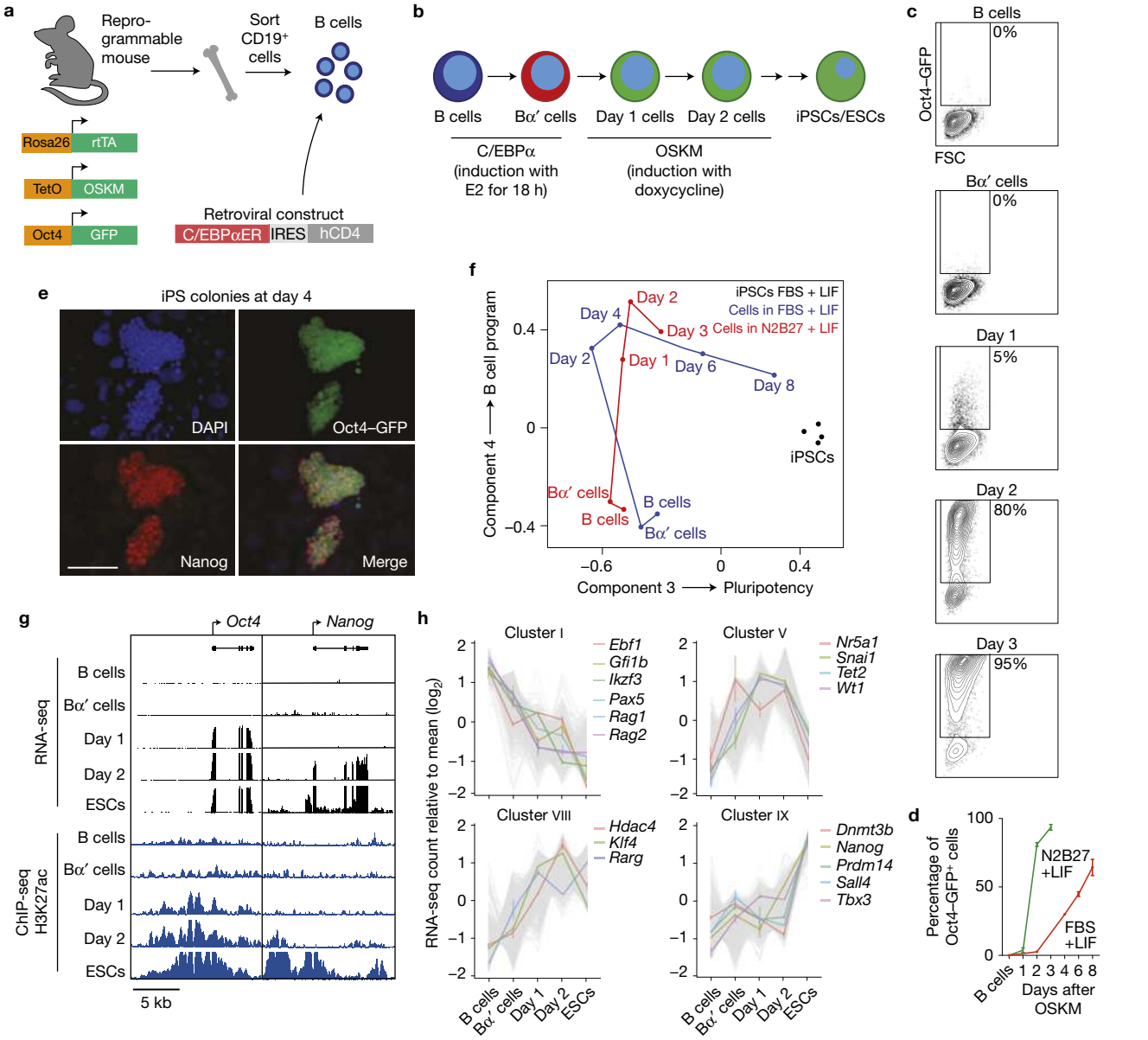


Figure 1 An improved ultrafast reprogramming system of B cells to pluripotency. (**a**) A schematic of the tools used. (**b**) An outline of the reprogramming strategy used. (**c**) Representative FACS analysis of Oct4–GFP expression during B to iPS cell reprogramming. (**d**) Comparison of Oct4–GFP kinetics of the data shown in **c** with data from ref. 23. Error bars indicate s.d. ($n=3$ biological independent experiments). (**e**) Representative immunofluorescence image showing Nanog expression in colonies of Bα′ cells treated for four days with doxycycline (scale bar, 100 μm). (**f**) Independent component analysis of Agilent microarray expression data obtained from cells during reprogramming in serum-free (N2B27+LIF) and serum-containing[23] (FBS+LIF) conditions compared with four iPSC lines. Each time point represents the average from duplicates. (**g**) RNA-seq and H3K27ac ChIP-seq data for the *Oct4* and *Nanog* loci at the indicated times. ESCs were used as controls. (**h**) Representative RNA-seq expression kinetics of four selected gene clusters (see Supplementary Fig. 1B for a complete overview).

these Bα′ cells—unlike control B cells—can be converted into iPSCs at nearly 100% efficiency, also showing that cell reprogramming is basically a deterministic process[23]. Similar efficiencies have been described for genetically modified MEFs and cells treated with specific compound combinations[24–26], including fibroblasts and granulocyte/macrophage progenitors[27,28] (GMPs).

Here we have studied the changes induced in B cells by a C/EBPα pulse as well as during the subsequent rapid onset of OSKM-induced reprogramming, by dynamically monitoring transcription-factor-induced changes in transcription, protein expression, enhancer activity and chromatin accessibility.

RESULTS

Changes in culture conditions result in an ultrafast reprogramming system

We have now improved our B cell reprogramming system by culturing the cells under conditions that favour naive pluripotency[29]. Briefly, pre-B cells from reprogrammable mice crossed with Oct4–GFP mice

© 2016 Macmillan Publishers Limited. All rights reserved

were infected with C/EBPαER retrovirus (Fig. 1a), seeded on MEFs in serum-free medium supplemented with LIF and treated for 18 h with β-oestradiol (E2) to activate C/EBPα, followed by doxycycline to activate OSKM (Fig. 1b). Two to five per cent of the cells became Oct4–GFP positive within 1 day and 95% after 3 days (Fig. 1c,d), and Nanog-positive iPSC colonies could already be detected at day 4 (Fig. 1e). A comparison of the Oct4–GFP kinetics with cells cultured in medium containing FBS and LIF showed a 4 to 5 day acceleration under the new culture conditions[23] (Fig. 1d), which was also confirmed by an independent component analysis of the transcriptome (Fig. 1f and Supplementary Table 1). The resulting iPSCs were found to contribute to chimaera formation after blastocyst injection (Supplementary Fig. 1A).

RNA-seq experiments showed the activation of *Oct4* within 1 day after OSKM induction and of *Nanog* within 2 days, correlating with the genes' decoration by the activating mark H3K27ac (Fig. 1g). The 12,781 genes whose expression changed >2-fold at any given time point were subdivided into nine groups by C-means clustering, and analysed for Gene Ontology (GO) annotations (Fig. 1h and Supplementary Fig. 1B,C). Genes in clusters I and II became downregulated following the C/EBPα pulse/OSKM expression and were enriched for the GO terms 'B cell activation' and 'immune response'. In contrast, genes in cluster III became transiently upregulated and were enriched for 'epigenetic regulation of gene expression' and 'chromatin organization', whereas cluster V was enriched for 'mesenchymal–epithelial transition (MET)' and 'cell adhesion'. Finally, genes in clusters VIII and IX were upregulated by OSKM and enriched for 'stem cell maintenance' and 'blastocyst formation', containing the pluripotency genes *Nanog* and *Klf4*.

Pluripotency genes were activated in three waves, initiated by *Oct4*, *Lin28a*, *Zfp296*, *Gdf3* and *Tdh* at day 1, followed by *Nanog* at day 2, and by *Sall4*, *Sox2*, *Esrrb* and *Zfp42* (*Rex1*) at day 4 (Supplementary Fig. 1D). Changes in expression of pluripotency genes were confirmed at the protein level (Supplementary Fig. 1E).

In the following, we will use this ultrafast reprogramming system to investigate changes at the protein and chromatin levels, both after the initial expression of C/EBPα and following OSKM induction.

The C/EBPα pulse elevates the levels of multiple proteins, without concomitant transcriptional changes

To analyse proteome dynamics during reprogramming, we performed shotgun proteomics by label-free quantification of samples from the four reprogramming time points and ESCs. We detected 7,497 proteins that included 520 transcription factors, 295 kinases, 141 phosphatases and 96 isomerases (Supplementary Fig. 2A), showing an excellent correlation between duplicates (Supplementary Fig. 2B). C-means clustering of the proteins that changed >2-fold at any given time point resulted in ten groups (Fig. 2a and Supplementary Fig. 2C). Downregulated proteins (clusters a,b) were highly enriched for the GO terms 'B cell immunity' and 'immune system' (Fig. 2b). The early upregulated proteins were found to be associated with the terms 'RNA splicing' and 'protein degradation' (Supplementary Table 2). In contrast, late upregulated proteins (clusters i,j) were enriched for 'stem cell maintenance' and 'stem cell development' and contained Oct4 and Sox2. Surprisingly, 439 out of 538 proteins elevated in Bα′ cells (clusters g,h) were not regulated at the RNA level (Fig. 2c). These belonged to the categories 'chromatin remodelling' and 'histone modifications' and included the epigenetic factors Lsd1, Hdac1, Brd4 and Med1. The observed accumulation at the protein level of the histone demethylase Lsd1 (Kdm1A) and the histone deacetylase Hdac1 (Fig. 3a) could also be confirmed by western blot (Supplementary Fig. 3A).

A protein complex containing Lsd1, Hdac1 and C/EBPα mediates B cell gene silencing

Lsd1 is known to demethylate H3K4me1/2 (ref. 30) and to be part of a complex that includes the histone deacetylase Hdac1 (ref. 31). Hdac1 in turn has been described to interact with C/EBPα (ref. 32). We therefore investigated whether C/EBPα can interact with both Lsd1 and Hdac1 and whether the complex is required to silence B cell gene expression (Supplementary Fig. 3B) during reprogramming. Gel-filtration chromatography showed that C/EBPα, Lsd1 and Hdac1 co-elute in a high-molecular-weight complex (Supplementary Fig. 3D). Furthermore, C/EBPα immunoprecipitation followed by mass spectrometry revealed an interaction with Lsd1 and Hdac1 (Fig. 3b and Supplementary Fig. 3E), a finding confirmed by western blot (Supplementary Fig. 3F). Our proteomic analyses also showed an interaction of C/EBPα with Wdr5 and other members of the MLL complex, as recently reported[33]. Finally, an antibody against Lsd1 co-immunoprecipitated Hdac1 and C/EBPα (Fig. 3c), but not the control proteins Pcna and Parp1 (Supplementary Fig. 3G), further supporting the proposed association between C/EBPα, Lsd1 and Hdac1.

As the Lsd1/Hdac1 complex has been shown to regulate the inactivation of super-enhancers[31] (SEs), we tested the effect of C/EBPα on B cell SEs (B-SEs), examining 514 regions with high H3K27ac levels[11,34]. This revealed a marked reduction in B-SE activity, as seen by the loss of H3K27ac and Brd4 (Fig. 3d and Supplementary Fig. 3H). We also found a significant decrease in the activation-associated marker H3K4me2 at regulatory elements of the B cell genes *Ebf1*, *Fox1*, *Gfi1b* and *Ikzf3* (ref. 35), as well as of the lymphoid *Rag1* and *Ciita* genes (Fig. 3e).

To test whether Lsd1 is required for the C/EBPα-mediated B cell silencing (Supplementary Fig. 3B) and decrease in B-SE activity, we tested the effect of S2101, a compound that specifically blocks the enzymatic activity of Lsd1 (ref. 36). We found that it indeed prevented the C/EBPα-induced decommissioning of B cell enhancers (Fig. 3e) and, at least in part, B cell silencing (Fig. 3f), as did Hdac1 inhibition (Supplementary Fig. 3I). Lsd1 was also found to be required for the fast reprogramming of B cells into iPS cells as S2101 treatment impaired Oct4–GFP upregulation and markedly reduced the number of iPS colonies (Fig. 3g–i). This was not due to an adverse effect of Lsd1 inhibition on B cell viability (Supplementary Fig. 4A) or proliferative capacity (Supplementary Fig. 4B). Knockdown experiments confirmed these observations, yielding reduced numbers of iPS colonies (Fig. 3j and Supplementary Fig. 4C–E). Interestingly, OSKM-induced reprogramming of B cells not exposed to C/EBPα was not inhibited by S2101 (Supplementary Fig. 4F), indicating that Lsd1 acts downstream of C/EBPα.

These results show that inhibition of Lsd1 and Hdac1 impairs enhancer decommissioning and B cell gene silencing in Bα′ cells.

© 2016 Macmillan Publishers Limited. All rights reserved

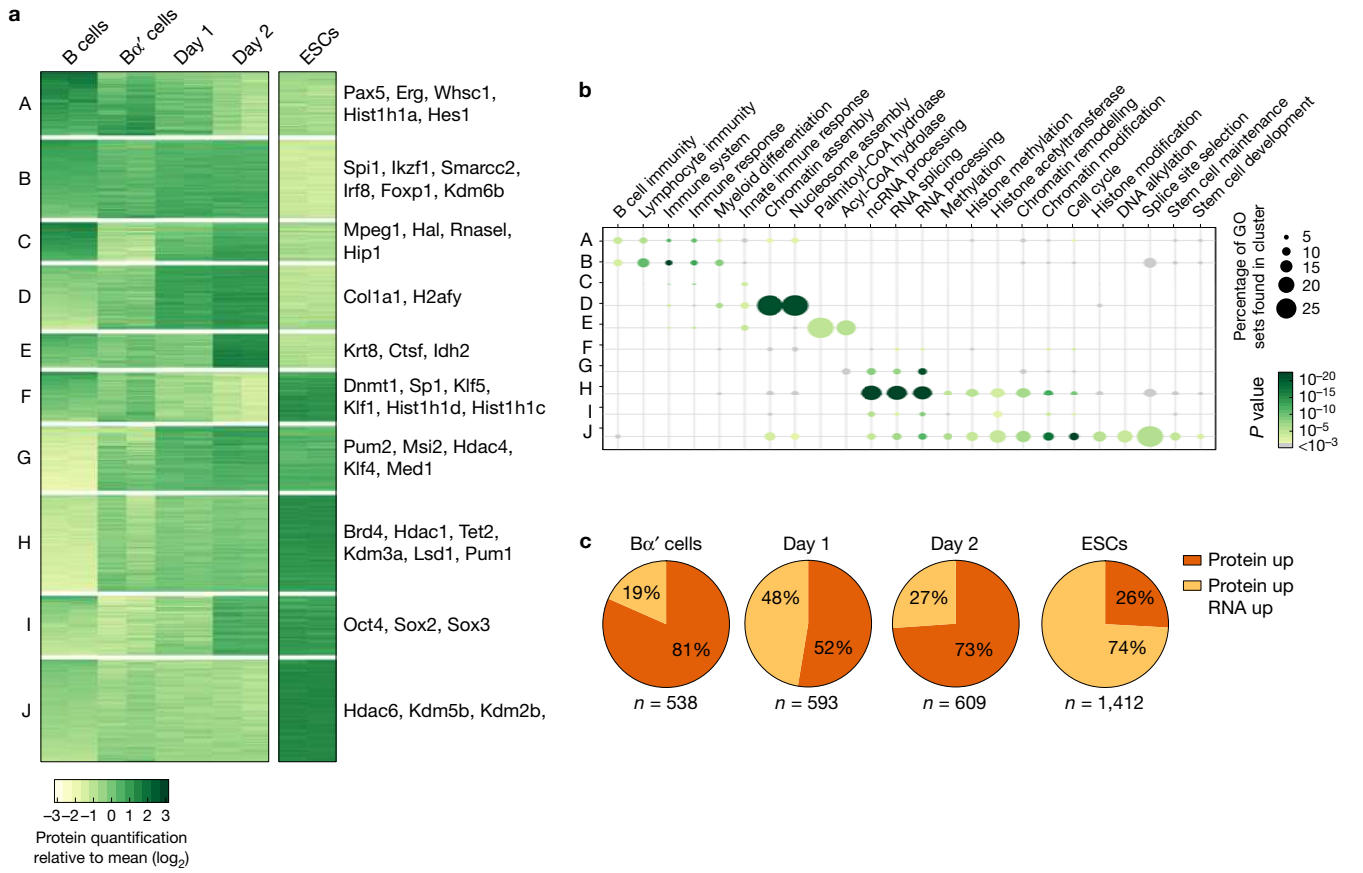


Figure 2 C/EBPα- and OSKM-induced proteome changes during reprogramming. (**a**) C-means clustering of proteins whose abundance changes >2-fold during reprogramming. Selected members of each cluster are indicated. (**b**) GO analysis of protein clusters shown in **a**. The size of each circle represents the proportion of GO sets found in each cluster; the intensity of the colour represents the *P* value as determined by a hypergeometric test. (**c**) Percentage of proteins upregulated during reprogramming for each time point with respect to the previous one; that is, the values shown for Bα′ cells are relative to B cells. In orange are proteins whose abundance increases without concomitant transcriptome change; in ochre are proteins that become upregulated at both the protein and transcriptional levels.

C/EBPα post-transcriptionally increases Brd4, Med1 and Cdk9 protein levels and Brd4 is required for reprogramming

The C/EBPα pulse also caused a marked and coordinated post-transcriptional increase of Brd4, Med1 and Cdk9 as detected by mass spectrometry and western blot (Fig. 4a and Supplementary Fig. 3A,C). This is intriguing as recent studies have shown that a complex containing Brd4, Med1 and Cdk9 is required for the self-renewal capacity and pluripotency of ESCs (ref. 12) and for reprogramming of MEFs (ref. 13), and that Brd4 inhibition results in the complex's disruption[12]. As Brd4 is also required for SE activity[37], we analysed the changes of H3K27ac decoration at ESC-SEs during reprogramming. Notably, some ESC-SEs became already activated in Bα′ cells (Fig. 4b), including the *Lefty* and *Rarg* enhancers, and 19/32 showed increased Brd4 binding. This finding raised the possibility that C/EBPα recruits Brd4 to acetylated histone tail residues[38]. To test this hypothesis, we performed ChIP-seq experiments with B cells and Bα′ cells and found that C/EBPα and Brd4 co-occupy approximately 25% of the newly bound sites, including ESC-SEs of *Id1*, *Iqgap1*, *Rarg* and *Egln3* genes (Fig. 4c,d and Supplementary Fig. 4G). Co-immunoprecipitations further confirmed that the two proteins form a complex (Fig. 4e), as was also recently described[38].

Next we tested whether Brd4 inhibition impairs iPS reprogramming by using the compound JQ1, known to effectively inhibit the function of bromodomain and extraterminal domain (BET) family proteins, including Brd4 (ref. 39). JQ1 was either administered only during the pulse or during a two-day doxycycline treatment. Cells were monitored for the expression of Oct4–GFP by fluorescence-activated cell sorting (FACS) or of key pluripotency genes. Strikingly, JQ1 treatment completely inhibited Oct4–GFP activation (Fig. 4f) and significantly reduced the upregulation of *Oct4*, *Lin28a*, *Zfp296*, *Lefty1* and *Cdh1* when administered during both the C/EBPα pulse and OSKM activation (Fig. 4g). Finally, we observed a threefold reduction of iPSC colonies when the drug was administered during the C/EBPα pulse and a tenfold impairment when provided together with doxycycline (Fig. 4h and Supplementary Fig. 4H), a finding confirmed with an short hairpin RNA (shRNA) against Brd4 (Fig. 4i and Supplementary Fig. 4C–E). Of note, the concentration of JQ1 used did not severely affect B cell survival or proliferation (Supplementary Fig. 4A,B).

We conclude that C/EBPα induces a post-transcriptional accumulation of the Brd4–Med1–Cdk9 complex and re-localizes Brd4 to ES-SEs. Our data also show that Brd4 is strictly required for iPS cell formation from B cells.

© 2016 Macmillan Publishers Limited. All rights reserved

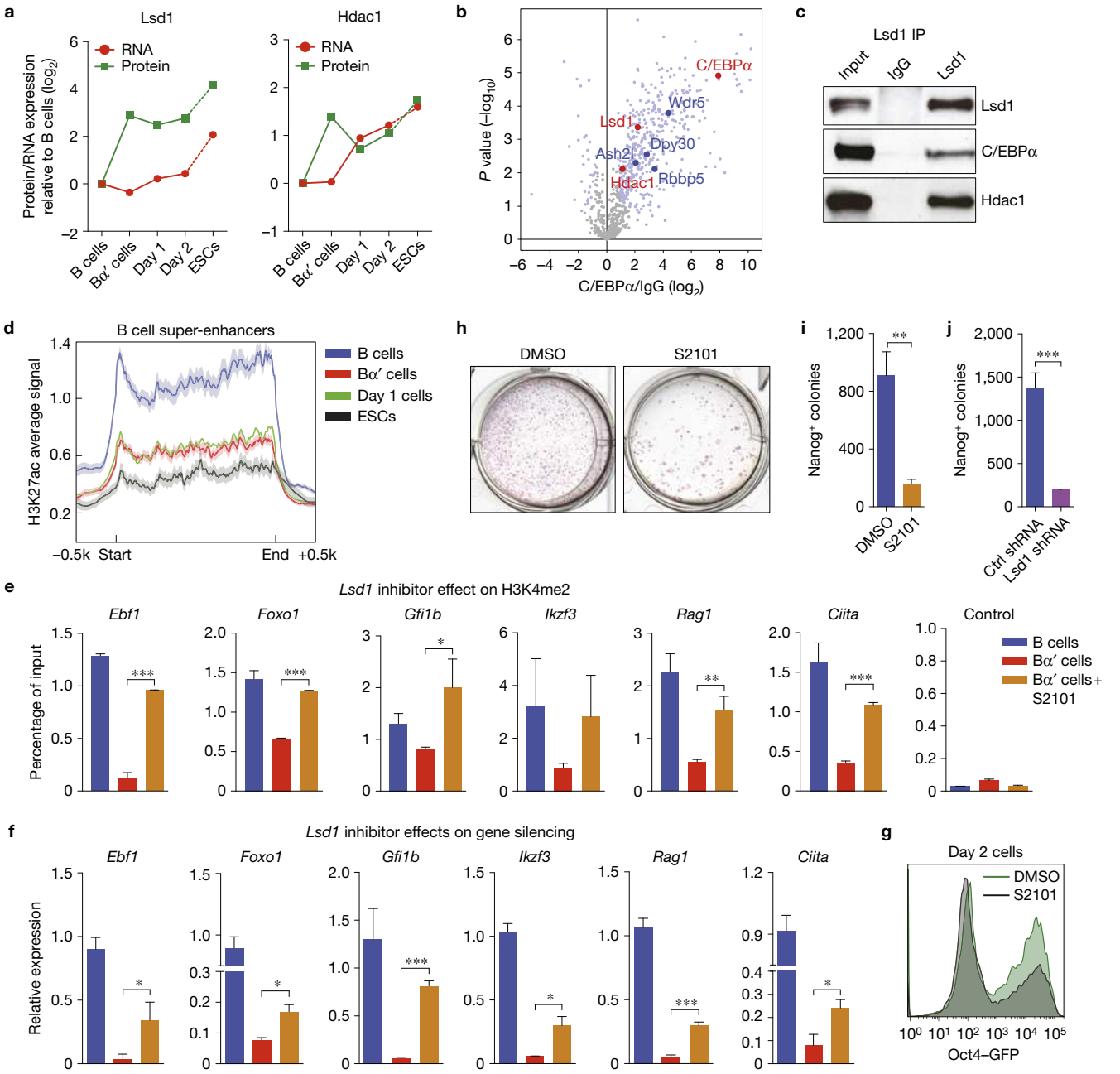


Figure 3 C/EBPα-induced Lsd1 and Hdac1 upregulation and requirement of Lsd1 for silencing of the B cell program. (**a**) Expression of Lsd1 and Hdac1 mRNA and proteins during reprogramming compared with ESCs. The data represent the average from two biologically independent samples. (**b**) C/EBPα immunoprecipitation mass-spectrometry data (based on three experiments using a specific antibody and three using IgG control serum). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($n = 3$ biologically independent samples). Small grey dots: non-significantly enriched proteins (FDR > 0.05). Small blue dots: significantly enriched proteins (FDR < 0.05). Large blue dots: proteins from the MLL complex and Top2a. Red dots: see main text. (**c**) Representative western blot of co-immunoprecipitation experiments. Extracts from Bα′ cells were probed with Lsd1, C/EBPα or Hdac1 antibodies. Unprocessed original scans of blots are shown in Supplementary Fig. 8. (**d**) H3K27ac levels at B cell super-enhancers in B cells, Bα′ cells, day 1 cells and ESCs. (**e**) H3K4me2 at selected B cell enhancers as measured by ChIP-qPCR in B cells, Bα′ cells and Bα′ cells treated with the Lsd1 inhibitor S2101 during the pulse. Error bars indicate s.d. ($n = 3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). (**f**) Gene expression by qRT-PCR of samples shown in **d**. Values were normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P < 0.05$, $^{***}P < 0.001$). (**g**) Representative Oct4–GFP FACS analysis of B cells treated with S2101 at day 2 of reprogramming. (**h**) Representative alkaline-phosphatase-positive iPS colonies obtained from reprogramming of Bα′ cells induced with OSKM and treated with S2101 or dimethylsulphoxide as a control. (**i**) Nanog^{+} iPSC colony counts at day 12 of reprogramming of cells treated with S2101 or dimethylsulphoxide as control. Error bars indicate s.d. ($n = 3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{**}P < 0.01$). (**j**) Nanog^{+} iPSC colony counts at day 12 of reprogrammed cells with a knockdown of Lsd1. Error bars indicate s.d. ($n = 3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P < 0.001$).

© 2016 Macmillan Publishers Limited. All rights reserved

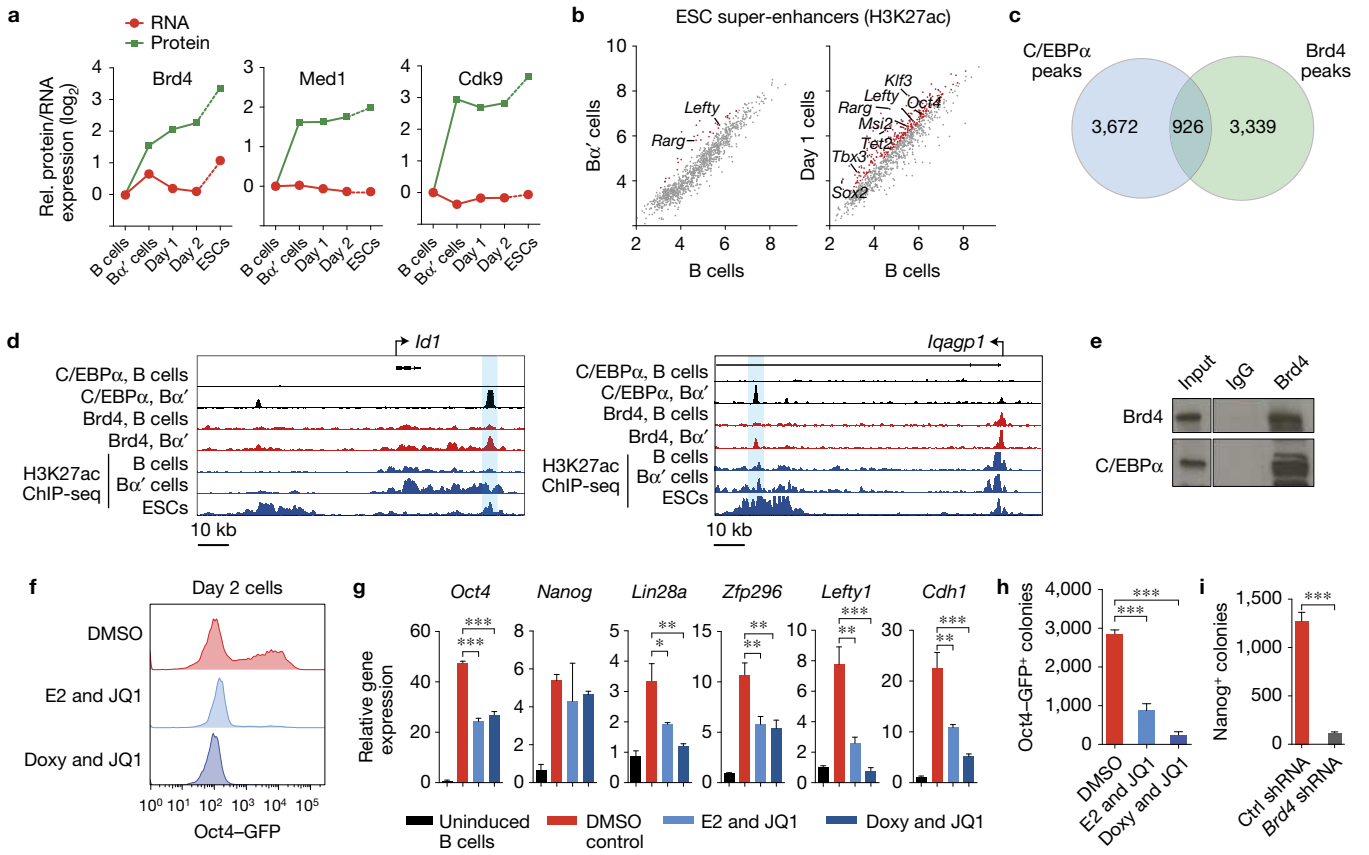


Figure 4 C/EBPα-induced Brd4, Med1 and Cdk9 upregulation and requirement of Brd4 for activation of the pluripotency program. (**a**) Expression of Brd4, Med1 and Cdk9 mRNA and proteins during reprogramming and in ESCs. The data represent the average from two biologically independent samples. (**b**) Activation of ESC-SEs during reprogramming as measured by H3K27Ac, comparing B with Bα′ cells and B with day 1 cells. Axes show the RPK score (log scale) for H3K27ac. Selected genes associated with super-enhancers showing a >1.5-fold increase in activity compared with B cells are highlighted in red. (**c**) Venn diagram showing the overlap between C/EBPα and Brd4 ChIP-seq peaks in Bα′ cells. (**d**) Representative genome browser screenshots of the *Id1* and *Iqagp1* genes showing C/EBPα, H3K27ac and Brd4 ChIP-seq data. (**e**) Representative western blot of co-immunoprecipitation experiment. Extracts from Bα′ cells were probed with Brd4 or C/EBPα antibodies. Unprocessed original scans of blots are shown in Supplementary Fig. 8. (**f**) Representative Oct4–GFP FACS analysis of B cells treated with JQ1 for 18 h or 2 days. (**g**) Effect of JQ1 on the expression of selected pluripotency genes by qRT-PCR. Values were normalized against *Pgk* expression. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$). (**h**) Oct4–GFP^{+} iPSC colonies at day 12 of reprogramming, after treatment of B cells with JQ1 during either C/EBPα or OSKM induction. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$). (**i**) Nanog^{+} iPSC colony counts after reprogramming of B cells with a knockdown for Brd4. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$).

C/EBPα induces chromatin accessibility at selected pluripotency loci

To study changes in chromatin accessibility, we used ATAC-seq[40] to identify accessible chromatin regions and reveal potential regulatory elements bound by transcription factors. After calling discrete ATAC peaks, we identified 6,319 regions with distinct dynamics during the time course, which were divided by C-means clustering into four groups (Fig. 5a). Cluster I contained 525 regions with ATAC peaks formed *de novo* after the C/EBPα pulse and that reach maximum accessibility in ESCs. Their association with GO categories 'developmental process' and 'embryo development' (Fig. 5a and Supplementary Fig. 5A) suggests that they mark sites associated with pluripotency. In contrast, most ATAC peaks induced by the C/EBPα pulse, represented in clusters II, III and IV, diminished in day 2 cells and were poorly represented in ESCs. GO analyses of genes associated with these regions revealed an enrichment for myeloid genes (Supplementary Fig. 5A) as exemplified by *Id1* and *Ifitm6*, genes that are upregulated by C/EBPα (Supplementary Fig. 5B). We then investigated the presence of transcription-factor-binding motifs within ATAC-seq peaks. As predicted, cluster I regions were strongly enriched for motifs associated with pluripotency transcription factors such as Klf4, Oct4, Sox2 and Essrb. In contrast, newly accessible regions in clusters II, III and IV were enriched for potential binding sites of the myeloid transcription factors PU.1 (Ets), C/EBPα, Irf8 and Runx1 (Fig. 5b, and Supplementary Fig. 5C).

C/EBPα directly regulates Klf4 expression and together they increase chromatin accessibility

Klf motif enrichment in ATAC cluster I suggested that Klf4 might already be expressed in Bα′ cells. We found that it was

© 2016 Macmillan Publishers Limited. All rights reserved

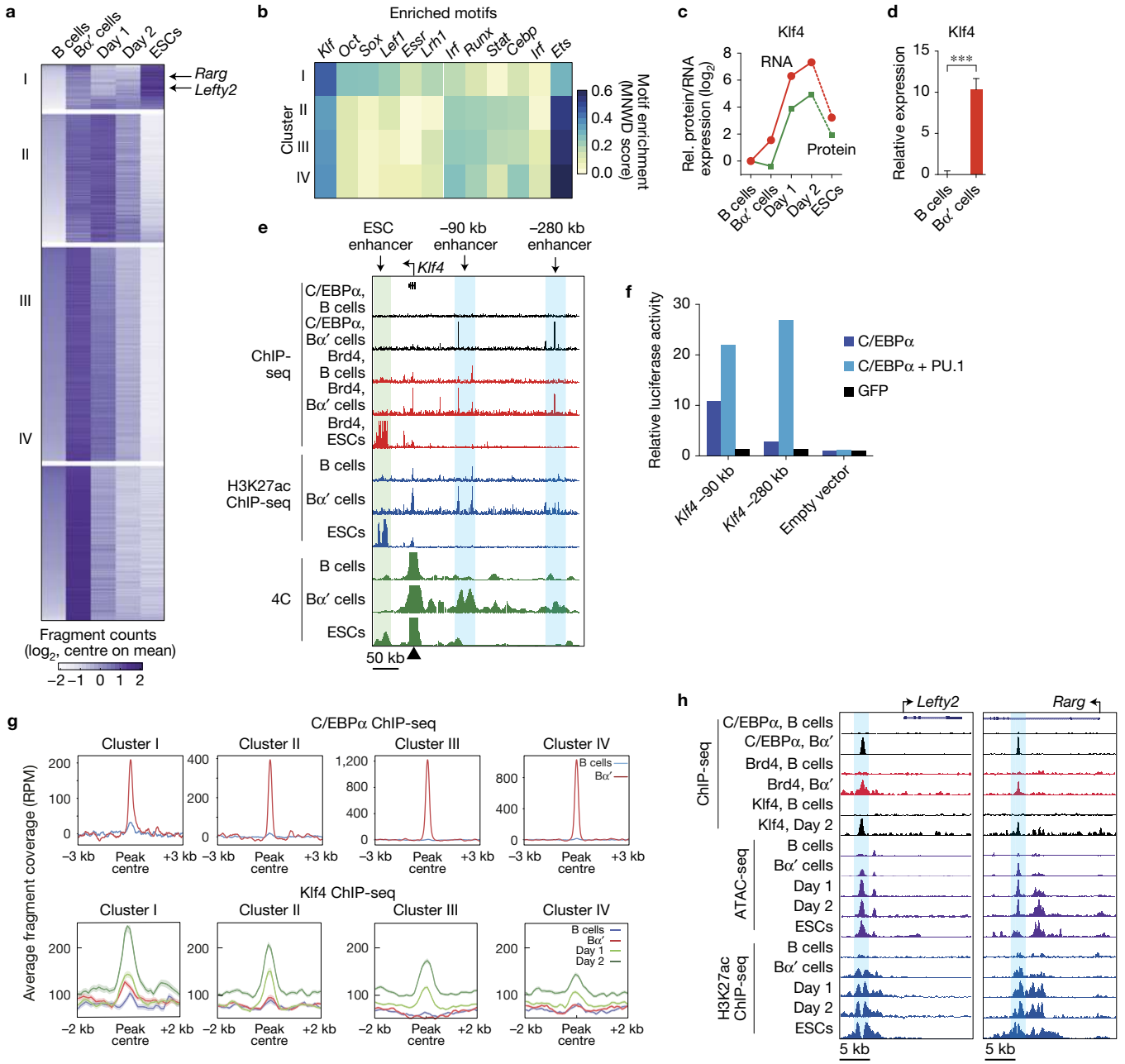


Figure 5 Analysis of chromatin accessibility changes and transcription factor binding in C/EBPα-pulsed B cells. (**a**) C-means clustering of newly formed accessible chromatin regions (as measured by ATAC-seq) in Bα′ cells compared with B cells, yielding 4 clusters. (**b**) Motif analysis of ATAC-seq peaks within the clusters identified in **a**. (**c**) mRNA and protein expression kinetics of Klf4 during reprogramming. The data represent the average from two biologically independent samples. (**d**) Comparison of Klf4 expression by qRT-PCR between B cells and Bα′ cells. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$). (**e**) Genome browser screenshot of the *Klf4* locus, showing ChIP-seq data for C/EBPα, Brd4 and H3K27ac, as well as 4C data using the *Klf4* promoter as the view point (black triangle at the bottom). (**f**) Luciferase reporter assay of 293T cells transfected with luciferase constructs containing only a minimal promoter (empty vector) or also one of the two putative *Klf4* enhancers at −90 kb and −280 kb predicted to be regulated by C/EBPα and PU.1. Plasmids encoding C/EBPα, C/EBPα and PU.1 or GFP were co-transfected. The data represent the average from two biologically independent samples. (**g**) C/EBPα and Klf4 ChIP-seq analysis in B, Bα′, Day 1 and Day 2 cells, showing average peak distribution. (**h**) Representative genome browser screenshots of *Rarg* and *Lefty2*, showing ChIP-seq data for C/EBPα, Klf4 and H3K27ac in Bα′ cells, day 2 cells and ESCs.

indeed transcriptionally upregulated >2-fold by C/EBPα (Fig. 5c,d), a finding confirmed by western blot (Supplementary Fig. 3A). Remarkably, C/EBPα binds to two regions upstream of Klf4 (−90 kb and −280 kb) that are also occupied by Brd4 and PU.1 and were enriched for H3K27ac, suggesting that these sites represent active enhancers (Fig. 5e and Supplementary Fig. 5D). Circularized chromosome conformation capture (4C-Seq) experiments support this interpretation, revealing that the *Klf4* promoter interacts with

© 2016 Macmillan Publishers Limited. All rights reserved

these two regions in Bα′ cells but not in B cells (Fig. 5e and Supplementary Fig. 5D). Interestingly, in ESCs these enhancers showed no activity and no looping with the *Klf4* promoter, which interacted instead with a +50 kb enhancer (Fig. 5e). In addition, co-transfection of the −90 kb enhancer linked to a luciferase reporter with C/EBPα enhanced reporter activity, whereas both C/EBPα and PU.1 were necessary to activate the −280 kb enhancer (Fig. 5f). Their differential responsiveness to the two transcription factors can be explained by the observation that in B cells the −280 kb, but not the −90 kb enhancer, is primed by endogenous PU.1 (Supplementary Fig. 5D), a transcription factor known to synergize with C/EBPα (ref. 41).

To explore the possibility of interplay between C/EBPα and Klf4, we compared the binding sites of the two factors and analysed their enrichment at the ATAC peaks within the four clusters. Supporting our motif analyses, Klf4 binding showed the strongest enrichment at cluster I sites, decreasing progressively towards cluster IV, whereas C/EBPα binding showed the reverse trend (Fig. 5g). In ESCs, cluster I sites were specifically enriched for Brd4 and Klf4 binding, highlighting the known interaction between the two factors[13] (Supplementary Fig. 5E). Our data also showed that approximately 5% of regions bound by C/EBPα in Bα′ cells are bound by Klf4 in pluripotent cells (Supplementary Fig. 5F), raising the possibility that here C/EBPα acts as a pioneer factor for subsequent Klf4 binding. Analysis of MNase-seq and C/EBPα ChIP-seq data obtained after C/EBPα activation[41], generated in an inducible pre-B cell line, support this hypothesis (Supplementary Fig. 5G). A pioneering role of C/EBPα is also suggested for the pluripotency-associated genes *Rarg* and *Lefty1/2*. C/EBPα binds to the enhancers of these genes in Bα′ cells, which then become activated, as seen by their decoration with H3K27ac and Brd4 (Fig. 5h). Following OSKM induction, these enhancers become subsequently bound by Klf4 and further activated. Moreover, both regions are bound by Oct4, Nanog, Sox2, Klf4 and Brd4 in ESCs and are also enriched for H3K27ac, confirming that they correspond to active pluripotency enhancers (Supplementary Fig. 6A). Consistent with this interpretation, *Rarg* and *Lefty1* become upregulated during reprogramming (Supplementary Fig. 6B).

In conclusion, C/EBPα creates a large number of newly accessible chromatin regions in B cells and it upregulates *Klf4* by binding to two haematopoietic enhancers. C/EBPα and Klf4 subsequently facilitate accessibility to other pluripotency-associated transcription factors, as exemplified for the *Lefty* and *Rarg* enhancers.

C/EBPα and Klf4 link the elite cell state of Bα′ cells to that of granulocyte/macrophage progenitors

GMPs are the cells within the haematopoietic system most susceptible to OSKM-induced reprogramming, with 25% to >90% % of single seeded cells converting into iPSCs (refs 25–28; summarized in Fig. 6a), thus representing an 'elite' or 'privileged' cell state. C/EBPα-deficient mice lack GMPs (ref. 42), showing that their formation strictly requires C/EBPα. To compare Bα′ cells and GMPs, we performed expression microarray and RNA-seq experiments. *Cebpa* was expressed >100-fold higher in GMPs than in B cells and ESCs (Fig. 6b), whereas *Klf4* was expressed about fivefold higher in Bα′ cells and GMPs compared with B cells (Fig. 6b). Moreover, C/EBPα binds to the −90 and −280 kb Klf4 enhancers in GMPs (ref. 43; Fig. 6c), raising the possibility that C/EBPα and Klf4 co-regulate a substantial proportion of genes in GMPs and Bα′ cells. Indeed, almost 30% of the genes expressed in GMPs—but not B cells—were also found to be expressed in Bα′ cells (Supplementary Fig. 6C). Furthermore, a canonical variate analysis showed that Bα′ cells are most similar to GMPs when compared with all other haematopoietic populations (Supplementary Fig. 6D). We then performed ATAC-seq in GMPs to test whether chromatin accessible regions in Bα′ cells are also accessible in these cells. Strikingly, of 525 newly accessible regions in Bα′ cells, 357 overlapped with ATAC peaks in GMPs (Fig. 6d and Supplementary Fig. 6E,F), which were not detectable in MEFs (ref. 44; Supplementary Fig. 6F). This is exemplified by the pluripotency genes *Lefty* and *Rarg* and the myeloid genes *Tet2*, *Ifitm6* and *Id1* (Supplementary Fig. 6G). Furthermore, the intersection between C/EBPα and ATAC-seq peaks in Bα′ cells shows that C/EBPα binding exhibits a similar enrichment in GMPs (Supplementary Fig. 6H and Fig. 5g).

Rapidly cycling GMPs express elevated levels of Klf4 and Tet2

It was recently reported that a rapidly cycling GMP sub-fraction can be reprogrammed into iPSCs at a >3-fold higher efficiency than the slowly cycling sub-fraction[28]. We therefore determined whether the elite-cell-associated factors Klf4 and Tet2, which have been shown to enhance the reprogramming efficiency of B cells and MEFs (refs 23,45), are differentially expressed in the two cell fractions. For this purpose, GMPs from reprogrammable mice were labelled with CFSE and separated into a CSFE-high fraction (slow-cycling cells) and a CSFE-low fraction (fast-cycling cells; Supplementary Fig. 7A). The fast-cycling GMPs expressed twofold more *Klf4* and 1.5-fold more *Tet2* than the slow-cycling cells, a finding confirmed by RNA-seq and quantitative real-time PCR (qRT-PCR) for *Klf4* (Fig. 6e,f and Supplementary Fig. 7B) whereas the slow-dividing cells expressed higher levels of the cell cycle inhibitor *Cdkn1a* (p21; Fig. 6e). In contrast, the two cell fractions did not differ in their expression of *Cebpa* and *Cebpb* nor of *Lsd1*, *Hdac1*, *Brd4*, *Med1* and *Cdk9* (Supplementary Fig. 7C), in line with the observation that C/EBPα regulates these genes only at the protein level. Finally, we examined the role of Tet2 in the reprogramming to iPSCs of both B cells and GMPs. After Tet2 knockdown, we observed a significant decrease in the number of iPSC colonies for both cell types (Fig. 6g,h and Supplementary Fig. 7D).

DISCUSSION

Our results describe molecular mechanisms by which a pulse of the transcription factor C/EBPα converts B cells into elite cells for reprogramming into iPSCs. C/EBPα initiates the reprogramming cascade by concomitantly boosting the levels of protein complexes required for silencing the B cell program and for establishing the pluripotency program. C/EBPα also controls Klf4 expression by binding to two haematopoietic specific enhancers, allowing the two factors to induce chromatin accessibility at regions that become fully accessible in pluripotent cells (Fig. 7a,b).

Our most surprising observation was that C/EBPα post-transcriptionally increases the levels of hundreds of proteins, including key chromatin-modifying factors. The finding that C/EBPα interacts with Lsd1 (Fig. 3b,c), Hdac1 (Fig. 3b,c and ref. 32) and Brd4 (Fig. 4e) raises the possibility that it could increase protein stability

© 2016 Macmillan Publishers Limited. All rights reserved

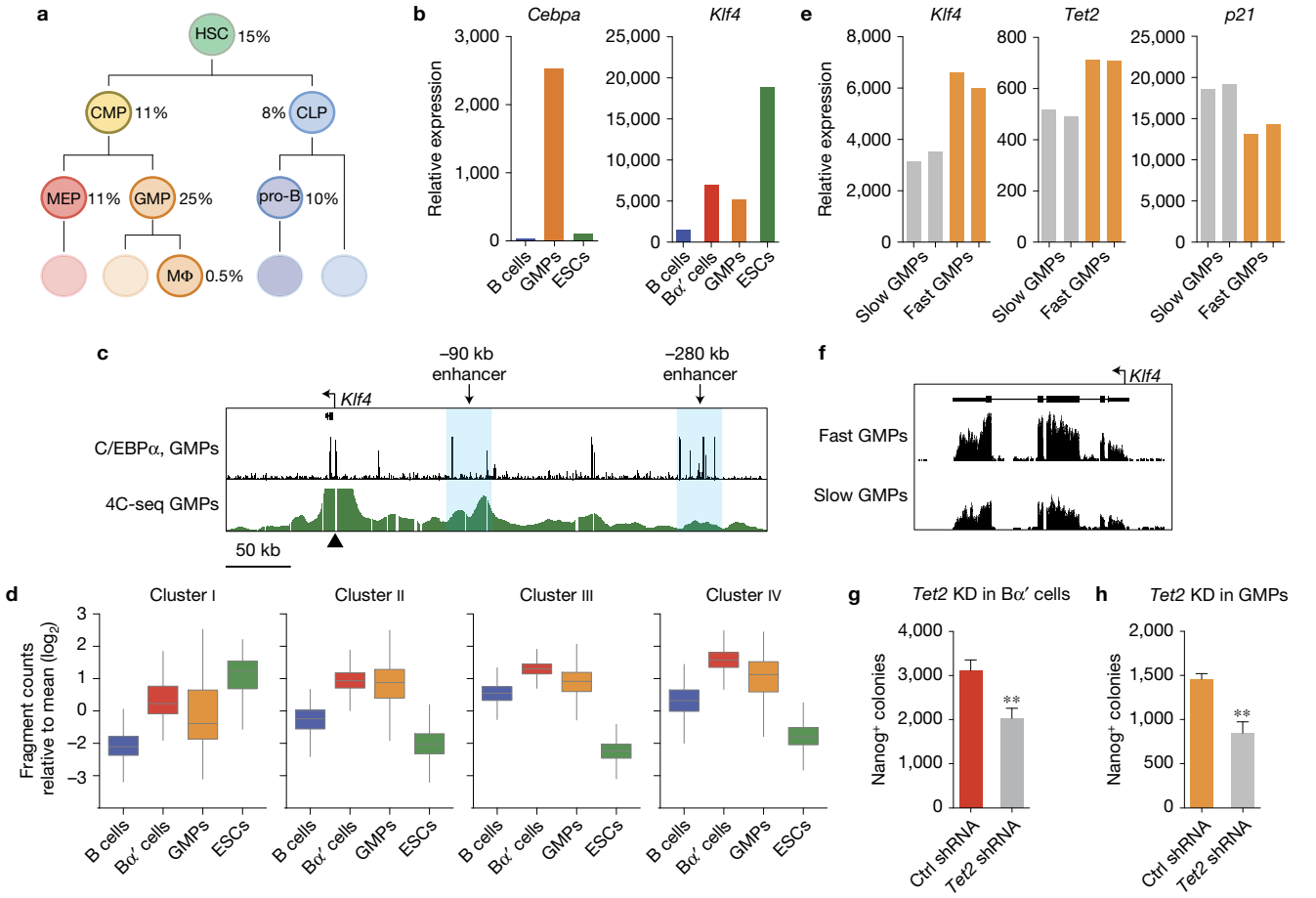


Figure 6 Comparison between molecular properties of Bα′ cells and GMPs. (**a**) A schematic of the haematopoietic lineage tree showing the percentage of single cells that could be reprogrammed into iPSCs as determined in ref. 27. (**b**) Comparison of *Cebpa* and *Klf4* expression in B cells, GMPs and ESCs (expression microarrays). The data represent the average from two biologically independent samples. (**c**) Genome browser screenshot of the *Klf4* locus, showing C/EBPα ChIP-seq in GMPs, and 4C using the *Klf4* promoter as the view point (black triangle, similar to Fig. 5e). (**d**) Boxplots of a comparison between the four clusters of chromatin accessible regions as determined by ATAC-seq (Fig. 5a), with ATAC-seq data obtained from GMPs. Central line represents the median, hinges represent the 25th and 75th percentiles, and whiskers represent the lowest and highest values within 5×IQR (the interquartile range of the hinges). Cluster I, $n=525$; cluster II, $n=1{,}497$; cluster III, $n=2{,}489$; and cluster IV, $n=1{,}808$. (**e**) Microarray gene expression values for *Klf4*, *Tet2* and *p21* in fast- and slow-cycling GMPs. The data are from two biologically independent samples. (**f**) Representative screenshots of *Klf4* gene expression by RNA-seq for fast- and slow-cycling GMPs. (**g**) Nanog$^+$ iPSC colony counts after reprogramming of B cells depleted of *Tet2*. Error bars indicate s.d. ($n=3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$, $^{**}P<0.01$). (**h**) Nanog$^+$ iPSC colony counts after reprogramming of GMPs with a knockdown of Tet2. Error bars indicate s.d. ($n=3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{**}P<0.01$).

by preventing their ubiquitin-mediated degradation. In support of this possibility, proteasome inhibition has indeed been shown to increase the levels of Lsd1 and Hdac1 (refs 46,47). The finding that Lsd1 is required for the inactivation of B-SEs and the silencing of the B cell program suggests that it has a similar role as in ESCs, where it is responsible for the decommissioning of ESC-SEs during differentiation[31]. The observation that the Brd4 inhibitor JQ1 blocks reprogramming supports recent reports demonstrating that the Brd4–Med1–Cdk9 protein complex is required for reprogramming of MEFs (ref. 13) and for the maintenance of pluripotency of ESCs (ref. 12). That JQ1 also impaired reprogramming when used to treat B cells during the C/EBPα pulse was unexpected, but could be explained by the ability of C/EBPα to recruit Brd4 to ESC-SEs required for activation of pluripotency genes. Together, these results suggest that Brd4 is generally required for cell fate conversions and Lsd1 acts in a more tissue-restricted manner. In line with this interpretation, we found that JQ1 impairs the TFIID-enhanced reprogramming of MEFs into iPSCs (ref. 48), whereas S2101 exhibited a slight acceleration (Supplementary Fig. 7E).

As summarized in Fig. 7b, several lines of evidence help to explain why GMPs, like Bα′ cells, behave as elite cells for reprogramming: (i) GMPs express high levels of C/EBPα and strictly require the gene for their formation[42]; (ii) Klf4 is more highly expressed in GMPs than in B cells. In addition, the rapidly cycling GMP subpopulation with higher reprogramming efficiency express elevated levels of Klf4 (Fig. 6e,f); (iii) Tet2 is upregulated by C/EBPα, highly expressed in GMPs and required for the efficient reprogramming of both cell types[23] (Fig. 6g,h); (iv) GMPs specifically express a large set of differentiation-related genes shared with Bα′ cells[49] (Supplementary Fig. 6C,D); (v) GMPs share a substantial cohort of

© 2016 Macmillan Publishers Limited. All rights reserved

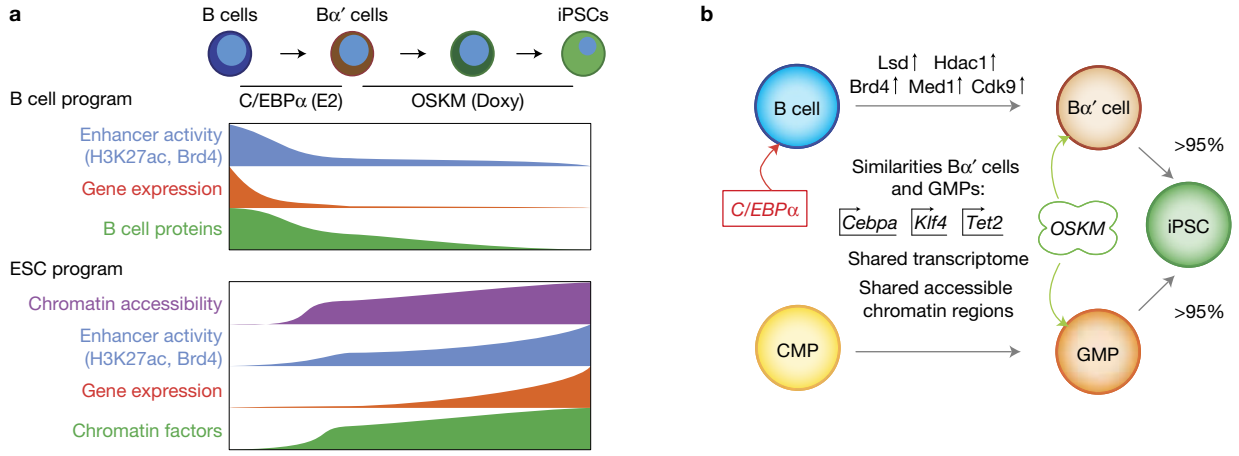


Figure 7 Summary diagrams. (**a**) A diagram summarizing the changes in enhancer activity, gene expression, proteome and chromatin accessibility described in this paper for both the B cell and the pluripotency gene expression program. (**b**) A diagram summarizing the main shared properties between Bα′ cells and GMPs, as elite cells for iPS reprogramming.

accessible chromatin regions (ATAC-seq cluster I) present in ESCs but not in B cells or MEFs.

Our data indicate that the interplay between C/EBPα and Klf4 induces chromatin accessibility to pluripotency factors as exemplified for the *Rarg* and *Lefty* enhancers where C/EBPα binds first followed by Klf4 (Fig. 5h). As described here and elsewhere[41], C/EBPα can act as a pioneer factor and Klf4 has also been shown to bind to closed chromatin[50]. However, it is possible that the predominant effect of C/EBPα is to activate *Klf4*, which subsequently binds to and activates other pluripotency genes (Fig. 5h).

Myeloid progenitors, including GMPs, also play an important role in blood cell malignancies because they have been shown to be key target cells in a mouse model for MLL–AF9-induced acute myeloid leukaemia[51] (AML). Strikingly, their oncogenic transformation strictly depends on the C/EBPα-driven differentiation to the GMP stage and/or on the transcriptional activity of C/EBPα (ref. 52), compatible with the notion that C/EBPα is responsible for chromatin remodelling events necessary for leukaemia[51–53]. It will now be interesting to determine whether the increase in the levels of chromatin-modifying factors described here also plays a role in AML formation, especially in view of the observation that human AML cells typically exhibit high levels of Lsd1 activity[54]. Our recent demonstration that C/EBPα can also poise human B cells for enhanced OSKM-induced reprogramming[55] underscores the factor's capacity to act across species.

Our findings have provided insights into the mechanism by which C/EBPα creates an elite cell state in B cells and the earliest events in reprogramming to iPSCs. It will now be interesting to determine whether upregulation of chromatin-related factors is a more general feature of somatic cell reprogramming, and whether this is also important for the GMP formation and their predisposition to malignancy. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank J. E. Bradner (Dana Farber Cancer Institute, Harvard Medical School, USA) for the JQ1 compound, J. Zuber (Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Austria) for the Brd4 shRNA construct, R. Levine (Memorial Sloan Kettering Cancer Center, USA) for the Tet2 shRNA construct, L. de Andres for help with GMP isolations, L. Batlle for help with the chimaeric mice, the Genomics Facility and the Biomolecular Screening & Protein Technologies Facility of the CRG for technical assistance and members of the Graf laboratory for discussions. This work was supported by Ministerio de Educacion y Ciencia, SAF.2012-37167, Fundacio La Marató TV3 120410, AGAUR SGR 1136 and European Research Council Synergy Grant (4D-Genome). R.S. was supported by an EMBO Long-term Fellowship (ALTF 1201-2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014). J.L.S. was supported by MINECO (IJCI-2014-21872). Work in the Porse laboratory was supported through a Centre grant from the NovoNordisk Foundation (Section for Stem Cell Biology in Human Disease and R179-A1513).

AUTHOR CONTRIBUTIONS

B.D.S. and T.G. conceived the study and wrote the manuscript. B.D.S. performed the cell culture, animal and molecular biology experiments and analysed the data. S.C., D.T. and M.F. performed the bioinformatics analysis, J.S.J. and B.P. the ChIP-seq experiments, J.L.S. and C.S.-M. the western blots and the protein immunoprecipitations, and A.L. the ATAC-seq experiments, M.W. and M.M. collected and analysed the proteomics data, R.S. performed and analysed the 4C-seq experiments; F.L. performed cell culture experiments.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://dx.doi.org/10.1038/ncb3326
Reprints and permissions information is available online at www.nature.com/reprints

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
3. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
4. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
5. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).
6. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
7. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
8. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).

© 2016 Macmillan Publishers Limited. All rights reserved

9. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
10. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
11. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
12. Di Micco, R. *et al.* Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* **9,** 234–247 (2014).
13. Liu, L. *et al.* Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15,** 574–588 (2014).
14. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* **14,** 427–439 (2013).
15. Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* **40,** 1478–1483 (2008).
16. Krizhanovsky, V. & Lowe, S. W. Stem cells: the promises and perils of p53. *Nature* **460,** 1085–1086 (2009).
17. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45,** 34–42 (2013).
18. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* **14,** 457–466 (2012).
19. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
20. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19,** 1718–1723 (2009).
21. Korpal, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.* **283,** 14910–14914 (2008).
22. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460,** 49–52 (2009).
23. Di Stefano, B. *et al.* C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506,** 235–239 (2014).
24. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).
25. Bar-Nur, O. *et al.* Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* **11,** 1170–1176 (2014).
26. Vidal, S. E., Amlani, B., Chen, T., Tsirigos, A. & Stadtfeld, M. Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Rep.* **3,** 574–584 (2014).
27. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat. Genet.* **41,** 968–976 (2009).
28. Guo, S. *et al.* Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156,** 649–662 (2014).
29. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
30. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).
31. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482,** 221–225 (2012).
32. Wang, G. L. *et al.* HDAC1 cooperates with C/EBPα in the inhibition of liver proliferation in old mice. *J. Biol. Chem.* **283,** 26169–26178 (2008).
33. Grebien, F. *et al.* Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. *Nat. Chem. Biol.* **11,** 571–578 (2015).
34. Adam, R. C. *et al.* Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521,** 366–370 (2015).
35. Northrup, D. L. & Allman, D. Transcriptional regulation of early B cell development. *Immunol. Res.* **42,** 106–117 (2008).
36. Laurent, B. *et al.* A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell* **57,** 957–970 (2015).
37. Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–334 (2013).
38. Roe, J. S., Mercan, F., Rivera, K., Pappin, D. J. & Vakoc, C. R. BET bromodomain inhibition suppresses the function of hematopoietic transcription factors in acute myeloid leukemia. *Mol. Cell* **58,** 1028–1039 (2015).
39. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).
40. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
41. van Oevelen, C. *et al.* C/EBPa activates pre-existing and *de novo* macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.* **5,** 232–247 (2015).
42. Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proc. Natl Acad. Sci. USA* **94,** 569–574 (1997).
43. Hasemann, M. S. *et al.* C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. *PLoS Genet.* **10,** e1004079 (2014).
44. Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33,** 769–774 (2015).
45. Chen, J. *et al.* Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* **45,** 1504–1509 (2013).
46. Canettieri, G. *et al.* Histone deacetylase and Cullin3-REN(KCTD11) ubiquitin ligase interplay regulates Hedgehog signalling through Gli acetylation. *Nat. Cell Biol.* **12,** 132–142 (2010).
47. Wu, Y. *et al.* The deubiquitinase USP28 stabilizes LSD1 and confers stem-cell-like traits to breast cancer cells. *Cell Rep.* **5,** 224–236 (2013).
48. Pijnappel, W. W. *et al.* A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495,** 516–519 (2013).
49. Di Tullio, A. *et al.* CCAAT/enhancer binding protein α (C/EBP(α))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108,** 17016–17021 (2011).
50. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).
51. Ye, M. *et al.* Hematopoietic differentiation is required for initiation of acute myeloid leukemia. *Cell Stem Cell* **17,** 611–623 (2015).
52. Ohlsson, E. *et al.* Initiation of MLL-rearranged AML is dependent on C/EBPα. *J. Exp. Med.* **211,** 5–13 (2014).
53. Krivtsov, A. V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27,** 852–860 (2013).
54. Harris, W. J. *et al.* The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell* **21,** 473–487 (2012).
55. Bueno, C. *et al.* Reprogramming human B-cells into induced pluripotent stem cells and its enhancement by C/EBPα. *Leukemia* http://dx.doi.org/10.1038/leu.2015.294 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved

DOI: 10.1038/ncb3326

METHODS

Mice. We used a cross between 'reprogrammable mice' containing a doxycycline-inducible OSKM cassette and tetracycline transactivator[56], and the Oct4–GFP reporter[57], as previously described[23] (Fig. 1a).

B cell isolation was routinely performed from 8- to 16-week-old male and female mice.

Mice were housed in standard cages under 12 h light–dark cycles and fed *ad libitum* with a standard chow diet. All experiments were approved by the Ethics Committee of the Barcelona Biomedical Research Park (PRBB) and performed according to Spanish and European legislation.

Cell cultures. ESCs (E14TG2) were cultured on gelatinized plates in N2B27 media (50% DMEM-F12, 50% Neurobasal medium, N2 (100×), B27 (50×)) supplemented with small-molecule inhibitors PD (1 μM, PD0325901) and CHIR (3 μM, CHIR99021), as well as LIF (10 ng ml^{-1}). CD19^{+} pre-B cells were obtained from bone marrow of reprogrammable mice with monoclonal antibody against CD19 (concentration of 0.2 μg per 10^6 cells; clone 1D3, BD Pharmingen, Cat. no. 553784) using MACS sorting (Miltenyi Biotech). Cell purity was confirmed by FACS using an LSR2 machine (BD) (>98%). After isolation, B cells were grown in RPMI medium supplemented with 10% FBS and 10 ng ml^{-1} IL-7 (Peprotech), 100× L-glutamine, 100× penicillin/streptomycin, 100× nonessential amino acids, 1,000× β-mercaptoethanol (Life Technologies). Lin^{-} c-Kit^{+} Sca-1^{-} CD16^{+}/CD32^{+} CD34^{+} GMP cells were isolated by FACS sorting using a BD INFLUX sorting machine and cultured in STEMSPAN medium (Stemcell Technologies) supplemented with 100 ng ml^{-1} SCF, 50 ng ml^{-1} IL3, 50 ng ml^{-1} Flt3L and 50 ng ml^{-1} mTPO (all from Peprotech). The following antibodies have been used for GMP isolations: eFluor780 anti-mouse c-Kit (0.5 μl for 5 × 10^6 cells; clone 2B8, eBioscience, Cat. no. 47-1171-80), PeCy7 anti-mouse Sca-1 (0.5 μl for 5 × 10^6 cells; clone D7, eBioscience, Cat. no. 25-5981-81), Alexa Fluor 647 rat anti-mouse CD34 (2 μl for 5 × 10^6 cells; clone RAM34, BD Pharmingen, Cat. no. 560230), anti-mouse CD16/CD32 FITC (1 μl for 5 × 10^6 cells; clone 93, eBioscience, Cat. no. 11-0161-82) and the lineage depletion kit (Miltenyi, Cat. no. 130-092-613). Mouse embryonic fibroblasts were isolated from E13.5 embryos of reprogrammable mice crossed with Oct4–GFP mice and cultured in DMEM supplemented with 10% FBS, 100× L-glutamine and 100× penicillin/streptomycin.

Reprogramming experiments. Pre-B cells were isolated from the bone marrow of reprogrammable mice crossed with Oct4–GFP mice, infected with C/EBPαER–hCD4 retrovirus, plated at 500 cells cm^{-2} in gelatinized plates (12 wells) on irradiated MEF feeders in RPMI medium. To activate C/EBPα, cultures were treated for 18 h with 100 nM β-oestradiol (E2), resulting in Bα′ cells.

After E2 washout, the cultures were switched to serum-free N2B27 medium supplemented with IL-4 10 ng ml^{-1}, IL-7 10 ng ml^{-1} and IL-15 2 ng ml^{-1}. To activate OSKM, the cultures were treated with 2 μg ml^{-1} of doxycycline. From day 2, onwards we supplemented the N2B27 medium with 20% KSR (Life Technologies), 3 μM CHIR99021 and 1 μM PD0325901.

OSKM Oct4–GFP MEFs were seeded in gelatinized plates and induced with 2 μg ml^{-1} of doxycycline in ES medium.

A step-by-step protocol describing the reprogramming procedure can be found at Nature Protocol Exchange[58]. All cell lines have been routinely tested for mycoplasma contamination.

Stable cell lines. The PlatE retroviral packaging cell line was obtained from Cell Biolabs (Cat. no. RV-101). The C10 pre-B cell line stably expressing C/EBPαER–GFP has been described previously[59].

Immunofluorescence. For Nanog staining, the cells were fixed with 4% paraformaldehyde, blocked and incubated with the primary antibody (1:500; Calbiochem, SC100) overnight at 4 °C. They were then stained with secondary anti-rabbit antibodies (1:1,000; Alexa Fluor 547, Life Technologies) at room temperature for one hour. Nuclear staining was performed with DAPI (Invitrogen).

RNA extraction. To remove the feeders, cells were trypsinized and pre-plated for 30 min before RNA isolation with the miRNeasy mini kit (Qiagen). RNA was eluted from the columns using RNase-free water and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystems).

FACS analyses. Oct4–GFP expression was analysed with an LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar).

Isolation of fast- and slow-cycling cells. CFSE (Life Technologies) was used to subdivide GMPs into fast- and slow-cycling cell populations. 3 × 10^5 cells were resuspended in 1 ml PBS with 0.2 μM of CSFE for 5 min and then rinsed twice with four volumes of PBS. After 24 h, cells were sorted on the basis of FITC brightness as indicated in Supplementary Fig. 8A.

Gene expression arrays. RNA samples with a RIN > 9 were subjected to transcriptional analyses using Agilent expression arrays. For hybridization, 500 ng of total RNAs were labelled using Agilent's QuickAmp labelling kit and hybridized to Agilent 8X60K expression arrays.

qRT-PCR analyses. qRT-PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystems) and primers as listed in Supplementary Table 3. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C.

Small-molecule inhibitors. Lsd1 inhibition was achieved by treating the cells with 50 μM S2101 (Calbiochem). Brd4 was inhibited with 100 nM JQ1 obtained from J. E. Bradner. Hdac1 inhibition was achieved by culturing the cells in the presence of 1 mM of valproic acid.

BrdU staining. To check cell proliferation, BrdU was added to the culture medium at a concentration of 10 μM for 6 h. Staining was performed using the BrdU APC Flow Kit from BD Pharmingen.

Cell viability. Cell viability was assessed using the Pacific Blue Annexin V/SYTOX AADvanced Apoptosis Kit for flow cytometry from Thermo Fisher Scientific.

Luciferase assay. The −90 kb and −280 kb enhancers of *Klf4* were amplified by qPCR using the following primers: −90 kb forward 5′-GCCCTCGAGCGGGTCTGGCCT TCAGTGATA-3′ and reverse 5′-GCCATGCATCCGGACTCCCTTTTGCTAGTG -3′; −280 kb forward 5′-GCCCTCGAGCTGGTATATGCACACATGCAC-3′ and reverse 5′-GCCCTGCAGCTCTCCCTGCATTGGCTTAGT-3′.

The PCR fragments were digested using either XhoI and PstI or XhoI and NsiI and cloned in the CSI–LUC2–minP vector in the XhoI and PstI sites.

Twelve-well plates with 293T cells were transfected with 0.1 μg of a β-galactosidase control plasmid, 0.5 μg of the respective enhancer plasmid or the empty vector and 0.5 μg of C/EBPα, or 0.5 μg of C/EBPα and PU.1 plasmids or 0.5 μg of GFP plasmid using the TransIT transfection reagent (Mirus). Luciferase and β-galactosidase activities were measured using the Luciferase Assay System (Promega) and the β-Galactosidase Enzyme Assay System (Promega) following the manufacturer's protocol. All experiments were carried out in triplicates and normalized to β-galactosidase activity to account for differences in transfection efficiencies.

Vectors and virus production and infection. The retroviral vectors LMN–Ctrl-shRNA–PGK–GFP and LMN–Lsd1-shRNA–PGK–GFP were purchased from Transomic Technologies. The retroviral vector containing the Brd4 shRNA (no. 552) has been described previously[60]. The pMX vectors of TAF4, TAF5, TAF6 and TBP (ref. 48) were obtained from Addgene. The retroviral vector Tet2-shRNA–PGK–GFP has been described previously[61]. The C/EBPαER–hCD4 retroviral vector has been used before[62]. Viral production and B cell infection were performed as described previously[62]. For the infection of GMPs, freshly sorted cells were infected by centrifugation for 90 min on Retronectin (Clontech)-coated plates.

4C. 4C-seq was performed as described previously[63]. Briefly, 0.5–1.0 million crosslinked nuclei were digested with Csp6I followed by ligation under dilute conditions. After decrosslinking and DNA purification, samples were digested overnight with DpnII and once more ligated under dilute conditions. Column-purified DNA was directly used as input for inverse PCR using primers (Supplementary Table 3) with Illumina adapter sequences as overhangs. Several PCR reactions were pooled, purified and sequenced on an Illumina HiSeq 2500.

RNA-seq. Libraries were prepared using the SMARTer Universal Low Input RNA Kit (Clontech Laboratories) according to the manufacturer's protocol. The libraries were analysed using Agilent DNA 1000 chip to determine the quantity and size distribution and then quantified by qPCR using the KAPA Library Quantification Kit (ref. KK4835, KapaBiosystems) before amplification with Illumina's cBot. Libraries were loaded at a concentration of 10 pM onto the flow cell and sequenced on Illumina's HiSeq 2000.

Chromatin immunoprecipitation (ChIP). The H3K4me2 ChIPs were performed using the True MicroChIP kit from Diagenode following the manufacturer instructions. Low-cell-number ChIP-seq experiments were performed essentially as previously described[64], with varying RIPA-buffer washes. Briefly, cells were fixed in 1% formaldehyde in 50% PBS/50% culture medium, with rotation for 10 min at room temperature. For Brd4 ChIP, cells were fixed with 2 mM DSG in PBS for

© 2016 Macmillan Publishers Limited. All rights reserved

20 min, followed by addition of 1% formaldehyde, 10 min, both steps with rotation at room temperature. Chromatin was sheared to 200-bp fragments using a Bioruptor sonicator and 0.5 ml siliconized tubes. H3K27ac-carrying histones were precipitated using four low-salt RIPA-buffer washes and protein A Sepharose beads, C/EBPα and Brd4 using six low-salt RIPA-buffer washes and protein G and protein A beads respectively, and Klf4 with two low- and two high-salt RIPA washes and protein G beads. See below for the references of the antibodies used.

Protein fractionation. Gel-filtration experiments were performed in the Biomolecular Screening & Protein Technologies Facility using a Superose 6 10/300 GL column (GE Healthcare). For analysis of endogenous Lsd1–Hdac1–C/EBPα complex, Bα′ cells were lysed in RIPA buffer, and 1 mg of cleared lysate was loaded onto a single Superose 6 10/300 GL column, calibrated using a mixture of molecular mass marker proteins (MWGF1000, Sigma-Aldrich). A 300 µl portion of lysate was loaded onto the column and collected into 250 µl fractions; fractions were processed for western blot analysis.

Co-immunoprecipitation experiments and western blots. Owing to the large number of cells required, the immunoprecipitation experiments were performed in an inducible, C/EBPαER-containing C10 pre-B cell line[59]; the western blots were performed with primary cells. Cell lysates were prepared using NET-2 buffer (200 mM NaCl, 50 mM Tris pH 7,5, 0.1% Triton and 1× protease inhibitors) for 20 min at 4 °C and centrifugated for 10 min at 16,000*g* to remove cellular debris. Supernatants were used for protein detection by western blot or for co-immunoprecipitation. For the latter 19/20 of the extract was pre-cleared with magnetic beads (Dynabeads, Invitrogen) by rotating for 2 h on a wheel at 4 °C. One-twentieth of the lysates was kept as the input. Separately, to crosslink the antibodies to beads, 50 µl of beads were incubated with 3 µg of antibody or IgG for 2 h at room temperature on a wheel and then washed once with PBS and twice with 0.2 M triethanolamine pH 8.2. The suspension was then incubated for 20 min at room temperature with 20 mM dimethyl pimelimidate, the reaction stopped with two 5 min washes with 50 mM Tris pH 8 followed by three washes with PBS and the beads incubated with 0.1 M citric acid pH 3 for 2 min to remove non-crosslinked antibodies followed by two washes with lysis buffer. Finally, the crosslinked antibodies were incubated with the pre-cleared protein extracts overnight at 4 °C on a wheel and the beads selected with a magnet, 1/20th of the supernatant kept as the unbound control fraction and the beads washed six times with lysis buffer. To elute, the beads were resuspended in 1× Laemmli buffer (without βME) and heated for 20 min at 60 °C. Magnetic beads were then separated with a magnet, the supernatant complemented with 5% βME and boiled before loading in an SDS–PAGE acrylamide gel to analyse bound proteins by western blotting. As controls, one lane was loaded with the original input, another lane with the unbound fraction and a third lane with beads coupled to either mouse or rabbit IgG. Finally, after transferring the proteins to a membrane, the blot was blocked with 5% milk in TBS–Tween, probed with an antibody to the homologous protein, and after developing stripped and probed with an antibody against the antigen tested for interaction.

C/EBPα immunoprecipitation for mass spectrometry. For affinity purification experiments of C/EBPα, we used our inducible, C/EBPαER-containing C10 pre-B cell line[59]. Immunoprecipitations were performed using three separate samples with antibody (Santa Cruz, SC-61) and three samples with IgG essentially as above but without pre-clearing and without crosslinking the antibodies to the beads. After overnight incubation with extracts, beads were washed three times with NET-1.5 buffer (150 mM NaCl, 50 mM Tris pH 7.5, 0.1% Triton) and twice with TBS. Precipitated proteins were eluted with elution buffer (6 M urea, 2 M thiourea, 10 mM Hepes pH 8.0) containing 2 mM dithiothreitol for 30 min at room temperature, followed by a second elution with elution buffer. Both elutes were combined, alkylated with 5 mM CAA, diluted 5 times with 50 mM ammonium bicarbonate and digested with 1 µg trypsin overnight at room temperature. The resulting peptide mixtures were acidified with 1% TFA, desalted with StageTips containing three layers of C18 material and analysed by mass spectrometry as described below.

Antibodies for ChIPs, western blots and immunoprecipitations. The antibodies used in this study were H3K4me2 (2 µg for 25 µg of chromatin) (Abcam, ab7766), Lsd1 (1:1,000) (Kdm1a, Abcam, ab17721), Brd4 (ChIP: 2.5 µg ml^{-1}) (Bethyl Laboratories, A301-985A); (western blot: 1:1,000) (Abcam, ab128874), H3K27ac (1 µg ml^{-1}) (Abcam, ab4729), Klf4 (1:200) (R&D, AF3158), Hdac1 (1:2,000) (Abcam, ab7028), C/EBPα (western blot: 1:300) (Santa Cruz, SC-61); (ChIP: 0.2 µg ml^{-1}) (Santa Cruz, SC-61), PU.1 (0.5 µg ml^{-1}) (Santa Cruz, SC-352X), ERα antibody (1:500) (Santa Cruz, SC-543), H3 (1:2,000) (Abcam, ab1791), Parp (1:1,000) (no. 9542, Cell Signaling), β-tubulin (1:5,000) (Sigma, T7816), GAPDH (1:1,000) (Santa Cruz, SC-32233), PCNA (1:400) (Santa Cruz, SC-56 (PC10)), Cdk9 (1:1,000) (Cell Signaling, rabbit mAb no. 2316), Oct4 (1:200) (Santa Cruz, SC-8628), Gdf3 (1.5 µg ml^{-1}) (R&D, AF958), Lin28a (1:1,000) (Cell Signaling, no. 8641), Sall4 (1 µg ml^{-1}) (Abcam, ab57577), Nanog (1:500) (Calbiochem, SC1000), Sox2 (1:1,000) (Santa Cruz, SC-17320), Tcfcp2l1 (1 µg ml^{-1}) (R&D, AF5726), hCD4 (0.5 µl per 10^6 cells) (clone RPA-T4, eBioscience, 13-0049-80), β-catenin (1:1,000) (BD Biosciences, 610153).

Chimaeric mice. For the chimaera formation assay, 10–15 iPS cells (Agouti colour coat) were injected into 3.5 dpc blastocysts of C57BL/6-Albino mice (white coat colour). Chimaerism of the transplanted offspring was assessed by the presence of agouti coat colour derived from the iPS cells.

Whole-cell proteomic analysis. Sample preparation for mass spectrometry was performed essentially as described previously[65]. Cells were lysed in lysis buffer (6 M guanidinium chloride, 10 mM TCEP), incubated for 10 min at 95 °C and sonicated for 15 min using a Bioruptor. Proteins were alkylated with 40 mM 2-chloroacetamide, diluted tenfold with digestion buffer (20 mM Tris-HCl, pH 8.5 and 10% ACN) and digested with 1:50 (w/w) Lys-C (Wako) and 1:50 (w/w) trypsin (Promega) at 37 °C overnight. The resulting peptide mixture was acidified by addition of 1% TFA and desalted on StageTips with three layers of SDB-RPS. Peptides were separated on 50-cm columns of ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH) packed in-house. Liquid chromatography was performed on an EASY-nLC 1000 ultrahigh-pressure system coupled through a nanoelectrospray source to a Q-Exactive mass spectrometer (all from Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1% formic acid) and separated applying a nonlinear gradient of 5–60% buffer B (0.1% formic acid, 80% acetonitrile) at a flow rate of 250 nl min^{-1} over 240 min. Data acquisition switched between a full scan and five data-dependent MS/MS scans. Multiple sequencing of peptides was minimized by excluding the selected peptide candidates for 45 s. For AP-MS experiments, a linear gradient of 5–30% buffer B was applied over 70 min, with dynamic exclusion set to 15 s.

ATAC-seq. ATAC-seq was performed as previously described[40]. A total of 100,000 cells were washed once with 100 µl PBS and resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.2% IGEPAL CA-630). The suspension of nuclei was then centrifuged for 10 min at 500*g* at 4 °C, followed by the addition of 50 µl transposition reaction mix (25 µl TD buffer, 2.5 µl Tn5 transposase and 22.5 µl nuclease-free H_2O) and incubation at 37 °C for 45 min. DNA was isolated using the MinElute Kit (Qiagen). Library amplification was done by two sequential PCR reactions (8 and 5 cycles, respectively). After the first PCR reaction, the library was selected for fragments below 700 bp with AmpureXP beads followed by a second PCR reaction. Libraries were purified with Qiaquick PCR (Qiagen) and integrity checked on a Bioanalyser before sequencing.

Bioinformatic analyses. All sequencing data were mapped onto the mouse genome assembly mm10 (Ensembl GRCm38.78) using STAR (-outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06; for RNA-seq: -sjdbOverhang 100 –outFilterType BySJout -alignSJoverhangMin 8 -alignSJDBoverhangMin 1 -alignIntronMin 20 -alignIntronMax 1000000 -alignMatesGapMax 1000000; for genomic sequencing: -alignIntronMax 1 -alignEndsType EndToEnd -alignMatesGapMax 2000) and analysed with R (3.1.0) using packages from the bioconductor suite (v3.0). For peak calling, regions overlapping the 'Encode blacklist' were removed. All clusterings were performed using the Rpackage Mfuzz (2.26.0). All GO enrichment analyses were performed using the Rpackage ReactomePA (1.10.1).

For RNA-seq, mapping was performed using Ensembl annotation (GRCm38.78). Genes expression quantification was performed using the Rpackage Rsubread (1.16.1). Sample scaling and statistical analysis were performed using the Rpackage DESeq2 (1.6.3) and vsd counts were used for further analysis. Genes changing significantly at any time point were identified using the nbinomLRT test (FDR < 0.01) and for >2-fold change between at least two time points (average of replicates, vsd values). Genes upregulated between two time points were selected using the nbinomWaldTest (FDR < 0.01) and for >1.5-fold change.

For ATAC-seq, duplicates reads were removed using Picard (http://picard.sourceforge.net) (MarkDuplicates, REMOVE_DUPLICATES=true). Bigwig tracks were made using DeepTools BamCoverage (1.5.9.1). Peak calling was performed using macs2 (2.1.0.20140616) (-f BAMPE -g mm -p 0.001 -nolambda). For quantitative analyses, peaks from all time points were merged as one set of non-overlapping regions using Bedops. Reads were counted on merged regions for each time point, using the Rpackage csaw (1.0.7). Counts were scaled on genome-wide fragments using the Rpackage csaw (function WindowCounts on 10 kb windows), and size factors were calculated using DESeq2. Scaled counts were transformed to a $\log_2$ scale (with pseudo-count of 1). Regions becoming more accessible in Bα′ cells were identified as showing a >1.5-fold increase after pulse and a difference of more than 10 fragments, and a difference of >2-fold changes and >40 fragments between any of the time points. Association of peaks with genes was performed with the Rpackage ChipPeakAnno, using Ensembl transcripts (GRCm38.78).

© 2016 Macmillan Publishers Limited. All rights reserved

DOI: 10.1038/ncb3326

Motif analyses were performed using RSAT. Repeats from the Ensembl annotation (GRCm38.75) were masked using bedtool maskfasta. Motif discovery was performed using peaks-motif and compared with the databases JASPAR (v.2015.03) and HOCOMOCO (v9). Motif enrichment in ATAC-seq clusters was performed with matrixQuality. We used as background a Markov model of order 1. The maximum-normalized-weight-difference score (MNWD) of matrix quality was used as enrichment score.

For the ChIP-seq, duplicate reads were removed using picard. Bigwig tracks were made using DeepTools BamCompare to subtract the input from the ChIP (-scaleFactorsMethod SES -ratio subtract -fragmentLength 200). H3K27ac peak calling was performed using macs2 on immunoprecipitation versus input (-broad -q 0.01, -broad-cutoff 0.01). Super-enhancers were called using ROSE (ref. 66) on H3K27ac data (-t 2500). Quantitative analyses were performed as for ATAC-seq. Average plots were obtained using DeepTools computeMatrix and profilers.

For proteomic, raw mass-spectrometry data were analysed with the MaxQuant workflow (1.5.1.6 and 1.5.3.29). Peak lists were searched against the Uniprot mouse FASTA database (2013_05) combined with 262 common contaminants by the integrated Andromeda search engine. FDR was set to 1% for both peptides (minimum length of 7 amino acids) and proteins. 'Match between runs' (MBR) with a maximum time difference of 0.7 min was enabled. Relative protein amounts were determined by the MaxLFQ algorithm, with a minimum ratio count of two peptides. Missing values were input from a normal distribution using the Perseus software package (width = 0.2, downshift = 1.8 s.d.).

For AP-MS, Student's *t*-test was applied for detection of differentially enriched proteins between triplicate immunoprecipitations of specific antibody and control IgG. For total protein analysis, proteins changing significantly at any time point were identified using an LRT test (FDR < 0.1) from the Rpackage msmsTests. Proteins changing between two time points were selected for 1.5>-fold change between each replicate (all four comparisons).

For the 4C analysis, the sequence of the reading primer was trimmed from the 5′ of reads using the demultiplex.py script from the R package fourCseq (version 1.0.0, allowing 4 mismatches). Reads in which this sequence could not be found were discarded. Reads were mapped using STAR and processed using fourCseq to filter out reads not located at the end of a valid fragment and to count reads per fragment. Tracks of signal were made after smoothing the RPKM counts per fragment with a running mean over five fragments.

Independent component analysis (ICA) and canonical correlation analysis (CCA) were performed on gene expression data using the R packages fastICA and CCA, respectively.

All codes for bioinformatic analysis are available on request.

Accession numbers. Microarray data obtained after reprogramming of B cells in FBS+LIF conditions[23]: GSE52397. H3K27ac ChIP-seq in ESC (ref. 67): GSE62380; Oct4, Nanog and Sox2 (ref. 66): GSE44286; Klf4 (ref. 68): GSE11431; Brd4: GSE36561. C/EBPα in GMPs: GSE43007 (ref. 43). ATAC-seq in MEFs (ref. 44): GSE67298. RNA-seq in different haematopoietic cell types[69]: GSE60101.

The RNA-seq, ChIP-seq and ATAC-seq data obtained for B cells, Bα′ cells, day 1 cells, day 2 cells and ESCs described in this paper have been deposited in GEO under the number GSE71218. Proteomic data have been deposited in proteome Xchange under the identifier PXD002769.

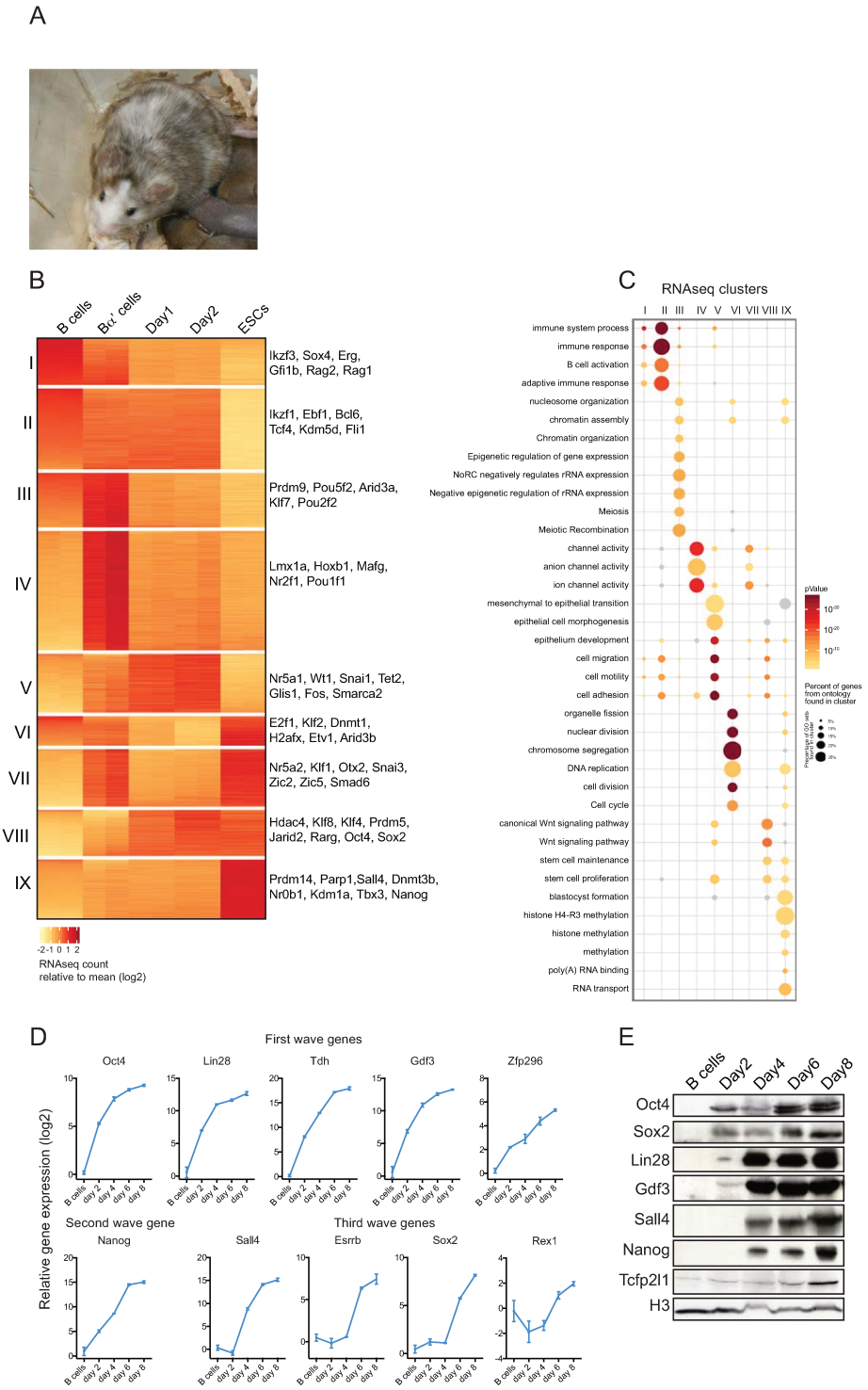
Statistics and reproducibility. Western blot, immunoprecipitation, ChIP, qRT-PCR, immunofluorescence, immunoprecipitation mass spectrometry, alkaline phosphate and FACS data presented are representative of at least three independent experiments that yielded similar results; microarray, RNA-seq and proteomic data were obtained from two independent experiments that yielded similar results. Statistical analyses were performed using Prism software (GraphPad). No statistical method was used to predetermine sample size and the experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

56. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* **7,** 56–59 (2010).
57. Boiani, M., Eckardt, S., Scholer, H. R. & McLaughlin, K. J. Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev.* **16,** 1209–1219 (2002).
58. Di Stefano, B. & Graf, T. Rapid generation of induced pluripotent stem cells from mouse pre-B cells. *Nat. Protoc. Exch.* http://dx.doi.org/10.1038/protex.2016.001 (2016).
59. Bussmann, L. H. *et al.* A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* **5,** 554–566 (2009).
60. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478,** 524–528 (2011).
61. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18,** 553–567 (2010).
62. Di Stefano, B. & Graf, T. Very rapid and efficient generation of induced pluripotent stem cells from mouse pre-B cells. *Methods Mol. Biol.* **1357,** 45–56 (2014).
63. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8,** 509–524 (2013).
64. Jakobsen, J. S. *et al.* Amplification of pico-scale DNA mediated by bacterial carrier DNA for small-cell-number transcription factor ChIP-seq. *BMC Genomics* **16,** 46 (2015).
65. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324 (2014).
66. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
67. Ji, X. *et al.* Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl Acad. Sci. USA* **112,** 3841–3846 (2015).
68. Chen, X., Xu, H., Yuan, P., Fang, F. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
69. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345,** 943–949 (2014).

© 2016 Macmillan Publishers Limited. All rights reserved

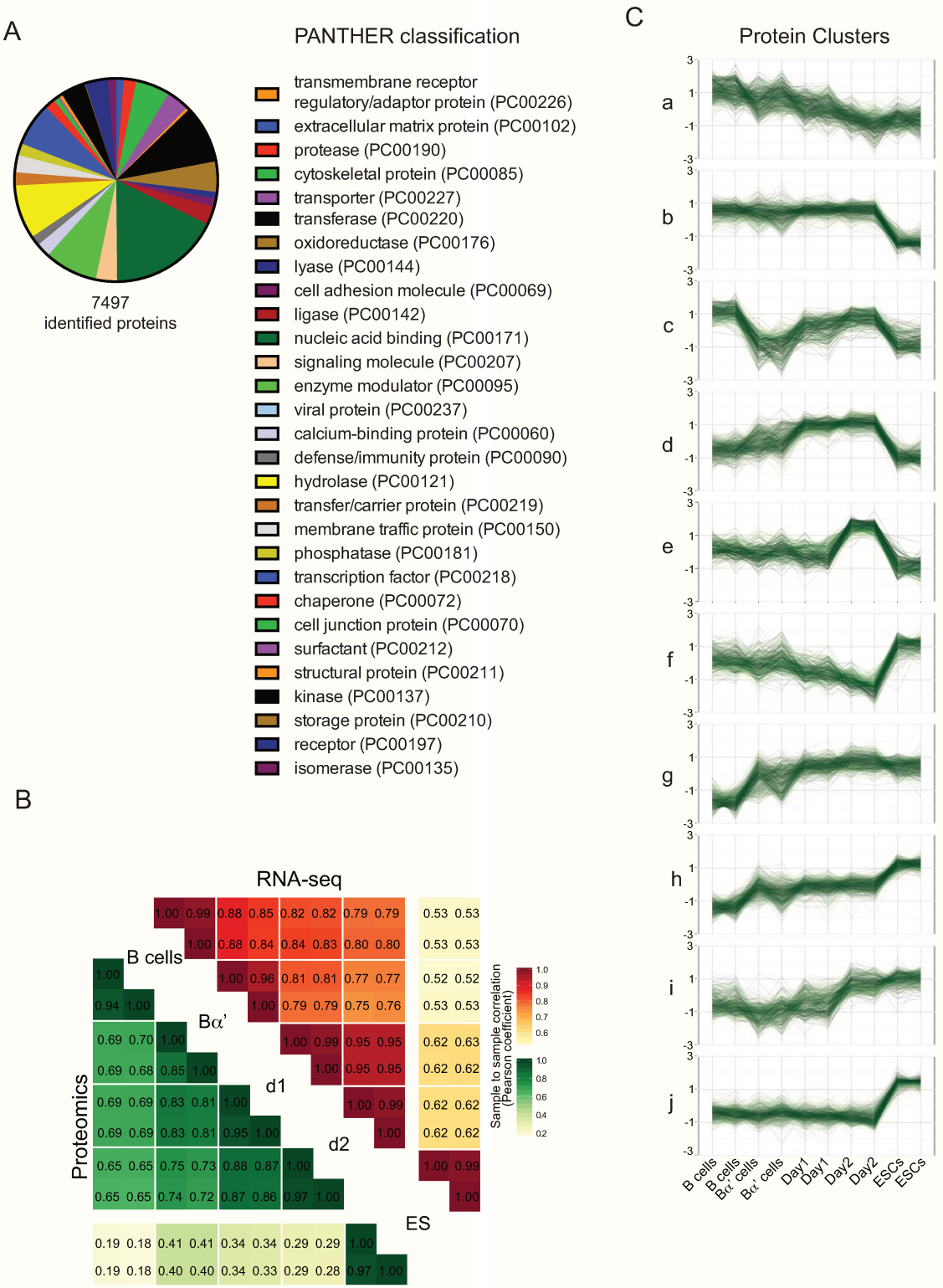
DOI: 10.1038/ncb3326

nature cell biology



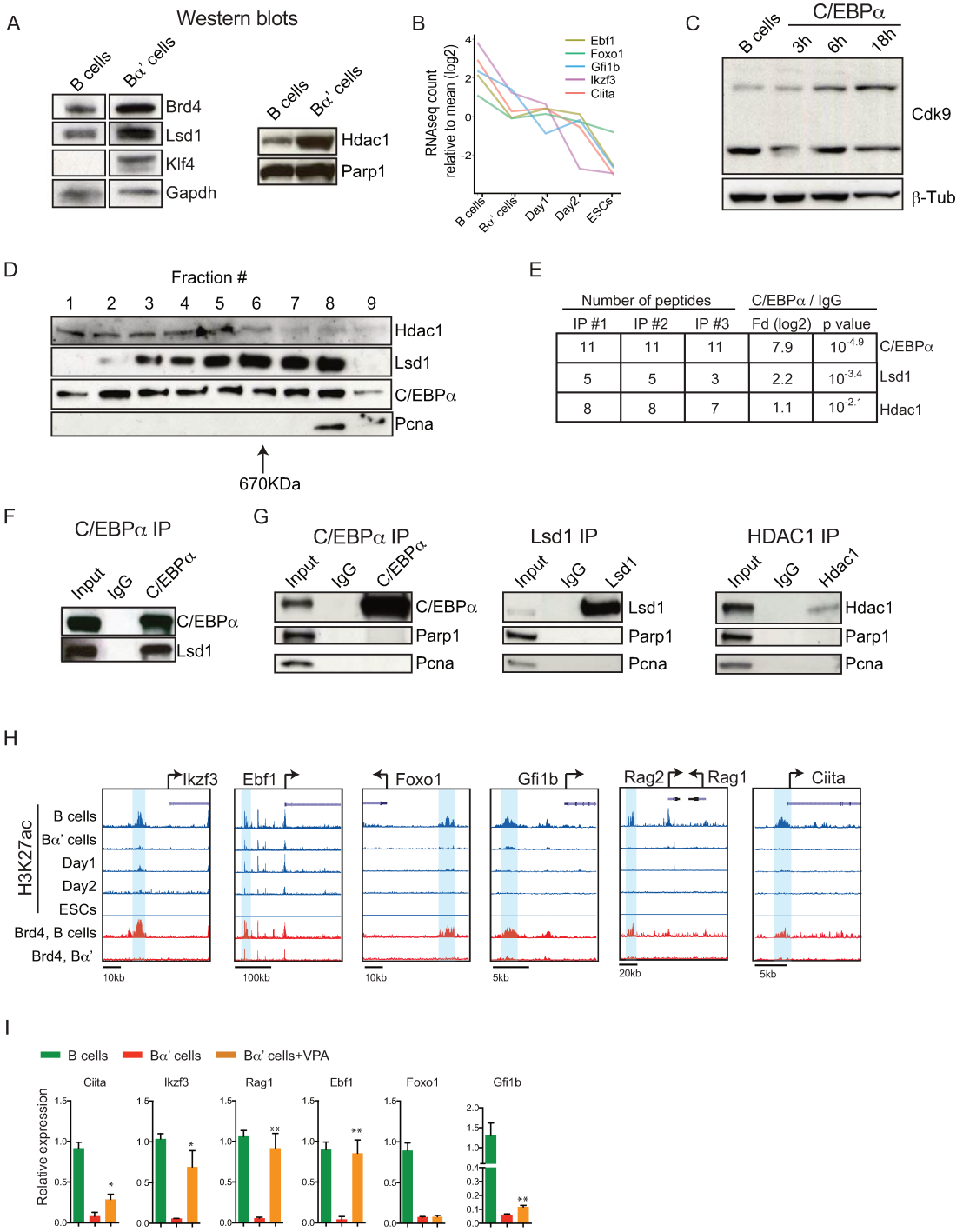
Supplementary Figure 1 Characterization of Bα' cell reprogramming into iPS cells. (A) Representative chimeric mouse obtained after blastocyst injection of aiPS clone. (B) Heatmap of RNA-seq data showing genes changing >2fold during reprogramming (FDR<1%, LRT test). (C) Gene Ontology (GO) analysis of protein clusters shown in panel A. The size of each circle represents the proportion of GO sets found in each cluster; the intensity of the color represents the *P*-value, determined by a hypergeometric test. (D) Gene expression (qRT-PCR) of selected pluripotency genes. Values were normalized against *Pgk* expression. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative western blots for selected pluripotency transcription factors. See Suppl. Fig. 8 for uncut gel images.

© 2016 Macmillan Publishers Limited. All rights reserved



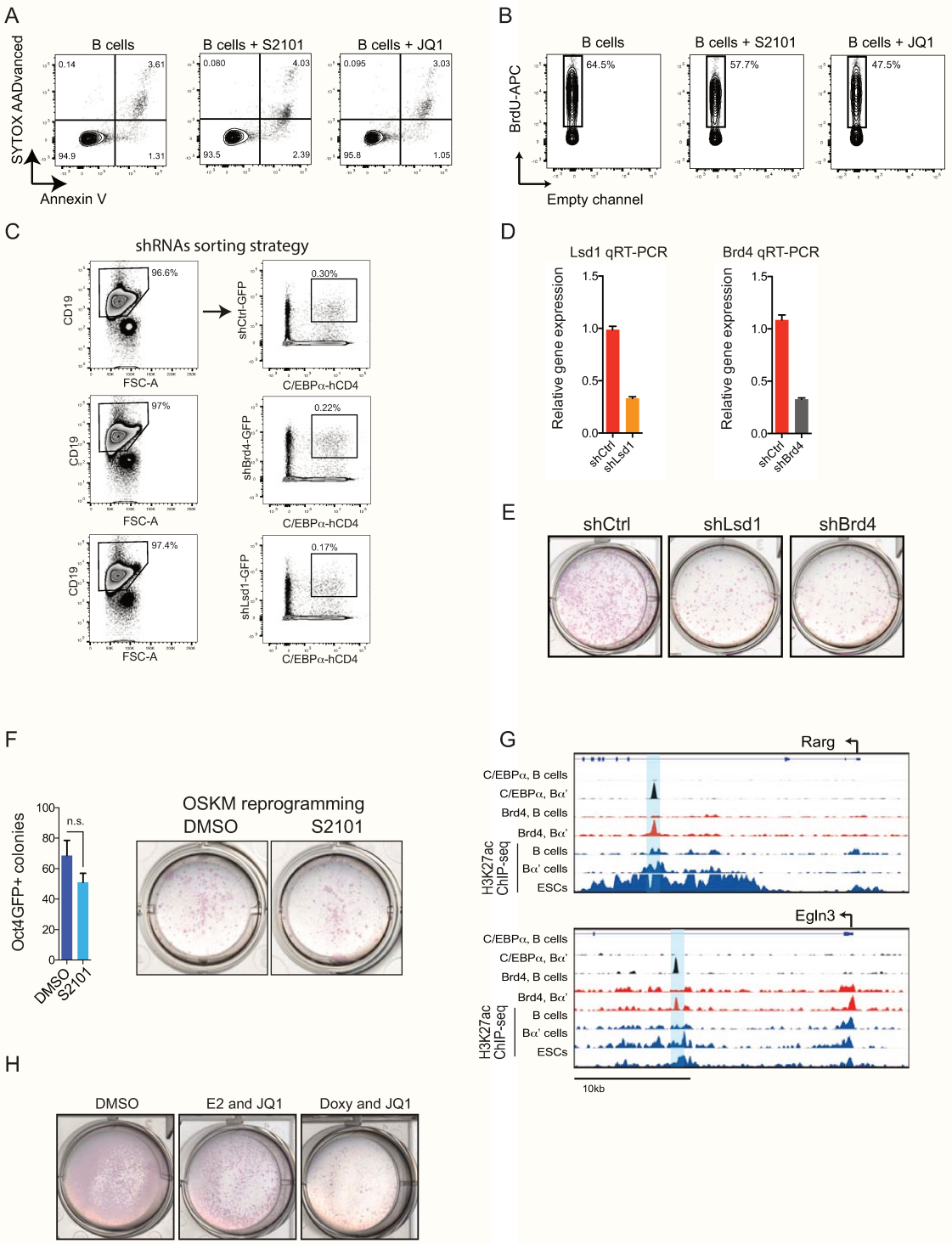
Supplementary Figure 2 Protein dynamics during reprogramming. (A) PANTHER classification for all the proteins identified by mass spectrometry in the samples tested. (B) Correlation between biological duplicates of RNA-seq and proteomic data. (C) C-means clustering of proteins changing >2 fold at any time points during reprogramming.

© 2016 Macmillan Publishers Limited. All rights reserved



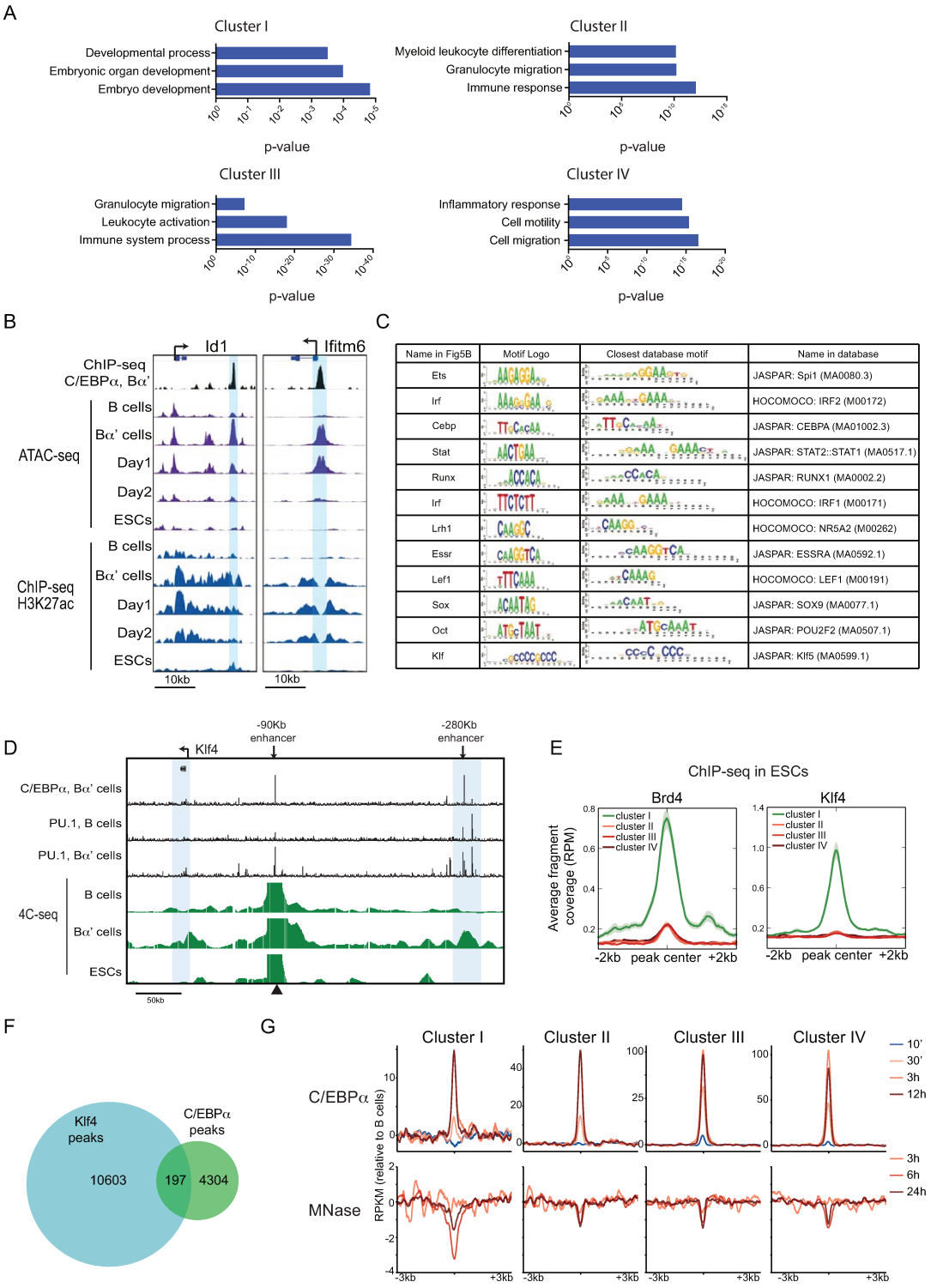
Supplementary Figure 3 Gene silencing induced by C/EBPα, protein interactions and B cell specific gene enhancer activities during reprogramming. (A) Representative western blots of Brd4, Lsd1, Klf4 and Hdac1 in B and Bα' cells. See Suppl. Fig. 8 for uncut gel images. (B) RNA-seq expression values for selected B cell specific genes. The data represent the average from two biologically independent samples. (C) Western blots of Cdk9 after induction of C/EBPα in B cells. See Suppl. Fig. 8 for uncut gel images. (D) Bα' cell extracts were fractionated on Superose 6 10/300 GL column and Hdac1, Lsd1 and C/EBPα were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (E) Peptide counts, P-value and enrichment over IgG of C/EBPα, Hdac1 and Lsd1, for the IP-mass spectrometry shown in Fig. 3B. (F) C/EBPα co-immunoprecipitation experiment. Lsd1 or C/EBPα were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (G) Co-immunoprecipitation of C/EBPα, Lsd1 and Hdac1. Parp1 and Pcna (negative controls) were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (H) Screenshots of H3K27ac histone decoration and Brd4 binding by ChIP-seq at enhancers of selected B cell transcription factors. (I) Gene expression of selected B cell genes as measured by qRT-PCR in B cells (data from Fig. 3F), B cells treated for 18h with E2 (Bα' cells) and B cells treated for 18h with both E2 and the Hdac1 inhibitor VPA. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (*P<0.05, **P<0.01).

© 2016 Macmillan Publishers Limited. All rights reserved



Supplementary Figure 4 Effect of Lsd1 and Brd4 inhibitions on iPS reprogramming. (A) Representative flow cytometry analysis of B cells treated with JQ1 or S2101 for 24 hours using Pacific Blue™ Annexin V/SYTOX® AADvanced™ Apoptosis Kit. (B) Representative BrdU (6h pulse) FACS staining of B cells treated with JQ1 or S2101 or DMSO as a control. (C) shRNA sorting strategy. (D) Gene expression by qRT-PCR of *Lsd1* and *Brd4* after specific knockdown in B cells. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative alkaline phosphatase positive iPS colonies obtained from reprogramming of B cells after Lsd1 and Brd4 knockdown. (F) Oct4-GFP and alkaline phosphatase positive iPS colonies obtained from reprogramming of B cells (OSKM alone without C/EBPα pulse) treated with S2101 or DMSO as a control. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (n.s. P>0.05). (G) Genome browser screenshots of *Rarg* and *Egln3* loci showing C/EBPα, Brd4 and H3K27ac ChIP-seq data. (H) Representative alkaline phosphatase positive iPS colonies obtained from reprogramming of Bα' cells induced with OSKM and treated with JQ1 during C/EBPα (E2) or OSKM (Doxy) induction.

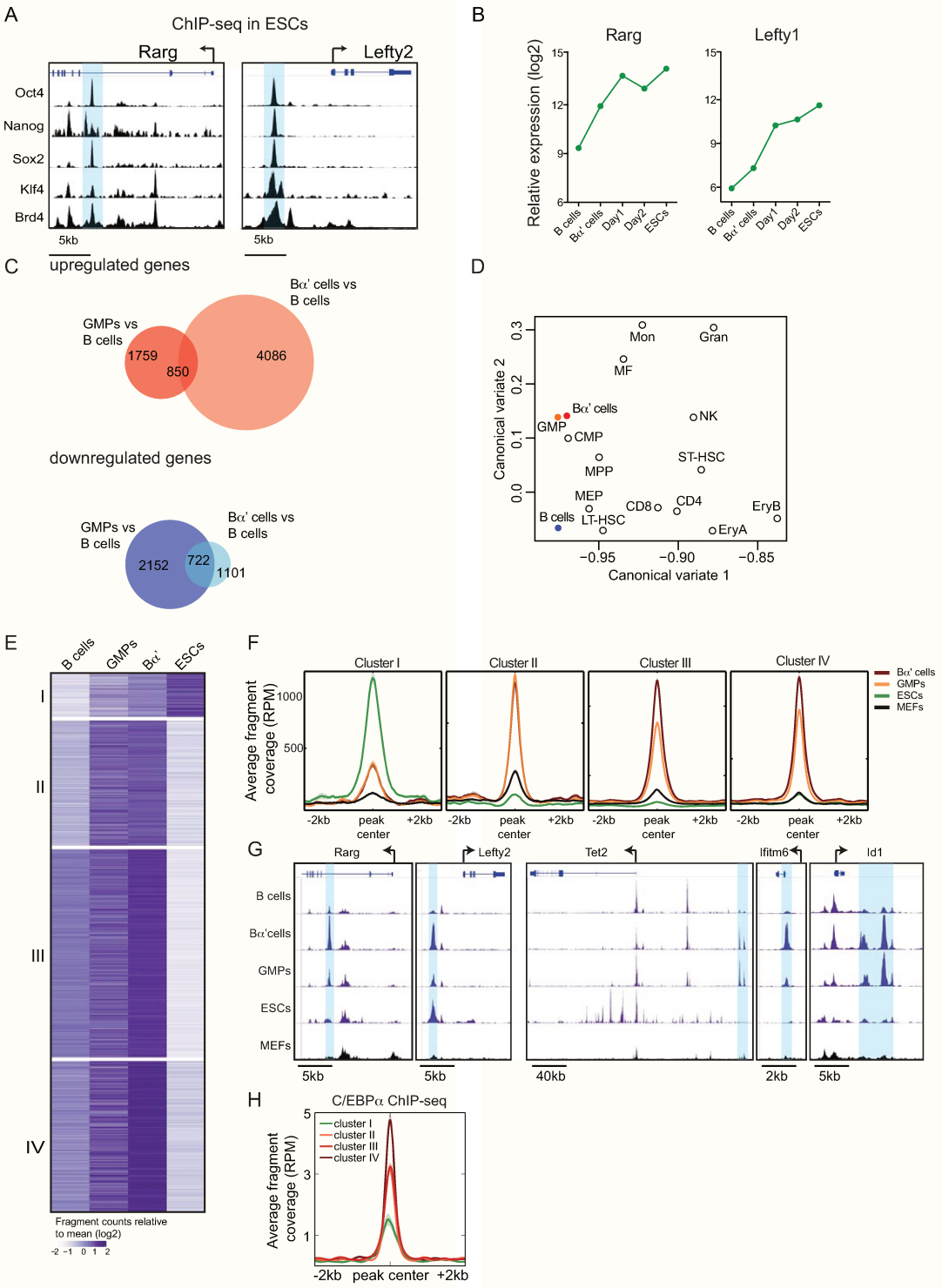
© 2016 Macmillan Publishers Limited. All rights reserved



Name in Fig5B	Motif Logo	Closest database motif	Name in database
Ets			JASPAR: Spi1 (MA0080.3)
Irf			HOCOMOCO: IRF2 (M00172)
Cebp			JASPAR: CEBPA (MA01002.3)
Stat			JASPAR: STAT2::STAT1 (MA0517.1)
Runx			JASPAR: RUNX1 (MA0002.2)
Irf			HOCOMOCO: IRF1 (M00171)
Lrh1			HOCOMOCO: NR5A2 (M00262)
Essr			JASPAR: ESSRA (MA0592.1)
Lef1			HOCOMOCO: LEF1 (M00191)
Sox			JASPAR: SOX9 (MA0077.1)
Oct			JASPAR: POU2F2 (MA0507.1)
Klf			JASPAR: Klf5 (MA0599.1)

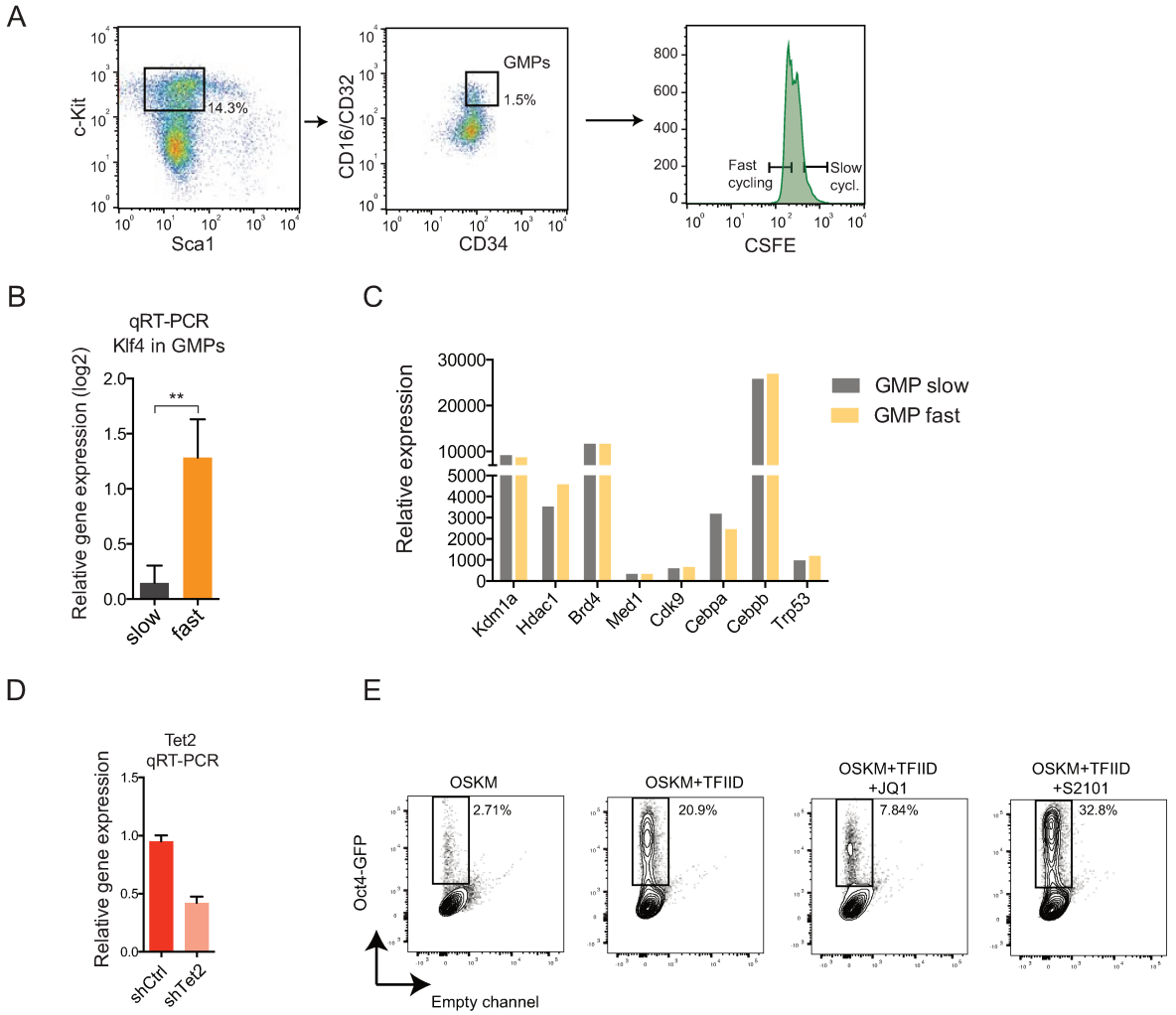
Supplementary Figure 5 ATAC-seq cluster analysis. (A) Gene ontology enrichment for genes associated with ATAC-seq peaks in each cluster shown in Figure 5A (nearest gene relative to the peak). P-values were determined by a hypergeometric test. (B) Genome browser screenshots of *Id1* and *Ifitm6* loci showing C/EBPα and H3K27ac ChIP-seq, as well as ATAC-seq data. (C) Selected over-represented DNA motifs shown in Figure 5B discovered (*de novo*) in ATAC-seq peaks, and similar motifs found in the JASPAR or HOCOMOCO database. (D) Genome browser screenshot of the *Klf4* locus showing C/EBPα and PU.1 ChIP-seq data, and 4C data using the newly discovered -90kb enhancer as view point (black triangle at the bottom). The second highlighted region (right) correspond to the second -280kb enhancer, as shown in Figure 5E. (E) Comparison of our ATAC-seq data (Fig. 5A), with Brd4 (GSE36561) and Klf4 (ref. 56) ChIP-seq data in ES cells. (F) Venn diagram showing the overlap between C/EBPα ChIP-seq peaks in Bα' cells and Klf4 ChIP-seq peaks in ES cells. (G) Average plots of C/EBPα ChIP-seq (top) and MNAse-seq signal (bottom) in the C10 pre-B cell line at different timepoints after induction of C/EBPα, for each ATAC-seq cluster (Fig. 5A). Profiles were normalized to B cells and centered on the median.

© 2016 Macmillan Publishers Limited. All rights reserved



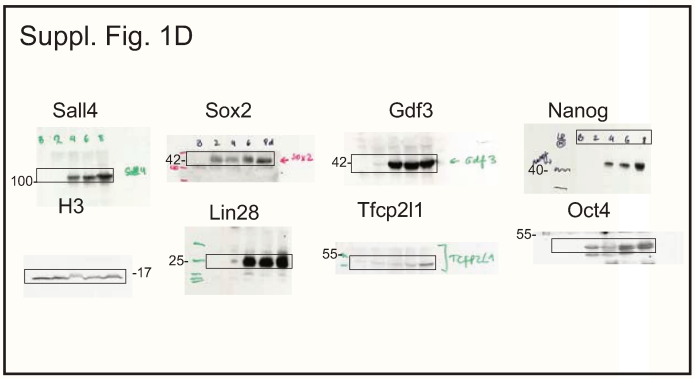
Supplementary Figure 6 C/EBPα induced changes in chromatin accessibility at myeloid and ES cell loci. (A) Genome browser screenshots of the *Rarg* and *Lefty2* loci showing ChIP-seq data for Oct4, Nanog, Klf4 and Brd4 in ESCs (ref. 56, 57 and GSE36561). (B) Gene expression profile by RNA-seq for *Rarg* and *Lefty1* during iPS reprogramming. The data represent the average from two biologically independent samples. (C) Comparison of GMPs and Bα' cells for the number of upregulated and downregulated genes (>2fold) between B and Bα' cells as well as between B cells and GMP, indicating the number of genes that overlap. (D) Canonical component analysis (CCA) of RNA-seq from B cells and Bα' cells, together with RNA-seq from different hematopoietic cell populations (ref. 58). (E) Heatmaps of ATAC-seq data from clusters I to IV of B cells, GMPs, Bα' cells, ESCs. (F) Average peak intensities of ATAC-seq data from clusters I to IV of GMPs, Bα' cells, ESCs and MEFs (ref. 44). (G) Genome browser screenshots of selected genomic loci displaying ATAC-seq data. (H) Average plot of C/EBPα ChIP-seq signal in GMPs for each ATAC-seq cluster.

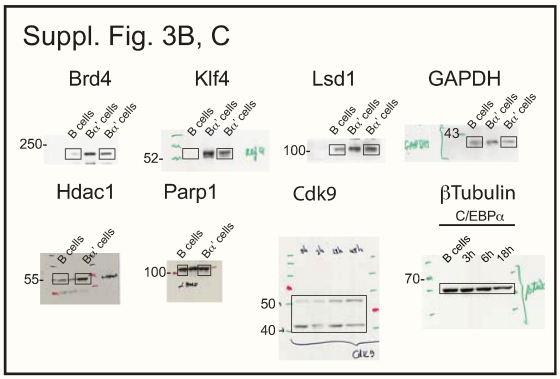
© 2016 Macmillan Publishers Limited. All rights reserved

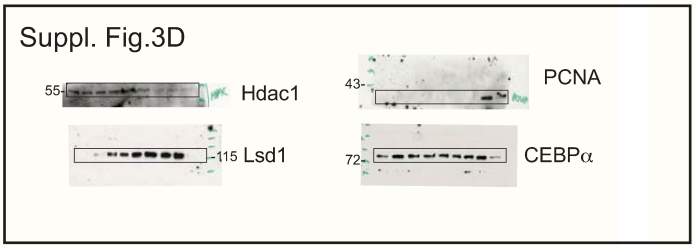


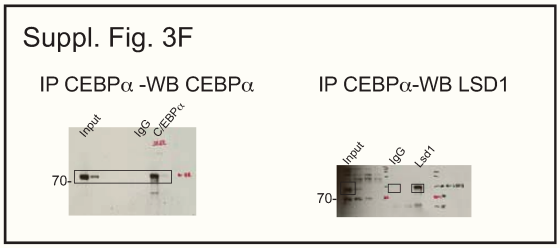
Supplementary Figure 7 Comparison of fast and slow cycling GMPs. (A) FACS plots showing sorting strategy to obtain GMPs and their separation into fast and slow cycling fractions after CSFE treatment. (B) *Klf4* expression as determined by qRT-PCR in fast and slow cycling GMPs. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (**P<0.01). (C) Array expression values for selected genes in fast and slow cycling GMPs. The data represent the average from two biologically independent samples. (D) Tet2 knockdown efficiency tested by qRT-PCR. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative Oct4-GFP FACS analysis of OSKM-induced MEFs overexpressing TFIID and treated with JQ1 or S2101.

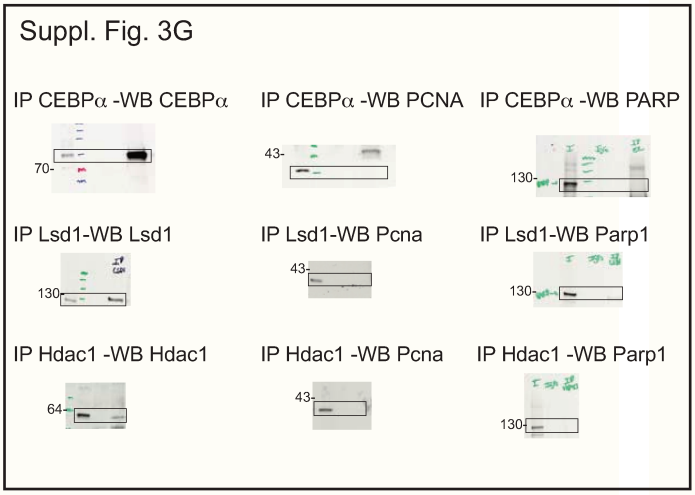
© 2016 Macmillan Publishers Limited. All rights reserved

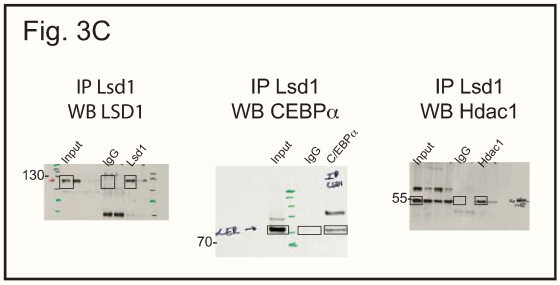


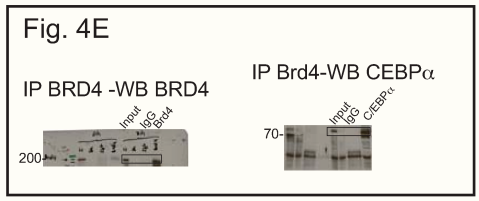












Supplementary Figure 8 Uncut gels.

© 2016 Macmillan Publishers Limited. All rights reserved

Di Stefano et al_Suppl. Table 2

Gene	B cells	Ba' cells	Day1	Day2	ESCs
Sirt1	24.70275	27.23866	25.46961	26.54176	28.35141
Stx8	26.8133	25.78702	25.43324	27.02448	25.54576
Dcaf13	25.02057	28.36722	28.79546	29.19405	30.42817
Mkrn2	25.2734	24.22111	24.44819	25.5344	26.42498
Atg3	28.54859	27.55685	28.53241	28.94388	27.5876
Huwe1	27.74928	29.57995	29.38354	30.64386	31.68534
Xrcc5	26.16631	27.92961	27.53529	28.66051	30.04875
Uhrf1	29.13824	30.59902	30.22031	29.73002	31.70755
Trim33	25.54421	27.52574	26.53638	27.20658	28.10385
Atg7	28.70343	27.14042	28.4678	29.6404	26.30808
Arrb1	28.12383	25.44546	26.19578	28.05901	24.23697
Plk1	24.74554	27.64426	27.57963	26.2345	29.81607
Nedd4	25.4834	29.35289	30.33005	30.18005	32.64503
Fbxo22	27.85522	25.60917	27.27744	28.01333	29.16994
Ube2q1	24.48448	27.64395	27.82395	27.28807	27.78896
Ltn1	24.97531	27.54043	27.92525	28.42498	28.48326
Bid	27.77276	25.6609	27.3468	28.1997	27.91658
Senp3	25.23601	27.31095	27.73338	28.47647	30.40767
Trim28	32.05651	32.93272	32.59655	32.56633	34.43291
Trim30a	26.36436	25.46102	24.4115	26.1353	24.21135
Rnf213	33.08924	29.53486	31.33844	33.43363	31.60491
Psmd4	28.5357	30.09909	30.02147	30.02265	31.02364
Ubox5	25.54592	24.13976	25.1692	25.20078	25.37965
Ubr5	25.97307	28.40239	29.09079	29.83672	29.27233
Eif4e2	24.71424	27.16535	27.44242	27.68783	28.28967
Mycbp2	25.65503	24.56877	25.71953	27.95303	24.91644
Sash1	27.54677	25.3172	27.50339	27.36829	24.05951
Casc3	24.90638	26.14075	26.24807	26.37039	27.42977

Supplementary Table 1 List of genes in the independent component analysis shown in Fig. 1F.

© 2016 Macmillan Publishers Limited. All rights reserved

Di Stefano et al _Supplementary Table 3

Primer Table

qPCR primers

Genes	Forward primer	Reverse primer
Tdh	CAGACTGAAGATAAAAGGCAG	GCATCTGTTCTTCTGATACC
Zfp296	CCATCTCAGAATCCAAAGAG	TATCTAGGTGTTGTGTGTCTGG
Nanog	CAGTTTTTCATCCCGAGAAC	CTTTTGTTTGGGACTGGTAG
Lin28a	TGTTCTGTATTGGGAGTGAG	CCATATGGTTGATGCTTTGG
Sall4	AAGAACTTCTCGTCTGCC	AGTGTACCTTCAGGTTGC
Gdf3	CGTCTTAAGGAAAATCATCCG	GGCAGACAAGTTAAAATAGAGG
Pou5f1	GTCCCTAGGTGAGCCGTCTTT	AGTCTGAAGCCAGGTGTCCAG
Sox2	ATGAGAGATCTTGGGACTTC	TCTATACATGGTCCGATTCC
Zfp42	GTTCGTCCATCTAAAAAGGG	TAGTCCATTTCTCTAATGCCC
Esrrb	AAAGCCATTGACTAAGATCG	AATTCACAGAGAGTGGTCAG
Pgk	ATGTCGCTTTCCAACAAGCTG	GCTCCATTGTCCAAGCAGAAT
18s	AACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG
Cdh1	CATGTTCACTGTCAATAGGG	GTGTATGTAGGGTAACTCTCTC
Ebf1	ACACAATTCATTCCCCGAAA	AAGTCAACGGTTTTGCATCC
Foxo1	AAGAGCGTGCCCTACTTCAA	CTCCCTCTGGATTGAGCATC
Gfi1b	TAATTCCTGGGCAAAAGAG	TGTTTGATTGTGTTCCAGCT
Ikzf3	CTTTTCTTCAGAACCCTGAC	CAATTGCTTGCTAATCTGTCC
Rag1	GAAGCTTCTGGCTCAGTCTACATCT	ACCTCATAGCGCTGCAGGTT
Ciita	CTGGACAAGAATGTCATCTG	TTGACTCTTATGGGCTATGG
Klf4	CATTAATTGTGTCGGAGGAAG	CCGTTTGGTACCTTTAGAAC
Lefty1	TGTGTGCTCTTTGCTTCC	GGGGATTCTGTCCTTGGTTT
Lsd1	TCATTCAGCTGCAAGAAAAG	TCCTCCTGAGTTTTCACTATC
Brd4	CTGATGTCCGATTGATGTTC	AGAGGACACTGTAACAACTG

ChIP primers

Genes	Forward primer	Reverse primer
Ebf1	CAGCAACCAAAACCTAGCAA	TCCCACTATTTATTCCCACA
Ciita	ACCTTGGGAGTATGCACTGG	AATTGGGTGACCACAGAAGC
Rag1	TCTCGCTCTCCTGTCAGTCA	CCGAGCAGAGACGTTAGCTT
Gfi1b	TCCCCAGAAATCATGTCAGA	GCTATTTCTGCCAAGGGTGA
Foxo1	CTGGTCAAGCTCTTGCCTGT	GGATTGCAAGTTCTCCTCCA
Ikzf3	GCCAAAGAAACACAGGCAAT	CCTCAAGAGCTGCTCACCTT
control (gene desert)	TCAGAAAGGAATCAATCAATCAAA	ATGCCCTCTTCTGGTGTGTC

4C primers

Genes	Reading primer	Non-reading primer
Klf4 promoter	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT GACAGGACAAGCGCGTAC	CAAGCAGAAGACGGCATACGAGAGATA CCTTTCACCAGGGAT
Klf4 enhancer	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT CGCTTTATGTTCTGCCAGTAC	CAAGCAGAAGACGGCATACGATGTCAC AGCCCCAGTAGTG

Supplementary Table 2 List of proteins annotated with the GO term "protein degradation".

© 2016 Macmillan Publishers Limited. All rights reserved

2.3 Discussion: on the importance of dynamical analyses

One of the key aspects of these analyses was the identification of **transient phenomena**, which was made possible by the use of **time series**.

In section 1, the identification of C/EBPa pioneering effect was discovered because we could identify time points where C/EBPa was already binding to the DNA, but PU.1 was not binding yet. As both factors are binding together in macrophages, end time points do not suffice to identify which factor opens the chromatin.

In section 2, we found that the regulation of Klf4 by C/EBPA during iPSCs reprogramming was due to the transient activity of its enhancer. Even the chromatin looping between the enhancer and the promoter was specific of the intermediate state, but present neither in B cell nor in ESCs.

Interestingly, we later found that the repression of B cell genes, involving KDM1A (Lsd1) and HDAC1, was mediated through a **transient binding of C/EBPa** at B cell enhancers, not found at later time point, neither in myeloid cells (this is described in another publication in section 3.2).

One hypothesis would be that this is an effect of C/EBPa over-expression, saturating the chromatin and competing with B cell factors; however this seems improbable as C/EBPa protein levels has been shown to be similar in our B cell system than in macrophages (Xie et al., 2004).

These enhancers become shutdown through modifications of histone marks, leading to the down-regulation of B cell factors, such as Ebf1, Pax5, E2a and Foxo1. It is then possible that the loss of these factors leads to the closing of the chromatin.

It would be interesting to study the role of KDM1A and HDAC1 in the closing of the chromatin, whether their effect is linked to nucleosome positioning or if they mostly affect the induction of transcription. It would also be interesting to see if such mechanisms are involved in other cell reprogramming system.

The **importance of these transient phenomena** highlights the complexity of cell fate conversion. Although some transient steps are known to be important, such as the mesenchymal to epithelial transition, reprogramming is often seen as a linear sequence of events. However, It seems clear that different processes overlap, and follow different time scales (e.g. in our system, the MET and the regulation of cell cycle, as shown in Figure 9). These can be seen as parallel events, but it is most probable that they are intertwined and co-regulated. Taking into account the **multiple dimensions** of such systems allows for a better understanding of the mechanism controlling them.

Section 3: Dynamical modelling of gene regulatory networks

3.1 Gene network controlling B cell and macrophage specification

3.1.1 How to integrate high-throughput data into gene regulatory network model

High-throughput data are very promising for the development of dynamical models. However, the **integration of omic data into regulatory networks** is still challenging. Two approaches should be distinguished here: the inference of networks from data, where a network is build from scratch based on data, and the integration of data in a pre-existent model.

The **inference of gene networks** from transcriptomic data has been a major challenge in systems biology. Classical approaches infer regulations between genes based on the correlation of their expression (of similar measures) (Liu, 2015). Many methods have been developed in that respect and this has been one of the major topics of the DREAM challenge (Seyednasrollah et al., 2017). These methods require extremely large and rich datasets, getting as close as possible to conditions exhibiting a single change in expression to infer each regulation, which is not possible to reach in time series during cell specification or reprogramming.

Other methods have been proposed, based on epigenetic data such as open chromatin or histones marks (Lundberg et al., 2016; Neph et al., 2012). These methods are used to study global characteristic of genomic network such as topology, conservation of regulations, etc.

A second approach is the **integration of high-throughput data** into pre-existent gene networks. Such networks are built through the manual curation of the literature, and the integration of experimental evidences of regulations. This has been extensively done for signalling pathways, which can be found in databases such as Reactome (Fabregat et al., 2016) or KEGG (Kanehisa et al., 2017).

Data integration can be achieved by **visualising omic data on a gene network**, for example by projecting gene expression levels onto nodes. This can be done with the software Cytoscape (Smoot et al., 2011), and can be coupled with static analysis of the graph such as the research of modules. Such analyses are useful to interpret data and identify some inconsistencies in the model.

A next step is the use of high-throughput data to **improve an existing network**. Starting by inferring regulations from the data (e.g. using the methods of network inference discussed in the previous paragraph), these predictions can be compared to the regulations present in the network, and discrepancies can be further studied. This can be useful to discover unsuspected regulations, but can be tedious when many new regulations are predicted.

Dynamical modelling can be used to identify key predictions from the data and **prioritize experimental validation**. Identifying caveats in a model

(missing cell type, wrong behaviour, etc.), one can consider regulations predicted from data to extend the model, and test if such additions improve it.

This approach is limited to systems where a relatively precise model has already been built. This approach has already been used in several cases (Schütte et al., 2016) but no systematic method has been provided, less of software implementation. In section 3.2 we propose a method to integrate ChIP-seq with a logical model and iteratively select predictions of regulation from the data to refine the model.

3.1.2 The logical formalism

The **logical formalism** is particularly interesting for the integration of heterogenous data. As a qualitative formalism, it does not require quantitative parameters (which are difficult to measure in mammalian cells). Rather, it offers a formal framework to translate the way regulations are often described in molecular biology, such as "this and this factor are required for activation but this inhibitor should be absent".

In this formalism, molecular entities (genes, proteins, signalling molecules, etc.) are represented by **logical variables**, which can take two or more levels, and are depicted graphically as nodes. Regulations are depicted by oriented edges (arrows) between the nodes, and integrated into logical formula, using the logical operators AND, OR and NOT. These formula (or rules) define the combinations of factors allowing the target factors to become active or to go to a higher level.

The different levels a variable can take have to be defined a priori. The lowest level (value 0) represents a negligible of activity of the factor (not necessarily its total absence), and higher levels represent its activity. In the case of factors which do not regulates other ones, such as surface markers, this can also represent a threshold of expression, similarly to the threshold defined in flow cytometry. A variable with only two levels is called a Boolean variable.

Multiple levels of activity are used when a factor has been shown to regulate different targets at different levels. For example, the Lambda phage factor Cro repress the factor Cl even when expressed at low level, but can repress itself only at higher concentration (Thieffry and Thomas, 1995). This has been shown to be also the case in some mammalian system (Laslo et al., 2006). Multi-level variables should be used parsimoniously, to limit the complexity and over-fitting of the model.

3.1.3 Model analysis and dynamical simulations

A first analysis that can be performed on a model is the **identification of stables states**. These are supposed to fit approximately with expected cell types, and the pattern of variable levels can be compared to gene expression data. This can be used to fit the model, and inconsistencies in the stable states can reveal caveats in the model (see application in section 3.2).

Dynamical simulations can then be performed to study cell behaviours, such as differentiation, reprogramming or the effect of gain- or loss-of-function. In the logical formalism, simulations are performed by starting from a given initial state, and iteratively update variables that tend towards a new value (given the combination of their regulators), defining the subsequent states of the network. This is repeated until stable states (or stable cycles) are reached, forming the state transition graph (**STG**).

The **updating of variables values** can be done **synchronously**, i.e. all variables tending to another value are updating together (and therefore each parent state has only one child state). Although simple, this approach is not biochemically realistic as it assumes that the rates of synthesis/degradation of every molecule in the model are identical, and this can lead to artefactual behaviours (Abou-Jaoudé et al., 2016).

In contrast, when using **asynchronous** updating, where only one variable can change in a transition, each parent state has *n* children state, one for each of the *n* variables tending toward another value.

When differences in rates are known, they can be taken into account in the model to define updating priorities. In the absence of such data, the asynchronous updating offers an over-approximation of the dynamical behaviour of the system.

Of particular interest is the identification of the necessary and sufficient **conditions for commitment**. Formally, this comes down to the identification and characterisation of the strict basin of attraction of a stable state, i.e. the set of all states leading only to this stable state. Methods have been developed to cluster the STG based on attractors' reachability, resulting in a **hierarchical transition graph** (HTG) (Bérenguier et al., 2013). This allows for the rapid identification of basins of attractions, and analysis of the transition between states (or clusters of states) can reveal key factors involved in commitment.

Finally, a logical model can be extended into a **stochastic** one, using methods such as continuous Markov chains implemented in the software MaBoSS (Stoll et al., 2012). In this formalism, the behaviour the network is simulated, following only one of the possible transitions at each step, The choice of the transition is done according to some probabilities, which can be kept equal or fitted to data. Such methods offer a way to study a system at the cell population level, evaluating the percentages of different cell types, and providing a more continuous view of the dynamic.

3.2 Publication: Logical modelling of lymphoid and myeloid cell specification and trans-differentiation (PNAS, 2017)

Contribution:

I built the gene network model, analysed all data, performed all the simulations and model analyses, designed the subsequent experiments and performed the experiments with help of Bruno di Stefano and Jose Luis Sardina Ortega from the laboratory of Thomas Graf.

Logical modeling of lymphoid and myeloid cell specification and transdifferentiation

Samuel Collombet[a,1], Chris van Oevelen[b,2], Jose Luis Sardina Ortega[b,2], Wassim Abou-Jaoudé[a], Bruno Di Stefano[b,3], Morgane Thomas-Chollier[a], Thomas Graf[b,c,1], and Denis Thieffry[a,1]

[a]Computational Systems Biology Team, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France; [b]Hematopoietic Stem Cells, Transdifferentiation, and Reprogramming Team, Gene Regulation, Stem Cells, and Cancer Program, Center for Genomic Regulation, Barcelona Institute for Biotechnology, 08003 Barcelona, Spain; and [c]Universitat Pompeu Fabra, 08002 Barcelona, Spain

Edited by Ellen V. Rothenberg, California Institute of Technology, Pasadena, CA, and accepted by Editorial Board Member Neil H. Shubin November 18, 2016 (received for review September 1, 2016)

Blood cells are derived from a common set of hematopoietic stem cells, which differentiate into more specific progenitors of the myeloid and lymphoid lineages, ultimately leading to differentiated cells. This developmental process is controlled by a complex regulatory network involving cytokines and their receptors, transcription factors, and chromatin remodelers. Using public data and data from our own molecular genetic experiments (quantitative PCR, Western blot, EMSA) or genome-wide assays (RNA-sequencing, ChIP-sequencing), we have assembled a comprehensive regulatory network encompassing the main transcription factors and signaling components involved in myeloid and lymphoid development. Focusing on B-cell and macrophage development, we defined a qualitative dynamical model recapitulating cytokine-induced differentiation of common progenitors, the effect of various reported gene knockdowns, and the reprogramming of pre-B cells into macrophages induced by the ectopic expression of specific transcription factors. The resulting network model can be used as a template for the integration of new hematopoietic differentiation and transdifferentiation data to foster our understanding of lymphoid/myeloid cell-fate decisions.

gene network | dynamical modeling | hematopoiesis | cell fate | cell reprogramming

Hematopoiesis is the process through which all blood cells are produced and renewed, starting from a common population of hematopoietic stem cells (HSCs) (1). HSCs differentiate into lineage-specific progenitors with restricted differentiation potential and expressing specific surface markers (Fig. 1*A*). Loss- or gain-of-function experiments targeting transcription factors (TFs) or signaling components have led to the identification of factors required for specific developmental steps. Some factors are required for the development of entire lineages (e.g., Ikaros for lymphoid cells), whereas others are needed only at late stages of cell-type specification (e.g., the requirement for the paired-box factor Pax5 after the pro–B-cell stage). These factors cross-regulate each other to activate one gene-expression program and silence alternative ones.

Although cell commitment to a specific lineage was long considered irreversible, recent reprogramming experiments emphasized the pervasive plasticity of cellular states. Indeed, the ectopic expression of various regulatory factors (mainly TFs and signaling components) can enforce the establishment of new gene-expression programs in many kinds of differentiated cells (2). Strikingly, pluripotency can be induced in somatic cells by forcing the expression of a handful of TFs, enabling further differentiation into any cell type (3). In the hematopoietic system, TF-induced transdifferentiation between erythroid and myeloid cells and between lymphoid and myeloid cells has been described (4).

In this study, we focus on B-cell and macrophage specification from multipotent progenitors (MPs) and on TF-induced transdifferentiation between these lineages. Ectopic expression of the myeloid TF C/EBPα (CCAAT/enhancer-binding protein alpha, encoded by the Cebpa gene) can induce B cells to transdifferentiate into macrophages (Fig. 1*A*, red arrows) (5). C/EBPα is also required for the transition from common myeloid progenitors (CMPs) to granulocyte-macrophage progenitors (GMPs), and mutation in this gene can result in acute myeloid leukemia (6). Understanding the molecular mechanisms by which such factors can induce cell-fate decisions is of primary importance and might help in the development of novel therapeutic strategies.

Computational modeling of regulatory networks is increasingly recognized as a valuable approach to study cell-fate decisions. Indeed, the integration of the available information about gene regulation into a common formal framework allows us to identify gaps in our current knowledge, as successfully shown in previous studies on the differentiation of hematopoietic cells (7). Dynamic analysis can reveal nontrivial properties, including transient phenomena, and can be used to identify key regulatory factors or interactions involved in the control of cell-fate commitment (8, 9). Furthermore, genome-wide approaches such as ChIP-sequencing (ChIP-seq) can unveil novel regulations to be further incorporated in a gene-network model (10). Here, we combined a logical multilevel formalism, capturing the main qualitative aspects of the dynamics of a regulatory network in the absence of quantitative kinetic data (11), with a meta-analysis of all available ChIP-seq datasets for a selection of TFs, revealing tens of previously unknown regulations. We then performed iterations of computational simulations, followed by comparisons with experimental data and adjustments of the model, to identify caveats in our model and to test the effect of putative regulations in silico before confirming them experimentally (Fig. 1*B*).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Gene Regulatory Networks and Network Models in Development and Evolution," held April 12–14, 2016, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Gene_Regulatory_Networks.

Author contributions: S.C., C.v.O., J.L.S.O., T.G., and D.T. designed research; S.C., C.v.O., J.L.S.O., and B.D.S. performed research; S.C., W.A.-J., B.D.S., M.T.-C., T.G., and D.T. analyzed data; and S.C., J.L.S.O., W.A.-J., M.T.-C., T.G., and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.V.R. is a guest editor invited by the Editorial Board.

Data deposition: The ChIP-seq data for EBF1 and Foxo1 in pre-B cell lines and during transdifferentiation have been deposited in the Gene Expression Omnibus (GEO) database (accession code GSE86420). The final model has been deposited in the BioModels database (accession no. 1610240000).

[1]To whom correspondence may be addressed. Email: denis.thieffry@ens.fr, samuel.collombet@ens.fr, or thomas.graf@crg.eu.

[2]C.v.O. and J.L.S.O. contributed equally to this study.

[3]Present addresses: Department of Molecular Biology, Center for Regenerative Medicine and Cancer Center, Massachusetts General Hospital, Boston, MA 02114; Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138; Harvard Stem Cell Institute, Cambridge, MA 02138; and Harvard Medical School, Cambridge, MA 02138.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1610622114/-/DCSupplemental.

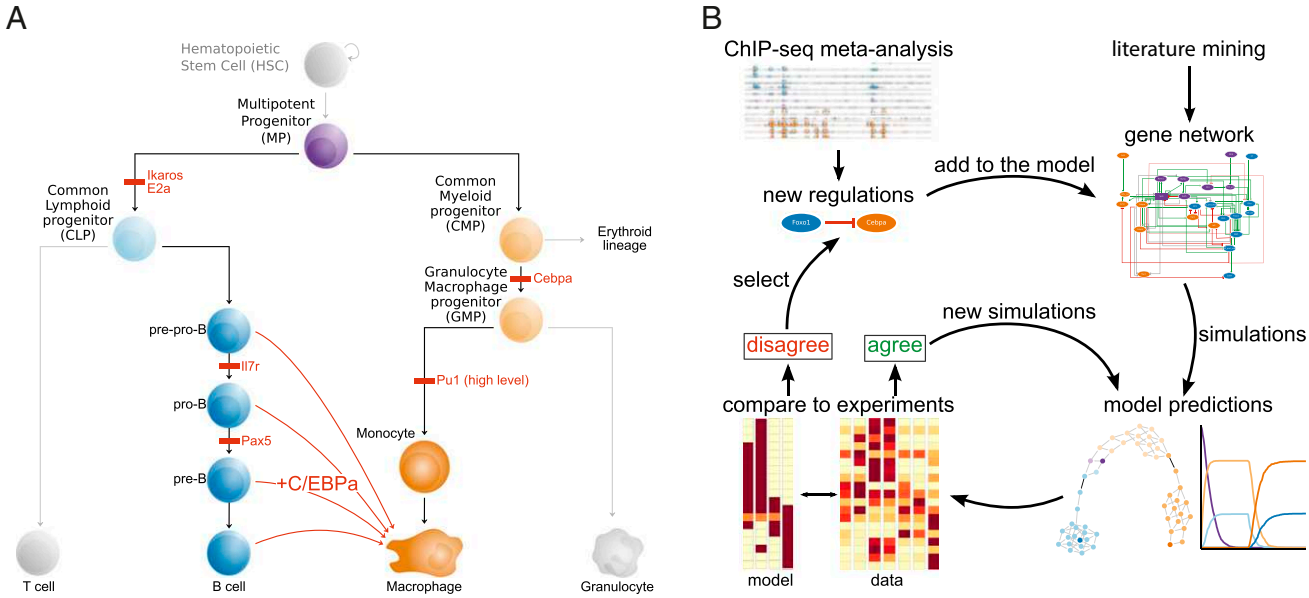


Fig. 1. (*A*) Schematic representation of hematopoietic cell specification. Genes in red are required for progression at the corresponding steps. C/EBPα-induced transdifferentiation is indicated by red arrows from B-lineage cells to macrophages. (*B*) Iterative modeling workflow. A model is first built based on the literature and is used to predict dynamical behaviors (cell phenotype, differentiation, reprogramming, and so forth). Predictions then are compared with experimental data; when the predictions and experimental data agree, further predictive simulations are performed; when they do not agree, further regulations are inferred from ChIP-seq data and are integrated into the model until simulations fully agree with data.

Results

Gene Network Controlling B-Cell and Macrophage Specification. To build a model of the gene-regulatory network controlling B-cell and macrophage specification from common progenitors, we first performed an extensive analysis of the literature to identify the TFs and signaling pathways controlling these events. The TF PU.1 (encoded by the *Spi1* gene) is required for the normal development of both lymphoid and myeloid cells (12). The development of common lymphoid progenitors (CLPs) depends on the TFs Ikaros (encoded by *Ikzf1*) and E2a (encoded by the transcription factor 3 gene *Tcf3*) (Fig. 1*A*) (13, 14). The B-cell lineage is further controlled by Mef2c, the interleukine 7 receptor (IL7r), Ets1, Foxo1, Ebf1, and Pax5 (15, 16, 17). The specification of the myeloid GMPs depends on C/EBPα (6), which is regulated by Runx1 (runt-related transcription factor 1) (18). Macrophage specification further relies on the macrophage colony-stimulating factor (M-CSF) receptor (CSF1r), on the up-regulation of PU.1, and on Cebpb and the Id proteins (including Id2) (19, 20). The TFs Egr and Gfi1 repress each other to specify macrophage versus granulocyte lineages (21); Gfi1 also is important for B-cell differentiation (22).

Finally, to distinguish among the different cell types, we further consider the B-cell marker CD19, the macrophage marker Mac1 (also called "Cd11b," encoded by the *Itgam* gene), and the cytokine receptor Flt3, which is expressed specifically on MPs and CLPs.

We then carried out an extensive review of the literature to collect information about cross-regulations between the selected factors and grouped these regulations into four classes, depending on the available evidence: (*i*) functional effect, e.g., an effect inferred from gain- or loss-of-function experiments (which could be either direct or indirect); (*ii*) physical interaction, e.g., TF binding at a promoter or enhancer; (*iii*) physical and functional evidence, suggesting a direct regulation; and (*iv*) fully proven regulation, e.g., evidence of functional effect and physical interaction along with reported binding-site mutations affecting the functional effect or reporter assays demonstrating *cis*-regulatory activity. Altogether, we gathered a total of 150 items of experimental evidence (Dataset S1) supporting 79 potential regulations (Fig. S1*A*).

Many of these regulations are sustained only by functional evidence. To assess whether they could correspond to direct regulations, we analyzed public ChIP-seq datasets targeting each of the TFs considered in our network, amounting to 43 datasets for 10 TFs in total (Dataset S2). We systematically looked for peaks in the "gene domain" (23) coding for each component involved in the network (*Materials and Methods*). This ChIP-seq meta-analysis confirmed 26 direct regulations (Fig. 2*A*, green or red cells with a star) and pointed toward 66 additional potential transcriptional regulations (gray cells with a star). For example, at the *Spi1* locus, we confirmed the binding of Ikaros at known enhancers, where it was previously reported to limit the expression of *Spi1* together with a putative corepressor (24). Because we also found that Pax5, Ebf1, and Foxo1 bind to the same sites (Fig. 2*B*), we suggest that these factors could act as corepressors. Ectopic expression of Foxo1 in macrophages induced a reduction of *Spi1* expression (Fig. S1*B*), further confirming this negative regulation.

C/EBPα Directly Represses B-Cell Genes. We have previously reported that C/EBPα can enforce B-cell TF silencing by increasing the expression of the histone demethylase Lsd1 (Kdm1a) and the histone deacetylase Hdac1 at the protein level and that these enzymes are required for the decommissioning of B-cell enhancers and the silencing of the B-cell program (25). Because key B-cell regulators such as Foxo1, Ebf1, and Pax5 are repressed after 3 h of C/EBPα induction (Fig. S1*C*), we wondered whether C/EBPα could be directly responsible for this early effect. To verify this hypothesis, we reanalyzed data from ChIP-seq for C/EBPα after 3 h of induction in a reprogrammable cell line (26). As expected, we detected binding of C/EBPα at the *cis*-regulatory elements of *Foxo1* (Fig. 2*C*), *Ebf1*, *Pax5*, *IL7r*, and *Mef2c* genes (Fig. S1*C*), supporting their direct repression by C/EBPα.

Furthermore, C/EBPβ also can induce transdifferentiation of pre-B cells (5), and it has been shown that C/EBPβ can rescue the formation of granulocytes in C/EBPα-deficient mice (27). Moreover, C/EBPβ almost always binds at C/EBPα-binding sites (Fig. 2*A*), as exemplified by the *Spi1* locus (Fig. 2*B*). These findings suggest a

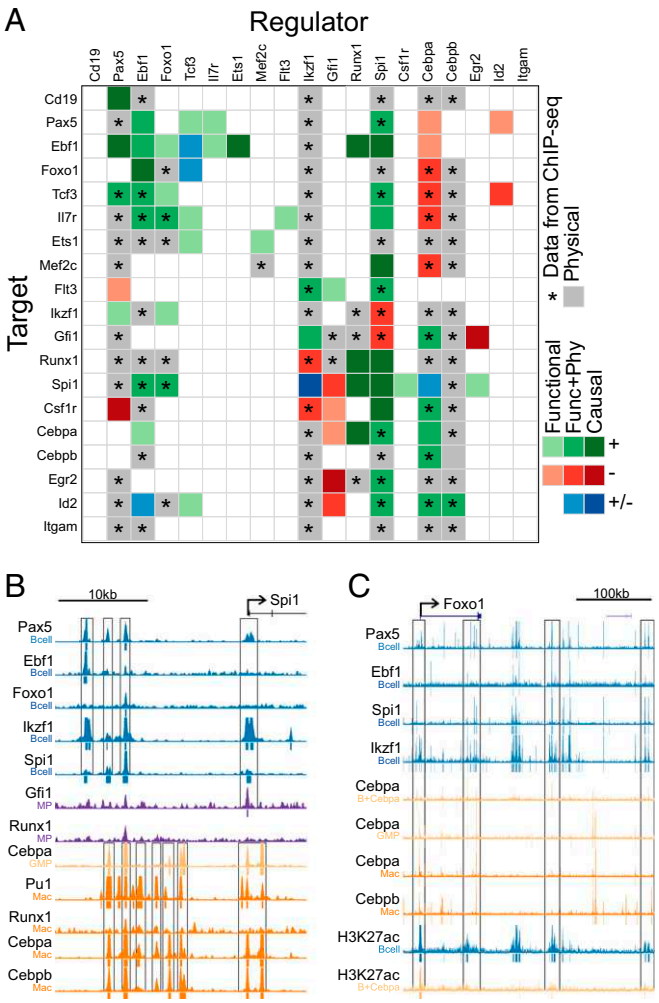


Fig. 2. (*A*) Heatmap showing the regulations inferred from the literature and from ChIP-seq meta-analysis. (*B*) ChIP-seq signals and peaks (under signal) at the *Spi1* locus. Black frames indicate known enhancers (24). The vertical axes represent reads per million (RPM) (maximum: 2 RPM for Ebf1 and Ikaros, 1.5 RPM for Foxo1, 1 RPM for Runx1 and Gfi1, 5 RPM for other TF). (*C*) ChIP-seq signals and peaks (under signal) at the Foxo1 locus. Black frames indicate B-cell enhancers in which C/EBPα binding is detected. The vertical axes represent RPM (maximum: 2 RPM for Ebf1, 5 RPM for other TFs, 3 RPM for H3K27ac). Note that Pax5 and Ikaros peaks are located downstream of the first exon and all other peaks are upstream of the TSS.

redundancy between these two factors in the regulation of their target genes (at least in those considered here), and we integrated this redundancy in our model.

Dynamical Modeling Using Multilevel Logic. The core components and regulations collected from our analysis of the literature and ChIP-seq datasets were assembled in a regulatory graph using the GINsim software (Fig. 3).

Validating all the predicted regulations (Fig. 2*A*, gray cells with a star) experimentally would be a daunting task. Instead, we focused on a selection of these regulations (depicted by the gray arrows in Fig. 3) and used dynamical modeling to assess their impact on cell specification.

To transform our regulatory graph into a predictive dynamical model, we took advantage of a sophisticated logical (multilevel) formalism. More precisely, we associated a discrete variable with each regulatory component. These variables usually take two values (0 or 1) but can be assigned more values whenever justified. Regulations are combined into logical rules using the Boolean operators NOT, AND, and OR, to define the conditions enabling the activation of each model component (*Materials and Methods*). This formalism relies essentially on qualitative information and allows the simulation of relatively large network models (encompassing up to a few hundred components). It should be noted that the value 0 does not necessarily imply that a factor is not expressed at all but rather that its level is insufficient to affect its targets significantly. PU.1 is the only factor for which we found clear evidence supporting a distinction between two functional (non-0) levels (21). Consequently, we assigned a ternary variable (taking the values 0, 1, or 2) to this node and assigned Boolean variables (i.e., taking the values 0 or 1) to the other nodes.

Regarding the definition of the logical rules, we first considered the regulations supported by both functional and physical evidence (depicted as green and red arrows in Fig. 3). As a default, we required that all activators but no inhibitor to be present to enable target activation and further adjusted the rules based on information gathered from the literature (see the rules in *Materials and Methods* and Dataset S3). As mentioned before, we then added selected regulations inferred from our ChIP-seq meta-analysis (depicted as gray arrows in Fig. 3) to refine our model.

Modeling Different Cell-Type Phenotypes. We first assessed whether our model properly accounts for progenitor, B-cell, and macrophage gene-expression patterns. Because stable states capture the long-term behavior associated with the acquisition of gene-expression patterns during cell specification, we computed all the stable states of our model using GINsim software (28) and compared them with gene-expression data (Fig. 4*A*) (29). We initially found that our stable states were largely inconsistent with known patterns of gene expression (Fig. S2*A*), revealing important caveats in the published data on which we based our model.

A first caveat concerned the regulation of Cebpa. Indeed, Cebpa is not expressed in lymphoid cells, although its well-known activators PU.1 (*Spi1*) and Runx1 are expressed in both B cells and

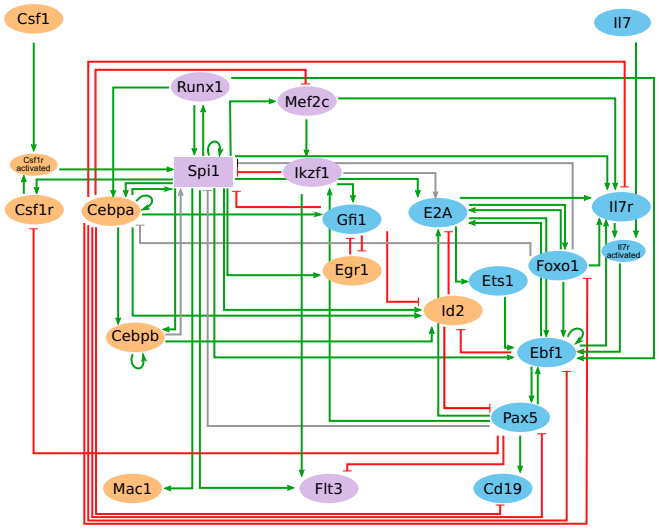


Fig. 3. A regulatory graph depicting the interactions inferred from the literature and ChIP-seq meta-analyses. Nodes represent genes (except for CSF1r_act and Il7r_act, which represent the activated forms of cytokine receptors), and arrows denote regulatory interactions. Orange nodes represent factors expressed in macrophages, purple nodes represent factors expressed in progenitors, and blue nodes represent factors expressed in B-lineage cells. Ellipses represent Boolean components; the rectangle emphasizes the ternary component *Spi1*. Green and red edges correspond to activations and inhibitions, respectively. Gray edges denote the regulations predicted by the ChIP-seq meta-analysis, which were included in the model to increase consistency with expression data.

macrophages. Therefore, our model exhibited only Cebpa$^+$ stable states (Fig. S2*A*), suggesting that an inhibitory regulation of Cebpa was missing during lymphoid specification. Foxo1, a factor controlling the early steps of B-cell commitment (30), stands as a relevant candidate. To test this hypothesis, we performed ChIP-seq for Foxo1 in our pre–B-cell line and observed binding at the Cebpa promoter, suggesting a physical interaction and potential direct regulation of Cebpa (Fig. S2*B*). To test if Foxo1 has a functional effect on Cebpa expression, we ectopically expressed it in a macrophage cell line (RAW) and found a significant down-regulation of Cebpa (Fig. S2*C*), suggesting a direct negative regulation of Cebpa by Foxo1. We therefore refined our initial model by including this additional regulation (see the rule associated with Cebpa in Dataset S3).

A second caveat revealed by our model analysis concerned the regulation of *Tcf3* (encoding E2a). Indeed, E2a was expressed in all the stable states, even after Cebpa repression by Foxo1 was included (Fig. S2*D*), although E2a has been shown to be expressed in MPs and in lymphoid cells but not in myeloid cells. Moreover, the only factor in our model expressed in MPs and regulating E2a is PU.1, which is also known to be expressed in myeloid cells, thus suggesting a missing regulation of E2a. However, despite our efforts, we could not find any evidence for a myeloid repressor of E2a in either the literature or our ChIP-seq data meta-analysis. Turning to putative activators of E2a, we focused on Ikaros. Indeed, like E2a, Ikaros is required for lymphoid development, and its knockout entails a loss of lymphocytes similar to that seen with E2a knockout. Interestingly, we found that Ikaros binds the E2a promoter in B cells (Fig. S2*E*), suggesting a direct activation of E2a by Ikaros. Hence, we further refined our model by including this regulation (see the rule associated with E2a in Dataset S3).

More surprising was the high expression of Egr2 observed in pro/pre-B cells. We also found expression of the related factor Egr1 in two different datasets (Fig. S2 *F* and *G*). It has been reported that Egr1/2 cross-inhibits Gfi1, the first favoring macrophage specification and the second favoring B lineage (21). However, although this study shows that Egr2 has an effect on the differentiation potential of MPs, it does not demonstrate that this factor is indeed not expressed in B cells or that it can antagonize the expression of B-cell genes. To assess the expression of Egr2, Egr1, and Gfi1 at the protein level, we performed Western blots for these proteins in B cells and macrophages. We were able to detect all three proteins in B cells (Fig. S2*H*), confirming the gene-expression data. We therefore propose that some late B-cell factors activate both Gfi1 and Egr2, overcoming their cross-inhibitions. Because Pax5 was the only B-cell factor found in our meta-analysis to bind to Gfi1 and Egr2 loci (Fig. 2*A*), we consider it to be an activator of both Gfi1 and Egr2 (see corresponding rules in Dataset S3).

When analyzing the resulting refined model, we found that its stable states correspond well to CLPs, GMPs, B-lineage cells, and macrophages, as defined by the known patterns of gene expression (Fig. 4*B*). For some genes, we obtained apparent discrepancies between expression data and stable state values; these discrepancies can be attributed to model discretization (see *SI Materials and Methods* for more details).

Our analysis points to previously unrecognized regulators of E2a and Cebpa that are important at the onset of lymphoid and myeloid specification and introduces refinements of the regulations of Egr2 and Gfi1. After incorporating these regulations in our model, we used it to study the dynamics of B-cell and macrophage specification.

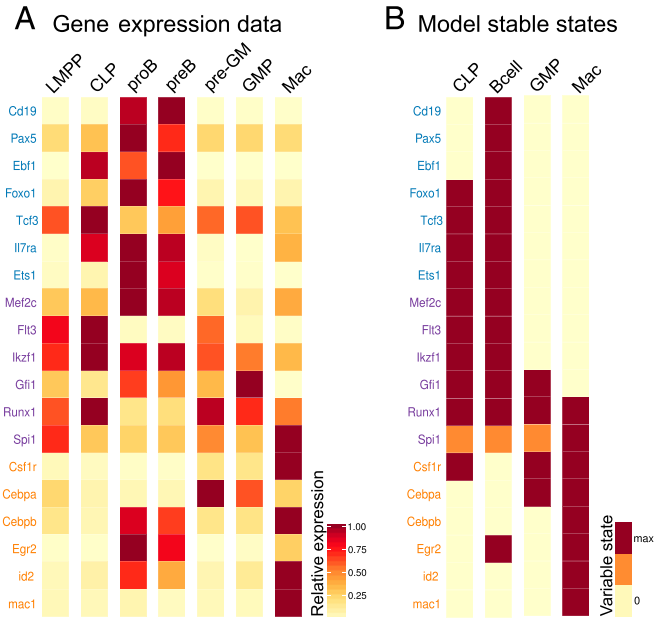


Fig. 4. (*A*) Gene-expression values (microarrays) in lymphoid/myeloid progenitors (LLPP), B cells, and macrophages (Mac) (29). These values are relative to the highest expression value. (*B*) Context-dependent stable states computed for the model. A yellow cell denotes the inactivation of the corresponding component, a red cell represents maximal activation (1 for Boolean components, 2 for *Spi1*), and an orange cell represents an intermediate level (1) for *Spi1*.

Specification of B-Cell and Macrophage Precursors from MPs. To improve our understanding of the transcriptional regulation of hematopoietic cell specification, we performed several iterations of hypothesis-driven simulations and comparisons with experimental data, followed by model modifications to solve remaining discrepancies.

First, using GINsim software, we simulated the specification of MPs, defined by the expression of *Spi1*, *Runx1*, *Ikzf1*, *Gfi1*, and *Flt3*. In the absence of environmental signals, we found that our model can lead to two different stable states corresponding to GMPs and CLPs (Fig. 5*A*). Upon stimulation with both CSF1 and IL7, the system tends to two new stable states, corresponding to macrophages and B lineage cells, respectively. These simulations thus recapitulate the commitment of cells to GMP- and CLP-associated states and their loss of potential for alternative lineages.

Next, using stochastic simulations (see *Materials and Methods* and ref. 31 for more details), we analyzed the evolution of the fraction of cells expressing distinct factors associated with specific cell lineages starting with the same initial state (MPs) and environmental conditions (initially no stimulation, followed by stimulation with Csf1 and Il7). Our results show two waves of gene activation for both myeloid and lymphoid factors. The first wave corresponds to the progenitor (GMP or CLP) expression programs, and the second one corresponds to terminally differentiated cells (macrophages or B cells) (Fig. 5*B*, *Top* and *Middle*). The evolution of the different cell populations (defined by the gene-expression signatures indicated in Dataset S4) was consistent with our logical simulations, with a rapid decrease of the MP population followed by the specification toward GMPs and CLPs and then by their differentiation into macrophages and B cells, respectively (Fig. 5*B*, *Bottom*). The proportions of myeloid and lymphoid cells were ∼75 and 25%, respectively, in qualitative agreement with the higher proportion of myeloid cells present in the bone marrow (32). Tentatively, this asymmetry could be encoded in the regulatory circuitry rather than merely being the result of differences in proliferation rates. A sensitivity analysis further revealed that the proportion of lymphoid and myeloid cells was affected only by changes in the up-regulation rates of Cebpa, Foxo1, and E2a (Fig. S2*I*), supporting the key function of Cebpa and Foxo1 in the commitment decision (E2a being required for Foxo1 expression).

To obtain more comprehensive insights into the alternative trajectories underlying myeloid and lymphoid lineage specification, we clustered the logical states (Fig. 5*A*) to generate a hierarchical (acyclic) graph (28) in which all the states with a similar potential (i.e., leading to the same attractors or differentiated states) are

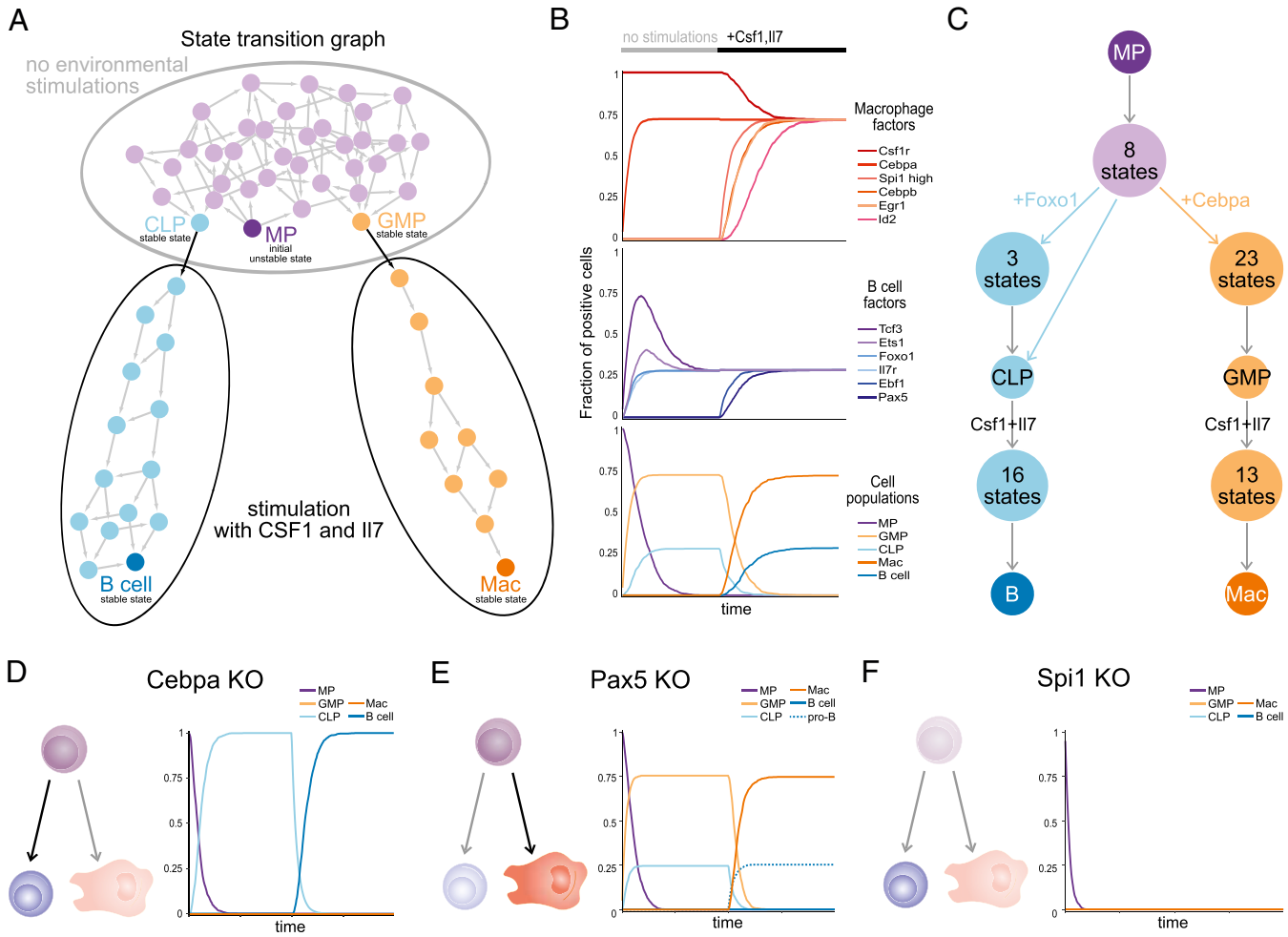


Fig. 5. (*A*) State transition graph generated by simulating the model starting from the unstable MP state in the absence of cytokine (*Upper*) and after the addition of CSF1 and IL7 (*Lower Left* and *Lower Right*). Nodes denote states, and arrows represent transitions between states. (*B*) Stochastic simulations showing the evolution over time, before and after cytokine exposition, of the fractions of cells expressing specific macrophage factors (*Top*), B-cell factors (*Middle*), and cell-type signatures (*Bottom*). The *x* and *y* axes represent time (in arbitrary units) and fractions of positive cells, respectively. (*C*) Hierarchical transition graph corresponding to the state transition graph in *A*. Nodes represent clusters of states, and arrows denote the possible transitions between the clusters. The labels associated with the edges highlight the crucial transitions involved in the decision between B-cell and macrophage specifications. (*D–F*) Schematic representations and stochastic simulations of the effects of Cebpa knockout (*D*), Pax5 knockout (*E*), or *Spi1* knockout (*F*) on the differentiation of MPs, compared with the wild-type situation in *A* and *B*. In the cartoons, the wild-type stable states (cell types) and transitions that are lost in each mutant are displayed using light gray arrows and shading. MP, B cells, and macrophages are represented in purple, blue, and red, respectively.

clustered in a single node (Fig. 5*C*). Interestingly, this analysis shows that the cell decision between GMPs and CLPs depends mainly on the concurrent activation of Cebpa and Foxo1, emphasizing the importance of these factors in early hematopoietic progenitor specification.

Simulation of Documented Genetic Perturbations. Next, we simulated the effects of well-documented gene loss-of-function experiments on progenitor cell specification. Our simulations faithfully recapitulated the effects of various published gene-ablation experiments (Dataset S5). For example, Cebpa knockout in MPs results in the loss of the stable states associated with GMPs and macrophages (Fig. 5*D*), in agreement with the reported impact in vivo (33). Pax5 knockout does not affect the formation of the progenitors but blocks the development of the B-cell lineage at the pro-B stage and prevents the acquisition of the terminal B-cell marker Cd19 (Fig. 5*E*), in agreement with published experimental data (34).

However, the simulation of *Spi1* knockout does not reproduce the reported viability of B cells in *Spi1*-knockout mice (35). This discrepancy arose because, in our model, *Spi1* is required for the expression of the B-cell factors E2a, Ebf1, and Il7r. Introducing additional cross-activations between the B-cell factors and releasing the requirement of Runx1 for Ebf1 up-regulation and of Mef2c for Il7r activation could rescue the expression of the B-cell factors. When we refined the corresponding rules accordingly (Dataset S3), the resulting model showed a stable state corresponding to B-cell patterns in the *Spi1*-knockout condition. However, such patterns cannot be reached from a $Spi1^{-/-}$ MP state, because the cells end up with a complete collapse of gene expression (Fig. 5*F*).

Dynamical Analysis of Transdifferentiation. Next, we analyzed in silico the transdifferentiation of pre-B cells into macrophages upon C/EBPα induction. We first simulated the behavior of B cells under a permanent induction of C/EBPα in the presence of CSF1 and IL7. The system converged toward a single stable state corresponding to macrophages, which does not further require induction of exogenous C/EBPα (Fig. S3*A*), in accordance with published reports (5).

We then focused on the effect of transient inductions of C/EBPα. We have previously shown with our β-estradiol–inducible pre–B-cell

line that a 24-h induction of C/EBPα followed by washout of the inducer was sufficient to trigger irreversible reprogramming (36). Shorter inducer exposure times led to the formation of two populations: one converting into macrophages, and the other initiating transdifferentiation but returning to a B-cell state. A simulation of this process testing all possible pulse durations at once (*Materials and Methods*) confirms that, depending on the duration of C/EBPα induction, B cells can be reprogrammed to macrophages or can go back to a B-cell state (the state transition graph for such simulations cannot be displayed because it contains more than 30,000 states).

Aiming at identifying the commitment point of reprogramming, we further analyzed the resulting hierarchical transition graph (Fig. 6*A*). Because endogenous Cebpa becomes activated very late during transdifferentiation (at about 48 h; see Fig. S3*B*), notably after the commitment point (~24 h), we focused on the Cebpa$^-$ states (i.e., with Cebpa = 0) leading to the sole macrophage stable state (Fig. 6*A*, *Lower*). Some of these states expressed Foxo1, suggesting that the inhibition of Cebpa by Foxo1 can be overcome, in contrast with what happens during the specification of GMPs and CLPs from MPs (Fig. 5*C*). Interestingly, we found that these Cebpa$^-$ states show low constraints on B-cell factors, because only Pax5 must be down-regulated. Furthermore, all Cebpa$^-$ states expressed Cebpb and *Spi1* at a high level, whereas Pax5 was the only B-cell factor required to be inactivated. Finally, some states were found to be Csf1r$^-$, but only when Gfi1 is silenced (along with its activator Ikaros, at least when its repressor Egr2 is not expressed), because Gfi1 can block high *Spi1* expression (21).

Turning to stochastic simulations, we observed the expected loss of B-cell and gain of macrophage phenotypes for both permanent and transient C/EBPα-induced expression (Fig. S3*C*). However, these more quantitative simulations also revealed some inconsistencies. (*i*) Cebpa is reactivated very rapidly; this discrepancy can be circumvented by lowering the kinetic rate of Cebpa up-regulation. (*ii*) The timing of the repression of B-cell genes and that of the loss of CD19 marker roughly coincide; however, we observed that B-cell genes are transcriptionally repressed very rapidly (after 3 h; see Fig. S1*C*), whereas CD19 protein is lost only after 24 h (36). (*iii*) Our model also does not properly capture the fact that short C/EBPα pulses result in the loss of CD19$^+$ cells, which are regained after Cebpa inactivation (36), suggesting that reversion of reprogramming is possible after short induction.

The last two points suggest that B-cell TFs are rapidly down-regulated at the transcriptional level but that the corresponding proteins are retained in transdifferentiating cells for longer times, facilitating reversion of the reprogramming. To address this possibility, we performed a ChIP-seq for Ebf1 at several time points upon permanent induction of Cebpa. Indeed, although Ebf1 RNA decreased by 50% after 3 h of C/EBPα induction (Fig. S1*C*), we observed that Ebf1 binding was lost only after 24 h of induction (Fig. 6*B*). We therefore added a delay in B-cell factor protein degradation to our model (*Materials and Methods*), resulting in a better fit with the observed timing of events during transdifferentiation for both permanent and transient C/EBPα induction (Fig. 6*C*).

In conclusion, our analysis suggests an important role for the Egr2-Gfi1-PU.1– and C/EBPβ-PU.1–positive loops in the irreversible commitment during transdifferentiation and emphasizes the importance of the balance between protein degradation and transcriptional regulation kinetics in the reversibility of the reprogramming.

Simulations of Combined Perturbations During Transdifferentiation. Finally, we analyzed the effects of various TF gain-/loss-of-functions on Cebpa-induced reprogramming, combining C/EBPα induction with a knockdown of *Spi1* or Cebpb or with a constitutive expression of E2a, Ebf1, Pax5, Foxo1, or Gfi1 (Fig. 7). As previously shown (26), only the *Spi1* knockdown is able to block

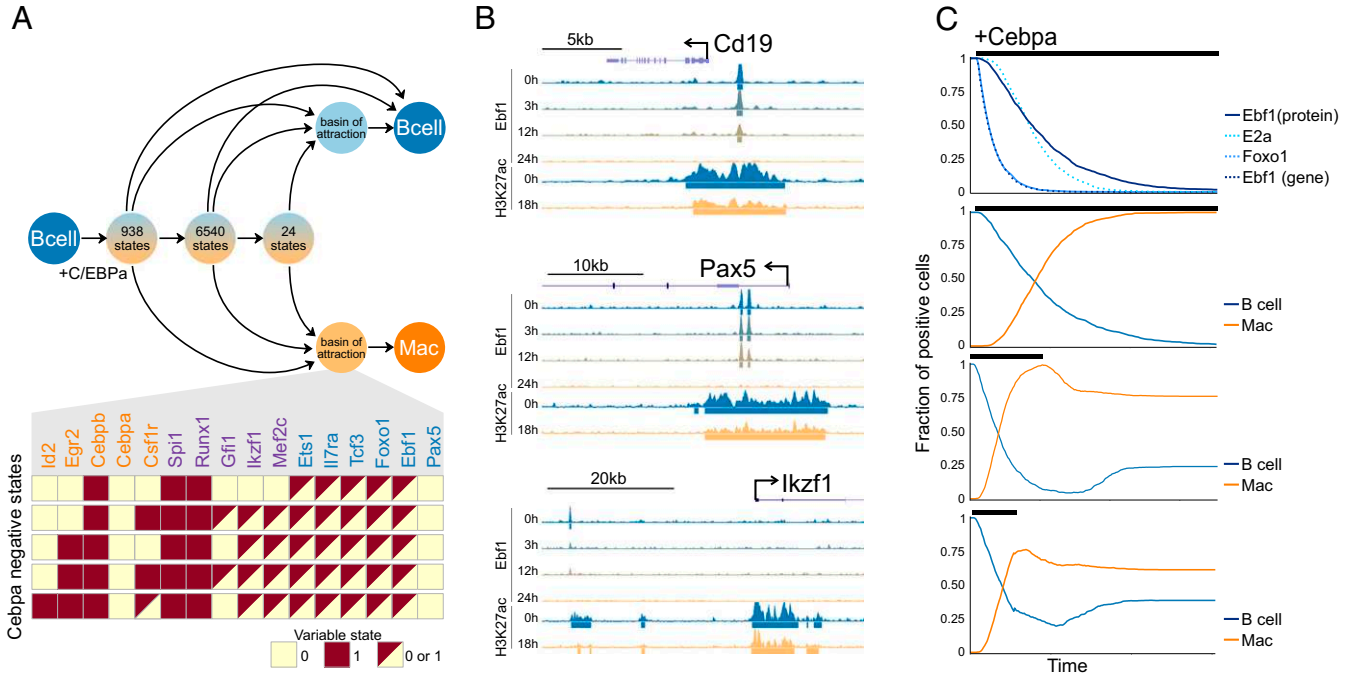


Fig. 6. (*A*, *Upper*) Hierarchical transition graph of the simulation of B-cell transdifferentiation upon transient C/EBPα expression, taking into account all possible C/EBPα pulse durations. Nodes represent clusters of states, and arcs correspond to transitions between these clusters. (*Lower*) Cebpa$^-$ states (rows) of the basin of attraction of the macrophage stable state. (*B*) ChIP-seq signals and peaks (under signal) in B cells (in blue, time point 0 h) and after induction of C/EBPα (at 3, 12, and 24 h). The vertical axes represent RPM (maximum, 5 RPM). (*C*) Stochastic simulations of the fraction of cells expressing different B-cell factors (*Top Panel*) and cell-population signatures (*Lower Three Panels*) during transdifferentiation upon permanent (*Upper Two Panels*) or transient (*Lower Two Panels*) C/EBPα ectopic expression. The corresponding induction durations (in arbitrary units) are indicated by the black lines above each panel.

transdifferentiation fully under permanent induction of C/EBPα. The analysis of the HTG obtained for transient C/EBPα induction in *Spi1*$^{-}$ cells indicates that, when Ebf1 is inhibited, the expression of all genes collapses (Fig. S3*D*). Cebpb knockdown does not block pulse-induced transdifferentiation, but then all committed cell states become Cebpa^{+}, suggesting that Cebpb knockdown could impair commitment if C/EBPα induction is stopped before the reactivation of endogenous Cebpa; this hypothesis remains to be tested. Interestingly, the simulation of the constitutive expression of both Foxo1 and Pax5 results in states expressing a mixture of myeloid and lymphoid genes, pointing toward aberrant reprogramming (Fig. S3*E*).

Discussion

Models of regulatory networks are classically built from detailed reviews of the literature. Despite the massive use of high-throughput assays in the last decade, taking advantage of such data for the construction of new models or to improve preexisting ones remains challenging. Here we combined a meta-analysis of ChIP-seq data with a dynamical model analysis to uncover important regulations. Such a meta-analysis requires an extensive manual curation of the datasets.

It would be tempting to explore the different logical rules in a more unsupervised way by building all possible models with all combinations of regulations and testing their accuracy in silico. Although this approach has been used previously (37), it can be applied only to a subset of possible combinations (e.g., testing the addition or removal of regulations under a general logical rule, such as requiring all activators but none of the inhibitors to enable the activation of a component) and impose certain technical constraints (e.g., limitation to Boolean variables or to synchronous updating). In this study, we first built a model based on published data and then used it to identify caveats in our current knowledge; these caveats then were addressed by exploiting relevant high-throughput datasets.

Our integrative modeling approach enabled us to clarify several aspects of the regulatory network controlling lymphoid and myeloid cell specification. First, although E2a was known to be a master regulator of lymphoid cell specification [required for both B- and T-cell specification (38)], the mechanism of its activation remained unclear, as did the mechanism of its repression in myeloid cells. In this respect, our analysis points to Ikaros as a main activator that is itself activated by Mef2c during lymphoid differentiation and repressed by Cebpa during myeloid differentiation.

In our model, Flt3 is considered a mere marker of multipotent/lymphoid progenitors. Although the Flt3 pathway has been shown to be required for lymphoid development and, more particularly, for the expansion of the CLP population, its impact on cell fate (i.e., beyond proliferation and cell survival) remains unclear. Likewise, ectopic Flt3 signaling has been shown to inactivate C/EBPα through posttranslational modifications (39), but it is unclear whether this inactivation occurs in physiological conditions.

The Egr2 and Gfi1 cross-inhibitory circuit has been shown to be important in the early decision between macrophage and B-cell fates (21). Our analysis suggests that this circuit becomes irrelevant after B-cell commitment, enabling high expression of Egr2 in both pre-B and mature B cells. We therefore proposed that Pax5 can act as an activator of both factors, allowing their coexpression, although it is possible that other factors are involved also.

Concerning the regulation of Cebpa, our work emphasizes the absence of known repressors in lymphoid cells. Ebf1 has been proposed to fulfill this function (40). However, the facts that CLPs lack myeloid potential and show no Cebpa expression and that a depletion of IL7R impedes the activation of Ebf1 but still allows B-cell specification until the pre-B stage (which is devoid of myeloid potential) suggest that another factor acting more upstream represses Cebpa. Mef2c has been shown to counteract myeloid potential (15), but we could not detect any binding at the Cebpa locus. We therefore proposed Foxo1 as a candidate repressor. Thus, according to our model, commitment during normal differentiation of MPs would be controlled mainly by the Cebpa–Foxo1 cross-inhibitory circuit. Hence, Foxo1$^{-/-}$ CLPs could show some myeloid potential. However, other factors could be involved also. In particular, the delay in Cebpa re-expression during reprogramming (long after Foxo1 inactivation) suggests an additional mechanism, possibly involving epigenetic modifications.

Constitutive expression | Pulse | Knock-down

Genotype								Phenotype		Ref
Cebpa	Spi1	Cebpb	Gfi1	E2a	Ebf1	Foxo1	Pax5	Model	Experiment	
								Mac	Mac	5
								B cell or Mac	Mac	26
								all 0	Dead	36
								B cell or all 0	B cell (24h pulse)	26
								Mac	Mac	36
								B cell or Mac	?	26
								Mac	?	
								B cell or Mac	?	
								Mac	Mac	26
								B cell or Mac	Mac (24h pulse)	26
								Mac	Mac	26
								B cell or Mac	Mac (24h pulse)	26
								Mac	?	
								B cell or mixed state	?	
								Mac	Mac	26
								Bcell or mixed state	B-cell (24h pulse)	26

Fig. 7. Table summarizing the impact of selected perturbations (knockin or knockout) on B-cell transdifferentiation into macrophages (Mac) upon either a permanent or a transient induction of C/EBPα. Orange boxes represent macrophages, blue boxes B cells, gray boxes all 0 stable states or cell death. Two-color boxes denote alternative outcomes (stable states).

Materials and Methods

ChIP-Seq Meta-Analysis. ChIP-seq data were collected from public databases (Gene Expression Omnibus), and SRR (sequenced reads run) accession numbers were gathered in Dataset S2 and were automatically downloaded using the Aspera Connect browser plug-in. SRA (Sequence Read Archive format) files were converted in FASTQ using fastq-dump and were mapped onto the mouse mm10 genome using STAR version 2.4.0f1 (41) (see parameters in *SI Materials and Methods*). Duplicated reads were removed using picard (broadinstitute.github.io/picard/). Bigwig tracks were made using Deeptools bamcoverage (42). Peak calling was performed using macs2 (43). Gene domains were defined as in ref. 23, and promoter regions were defined as the TF start site −5 kb/+1 kb, extended up to the next promoter regions or up to 1 Mb in the absence of other promoter regions. Peaks to gene domain associations were performed using R.

Gene Network Modeling and Simulations. The logical model of hematopoietic cell specification was built using GINsim version 2.9 software (44), which is freely available from ginsim.org. All logical simulations (leading to state transition graphs and hierarchical transition graphs) and computation of stable states were performed with GINsim. Stochastic simulations of cell populations were performed using MaBoSS (31). More detailed information can be found in *SI Materials and Methods*. The model can be downloaded from the BioModels database under accession number 1610240000 and from the logical model repository on GINsim website (ginsim.org).

Cell Culture. HAFTL (pre-B) cells and the C/EBPα-ER–containing cell derivative C10 were grown in Roswell Park Memorial Institute (RPMI) medium with L-glutamine supplemented with 10% (vol/vol) FBS, 1× penicillin/streptomycin, and 50 μM β-mercaptoethanol. The RAW 264.7 (ATCC TIB-71) macrophage cell line was grown in DMEM with L-glutamine supplemented with 10% FBS and 1× penicillin/streptomycin.

Western Blot. Western blots were performed using C10 cells and RAW cells as previously described (22). More information can be found in *SI Materials and Methods*. The following antibodies were used at dilution of 1:1,000: Gfi1 (6C5 ab21061; Abcam), Egr2 (EPR4004 ab108399; Abcam), Egr1 (s-25, sc-101033; Santa Cruz), and GAPDH (6C5 sc-32233; Santa Cruz).

ChIP-Seq. ChIP-seq experiments were performed as described previously (45). DNA libraries were prepared using Illumina reagents and instructions and were sequenced on an Illumina Hi-Seq 2000 system. Data are available on the Gene Expression Omnibus (GEO) database under accession codes GSE86420 (Ebf1 and Foxo1 ChIP-seq) and GSM1290084 (previously published Cebpa ChIP-seq in Cebpa-induced B cells).

Ectopic Expression of TFs and Gene-Expression Quantitative PCR. Forced expression of the B-cell TF Foxo1 in RAW cells was performed using retrovirus. More information can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank the staff of the computing platform at the Institut de Biologie de l'Ecole Normale Supérieure for support in hardware and software maintenance; the flow cytometry facility at the Center for Genomic Regulation (CRG)/Universitat Pompeu Fabra for help with cell sorting; the genomics facility of the CRG for sequencing; and Anna Niarakis, Ralph Stadhouders, and Tian Tian for helpful comments regarding earlier versions of this paper. S.C. is supported by a scholarship from the French Ministry of Superior Education and Research. J.L.S.O. was supported by a grant from the Ministry of Economy, Industry, and Competitiveness (MINECO) (IJCI-2014-21872). B.D.S. was supported by a long-term fellowship from the European Molecular Biology Organization (EMBO) (#ALTF 1143-2015). The T.G. laboratory was supported by Grant 282510 from the European Union Seventh Framework Program BLUEPRINT and Fundacio la Marato TV3. This work also was supported by the Spanish Ministry of Economy and Competitiveness, Centro de Excelencia Severo Ochoa 2013–2017 and the Centre de Rerserca de Cataluna (CERCA) Programme, Generalitat de Catalunya.

1. Orkin SH, Zon LI (2008) Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* 132(4):631–644.
2. Graf T, Enver T (2009) Forcing cells to change lineages. *Nature* 462(7273):587–594.
3. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–676.
4. Laiosa CV, Stadtfeld M, Graf T (2006) Determinants of lymphoid–myeloid lineage diversification. *Annual Rev Immunol* 24:705–738.
5. Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. *Cell* 117(5):663–676.
6. Zhang P, et al. (2004) Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity* 21(6):853–863.
7. Bonzanni N, et al. (2013) Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics* 29(13):i80–i88.
8. Abou-Jaoudé W, et al. (2015) Model checking to assess T-helper cell plasticity. *Front Bioeng Biotechnol* 2:86.
9. Krumsiek J, Marr C, Schroeder T, Theis FJ (2011) Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS One* 6(8):e22649.
10. Schütte J, et al. (2016) An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *eLife* 5:e11469.
11. Naldi A, et al. (2009) Logical modelling of regulatory networks with GINsim 2.3. *Biosystems* 97(2):134–139.
12. McKercher SR, et al. (1996) Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *EMBO J* 15(20):5647–5658.
13. Yoshida T, Ng SY-M, Zuniga-Pflucker JC, Georgopoulos K (2006) Early hematopoietic lineage restrictions directed by Ikaros. *Nat Immunol* 7(4):382–391.
14. Ikawa T, Kawamoto H, Wright LYT, Murre C (2004) Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* 20(3):349–360.
15. Stehling-Sun S, Dade J, Nutt SL, DeKoter RP, Camargo FD (2009) Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor Mef2c. *Nat Immunol* 10(3):289–296.
16. Zandi S, et al. (2008) EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J Immunol* 181(5):3364–3372.
17. Horcher M, Souabni A, Busslinger M (2001) Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity* 14(6):779–790.
18. Guo H, Ma O, Speck NA, Friedman AD (2012) Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood* 119(19):4408–4418.
19. Krysinska H, et al. (2007) A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c-fms gene. *Mol Cell Biol* 27(3):878–887.
20. Mossadegh-Keller N, et al. (2013) M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature* 497(7448):239–243.
21. Laslo P, et al. (2006) Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* 126(4):755–766.
22. Spooner CJ, Cheng JX, Pujadas E, Laslo P, Singh H (2009) A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. *Immunity* 31(4):576–586.
23. McLean CY, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28(5):495–501.
24. Zarnegar MA, Rothenberg EV (2012) Ikaros represses and activates PU.1 cell-type-specifically through the multifunctional Sfpi1 URE and a myeloid specific enhancer. *Oncogene* 31(43):4647–4654.
25. Di Stefano B, et al. (2016) C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat Cell Biol* 18(4):371–381.
26. van Oevelen C, et al. (2015) C/EBPα activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep* 5(2):232–247.
27. Jones LC, et al. (2002) Expression of C/EBPbeta from the C/ebpalpha gene locus is sufficient for normal hematopoiesis in vivo. *Blood* 99(6):2032–2036.
28. Bérenguier D, et al. (2013) Dynamical modeling and analysis of large cellular regulatory networks. *Chaos* 23(2):025114.
29. Di Tullio A, et al. (2011) CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc Natl Acad Sci USA* 108(41):17016–17021.
30. Mansson R, et al. (2012) Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc Natl Acad Sci USA* 109(51):21028–21033.
31. Stoll G, Viara E, Barillot E, Calzone L (2012) Continuous time Boolean modeling for biological signaling: Application of Gillespie algorithm. *BMC Syst Biol* 6(1):116.
32. Ye M, et al. (2003) Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. *Immunity* 19(5):689–699.
33. Zhang DE, et al. (1997) Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci USA* 94(2):569–574.
34. Nutt SL, Heavey B, Rolink AG, Busslinger M (1999) Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* 401(6753):556–562.
35. Ye M, Ermakova O, Graf T (2005) PU.1 is not strictly required for B cell development and its absence induces a B-2 to B-1 cell switch. *J Exp Med* 202(10):1411–1422.
36. Bussmann LHL, et al. (2009) A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* 5(5):554–566.
37. Dunn S-J, Martello G, Yordanov B, Emmott S, Smith AG (2014) Defining an essential transcription factor program for naïve pluripotency. *Science* 344(6188):1156–1160.
38. Xu W, et al. (2013) E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. *Blood* 121(9):1534–1542.
39. Radomska HS, et al. (2006) Block of C/EBP α function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. *J Exp Med* 203(2):371–381.
40. Pongubala JMRJ, et al. (2008) Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat Immunol* 9(2):203–215.
41. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.
42. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014) deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42(Web Server issue, W1):W187–91.
43. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
44. Chaouiya C, Naldi A, Thieffry D (2012) Logical modelling of gene regulatory networks with GINsim. *Methods Mol Biol* 804:463–479.
45. van Oevelen C, et al. (2008) A role for mammalian Sin3 in permanent gene silencing. *Mol Cell* 32(3):359–370.
46. Roessler S, et al. (2007) Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Mol Cell Biol* 27(2):579–594.
47. Lara-Astiaso D, et al. (2014) Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345(6199):943–949.
48. Chen X, et al. (1997) Impaired generation of bone marrow B lymphocytes in mice deficient in C/EBP β. *Blood* 90(1):156–164.
49. Didier G, Remy E, Chaouiya C (2011) Mapping multivalued onto Boolean dynamics. *J Theor Biol* 270(1):177–184.

Supporting Information

Collombet et al. 10.1073/pnas.1610622114

SI Materials and Methods

Parameters for Bioinformatic Analysis.

STAR:–outFilterMultimapNmax 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.1–alignIntronMax 1–alignEndsType EndToEnd

Deeptools bamcoverage:–binSize 1–normalizeUsingRPKM–extendReads 200

macs2: -g mm–bw 300 -q 0.01–keep-dup auto –call-summits

Building the Rules of the Logical Model Based on the Literature. Several lines of evidence point toward direct positive regulations of *Spi1* and Ebf1 by the CSF1 and IL7 pathways, respectively (20, 46). Although both pathways are required for the survival of differentiated cells, we believe that their blockage in differentiated cells should not cause dedifferentiation. This hypothesis is supported by reports of cytokine-independent cell lines of macrophages (the RAW 264.7 cell line) and B cells (the HAFTL cell line). Therefore, the rules for *Spi1* and Ebf1 should allow their up-regulation by the corresponding pathways while enabling sustained expression in the absence of pathway ligand.

We thus refined the rule of *Spi1* by enabling it to reach the level 2 in the presence of CSF1r activity, even in the presence of inhibitory complexes (see Dataset S3 for complete logical rules). Regarding Ebf1, we assume that IL7R activity is necessary in the presence of factors known to be expressed early before B-cell specification (i.e., *Spi1*, Runx1, *Ikzf1*, E2a, Ets1, and Foxo1) but becomes dispensable after the activation of Ebf1 itself and Pax5, which then contributes to the maintenance of Ebf1 expression. High levels of PU.1 have been shown to be required for the activation of myeloid genes (21). We therefore set high levels of PU.1 as a requirement for Cebpb but not for CSF1R and Cebpa, because both Csf1r and Cebpa can be activated before *Spi1* (PU.1) up-regulation (6, 20).

It has been shown thatEgr2 and Gfi1 cross-inhibit each other in MPs and that high levels of PU.1 can overcome Gfi1 activity (21). We therefore set Egr2 expression to require PU.1 at a low level in the absence of Gfi1 or to require *Spi1* at a high level in presence of Gfi1.

For *Ikzf1*, the two activators considered are Mef2c and Pax5. Because the second is expressed only in late B-cell differentiation, whereas *Ikzf1* is already expressed in MPs, we set the requirement that one or the other activator is sufficient for *Ikzf1* activation. Similarly, Gfi1 is expressed in B cells and in GMPs, and we therefore set the requirement that C/EBPα or Ikaros is sufficient for its activation.

Given the order of the activation of the B-cell master factors during normal development (E2a l7IL7r → Foxo1 → Ebf1 → Pax5), we did not take into account all their cross-regulations that go back from this order for E2a, Foxo1, and Pax5 (see the rule in Dataset S3).

To test if differences in protein decay for B-cell factors could be important for transdifferentiation irreversibility (Fig. 6 *B* and *C*), we included additional nodes for Ebf1, Pax5, Foxo1, and E2a to distinguish mRNA and protein expression. Because Id2 represses Pax5 and E2a at the protein level (by interacting with them and inhibiting their DNA binding), we also added two nodes to denote E2a and Pax5 proteins specifically in their active forms, on which Id2 acts as a repressor. All these additional nodes can be stripped out automatically in GINsim using the reduction function to obtain a model with only one node per gene or protein but can also be retained for more in depth analysis.

Despite some evidence supporting an auto-activation, we did not include Runx1 explicitly in our model, although Runx1 is involved in several indirect positive loops. Indeed, because Runx1 is regulated only by *Spi1* (PU.1) in our model, adding an auto-regulation would have led us to the rule [*Spi1* & Runx1], meaning that Runx1 cannot go back up after going down, or to the rule [*Spi1* | Runx1], meaning that *Spi1* expression is not strictly necessary for Runx1 expression. Because we had no evidence for the latter rule, we kept *Spi1* expression as necessary for Runx1 expression and therefore did not explicitly consider Runx1 auto-activation.

Discrepancies Caused by Model Discretization. Some discrepancies between gene-expression data and the model stable states can be attributed to a discretization effect. For example, E2a is expressed at a much higher level in CLPs than in B-lineage cells, but we know that it is still active in differentiated B cells. Cebpa and Gfi1 are more highly expressed in GMPs, but although this difference is significant in the context of macrophage versus neutrophil differentiation (21), it is not significant in the case of macrophage versus B-cell specification.

Ebf1 is highly expressed in CLPs in the data from Di Tullio et al. (29) but not in another dataset for hematopoietic progenitors (47), which better fits with our model stable states. This discrepancy might be caused by differences in the exact definition of CLPs, which in ref. 29 presumably includes more B-lineage–primed progenitors.

Surprisingly, Cebpb is expressed in pro- and pre-B cells (Fig. 4*A*) (but not in splenic B cells; see Fig. S2*G*), but it is inactive in the B-cell stable state. Although it is possible that C/EBPβ has a significant role in B cells (48), we consider here that the value 1 corresponds to the higher functional level reached in macrophages.

Finally, Ets1 is much more highly expressed in B cells than in CLPs, although in our model it is already active in CLPs. This result reflects the fact that the only potential regulator of Ets1 for which we found both physical and functional evidences is E2a, which is expressed in CLPs. However, it is possible that Ets1 is already functional at its low level of expression in CLPs.

Logical Simulations. All dynamical simulations were performed using GINsim after using a reduction for factors for which multiple nodes were considered (stripping the following nodes: Mac1_gene, Cebpa_gene, Csf1r_act, Il7r_act, E2a_gene, E2a_prot_active, Foxo1_gene, Ebf1_gene, Pax5_gene, Pax5_prot_active, Cd19_gene) and stripping outputs (i.e., genes that do not regulates any other genes in the model; here, Cd19, Mac1, and Flt3). All simulation parameters are saved in the GINsim file.

Stochastic Simulations Using MaBoSS. All stochastic simulations where performed using MaBoSS (31). Because MaBoSS handles only Boolean variables, the model was first processed with the GINsim function "booleanize model." This process split the *Spi1* variable into two variables, one corresponding to the first *Spi1* threshold and the other corresponding to second *Spi1* threshold. This function automatically recomputes the logic rules for all variables and ensures a complete correspondence in the dynamics of the model (49). The model with all nodes (including gene and multiple protein nodes; see above) was then exported from GINsim into a MaBoSS format file. Simulations were performed using 1,000 simulated cells (-n parameter) and a time tick of 0.1 (-t parameter). All other parameters were left as defaults. Increase and decrease variable rates were all set to 1, except for variables representing proteins (Cd19, Cebpa_ER, Cebpa, E2a, E2a_prot_Active,

Ebf1, Foxo1, Mac1, Pax5, Pax5_prot_active, Csf1r_act, Flt3_act, Il7r_act), which were set at 100,000 (i.e., virtually instantaneous transitions). For the simulations in Fig. 6*C*, the rate of up-regulation of Cebpa_gene (endogenous) was set at 0.1 to account for late Cebpa reactivation, and the rates of down-regulation of Ebf1, Cd19, and Mac1 were set at 0.3 to account for the delay in protein degradation.

For sensitivity analysis, the up and down rate parameters of each component were modified individually by multiplying the rate by various factors (from $1e^{-3}$ to $1e^{3}$), and a simulation of MP to CLP/GMP differentiation was performed (one independent simulation was performed for each variable and for each rate, i.e., 30 simulations for each multiplying factor tested). For all simulations, the time range was set to 10,000 to ensure that the system has reached equilibrium.

Ectopic Expression of TFs and Gene-Expression Quantitative PCR. Forced expression of the B-cell TF Foxo1 in RAW cells was performed as follows. cDNA coding for Foxo1 was cloned in the retroviral plasmid pMSCV-IRES-GFP. pMSCV-Foxo1-IRES-GFP was transfected into the Platinum E retroviral packaging cell line. Supernatants containing the viruses were collected at 24 h and 48 h after transfection and were used to infect RAW cells by spin infection. RNA samples from sorted live GFP-expressing cells were prepared with the RNeasy micro kit (Qiagen). RNA was eluted from the columns using RNase-free water and was quantified by a NanoDrop spectrophotometer (Thermo Scientific). cDNA was produced with the High-Capacity RNA-to-cDNA Kit (Applied Biosystems). RT-quantitative PCR (qPCR) reactions were set up in triplicate with the SYBR Green qPCR Master Mix (Applied Biosystems). Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 58 °C, and 30 s at 72 °C.

The following primers were used:

Foxo1:

AAGAGCGTGCCCTACTTCAA

CTCCCTCTGGATTGAGCATC

Cebpa:

GGCTCCTAATCCCTTGCTTT

TGAACTCACCCAGGAAACCT

Spi1:

GGAGAAAGCCATAGCGATCA

TCTGCAGCTCTGTGAAGTGG

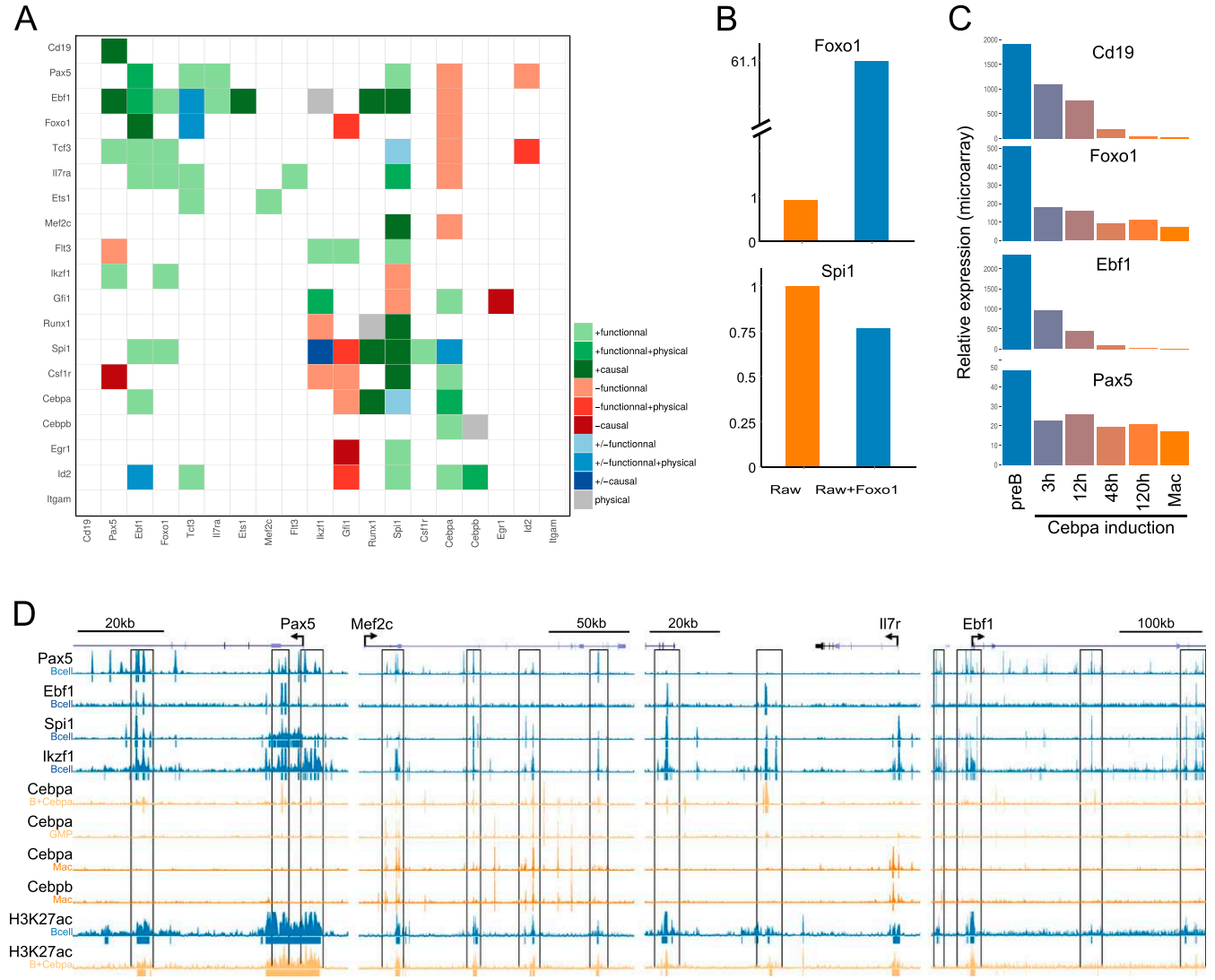


Fig. S1. (*A*) Heatmap showing the regulations inferred from the literature and from ChIP-seq meta-analysis. Green cells denote activations, red cells denote repressions, blue cells denote context-dependent regulations, and gray cells denote regulations supported only by physical evidence. The color intensity denotes the level of confidence of the regulation. (*B*) Gene-expression analysis by qPCR for Foxo1 (*Upper*) and *Spi1* (*Lower*) in macrophages (RAW cell line) before and after ectopic expression of Foxo1. Expression was normalized with respect to GAPDH and to wild-type RAW cells. (*C*) Expression profiles of B-cell genes during trans-differentiation (measured by Affymetrix microarrays) (29). (*D*) Signal from ChIP-seq data targeting different TFs and the histone marker H3K27ac in B cells, in B cells after C/EBPα induction (B+C/EBPα), in GMPs and in macrophages at Pax5, MEf2c, and IL7r loci. Black frames indicate B-cell enhancers presumably inactivated upon Cebpa binding after its induction. The vertical axes represent RPM (maximum: 2 RPM for Ebf1, 5 RPM for other TFs, 3 RPM for H3K27ac).

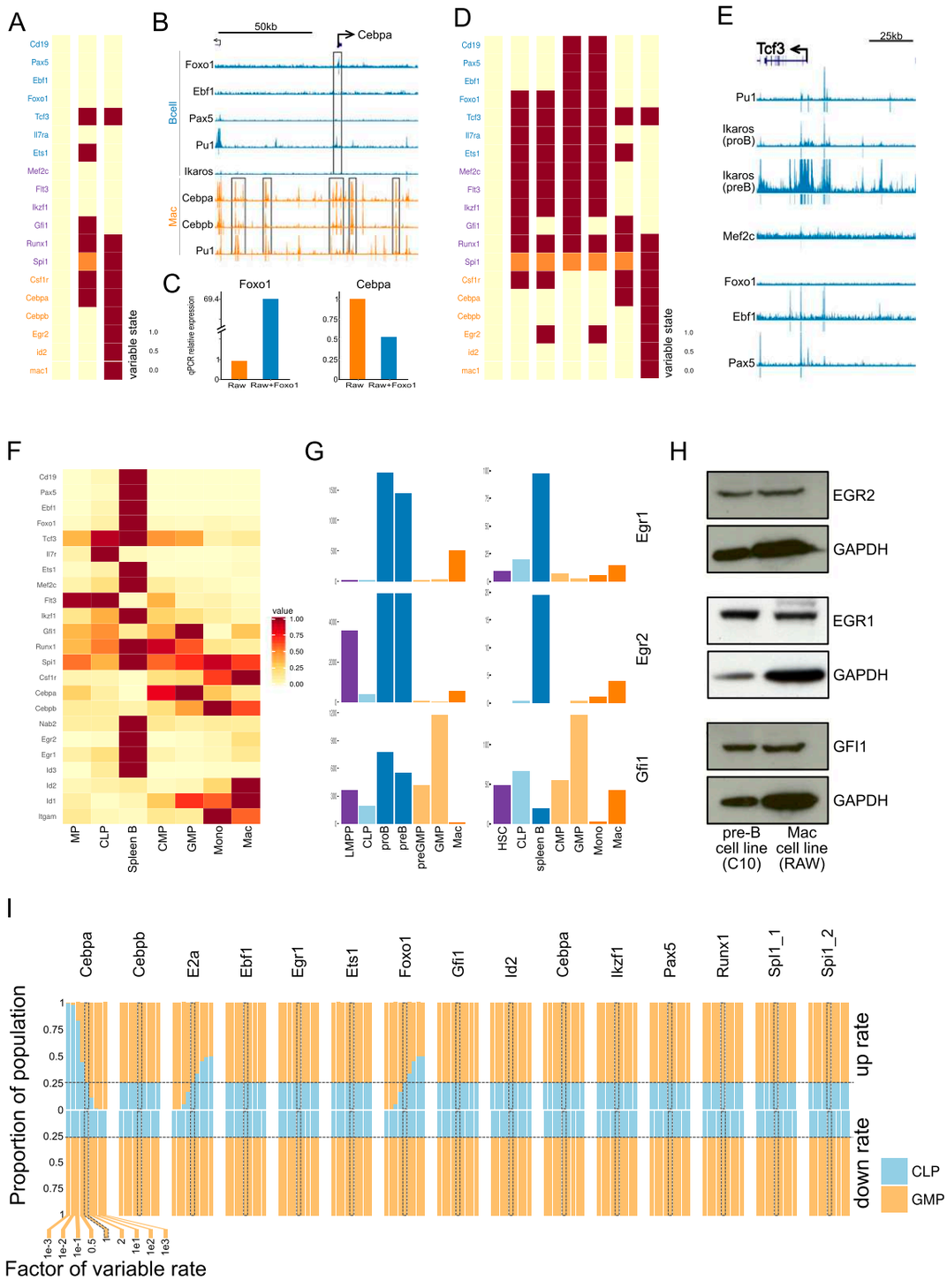


Fig. S2. (*A*) Context-dependent stable states obtained for our first literature-based model computed using GINsim software. The columns list the stable states. A yellow cell denotes the inactivation of the corresponding component (row), a red cell represents a maximal activation (1 for Boolean components, 2 for *Spi1*), and an orange cell represents an intermediate level (1) for *Spi1*. Note that the values of the input nodes (Csf1, Il7) are omitted. (*B*) Signal and peaks (under signal) from ChIP-seq data targeting different TFs in B cells and macrophages at the Cebpa locus. Black frames indicate loci of Foxo1 binding (*Upper*) or of myeloid factors binding (*Lower*). The vertical axes represent RPM (maximum 2 RPM for Ebf1 and Ikaros, 1.5 RPM for Foxo1, 5 RPM for other TFs). (*C*) Gene-expression analysis by qPCR for Foxo1 (*Left*) and Cebpa (*Right*) in macrophages (RAW cell line) before and after ectopic expression of Foxo1. Expression was normalized with respect to GAPDH and to wild-type RAW cells. (*D*) Context-dependent stable states of our model, as in *A*, after the consideration of Cebpa repression by Foxo1. (*E*) Signal and peaks (under signal) of ChIP-seq data targeting different B-cell factors in B cells. The vertical axes represent RPM (maximum: 2 RPM for Ebf1 and Ikaros, 1.5 RPM for Foxo1, 5 RPM for other TFs). (*F*) RNA-seq data (47) for hematopoietic cells corresponding to the stable states of our model. Values are normalized with respect to the highest expression value. (*G*) Gene expression for Egr1, Egr2, and Gfi1 as measured with microarrays (*Left*) (29) and RNA-seq (*Right*) (47). The *y* axis represents normalized average probe intensity for microarray and reads per kilobase of transcript per million reads mapped (RPKM) for RNA-seq. (*H*) Western blot for EGR2, EGR1, GFI1, and GAPDH proteins in pre-B and macrophage cell lines (C10 and RAW, respectively). Note that EGR1 and GFI1 were detected on the same blot, but pictures were taken with different exposure times, so their respective GAPDH control lanes are similar. (*I*) Sensitivity analysis of the proportion of lymphoid and myeloid cells after the differentiation of MPs in the absence of cytokine stimulation. Each bar represents a single simulation in which the up-regulation (*Upper*) or down-regulation (*Lower*) rate of a single variable was changed by more than six orders of magnitudes. The blue section of the bar represents the proportion of CLPs, and the orange sections represent the proportion of GMPs. Simulations with default rates equal to 1 (as in Fig. 5*B*) are highlighted with a gray frame. Gray horizontal lines show the proportions of CLPs and GMPs with default parameters.

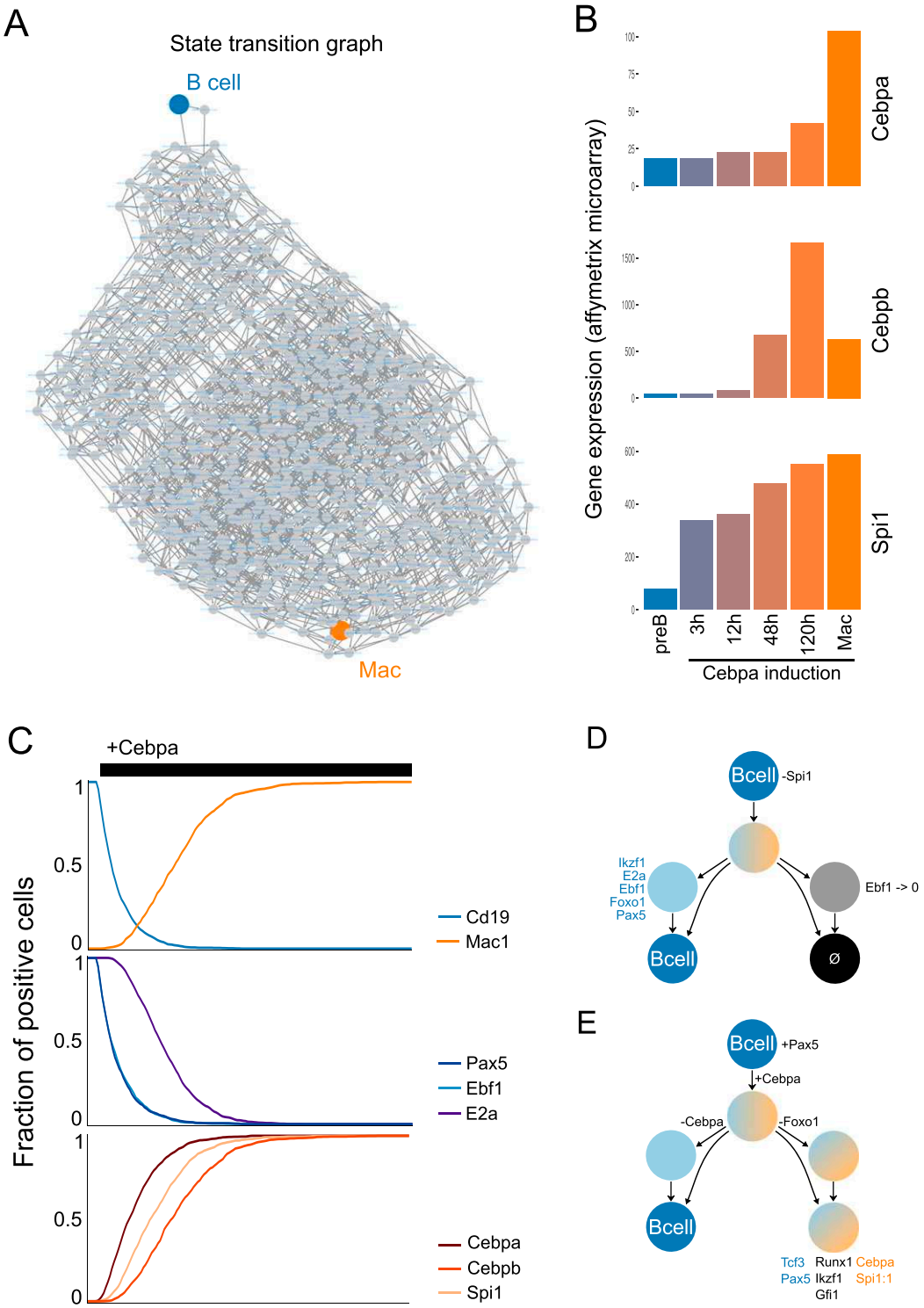


Fig. S3. (*A*) State transition graph obtained by simulating a permanent C/EBPα induction in B cells. Nodes represent states (i.e., vectors of values for all model components), and edges denote enabled transitions between states. The blue node corresponds to B cells, and the orange node corresponds to macrophages. (*B*) Expression profile for Cebpa, Cebpb, and *Spi1* during transdifferentiation as measured by Affymetrix microarrays (29). (*C*) Stochastic simulations of B-cell transdifferentiation upon permanent Cebpa induction showing the temporal evolution of the fractions of cells expressing CD19 or Mac1 markers (*Top*), B-cell TFs (*Middle*), and macrophage TFs (*Bottom*). The *x* and *y* axes represent time (in arbitrary units) and fractions of positive cells, respectively. (*D*) Hierarchical transition graph obtained by simulating a C/EBPα transient induction in *Spi1*-KO B cells. The gray node on the right depicts the basin of attraction of the state in which all the components are inactivated (black node). Ebf1 is inactivated in all states of this basin of attraction and cannot be reactivated. (*E*) Hierarchical transition graph obtained by simulating a C/EBPα transient induction in B cells constitutively expressing Pax5.

Supplementary Table S3: logical rules of the model

Variable	Logical rule
Mac1_gene	Spi1:2
Mac1	Mac1_gene
Id2	Cebpa & Cebpb & Spi1 & !Ebf1 & !Gfi1
Egr1	(Spi1:1 & !Gfi1) | (Spi1:2) | (Pax5_prot_active)
Cebpb	Spi1:2 & (Cebpa | Cebpb)
Cebpa_gene	Spi1 & Runx1 & !Foxo1
Cebpa	Cebpa_gene | Cebpa_ER
Csf1r	Spi1 & !Pax5_prot_active
Csf1r_act	Csf1r & Csf1
Spi1:1	(Spi1 & Runx1 & !(((Cebpa | Cebpb) & Csf1r_act) | (Foxo1 & Ebf1 & !Ikzf1) | !Gfi1)) | (Foxo1 & Ebf1 & Ikzf1 & !(Spi1 | Runx1))
Spi1:2	(Runx1 & Spi1 & (Cebpa | Cebpb) & Csf1r_act) | (Runx1 & Spi1 & Foxo1 & Ebf1 & !Ikzf1) | (Runx1 & Spi1 & !Gfi1)
Runx1	Spi1
Gfi1	((Ikzf1 | Cebpa) & !Egr1) | (Pax5_prot_active)
Ikzf1	Mef2c | Pax5_prot_active
Flt3	Ikzf1 & Spi1 & !Pax5_prot_active
Mef2c	Spi1 & !(Cebpa | Cebpb)
Ets1	E2a_prot_Active
Il7r	(Spi1 & Mef2c & !(Cebpa | Cebpb)) | (Ebf1 & Foxo1 & !(Cebpa | Cebpb))
Il7r_act	Il7r & Il7
E2a_gene	Ikzf1 & (Spi1 | (Ebf1 & Pax5_prot_active))
E2a	E2a_gene
E2a_prot_Active	E2a & !Id2
Foxo1_gene	E2a_prot_Active & !(Cebpa | Cebpb)
Foxo1	Foxo1_gene
Ebf1	Ebf1_gene
Ebf1_gene	(E2a_prot_Active & Foxo1 & Ets1 & Runx1 & Spi1 & Il7r_act & !(Cebpa | Cebpb)) | (E2a_prot_Active & Foxo1 & Ets1 & Pax5_prot_active & Ebf1 & !(Cebpa | Cebpb))
Pax5_gene	Ebf1 & !(Cebpa | Cebpb)
Pax5	Pax5_gene
Pax5_prot_active	Pax5 & !Id2
Cd19_gene	Pax5_prot_active & !Cebpa
Cd19	Cd19_gene

The other supplementary tables are available online.

3.3 Discussion

Although gene network modelling has become more and more popular over the last years, its applications and added value remain obscure to most experimental biologists. Gene network modelling should first be seen as a way to **integrate hypotheses on molecular regulations with data from different sources**, and formally test their consistency.

In that respect, gene network modelling is particularly powerful to **drive the interpretation of high-throughput data,** such as shown for ChIP-seq in our publication. Indeed, omic techniques are very powerful to identify potential factors (or regulatory elements) involved in cell fate, but with poor specificity as they are based on correlation or association. Testing their functionality by experimental approaches would not be realistic given their number. Dynamical modelling can therefore be used to guide data analysis and prioritise experiments.

This is particularly interesting given the explosion of the number of **publicly available dataset** in recent years, which will continue to increase. These datasets often contain numerous evidences for molecular regulations, which have not been described yet. Although some meta-analyses and integration studies have been performed (Lundberg et al., 2016), they remained at a relatively descriptive level, looking at global correlations and not at causal effects. Modelling offers a way to extract more precise information about molecular regulations.

Dynamical modelling can also lead to the identification of **circuits controlling a specific behaviour**. Although it is often tempting to identify minimal circuits, such as cross-inhibitions, it should be remembered that circuits are often intertwined and cannot always be disentangled from each other. For example, we have shown that the commitment toward trans-differentiation, when C/EBPa is withdrawn before activation of the endogenous one, rely on the relations between several circuits: cross-activation between C/EBPb, PU.1 and Runx1, the cross inhibition between Egr2 and Gfi1, and the negative circuit between PU.1, Mef2c, Ikaros and Gfi1.

Predicting higher level behaviour, such as novel **protocols for cell reprogramming**, is also of great interest. Using our model, we have also analysed the reverse reprogramming of macrophages into B cells, and predicted that the ectopic expression of Ebf1, Foxo1 and Gfi1 would allow it (Table 5). However, due to the size of the genes and the difficulties to infect primary macrophages, we were not able assess this prediction experimentally.

It is important to remember that **these predictions are limited by the assumptions of the model**. For example, only a number of factors are selected and many are excluded from a model. In our analysis, we predicted that the ectopic expression of Gfi1 and Foxo1 in macrophages should allow their de-differentiation toward a common progenitor state, which could be re-differentiated toward B cell or macrophages (Table 5). Such prediction seems quite unrealistic in view of the role of Gfi1 in granulocyte differentiation, and is probably due to the absence of other granulocyte factors in our model.

Genotype						Input	Reprogramming
E2a	Ets1	Pax5	Ebf1	Foxo1	Gfi1	Il7	
EE	EE	EE	EE	EE	EE	*	B cell
	EE	EE	EE	EE	EE	*	B cell
EE		EE	EE	EE	EE	*	B cell
EE	EE		EE	EE	EE	*	B cell
EE	EE	EE		EE	EE	1	B cell
EE	EE	EE		EE	EE	*	B cell - like
EE	EE	EE	EE		EE	*	X
EE	EE	EE	EE	EE		*	B cell – PU.1 high
			EE	EE	EE	*	B cell
			EE	EE		*	B cell – PU.1 high
			EE		EE	*	X
				EE	EE	1	de-differentiation re-differentiation
				EE	EE	*	partial de-differentiation
				EE		*	partial de-differentiation
					EE	*	X

Table 5: predictions for the reprogramming of macrophages into B cells

Table summarizing simulations of macrophage reprogramming into B cells. EE: ectopic expression. For Il7, * indicates that the outcome of the simulation was independent of the presence or absence of Il7.

Section 4: Conclusion and perspectives

4.1 Mechanisms of cell reprogramming and cell fate specification

By combining many genomic assays to study epigenetic regulations during the reprogramming of B cells into both macrophages and iPSCs, we have identified some key factors regulating these processes.

We have found that the **shutting down of the B cell program relies on the recruitment of the histone modifiers LSD1 and HDAC1 by C/EBPa** at lymphoid enhancers, leading to their decommissioning.

We also found that the activation of the **macrophage program is controlled by the interplay between C/EBPa, PU.1 and C/EBPb**. In accordance with previous reports of PU.1 acting as a pioneer factor, we found that a number of regulatory elements targeted by C/EBPa were "primed" by PU.1. On the other hand, C/EBPa pioneers other sites, before the recruitment of PU.1. Hence, **different TFs can act as pioneer factors in different DNA or chromatin contexts**.

Finally, we found that C/EBPa also acts as a pioneer factor at key enhancers for subsequent reprogramming into iPSCs by OSKM, **directly activating Klf4**. C/EBPa further **induces the up-regulation of BRD4 and the P-TEFb complex**, which are then recruited by Klf4 to enhancers of pluripotency genes and activates them.

Interestingly, we found that these mechanisms were not just happening in our synthetic system but **occur also in some physiological contexts**. Indeed, myeloid enhancers show a similar dynamic of activation during hematopoiesis. Regions primed by PU.1 in B cell exhibit an active/primed chromatin in hematopoietic stem cells, whereas the regions pioneered by C/EBPa becomes active only when it is up-regulated, in CMP and GMP. The dynamics of activity of these enhancers was also reflected by the expression of their target genes during hematopoiesis.

GMPs have also been shown to be highly susceptible to induction of pluripotency (Guo et al., 2014). We found that **the effects of C/EBPa in pulsed B cell was also happening in GMPs.** Pioneered pluripotency enhancers were also more accessible in GMPs, and the same enhancer of Klf4 was bound by C/EBPa and in contact with the promoter of Klf4. These results point to a common explanation for the elite states of both pulsed B cell and GMPs.

Whether this effect of C/EBPa is a coincidence due to some shared factors between myeloid progenitors and pluripotent cells, or **whether it recapitulates a role for C/EBPa in early development** remains to be elucidated. C/EBPa and C/EBPb have been shown to be involved in placental development (Begay et al., 2004) and it is therefore possible that they play a role in early embryogenesis. On the other hand, the binding motifs of C/EBPa and Oct4 share some similarities (Figure 10), and its

effect could simply be due to shared targets, in the presence of the right co-factors at the right concentration (mainly PU.1).

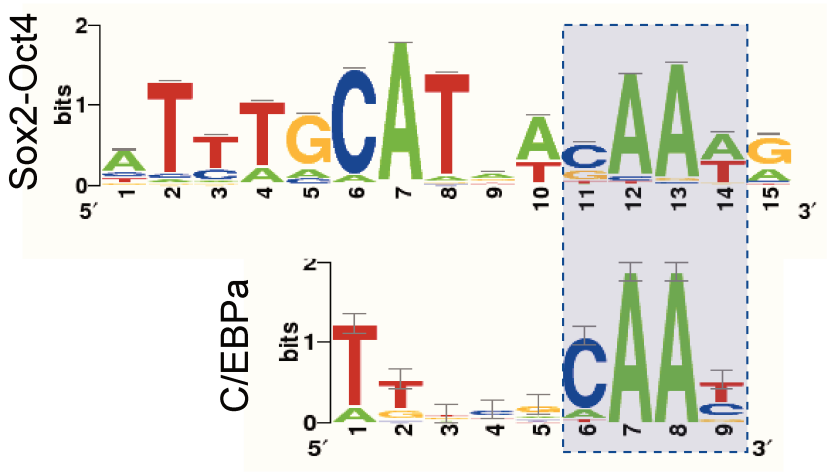


Figure 10: comparison of the binding motif of C/EBPa and Oct4

Comparison of the DNA binding motifs for Oct4 (ID: MA0142.1) and C/EBPa (ID: MA0102.3) from JASPAR (v. 2016).

4.2 Workflows for the integration of epigenetic data

Although large projects combining many different omic assays are becoming more common, there is **no standard methodology for data integration** and each analysis is often performed using dedicated, home-made scripts. During my PhD, I have evaluated the available tools for each step of data analysis (processing, comparison, GO analysis, motif enrichment, etc.) and developed workflows for the integration of omic data. This is summarised in Figure 11.

One of the most popular used method in the study of time course (or in the comparison of many cell types) is **unsupervised clustering**, which enable a global view of the different dynamics. It also offers direct internal control, as functional analyses can be done comparatively between clusters. For example, GO terms enrichment in the different clusters can be compared, to find specific functions associated with each time profile (Yu et al., 2012). Similarly, for ChIP-seq or ATAC-seq, DNA binding motifs specific for different clusters can be discovered by performing differential motif enrichment (Thomas-Chollier et al., 2012).

These functional comparisons are also useful to select the **number of clusters**. Although this can be "optimized" *a priori* (e.g. using the silhouette index (Rousseeuw, 1987)), it is often valuable to use functional biological information. Two clusters might have relatively similar profiles but show distinct enrichment for GO terms. This is particularly true as K-mean (or similar) methods are not very sensitive to detect very small subsets of profiles with particular profile, which might be associated with interesting functions. This was the case in our publication in section 2.2.2, Figure 5a, were the first cluster of regions pioneered by C/EBPa and open in ESCs could only be well isolated when making 4 clusters, as it is much smaller

than the 3 other clusters. Although algorithms such as K-medioids might be less sensitive to such problems, automatic methods cannot be perfect and manual iterations are always valuable.

Finally, although methods such as hierarchical clustering of K-mean are very popular, methods of fuzzy clustering such as C-mean, which we have used in our studies, can be more flexible (Futschik and Carlisle, 2005), where probabilities to belong to each cluster are attributed to genes. Such methods allow testing for different levels of stringency to define clusters, and are less sensitive to outliers by using weighted means (where the contribution of each gene to the mean is weighted by its probability).

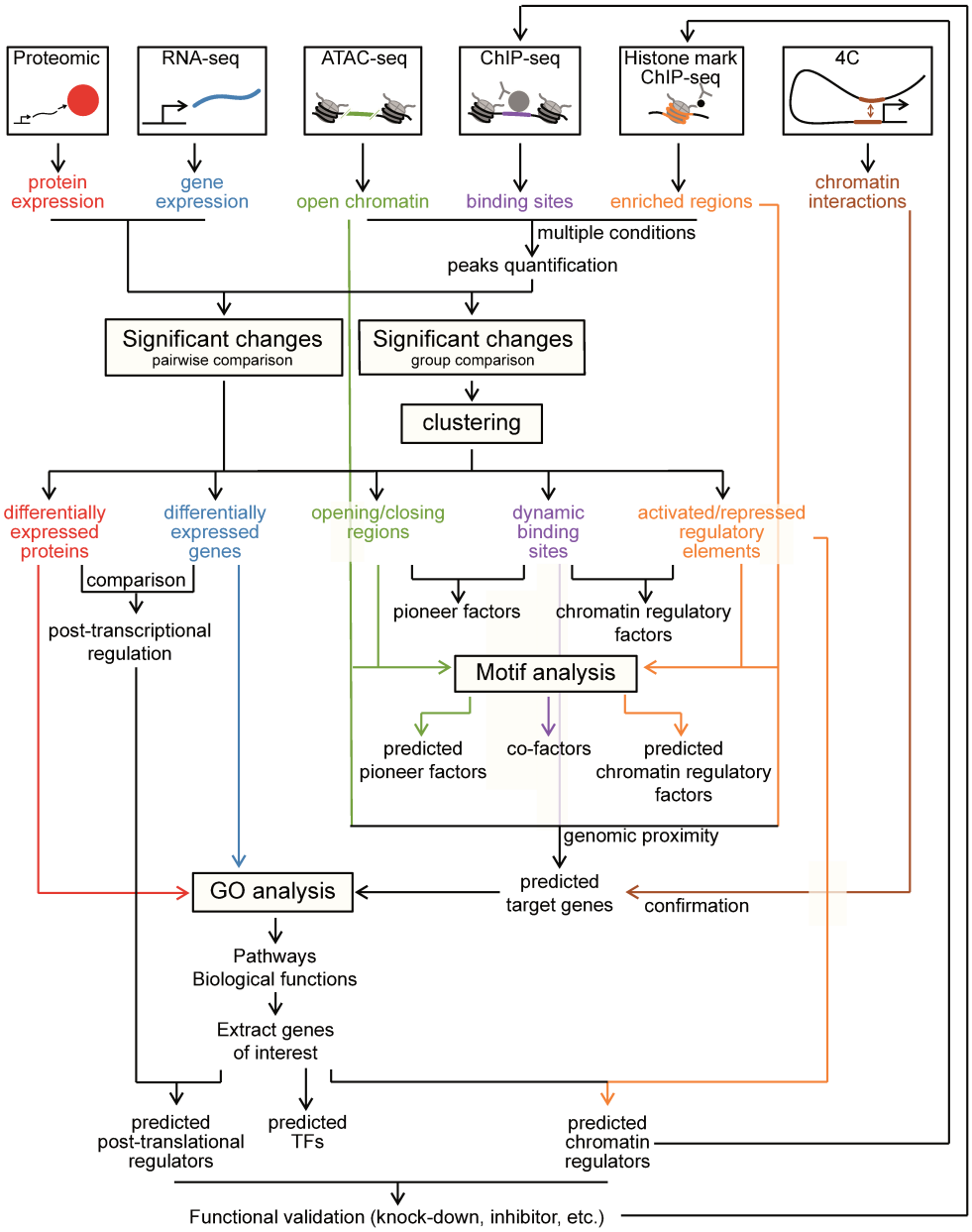


Figure 11: workflows for the integration of omics data

Schematic representation of the workflow I developed for the analysis and integration of omics data.

One major limitation of the integration of multi-omic data is the **classification of different profiles for different features at once** (for example, performing clustering for changes in H3K27ac, H3K4me2 and ATAC-seq altogether). This is hampered by the differences in the size of enriched regions (sharp peaks, broad peaks, domains, etc.) and the large difference in signal amplitude between different features (i.e. the scales of values of quantification). This often leads classification methods to not capture changes in some features, and this problem is not resolved by partitioning methods that usually perform crude normalisation. There is, up to now, no general normalisation methods allowing such comparison.

In our analyses, we have circumvented this problem by performing clustering for each feature separately (e.g. ChIP-seq for C/EBPa), followed by the analysis of the enrichment profiles for the other features (e.g. histone marks) on these clusters. However, this may not capture more subtle combinations of features. Developing more refined methods of classification would be very valuable for future analysis.

4.3 Contribution of omic assays to the deciphering of molecular regulation in cell fate

One challenge of the analysis of omic data is the **extraction of meaningful information** regarding causal molecular effects. In our work presented in section 2.1.2 and 2.2.2, we have tried to use global, genomic correlations to predict molecular mechanism of regulation for specific genes/loci, and further confirm them experimentally using low-throughput assays.

This goal has to be considered already in the **design of experiments and data analysis**, which should be able to lead predictions about specific molecular mechanisms, which can be validated experimentally. This also implies to focus on “high quality” predictions, sometimes to the detriment of global results.

Another major contribution of high-throughput data is the **large number of publicly available datasets** that can be re-analysed. These datasets can be used to compare the mechanisms operating in different systems, similarly to what we have done in section 2.1.2, where we compared activation of myeloid enhancers in B cell trans-differentiation and during haematopoiesis.

Such comparisons are greatly facilitated by **database of processed data** like CODEX, which contains most of the available datasets for gene expression, protein binding and chromatin states in blood cells and embryonic stem cells (Sanchez-Castillo et al., 2015), processed for visualisation. Other tools such as ReMap propose a higher level of analysis, allowing one to predict which TFs can bind to a set of genomic regions based on several hundreds of datasets (Griffon et al., 2015).

Publicly available data can be used to confirm hypothesis about molecular mechanism, e.g. to find potential co-factors by crossing novel and published ChIP-seq for TFs. Another application is the generalisation of a mechanism to other systems, such as we did to compare chromatin changes during reprogramming and hematopoiesis in section 2.1.2.

4.4 Integrating high-throughput data into gene networks

Although the use of **omic data to develop gene network** is very promising, it remains a challenge so far. In section 3.2 we have proposed a method for the integration of ChIP-seq meta-analysis with pre-existing gene network, and how to use the data to refine the model. Similar methods have also been proposed by other groups (Schütte et al., 2016).

The **automatic association of putative enhancers with genes** usually rely upon the assumption of genomic proximity. In that respect, the development of genome-wide 3C assays with enhancer resolution (Sahlén et al., 2015; Schoenfelder et al., 2015) is very promising, but they still lack sensitivity. Methods based on microscopy could circumvent this problem when focusing on a restricted set of genes (Wang et al., 2016). Recently, enhancer knock-out screen using CRISPR-CAS-9 has also been used to identify target genes with high-throughput (Xie et al., 2017).

Single cell approaches are also very promising for the inference of gene regulation. In particular, using **single cell RNA-seq** (scRNA-seq), it is possible to measure the level of expression of a few thousands of genes in several thousands of cells, allowing a very precise estimation of gene partial correlation and therefore to infer possible regulators (Moignard et al., 2015).

However, these techniques are still limited by their sensitivity, which is even more problematic for regulatory network reconstruction as TFs are often expressed at relatively low levels compared to other proteins. Single-cell quantitative PCR (scqPCR) can be used on a selection of genes of interest, to measure with high precision and sensitivity their expression in single cells, with still a relatively high throughput (Hamey et al., 2017).

4.5 Inferring gene regulation and factor combinations

A limiting step in the building of logical model is the **definition of the logical rules**. This is, in fact a general problem in all formalism (like the definition of function shape in ODE). In systems where many experimental evidences are available (as it is the case in haematopoiesis), it is in fact relatively simple to manually infer the rules, based on the literature. Caveats can then be identified with simulations, and several rounds of refinements allow one to define the logical formulas.

Various efforts have been made to develop methods for the **automatic definition of rules**. Some methods use **constraints** (e.g. defining what should be the stable states based on patterns of genes expression in cell type) and identify all the possible **random models** (i.e. different rules combining all the regulators of a variable) that respect these constraints (Dunn et al., 2014). However these methods rely on very strong assumptions (e.g. requiring all activators and no inhibitor to be active, and testing all sets of regulators) and often explore only part of the possibilities (like testing all combinations for one variable, the functions for all other variables being invariants).

Such methods can still be useful when available information does not allow the definition of definitive rules, in particular using hybrid methods where **systematic variations around a set of rules** are analysed.

4.7 Extending our model to other cell types

Our model covers the specification of B cells and macrophages, and partly the regulation of multipotent lympho-myeloid progenitors. It would be interesting to **extend this model to cover other cell types**, in particular T cells, as they can also be reprogrammed into macrophages upon ectopic expression of C/EBPa and PU.1.

Together with a former master student in the lab, we have developed a **preliminary dynamical model of the regulation of T cell specification**, from early progenitors up to the double negative stage. This model is still under development but already recapitulates several key steps of T cell specification. Integrating it, or at least its core network, with our model of B cell / Mac specification, combined with a comparative analysis of chromatin states in trans-differentiation and T cell, should lead to a better understanding of cell fate exclusion between B, T cells and macrophages.

Another cell type of interest would be the **Granulocytes**, as C/EBPa ectopic expression in myeloid progenitors leads to granulocyte and not macrophage differentiation (Zhang et al., 2002). This could be due to the high level of PU.1 in B cells, and incorporating more myeloid factors in our model could help to address this question.

Several models have been developed for the **regulation of pluripotency** (Dunn et al., 2014; Herberg and Roeder, 2015; Xu et al., 2014), and it would be of great interest to integrate the corresponding networks with our B cell network. However, besides the targets of C/EBPa, the possible regulations between pluripotency and hematopoietic factors have never been studied, and more data are required before trying to integrate these two networks.

4.6 Integrating pathways, TFs and chromatin into gene networks

Models of regulatory networks controlling cell fate have been mostly focused on transcription factors, and **external signal** are often integrated in a more phenomenological way rather than a precise modelling of molecular regulations. Although signalling pathways have been studied extensively and many maps are available in databases, integrating them with core networks of TFs remains challenging. This is in part due to the fact that pathways' maps are built from data coming from many different cells types, and are therefore too generic for cell type specific models.

Chromatin regulation is also important for **epigenetic memory**. Indeed, changes in histone modifications can be maintained after transient stimulations, such as inflammation or infection (Ostuni et al., 2013). We have also shown that such memory effect can be maintained through several divisions in mammal cells, with no change of gene expression but only

modifications at the chromatin level (see annexe 5.3). As these processes are intrinsically dynamic, it would be interesting to be able to model them.

This would require an **explicit modelling of enhancers**, which have been so far implicitly incorporated in the logical rules. Furthermore, enhancers not only regulate gene transcription but also the activity of other enhancers. For example, we have shown that an enhancer of Krox20, a TF involved in hindbrain development, is required for the chromatin opening of another enhancer regulating the same gene, and this was required for Krox20 proper expression and brain development (this is described in the annex 5.4). Such phenomena cannot be captured without formally incorporating enhancers in models.

Ultimately, network models should account for **chromatin state information**, which participates to the definition of accessible enhancers for the effectors of signalling pathway. With an increasing knowledge of histone modifications and their regulators, it becomes possible to delineate the potential of signalling pathways in different cellular contexts, which could be integrated into models.

In conclusion, our work has allowed for the identification of novel factors involved in blood cell specification and reprogramming, revealed the mechanism of their regulation during cell reprogramming, and foster our understanding of the role of regulatory circuits in cell fate. We have also developed workflows for the integration of high-throughput data from functional genomic assays, and to combine them with dynamical modelling, methods which could be applied to the study of other biological systems.

References

Abou-Jaoudé, W., Monteiro, P.T., Naldi, A., Grandclaudon, M., Soumelis, V., Chaouiya, C., and Thieffry, D. (2015). Model Checking to Assess T-Helper Cell Plasticity. Front. Bioeng. Biotechnol. *2*, 1–13.

Abou-Jaoudé, W., Traynard, P., Monteiro, P.T., Saez-Rodriguez, J., Helikar, T., Thieffry, D., and Chaouiya, C. (2016). Logical modeling and dynamical analysis of cellular networks. Front. Genet. *7*, 1–20.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L. a M., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295–306.

Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., et al. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell *8*, 376–388.

Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. Sci. Adv. *1*.

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J. (2013). Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. PLoS Comput. Biol. *9*, e1003326.

Barker, N., van Oudenaarden, A., and Clevers, H. (2012). Identifying the Stem Cell of the Intestinal Crypt: Strategies and Pitfalls. Cell Stem Cell *11*, 452–460.

Becker, A.J., McCulloch, E.A., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452–454.

Becskei, A., Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J. *20*, 2528–2535.

Begay, V., Smink, J., and Leutz, A. (2004). Essential Requirement of CCAAT / Enhancer Binding Proteins in Embryogenesis. Mol. Cell. Biol. *24*, 9744–9751.

Bérenguier, D., Chaouiya, C., Monteiro, P.T., Naldi, a, Remy, E., Thieffry, D., and Tichit, L. (2013). Dynamical modeling and analysis of large cellular regulatory networks. Chaos *23*, 25114.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Briggs, R., and King, T.J. (1957). Changes in the nuclei of differentiating endoderm cells as revealed by nuclear transplantation. J. Morphol. *100*, 269–311.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bussmann, L.H.L., Schubert, A., Manh, T.V., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., et al. (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell *5*, 554–566.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Chickarmane, V., Enver, T., and Peterson, C. (2009). Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. PLoS Comput. Biol. *5*, e1000268.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Stefano, B. Di, Brumbaugh, J., Kim, S., et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *1*, 1–7.

Cooper, S., and Brockdorff, N. (2013). Genome-wide shRNA screening to identify factors mediating Gata6 repression in mouse embryonic stem cells. Development *140*, 4110–4115.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., and Scacheri, P.C. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate

levels of gene expression to confer susceptibility to common traits. Genome Res. *24*, 1–13.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. Science (80-.). *295*, 1306–1311.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, a G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

Eguchi, A., Wleklinski, M.J., Spurgat, M.C., Heiderscheit, E.A., Kropornicka, A.S., Vu, C.K., Bhimsaria, D., Swanson, S.A., Stewart, R., Ramanathan, P., et al. (2016). Reprogramming cell fate with a genome-scale library of artificial transcription factors. Proc. Natl. Acad. Sci. U. S. A. *113*, E8257–E8266.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). The Reactome pathway Knowledgebase. Nucleic Acids Res. *44*, D481–D487.

Flobak, Å., Baudot, A., Remy, E., Thommesen, L., Thieffry, D., Kuiper, M., and Lægreid, A. (2015). Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. PLOS Comput. Biol. *11*, e1004426.

Fullwood, M.J., Han, Y., Wei, C.-L., Ruan, X., and Ruan, Y. (2010). Chromatin Interaction Analysis Using Paired-End Tag Sequencing. In Current Protocols in Molecular Biology, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. Unit 21.15.1-25.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. *7*, 703–713.

Gehring, W.J., and Ikeo, K. (1999). Pax 6: mastering eye morphogenesis and eye evolution. Trends Genet. *15*, 371–377.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grieshammer, U., Minowada, G., Pisenti, J.M., Abbott, U.K., and Martin, G.R. (1996). The chick limbless mutation causes abnormalities in limb bud dorsal-ventral patterning: implications for the mechanism of apical ridge formation. Development *122*, 3851–3861.

Griffon, A., Barbier, Q., Dalino, J., van Helden, J., Spicuglia, S., and Ballester, B. (2015). Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. Nucleic Acids Res. *43*, e27.

Guo, S., Zi, X., Schulz, V.P., Cheng, J., Zhong, M., Koochaki, S.H.J., Megyola, C.M., Pan, X., Heydari, K., Weissman, S.M., et al. (2014). Nonstochastic reprogramming from a privileged somatic cell state. Cell *156*, 649–662.

Gurdon, J.B. (1962). Adult frogs derived from the nuclei of single somatic cells. Dev. Biol. *4*, 256–273.

Hackett, J.A., and Azim Surani, M. (2014). Regulatory principles of pluripotency: From the ground state up. Cell Stem Cell *15*, 416–430.

Haeckel, E. (1877). Anthropogenie (Wilhelm Engelmann, Leipzig).

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell *5*, 597–609.

Hamey, F.K., Nestorowa, S., Kinston, S.J., Kent, D.G., Wilson, N.K., and Göttgens, B. (2017). Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. Proc. Natl. Acad. Sci. U. S. A. *114*, 5822–5829.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., et al. (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250–264.

Herberg, M., and Roeder, I. (2015). Computational modelling of embryonic stem-cell fate control. Development *142*, 2250–2260.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A. a, Hoke, H. a, and Young, R. a (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hutchinson, E. (2008). Cancer stem cells: Developmental block. Nat. Rev. Cancer *8*, 160–161.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science (80-.). *316*, 1497–1502.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat. Genet. *49*, 213–222.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.

Kaneko, S., Feldman, R.I., Yu, L., Wu, Z., Gritsko, T., Shelley, S.A., Nicosia, S. V., Nobori, T., and Cheng, J.Q. (2002). Positive feedback regulation between Akt2 and MyoD during muscle differentiation: Cloning of Akt2 promoter. J. Biol. Chem. *277*, 23230–23235.

Kauffman, S. (1969). Homeostasis and Differentiation in Random Genetic Control Networks. Nature *224*, 177–178.

Kleinsmith, L.J., and Pierce, G.B. (1964). Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Res. *24*.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. Stem Cells *26*, 903–911.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

Laiosa, C. V, Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity *25*, 731–744.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Sciammas, R., Gantner, B.N., Dinner, A.R., Singh, H., and Lee, H.-J. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell *126*, 755–766.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Cell Stem Cell *7*, 51–63.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). *326*, 289–293.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., et al. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat. Immunol. *11*, 635–643.

Liu, Z.-P. (2015). Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. Curr. Genomics *16*, 3–22.

Loh, K.M., and Lim, B. (2011). A Precarious Balance: Pluripotency Factors as Lineage Specifier. Cell Stem Cell *8*, 363–369.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lun, A.T.L., and Smyth, G.K. (2015). Csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. *44*.

Lundberg, S.M., Tu, W.B., Raught, B., Penn, L.Z., Hoffman, M.M., and Lee, S.-I. (2016). ChromNet: Learning the human chromatin network from all ENCODE ChIP-seq data. Genome Biol. *17*, 82.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110-5.

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat. Genet. *49*, 296–302.

Matys, V., Kel-Margoulis, O. V, Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res. *34*, D108–D110.

Mauro, A. (1961). Satellite cell of skeletal muscle fibers. J. Biophys. Biochem. Cytol. *9*, 493–495.

Maximow, A.A. (2009). The lymphocyte as a stem cell, common to different blood elements in embryonic development and during the post-fetal life of mammals: Lecture with a demonstration, held at a special meeting of the Berlin Hematological Society on 1 June 1909. Cell. Ther. Transplant. *1*, 14–24.

McKercher, S., Torbett, B., Anderson, K., Henkel, G., Vestal, D., Baribault, H., Klemsz, M., Feeney, A., Wu, G., Paige, C., et al. (1996). Targeted disruption of the PU. 1 gene results in multiple hematopoietic abnormalities. EMBO J. *15*, 5647–5658.

Mehta, A., and Baltimore, D. (2016). MicroRNAs as regulatory elements in immune system logic. Nat. Rev. Immunol. *16*, 279–294.

Meng, F., Du, Z., Federation, A., Hu, J., Wang, Q., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., Casellas, R., et al. (2014). Article Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability. Cell *159*, 1538–1548.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Monod, J., and Jacob, F. (1961). General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. Cold Spring Harb. Symp. Quant. Biol. *26*, 389–401.

Mossadegh-keller, N., Sarrazin, S., Kandalla, P.K., Espinosa, L., Stanley, E.R., Nutt, S.L., Moore, J., and Sieweke, M.H. (2013). M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. Nature *497*, 239–243.

Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227.

Naldi, A., Carneiro, J., Chaouiya, C., and Thieffry, D. (2010). Diversity and plasticity of Th cell types predicted from regulatory network modelling. PLoS Comput. Biol. *6*, e1000912.

Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J. a (2012). Circuitry and dynamics of human transcription factor regulatory networks. Cell *150*, 1274–1286.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20-31.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917–929.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795–801.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631–644.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell *152*, 157–171.

Panovska-griffiths, J., Page, K.M., and Briscoe, J. (2013). A gene regulatory motif that generates oscillatory or multiway switch outputs A gene regulatory motif that generates oscillatory or multiway switch outputs.

Parnas, O., Jovanovic, M., Eisenhaure, T.M.M., Herbst, R.H.H., Dixit, A., Ye, C.J.J., Przybylski, D., Platt, R.J.J., Tirosh, I., Sanjana, N.E.E., et al. (2015). A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell *162*, 1–12.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Pott, S., and Lieb, J.D. (2014). What are super-enhancers? Nat. Genet. *47*, 8–12.

Ptashne, M. (1967). Specific Binding of the λ Phage Repressor to λ DNA. Nature *214*, 232–234.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.-R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity. Cell 1524–1537.

Ramalho-Santos, M., and Willenbring, H. (2007). On the Origin of the Term "Stem Cell." Cell Stem Cell *1*, 35–38.

Remy, E., Ruet, P., and Thieffry, D. (2008). Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework. Adv. Appl. Math. *41*, 335–350.

Ring, A., Kim, Y.-M., and Kahn, M. (2014). Wnt/catenin signaling in adult stem cell physiology and disease. Stem Cell Rev. *10*, 512–525.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell *155*, 39–55.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65.

Sahlén, P., Abdullayev, I., Ramsköld, D., Matskova, L., Rilakovic, N., Lötstedt, B., Albert, T.J., Lundeberg, J., and Sandberg, R. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome Biol. *16*, 156.

Sanchez-Castillo, M., Ruau, D., Wilkinson, A.C., Ng, F.S.L., Hannah, R., Diamanti, E., Lombard, P., Wilson, N.K., and Gottgens, B. (2015). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. Nucleic Acids Res. *43*, D1117–D1123.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-cadete, F., Furlan-magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N.K., Riepsaame, J., Calero-Nieto, F.J., Moignard, V., Basilico, S., Kinston, S.J., et al. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. Elife *5*, 1–27.

Seyednasrollah, F., Koestler, D.C., Wang, T., Piccolo, S.R., Vega, R., Greiner, R., Fuchs, C., Gofer, E., Kumar, L., Wolfinger, R.D., et al. (2017). A DREAM Challenge to Build Prediction Models for Short-Term Discontinuation of Docetaxel in Metastatic Castration-Resistant Prostate Cancer. JCO Clin. Cancer Informatics 1–15.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). Nat. Genet. *38*, 1348–1354.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Soulé, C. (2004). Graphic requirements for multistationarity. Arxiv Prepr. Q-bio/0403033 1–19.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

Stoll, G., Viara, E., Barillot, E., and Calzone, L. (2012). Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. BMC Syst. Biol. *6*, 116.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P., Weintraub, H., and Lassar, A.B. (1988). MyoD1: A Nuclear Phosphoprotein Reqiring a Myc Homology Region to Convert Fibroblasts to Myoblasts. *295*, 405–412.

Temple, S. (1989). Division and differentiation of isolated CNS blast cells in microculture. Nature *340*, 471–473.

Teo, A.K.K., Arnold, S.J., Trotter, M.W.B., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. *25*, 238–250.

The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Thieffry, D., and Thomas, R. (1995). Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda. Bull. Math. Biol. *57*, 277–297.

Thomas, R. (1981). On the Relation Between the Logical Structure of Systems and Their Ability to Generate Multiple Steady States or Sustained Oscillations. Numer. Methods Study Crit. Phenom. *9*, 180–193.

Thomas-Chollier, M., Herrmann, C., Defrance, M., Sand, O., Thieffry, D., and van Helden, J. (2012). RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res. *40*, e31.

Di Tullio, A., and Graf, T. (2012). C/EBPa bypasses cell cycle-dependency during immune cell transdifferentiation. Cell Cycle *11*, 2739–2746.

Di Tullio, A., Manh, T.P.V., Schubert, A., Månsson, R., Graf, T., Di, A., Phong, T., Manh, V., Vu Manh, T.P., Castellano, G., et al. (2011). CCAAT/enhancer binding protein {alpha} (C/EBP{alpha})-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc. Natl. Acad. Sci. U. S. A. *108*, 1–6.

Turatsinze, J.-V., Thomas-Chollier, M., Defrance, M., and van Helden, J. (2008). Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat. Protoc. *3*, 1578–1588.

Waddington, C.H. (1957). The strategy of the genes - a discussion of some aspects of theoretical biology (Routledge Library Editions: 20th Century Science).

Wang, S., Su, J., Beliveau, B.J., Bintu, B., Moffitt, J.R., and Wu, C. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. Science (80-.). *353*, 598–602.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Weedon, M.N., Cebola, I., Patch, A.-M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H.-H., Allen, H.L., et al. (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. *46*, 61–64.

Weismann, A. (1893). The germ-plasm: a theory of heredity.

Whyte, W. a, Orlando, D. a, Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R. a (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517–534.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676.

Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G.C. (2017). Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. Mol. Cell *66*, 285–299.e5.

Xu, H., Ang, Y.-S., Sevilla, A., Lemischka, I.R., and Ma'ayan, A. (2014). Construction and Validation of a Regulatory Network for Pluripotency and Self-Renewal of Mouse Embryonic Stem Cells. PLoS Comput. Biol. *10*, e1003777.

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., and Yamada, Y. (2017). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. Nat. Publ. Gr. 20–22.

Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature *465*, 704–712.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors : establishing competence for gene expression.

Genes Dev. *25*, 2227–2241.

Zeineddine, D., Papadimou, E., Chebli, K., Gineste, M., Liu, J., Grey, C., Thurig, S., Behfar, A., Wallace, V.A., Skerjanc, I.S., et al. (2006). Oct-3/4 Dose Dependently Regulates Specification of Embryonic Stem Cells toward a Cardiac Lineage and Early Heart Development. Dev. Cell *11*, 535–546.

Zhang, P., Nelson, E., Radomska, H.S., Iwasaki-Arai, J., Akashi, K., Friedman, A.D., and Jenen, D.G. (2002). Induction of granulocytic differentiation by 2 pathways. Blood *99*, 4406–4412.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J. a, et al. (2004). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity *21*, 853–863.

Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., et al. (2016a). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell *166*, 1147–1162.e15.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Y., An, L., Yue, F., and Hardison, R.C. (2016b). Jointly characterizing epigenetic dynamics across multiple human cell types. Nucleic Acids Res. gkw278-.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

Ziller, M.J., Edri, R., Yaffe, Y., Gifford, C. a, Xing, J., Gu, H., Kohlbacher, O., and Gnirke, A. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature *518*, 355–359.

Section 5: Annexes

5.1 Publication: C/EBPa poises B cells for rapid reprogramming into induced pluripotent stem cells (Nature, 2014).

Contribution:

I analysed the RNA-seq and gene expression microarray data.

doi:10.1038/nature12885

C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells

Bruno Di Stefano[1,2], Jose Luis Sardina[1,2]*, Chris van Oevelen[1,2]*, Samuel Collombet[3,4,5], Eric M. Kallin[1,2]†, Guillermo P. Vicent[1,2], Jun Lu[6], Denis Thieffry[3,4,5], Miguel Beato[1,2] & Thomas Graf[1,2,7]

CCAAT/enhancer binding protein-α (C/EBPα) induces transdifferentiation of B cells into macrophages at high efficiencies and enhances reprogramming into induced pluripotent stem (iPS) cells when co-expressed with the transcription factors Oct4 (Pou5f1), Sox2, Klf4 and Myc (hereafter called OSKM)[1,2]. However, how C/EBPα accomplishes these effects is unclear. Here we find that in mouse primary B cells transient C/EBPα expression followed by OSKM activation induces a 100-fold increase in iPS cell reprogramming efficiency, involving 95% of the population. During this conversion, pluripotency and epithelial–mesenchymal transition genes become markedly upregulated, and 60% of the cells express Oct4 within 2 days. C/EBPα acts as a 'path-breaker' as it transiently makes the chromatin of pluripotency genes more accessible to DNase I. C/EBPα also induces the expression of the dioxygenase Tet2 and promotes its translocation to the nucleus where it binds to regulatory regions of pluripotency genes that become demethylated after OSKM induction. In line with these findings, overexpression of Tet2 enhances OSKM-induced B-cell reprogramming. Because the enzyme is also required for efficient C/EBPα-induced immune cell conversion[3], our data indicate that Tet2 provides a mechanistic link between iPS cell reprogramming and B-cell transdifferentiation. The rapid iPS reprogramming approach described here should help to fully elucidate the process and has potential clinical applications.

Reprogramming of somatic cells into induced pluripotent stem (iPS) cells by the OSKM transcription factors (also called Yamanaka factors) can be divided into a stochastic and a deterministic phase[4,5]. During this process cells undergo a mesenchymal–epithelial transition (MET) and activate endogenous pluripotency genes, paralleled by changes in histone marks, nucleosome positioning and chromatin accessibility[6]. In addition, their promoters become de-methylated, preceded by oxidation of methylated CpGs through the dioxygenases Tet2 and Tet1, both of which have been implicated in the establishment of pluripotency[7,8]. Earlier work showed that co-expression of C/EBPα with OSKM increases the reprogramming efficiency of B cells ~15-fold, reaching ~3% of the population[1]. Here we describe that a pulse of C/EBPα followed by OSKM overexpression permits the rapid reprogramming of B cells into iPS cells by activating Tet2 and facilitating accessibility of pluripotency gene promoters to Oct4 binding. Highly efficient reprogramming of somatic cells has recently also been reported with a loss-of-function approach[9].

Committed B-cell precursors (hereafter referred to as B cells) can be induced to transdifferentiate into macrophages at 100% efficiency by forced C/EBPα expression, deregulating ~7,500 genes[10]. Reasoning that chromatin of cells in transition might be more 'open' than that of end stages, we tested the effect of transiently exposing B cells to C/EBPα, followed by OSKM expression. B cells were isolated from the bone marrow of reprogrammable mice (containing a tetracycline-controlled transactivator (rtTA) and a doxycycline-responsive OSKM cassette[11]), infected them with C/EBPα-ER-hCD4 retrovirus, sorted human $CD4^+$ cells 4 days later and incubated them for different times with β-estradiol (E2) followed by a wash-out (Fig. 1a). Subsequently, OSKM was induced by doxycycline treatment and $Nanog^+$ colonies scored 12 days post induction (d.p.i.). B cells continuously co-expressing C/EBPα with OSKM showed an 11-fold enhancement in reprogramming efficiency compared to cells induced with OSKM alone (B+OSKM cells), confirming earlier reports[1,12]. In contrast, cells treated for 18 h with E2 and then treated with doxycycline (Bα′+OSKM cells) exhibited a 103-fold colony increase, with cells pulsed for 6 h already showing a 74-fold increase (Fig. 1b–d). Clonal assays showed that 92–94% of viable colonies were $Nanog^+$ after 12 days (Extended Data Fig. 1a–c). OSKM induction of B cells pre-treated with E2 did not increase $Nanog^+$ colony numbers nor *Oct4* expression levels (Extended Data Fig. 2a). A mutant of C/EBPα (BRM-2 (ref. 13)) defective for DNA binding failed to enhance iPS cell generation (Extended Data Fig. 2b). Induced Bα′+OSKM cells remained >50% viable (Extended Data Fig. 2c). C/EBPα pulses after OSKM induction had no effect on reprogramming efficiency (Extended Data Fig. 2d), indicating that C/EBPα acts as a path-breaker for OSKM-induced reprogramming. Stable iPS cell lines derived from Bα′+OSKM cells (αiPS cells) displayed similar gene expression profiles as embryonic stem (ES) cells, differentiated into all three germ layers *in vitro* and *in vivo*, and efficiently contributed to coat colour chimaerism (Fig. 1e and Extended Data Fig. 3a–e).

Bα′+OSKM-derived iPS cell colonies could be identified as early as 4 d.p.i. and their numbers increased modestly after 8 days (Fig. 1f, g), whereas B+OSKM colonies continued to increase 8–10 d.p.i. (Fig. 1f, g). Retroviral hCD4 expression was found to be silenced within 2–4 days in Bα′+OSKM cells compared to ~8 days in B+OSKM cells (Fig. 1h). To test the effect of C/EBPα on OSKM-induced transgene independence, Bα′ and B cells were doxycycline-treated for different times (Fig. 1i). The first transgene-independent iPS cell colonies were observed after 4 days for Bα′+OSKM cells, compared to 9–10 days for B+OSKM controls (Fig. 1j). Together these results show that the C/EBPα pulse accelerates iPS cell reprogramming by 4–6 days.

Gene expression analyses of Bα′+OSKM cells 8 d.p.i. showed the upregulation of 764 out of 1,668 genes expressed more highly in ES cells than in B cells, including all well described pluripotency genes (Fig. 2a). Unsupervised hierarchical clustering analysis revealed that 8 d.p.i. Bα+OSKM cells clustered with ES/iPS cells (Extended Data Fig. 4a). A large part of known pluripotency genes were activated 2–6 d.p.i. (Fig. 2b, c and Extended Data Fig. 4b), reaching levels comparable to αiPS cells and ES cells within ~1 week (Fig. 2c and Extended Data Fig. 4c). In contrast, they remained essentially silent in B+OSKM cells (Fig. 2b, c). In addition, five genes described to be activated very late (21 d.p.i.) during fibroblast reprogramming[14] became upregulated within 2–4 days in Bα′+OSKM cells (Extended Data Fig. 4d). Using Oct4–GFP reporter mice[12] crossed with reprogrammable mice[11], Bα′+OSKM

[1]Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003 Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain. [3]Ecole Normale Supérieure, Institut de Biologie de l'ENS, 45 Rue d'Ulm, Paris F-75005, France. [4]Inserm, U1024, Paris F-75005, France. [5]CNRS, UMR 8197, Paris F-75005, France. [6]Yale Cancer Center, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. [7]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg Lluis Companys 23, 08010 Barcelona, Spain. †Present address: Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York 10065, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

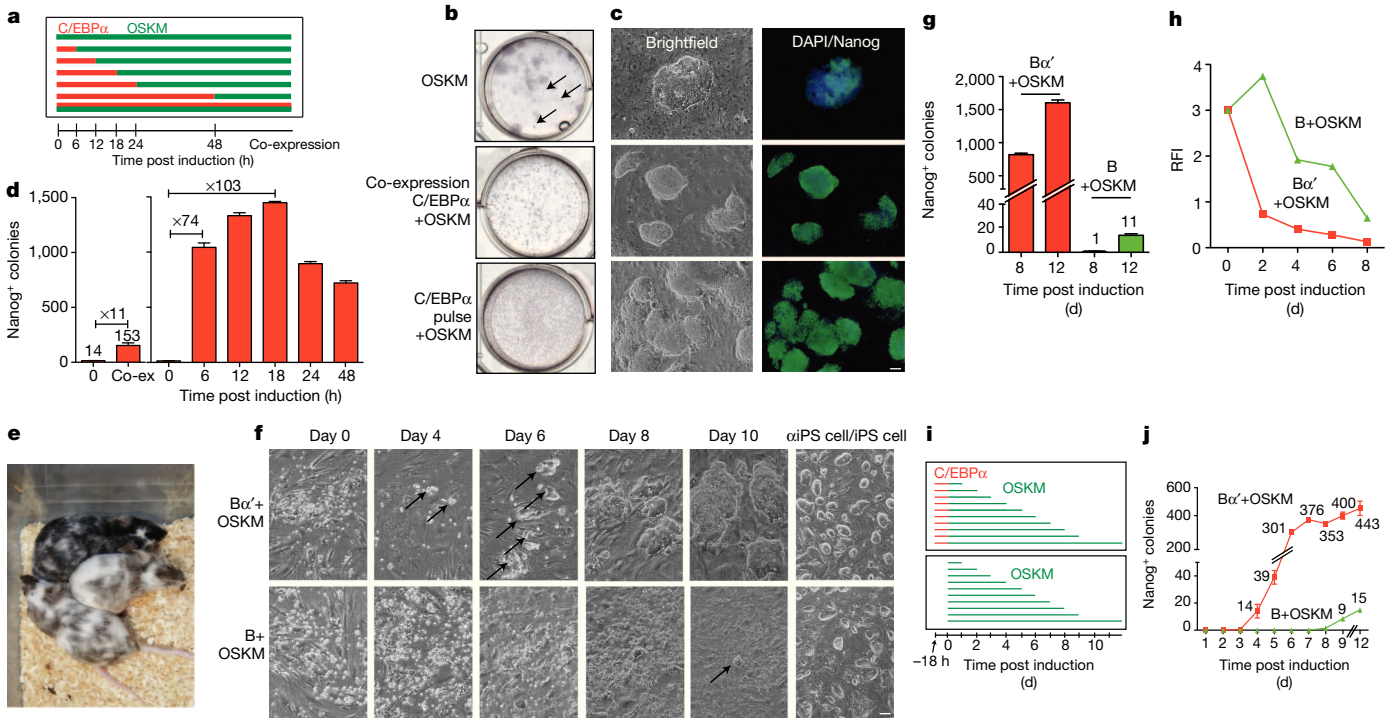


Figure 1 | Effect of transient C/EBPα expression on OSKM-induced iPS cell reprogramming of B cells. **a**, Experimental strategy. **b**, Six-well plates with alkaline-phosphatase-positive iPS cell colonies at 12 d.p.i. obtained from B-cell cultures treated with doxycycline to induce OSKM only (B+OSKM), cells co-expressing C/EBPα with OSKM, or cells treated for 18 h with E2 to activate C/EBPα followed by OSKM induction (Bα′+OSKM). **c**, Nanog$^+$ iPS cell colonies at 12 d.p.i. (scale bar, 100 μm). **d**, Quantification of 12 d.p.i. Nanog$^+$ colonies after OSKM induction in B cells pre-treated with C/EBPα for the indicated pulse lengths. Error bars indicate s.d. ($n = 3$). **e**, Chimaeric mice obtained after blastocyst injection of αiPS cell clone 12. **f**, Doxycycline-induced Bα′+OSKM cultures with arrows indicating earliest detectable ES-cell-like colonies (scale bar, 100 μm). **g**, Nanog$^+$ colonies of Bα′+OSKM and B+OSKM cultures at 8 and 12 d.p.i. Error bars indicate s.d. ($n = 3$). **h**, Retroviral silencing after induction of Bα′+OSKM and B+OSKM cells measured by hCD4 expression. RFI, relative fluorescence intensity. **i**, **j**, Experimental strategy to determine OSKM transgene independence (**i**) and formation of Nanog$^+$ iPS cell colonies (**j**) at 12 d.p.i. Error bars indicate s.d. ($n = 3$).

cells became 30% GFP$^+$ at 4 d.p.i. and 60% GFP$^+$ at 8 d.p.i., whereas B+OSKM cells remained negative (Extended Data Fig. 4e). Serum-free conditions further accelerated the process with 62% of the Bα′+OSKM cells becoming GFP$^+$ already at 2 d.p.i. and 95% at 4 d.p.i. (Fig. 2d, e), with a similar viability as in serum containing medium. Induced B+OSKM cells died without serum.

We next scored the expression of MET genes[15,16]. Bα′+OSKM cells upregulated all epithelial genes tested and expressed E-cadherin 2–4 d.p.i., again with B+OSKM cells remaining negative (Extended Data Fig. 5a, b). In line with the possibility that the C/EBPα pulse is capable of initiating an epithelial–mesenchymal transition (EMT)[17], mesenchymal genes encoding TGF-β pathway members, transcription factors and collagens were first upregulated by the C/EBPα pulse and then switched off after OSKM induction (Extended Data Fig. 5a, e). To compare gene expression changes during OSKM-induced reprogramming of Bα′ cells with C/EBPα-induced transdifferentiation of B cells we selected genes selectively expressed in macrophages (263) and B cells (83) (Supplementary Table 1). Most of these genes were already up- or downregulated in Bα′ cells, respectively, as expected from the induced conversion (Extended Data Fig. 6a). After OSKM induction, B-cell genes became further silenced whereas macrophage genes elicited a more heterogeneous response (Extended Data Fig. 6b, c). Bα′+OSKM cells at 2–4 d.p.i. co-expressed E-cadherin and the macrophage markers CSF-1R and Mac1, whereas B+OSKM cells remained negative and αiPS cells expressed exclusively E-cadherin (Extended Data Fig. 6d). Sorted Mac1$^+$E-cadherin$^+$ cells and Mac1$^-$E-cadherin$^+$ cells yielded similar proportions of iPS cell colonies, indicating that expression of Mac1 has no role in iPS cell reprogramming (Extended Data Fig. 6e). Analysis of the 18 h pulsed B cells by RNA-seq revealed 1,418 upregulated and 552 downregulated genes (Extended Data Fig. 7a). Gene set enrichment analysis (GSEA) showed a strong correlation between genes downregulated in Bα′ cells with sets of B-cell-specific genes, and upregulated genes with genes related to EMT, collagen expression and the TGF-β pathway (Extended Data Fig. 7b), confirming the array data. C/EBPα chromatin immunoprecipitation sequencing (ChIP-seq) analysis revealed 1,766 DNA regions bound by the protein in Bα′ cells of which ~20% were within 50 kb of genes whose expression changed (Extended Data Fig. 7c–g). These included two regulatory sites in *Tet2* (Fig. 3a, ref. 3) as well as sites close to lymphoid, MET and TGF-β pathway genes. *Tet2* became upregulated ~1.6-fold after the 18-h pulse and was further activated by OSKM, reaching levels comparable to ES cells within 2 days (Fig. 3b, c). In contrast, mouse embryonic fibroblasts (MEFs) pulsed with C/EBPα did not upregulate the gene (Extended Data Fig. 8a). Tet2 overexpression enhanced iPS cell reprogramming of B cells ~3-fold (Fig. 3d), supporting the reported role of Tet2 in iPS cell formation[7,8]. Because the Tet2-induced 5-hydroxymethylcytosine (5hmC) modification has been described to participate in the maintenance of pluripotency, we determined 5hmC levels by hydroxymethylated DNA immunoprecipitation followed by qPCR (hMeDIP-qPCR) in Bα′ cells at regulatory regions of nine pluripotency genes for which hydroxymethylation levels change during reprogramming[18]. 5hmC levels were significantly increased for all genes tested (Fig. 3e and Extended Data Fig. 8b) and Tet2 binding was detected at the same regions (Fig. 4f and Extended Data Fig. 8c). Unexpectedly, Tet2 was predominantly cytoplasmic in uninduced B cells and nuclear after the C/EBPα pulse (Fig. 3g). The nuclear localization was maintained after sustained expression of C/EBPα (72 h) (Extended Data Fig. 8d). No protein translocation was observed in B cells treated with E2 alone or infected with a control vector (Extended Data Fig. 8e) and MEFs exhibited predominantly nuclear Tet2 localization (Extended Data Fig. 8f). Because hydroxymethylation is required

©2014 Macmillan Publishers Limited. All rights reserved

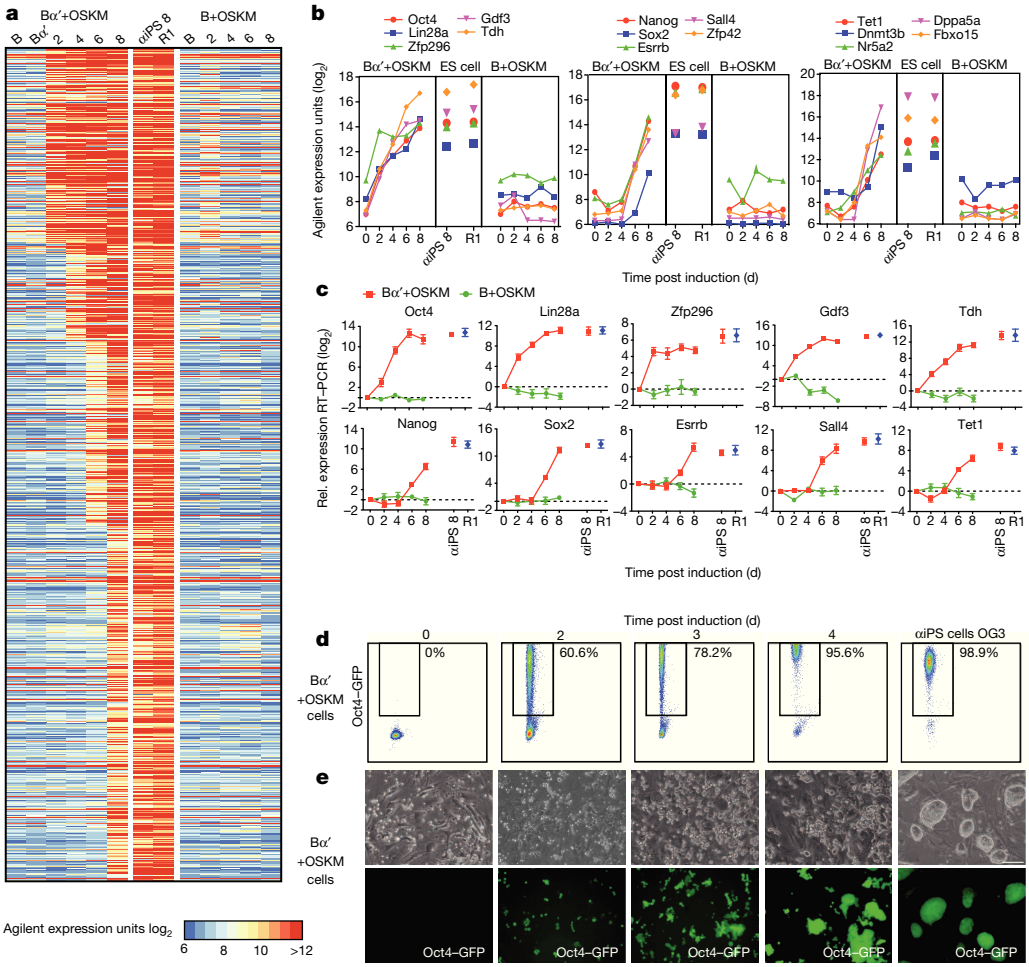


Figure 2 | Upregulation of pluripotency genes. **a**, Heat map of Agilent expression arrays showing upregulation of genes expressed >4-fold higher in ES cells than in B cells at the indicated time points after doxycycline treatment of Bα′+OSKM cells and B+OSKM cells, with αiPS cells (clone 8) and ES cells (R1) as controls. **b**, Array expression kinetics of three groups of pluripotency genes from Bα′+OSKM cells, αiPS cells and ES cells and B+OSKM cells. **c**, qRT–PCR validation of selected genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, **e**, Oct4–GFP expression analysed by FACS and microscopy in Bα′+OSKM cells (scale bar, 100 μm).

for active DNA demethylation, we analysed the DNA methylation status of the *Oct4* and *Nanog* promoters using bisulphite sequencing (Fig. 3h). Whereas the *Oct4* promoter became partially demethylated in Bα′+OSKM as early as 2 d.p.i. (Fig. 3h), the *Nanog* promoter only became demethylated at 8 d.p.i., consistent with the observed delayed upregulation of the gene (Fig. 2b, c).

The 5hmC changes at pluripotency genes raised the possibility that the C/EBPα pulse eliminates epigenetic barriers, increasing chromatin accessibility to the OSKM transcription factors. To test this, we performed quantitative DNase I sensitivity assays on selected regulatory regions of pluripotency genes. This revealed increased accessibility to the enzyme of all sites tested in the 18-h pulsed B cells whereas control regions and pluripotency sites in control B+OSKM cells showed no difference (Fig. 4a). Interestingly, after prolonged induction of C/EBPα (72 h) the chromatin of pluripotency genes returned to the closed state as seen in untreated B cells (Extended Data Fig. 9a). In addition, no demethylation of the *Nanog* and *Oct4* promoters was observed (Extended Data Fig. 9b). We next tested Oct4 occupancy in OSKM-induced Bα′ cells at sites within enhancers and promoters of pluripotency genes bound by Oct4 in ES cells[19] (Fig. 4b). At 2 d.p.i. all of these sites showed significant Oct4 occupancy, whereas they remained unbound in B+OSKM cells (Fig. 4b). Together, these findings show that the C/EBPα pulse transiently induces chromatin remodelling at regulatory regions of pluripotency genes, making them rapidly accessible to Oct4 binding.

Our results show that C/EBPα poises B cells for highly efficient and almost immediate reprogramming into iPS cells by OSKM transcription factors, with MET and pluripotency network activation occurring concomitantly (summarized in Extended Data Fig. 9c). In addition, our data indicate that the C/EBPα pulse already initiates an EMT, reminiscent of the recently described EMT–MET transition during reprogramming of fibroblasts[17]. The single cell cloning and Oct4–GFP reporter experiments suggest that essentially 100% of the poised cells can be reprogrammed into iPS cells. Several lines of evidence suggest that the rapid demethylation of the *Oct4* promoter involves Tet2: the C/EBPα pulse directly binds to and upregulates the *Tet2* gene and shuttles the protein into the nucleus. Then Tet2 binds to regulatory sites of *Oct4* and other pluripotency genes, converting 5mC residues into 5hmC (summarized in Fig. 4c). This modification might directly cause the gene's de-repression or be linked to demethylation and/or induction of chromatin remodelling[8]. The effect of C/EBPα is highly specific as only the closely related factor C/EBPβ, but not the lineage-instructive factors GATA1 (ref. 20), nor Mash1 or MyoD expressed in fibroblasts, was active (Extended Data Fig. 10a–i). C/EBPα also sensitized T-cell precursors[21] but not fibroblasts (Extended Data Fig. 10j, k). However, when C/EBPα was co-expressed with PU.1, a condition that mediates transdifferentiation to macrophage-like cells[22], it induced a ~6-fold enhancement of iPS cell reprogramming, whereas PU.1 alone had no effect (Extended Data Fig. 10k). This indicates that a myeloid determinant sensitizes somatic cells for iPS cell reprogramming, in line with

©2014 Macmillan Publishers Limited. All rights reserved

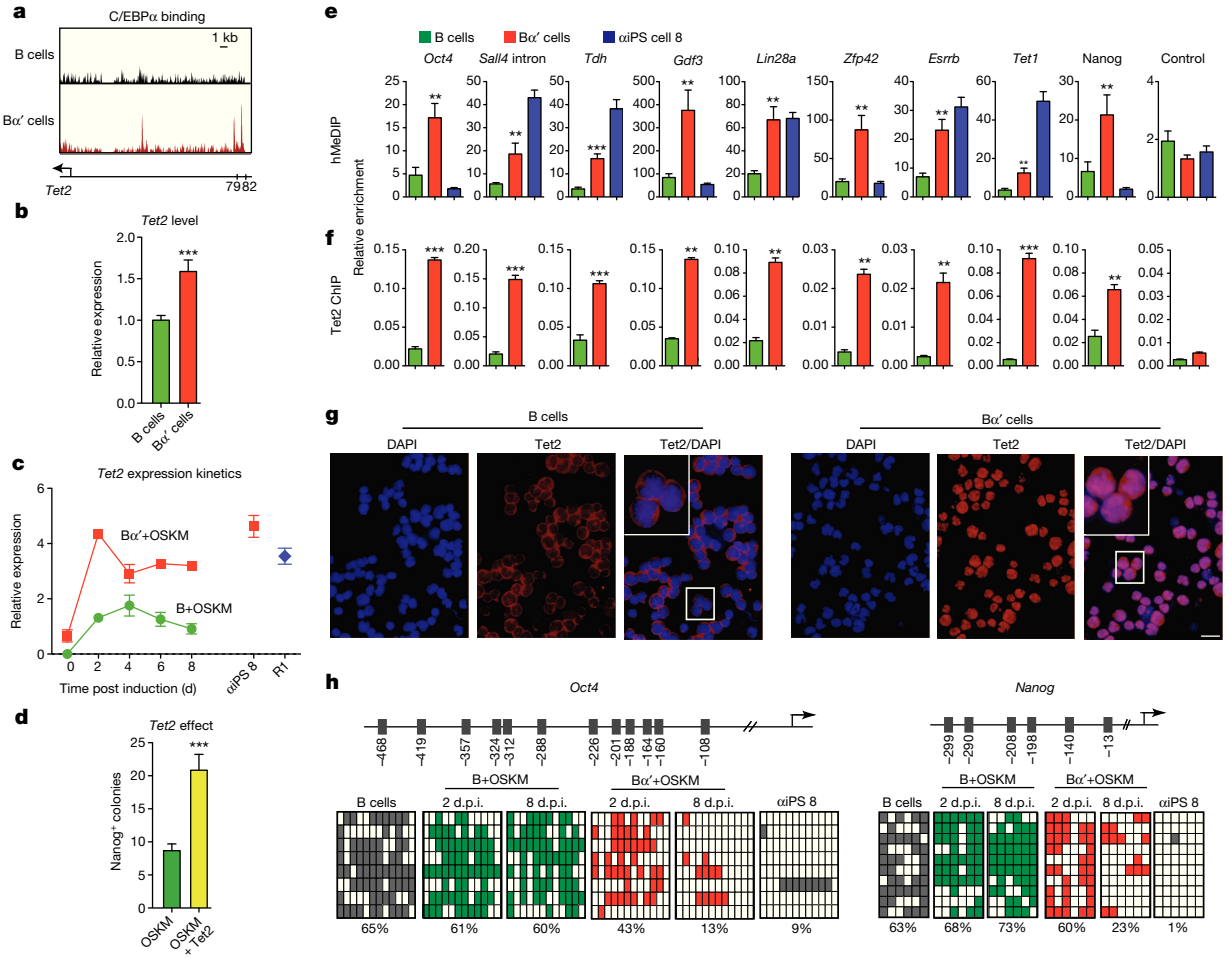


Figure 3 | Tet2 activation, cytosine hydroxymethylation and DNA methylation. **a**, C/EBPα binding at two enhancers upstream of the *Tet2* gene. **b**, **c**, qRT–PCR for Tet2 expression. Error bars indicate s.d. ($n = 6$). Student's *t*-test $***P < 0.001$ relative to control. **d**, Tet2 overexpression effect on B-cell reprogramming. Error bars indicate s.d. ($n = 6$). Student's *t*-test $***P < 0.001$ relative to control. **e**, hMeDIP followed by qPCR at regulatory regions of selected pluripotency genes. Error bars indicate s.d. ($n = 3$). Student's *t*-test $**P < 0.01$, $***P < 0.001$ relative to B cells. **f**, Tet2 binding sites determined by ChIP of the regions in **e**. Error bars indicate s.d. ($n = 3$). Student's *t*-test $**P < 0.01$, $***P < 0.001$ relative to B cells. **g**, Immunofluorescence of Tet2 protein in B cells and Bα′ cells (scale bar, 30 μm). **h**, CpGs analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters, with their position indicated on the top. Filled rectangles represent methylated CpGs; empty rectangles represent unmethylated residues.

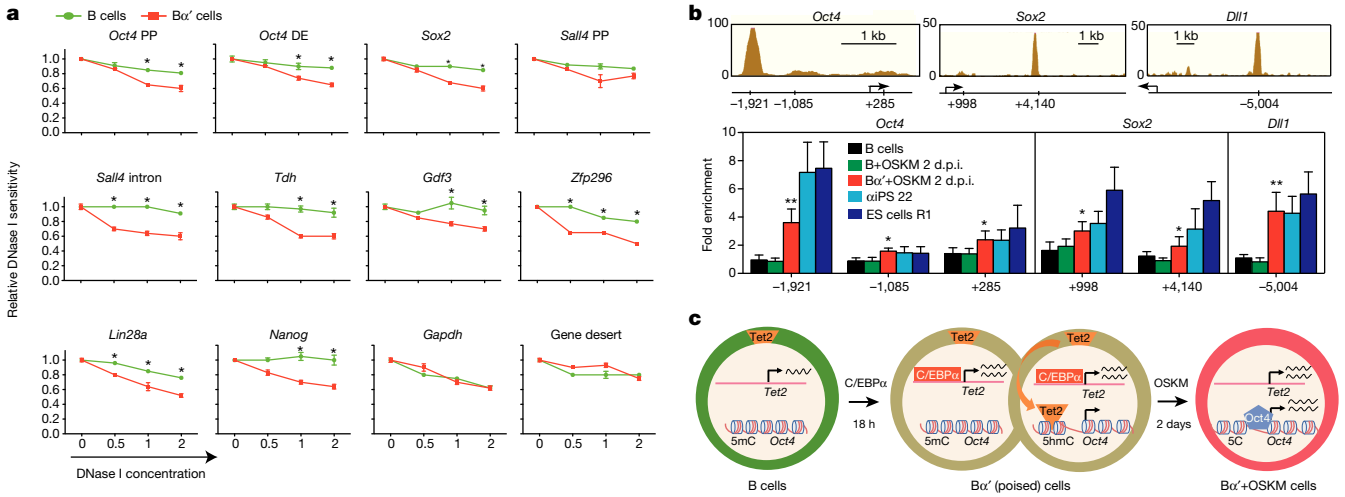


Figure 4 | Chromatin accessibility and summary schemes. **a**, DNase I sensitivity at regulatory regions in B cells and Bα′ cells determined by enzyme titration. Error bars indicate s.d. ($n = 3$). Student's *t*-test $*P < 0.05$ relative to control. DE, distal enhancer; PP, proximal promoter. **b**, Oct4 binding sites examined at the *Pou5f1*, *Sox2* and *Dll1* loci, with Oct4 peaks corresponding to ES cells[19]. Oct4 binding by ChIP-qPCR analysis. Student's *t*-test $*P < 0.05$, $**P < 0.01$ relative to control. Error bars indicate s.d. ($n = 3$). **c**, Model of reprogramming mechanism involving Tet2.

©2014 Macmillan Publishers Limited. All rights reserved

the finding that macrophage/granulocyte progenitors, the formation of which is driven by C/EBPα[23], show the highest described reprogramming efficiency of somatic cells (25%)[12]. The C/EBPα effect might recapitulate an embryonic function of the gene as it is required, in combination with C/EBPβ, for the formation of functional trophoblasts[24], cells with intriguing similarities to macrophages[25].

METHODS SUMMARY

Somatic cells used for iPS cell reprogramming were derived from the reprogrammable mouse[11]. ES cells and iPS cells were cultured on mitomycin-C-treated MEFs in ES cell medium containing 15% FBS and 1,000 U ml^{-1} of LIF. For B- and T-cell reprogramming experiments the medium was supplemented with cell-type-specific cytokines[12]. OSKM expression was induced with 2 ng ml^{-1} of doxycycline. αiPS cells and iPS cells correspond to individual iPS cell clones obtained after expansion into stable lines of OSKM induced Bα′ and B cells, respectively. The total number of iPS cell colonies was scored after alkaline phosphatase or Nanog staining. For qRT–PCR and microarray analyses, B cells at different time points during reprogramming, as well as ES cells and iPS cells, were trypsinized and FACS sorted to remove feeder cells and dead cells. Total RNA was extracted using the miRNeasy kit (Qiagen). Microarrays were done using Agilent 8X60K expression arrays. Details on injections into blastocysts to test for chimaerism are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 April; accepted 15 November 2013.
Published online 15 December 2013.

1. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
2. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117,** 663–676 (2004).
3. Kallin, E. M. *et al.* Tet2 facilitates the derepression of myeloid target genes during CEBPα-induced transdifferentiation of pre-B cells. *Mol. Cell* **48,** 266–276 (2012).
4. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150,** 1209–1222 (2012).
5. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).
6. Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145,** 835–850 (2011).
7. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495,** 370–374 (2013).
8. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488,** 652–655 (2012).
9. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).
10. Di Tullio, A. *et al.* CCAAT/enhancer binding protein α (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108,** 17016–17021 (2011).
11. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods* **7,** 56–59 (2010).
12. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41,** 968–976 (2009).
13. Porse, B. T. *et al.* E2F repression by C/EBPα is required for adipogenesis and granulopoiesis *in vivo*. *Cell* **107,** 247–258 (2001).
14. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11,** 769–782 (2012).
15. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
16. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
17. Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nature Cell Biol.* **15,** 829–838 (2013).
18. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12,** 453–469 (2013).
19. Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145,** 183–197 (2011).
20. Kulessa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev.* **9,** 1250–1262 (1995).
21. Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* **25,** 731–744 (2006).
22. Feng, R. *et al.* PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105,** 6057–6062 (2008).
23. Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proc. Natl Acad. Sci. USA* **94,** 569–574 (1997).
24. Begay, V., Smink, J. & Leutz, A. Essential requirement of CCAAT/enhancer binding proteins in embryogenesis. *Mol. Cell. Biol.* **24,** 9744–9751 (2004).
25. Guilbert, L., Robertson, S. A. & Wegmann, T. G. The trophoblast as an integral component of a macrophage-cytokine network. *Immunol. Cell Biol.* **71,** 49–57 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Batlle for generating the chimaeric mice, M. Stadtfeld, D. Egli, P. Cosma, G. Filion, B. Lehner and J. Valcarcel for critical reading of the manuscript, V. Broccoli, P. Muñoz and S. Aznar Benitah for constructs and antibodies and H. Schoeler for the Oct4 reporter mice. This work was supported by the Ministerio de Educacion y Ciencia, SAF.2007-63058 and AGAUR 2009 SGR768. B.D.S. holds a La Caixa International PhD Fellowship.

Author Contributions T.G. and B.D.S. conceived the idea for this project, designed the experiments and wrote the paper. B.D.S. performed the cell culture, animal experiments and molecular biology and B.D.S. and J.L.S. the methylation analyses. J.L.S. performed the studies on Tet2 and J.L.S and C.v.O. the ChIP experiments. S.C., E.M.K. and D.T. were responsible for the bioinformatics and G.P.V. and M.B. for the DNase I experiments and analyses. J.L. provided the Tet2 overexpressing plasmids.

Author Information The array data are available from the Gene Expression Omnibus under the accession number GSE52397. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.G. (Thomas.Graf@crg.eu).

©2014 Macmillan Publishers Limited. All rights reserved

METHODS

Mice. The reprogrammable mouse containing a tetracyclin-inducible OSKM cassette as well as rtTA has been described previously[11].

Cell cultures. ES cells and iPS cells were cultured on mitomycin-C-treated MEF feeder cells in KO-DMEM (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, non-essential amino acids, β-mercaptoethanol, 1,000 U ml^{-1} LIF (ES cell medium) and 15% fetal bovine serum (FBS, Invitrogen).

MEF cultures were established by trypsin digestion of mouse embryos (embryonic day 13.5) and the resulting cells cultured in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin. $CD19^+$ pro-B and pre-B cells were isolated from bone marrow using monoclonal antibodies to CD19 (1D3), obtained from BD Pharmingen, using MACS (Miltenyi Biotech). Double-negative (DN) T cells were prepared from thymii of 4-week-old mice by lineage depletion with Streptavidin microbeads. The following biotin-conjugated antibodies were used to label lineage (lin)-positive thymocytes before lineage depletion: CD3 (145-2C11), CD4 (GK1.5), CD8 (53-6.7), B220 (RA3-6B2), CD19 (1D3), Mac1/CD11b (M1/70), Gr1 (RB6-8C5), Ter119 (TER-119), I-A/I-E (2G9). The purity of the sorted cell fractions was confirmed by FACS using an LSR II machine (BD). After isolation, B cells were expanded in RPMI medium supplemented with 10% FBS and IL-7 for 1 week before inducing reprogramming. T cells were grown RPMI medium supplemented with 10% FBS supplemented with IL-7, Flt3 and SCF.

Reprogramming. Reprogramming experiments were conducted in gelatinized plates seeded with a feeder layer of the OP9 stromal cell line, using ES cell medium supplemented with 2 μg ml^{-1} of doxycycline and 15% FBS. For the reprogramming of B cells, IL-4 (10 ng ml^{-1}), IL-7 (10 ng ml^{-1}) and IL-15 (2 ng ml^{-1}) were added to the medium (ref. 13). B cells were seeded at a density of 500 cells cm^{-2} in 6-well plates. For the conditioning of Bα′ cells, C/EBPα-infected sorted B cells were exposed for 18 h to 100 nM of β-estradiol (E2) followed by inducer wash-out and addition of ES cell medium supplemented with doxycycline and cytokines. Serum-free iPS reprogramming was performed in serum-free ES cell medium (KO-DMEM supplemented with 15% knockout serum replacement, L-glutamine, penicillin/streptomycin, non-essential amino acids, β-ME, 1,000 U ml^{-1} LIF, N2 (100X), B27 (100X))[9] supplemented with 2 μg ml^{-1} doxycycline to activate OSKM factors. 2i inhibitors were added to the medium 2 days after doxycycline induction.

For expansion of iPS cell lines, colonies with ES cell morphology were picked after doxycycline withdrawal at 12 d.p.i. OSKM MEFs were seeded on mitomycin-C-treated MEF feeders in ES cell medium containing FBS and induced with doxycycline.

Vectors and virus production and infection. C/EBPα-ER-GFP, C/EBPα-ER-CD4, C/EBPβ, PU.1, C/EBPα-BRM-2 retroviruses and shPax5 lentivirus have been described previously[1,2,13]. *Gata1* cDNA was cloned by PCR in the BglII/XhoI restriction sites of the MIG vector. The MyoD-ER plasmid was obtained from P. Muñoz, Mash1 from V. Broccoli. The retroviral vector for murine Tet2 overexpression was derived from MSCV-based pMIRWAY-puro-TET2[26] by replacing puro with EGFP. Virus production has been described previously[27]. Briefly, HEK293T cells were co-transfected with vector plasmid and packaging plasmids using calcium phosphate transfection. Viral supernatants were collected 48–72 h later and concentrated by ultracentrifugation at 20,000*g* for 2 h at 20 °C. Viral concentrates were re-suspended in 1 × PBS and stored at −80 °C. Infection of MEFs was carried out in medium containing 5 μg ml^{-1} polybrene, and infection of B cells by centrifugation with concentrated virus for 2 h at 32 °C at 1000*g* in B-cell medium.

Alkaline phosphatase staining. For alkaline phosphatase staining, the cells were fixed in 2% paraformaldehyde and then incubated for 15 min in NTMT solution (NaCl 100 mM, Tris A 1 M, Tris B 1 M, $MgCl_2$ 50 mM, Tween 0.1%) supplemented with BCIP e NBT (Roche).

Immunofluorescence assays. The cells were fixed with 4% paraformaldehyde, blocked and incubated with primary antibodies overnight at 4 °C. On the next day, the cells were exposed to secondary antibodies (all Alexa Fluor from Invitrogen) at room temperature for 1 h. The primary antibodies used were Nanog (Calbiochem), Oct4 (SC-5279), SMA (obtained from P. Muñoz), Tuj1 (obtained from V. Broccoli), Foxa2 (obtained from S. Aznar Benitah) and Tet2 (SC-61). Nuclear staining was performed with DAPI (Invitrogen).

Differentiation of iPS cells. Embryoid bodies (EBs) were derived by plating iPS cells at a concentration of 1.3×10^6 cells ml^{-1} in non-adherent dishes in ES medium without LIF. After 4 days in suspension, cell aggregates were plated on gelatine-coated dishes in DMEM medium supplemented with 10% FBS for 10 days.

Teratoma assay. One million iPS cells were injected subcutaneously into SCID BEIGE mice. Three weeks after the injection, the tumours were surgically removed and embedded in paraffin. Teratomas were sectioned into 10-μm-thick slices using a cryostat (Leica, CM1850 UV) and sections were stained with haematoxylin and eosin.

Chimaeric mice. For the chimaera formation assay, 10–15 iPS cells (C57BL/6J background, black coat colour) were injected into a 3.5 days post coitum (d.p.c.) blastocyst of CD1 mice (white coat colour) and transferred into pseudo-pregnant 2.5 d.p.c. CD-1 recipients for chimaera generation. Chimaerism was ascertained after birth by the appearance of black coat colour (from the C57BL/6J iPS cells) over the white background from the host pups.

FACS. Cell suspensions were stained with various antibodies from either BD Pharmingen or eBiosciences. Cells were analysed with an LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar). Primary antibodies used were CD19, Mac1, B220, Cdh1, CD115 (all from BD Biosciences), hCD4 (eBioscience) and Nanog (Calbiochem).

RNA isolation and quantification. RNA isolation of MEFs and B cells was done with the miRNeasy mini kit (Qiagen). RNA was collected from sorted live cells at 0, 2, 4, 6 and 8 d.p.i. To remove the feeders, ES cells and iPS cell clones at passage 2 or higher were seeded on gelatinized plates and processed with the above kits. RNA was eluted from the columns using RNase-free water or TE buffer and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystem).

Gene expression arrays and qRT–PCR. RNA samples (with an RNA integrity number (RIN) greater than 9) were subjected to transcriptional analyses using Agilent expression arrays. For hybridization, 500 ng of total RNAs were labelled using Agilent's QuickAmp labelling kit following manufacturer instructions and analysed using Agilent 8X60K expression arrays. For the expression data with multiple probes annotated to the same gene only the most dynamic probe as defined by all array hybridizations was considered. Genes with expression changes during reprogramming were defined as showing >4-fold differences between ES cells and uninduced B cells, and having a Student's *t*-test P value <0.05 between technical duplicates. B-cell- and macrophage-specific genes were defined by combining array data from a previous study[28]. Lineage-specific genes were defined as >4-fold upregulated in one lineage over all of the others. Macrophage-specific genes were selected on the basis of >4-fold upregulation in primary macrophages and 5-day C/EBPα-induced macrophages as compared to B cells.

Unsupervised clustering was performed on Pearson correlation coefficients calculated between the indicated array samples using all genes that changed expression at least 2-fold between B cells and ES cells ($n = 10{,}982$). Data were analysed and visualized using R v2.15.1.

The array data are available from the Gene Expression Omnibus under the accession number GSE52397.

qRT–PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystem) and primers as listed in Supplementary Table 2. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 58 °C and 30 s at 72 °C.

Chromatin immunoprecipitation. ChIP experiments were performed as described previously[29]. Antibodies against Oct4 (SC-8626) and C/EBPα (SC-61) were purchased from Santa Cruz Biotechnologies. Tet2 ChIP was performed as in ref. 3. Data were obtained from three biological replicates, three independent immunoprecipitations and three technical triplicates.

For ChIP-seq analysis, reads were mapped onto mouse mm9 genome using bowtie2 (v2.1.0) (parameter:–very-sensitive). Reads filtering was done using SAMtools to keep reads that map only once, with a quality score of 10 or more, and to remove duplicates. Peaks were called using HOMER (v4.3) (parameter: style factor) and peaks were selected to control the false discovery rate at 0.001. Peaks annotation and genes association (Extended Data Fig. 7b–d) was performed using HOMER and BedTools (v 2.17).

Circular plot (Extended Data Fig. 7e) was done using Circos.

The sequencing data are available under the Gene Expression Omnibus number GSE52397.

DNA methylation. DNA was extracted using the Blood & Cell Culture DNA Mini kit (Qiagen). Bisulphite treatment of DNA was achieved using the EpiTect bisulphite kit (Qiagen) according to the manufacturer's instructions. The resulting modified DNA was amplified by polymerase chain reaction (PCR) using primer listed in Supplementary Table 2. The resulting amplified products were gel-purified (Qiagen), subcloned into the pGEM-T Vector Systems (Promega), and sequenced using the T7 and SP6 primers.

DNase I assay. Chromatin samples obtained as described before from two biological replicates were subjected to DNase I digestion. Briefly, 2 μg of chromatin were treated with 0.5, 1, and 2 units of DNase I (Roche) for 3 min at 37 °C in 1× DNase incubation buffer. Control samples were incubated in the absence of DNase I. Reactions were terminated by adding EDTA (40 mM final concentration) and the crosslinking was reversed by incubating the samples at 65 °C. After 6 h, proteinase K (40 μg ml^{-1} final concentration) was added to each reaction and incubated overnight at 37 °C. After careful phenol-chloroform extractions, the DNA was quantified and used as template for Real Time-PCR reactions using specific primers.

hMeDIP. DNA extraction and IP (DIP) was carried out by a previously published protocol (http://www.epigenome-noe.net/, PROT33) with the following alterations.

©2014 Macmillan Publishers Limited. All rights reserved

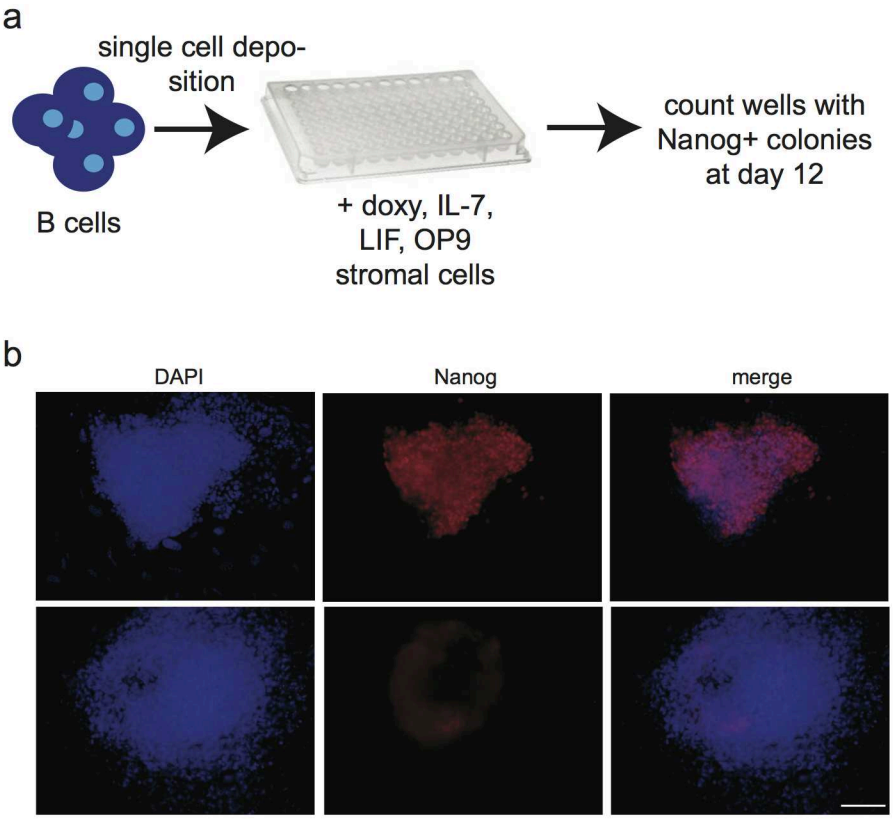
Genomic DNA was fragmented to a mean size of 350 bp using a Bioruptor (Diagenode) for 7×30 s cycles. DNA (re-suspended in 1× IP buffer) was incubated with 5 μg of anti-OHMeC antibody (Active Motif) at 4 °C for 2 h, and 40 μl of a 1:1 mixture of BSA-blocked Protein A (Millipore 16-125) and Protein G (Millipore 16-266) agarose beads were added, followed by an additional 2 h incubation at 4 °C. After washing 3 times in 1× IP buffer, immune complexes were released by incubation with 70 μg Proteinase K, and DNA was extracted once with phenol, extracted once with chloroform, and recovered by EtOH precipitation for qPCR analysis. The 5hmC enrichment in the samples was assessed by qPCR and the values were normalized against *Gapdh* promoter.

RNA-seq. RNA-seq was performed as described previously[30]. For RNA-seq analysis, reads were aligned onto mouse mm9 genome using STAR and Refseq mm9 annotation for splicing. Reads were filtered to keep only uniquely mapped reads with a maximum of two mismatches. Gene read count was performed using HTseq-count (parameter: mode = union, stranded, features = exons, attribute = gene_id) with Refseq mm9 annotation, and data were normalized using Bioconductor DEseq library (parameter: method = blind, sharingMode = fit-only, fitType = local). Genes with a read count above 300 in all conditions (set as a minimal threshold for functional expression) and genes with a 1.5-fold difference in normalized reads count between B cells and Bα′ cells were considered as differentially expressed. The sequencing data are available under the Gene Expression Omnibus number GSE52397.

Statistical analysis and replicates. All data presented are representative of at least three independent experiments that yielded similar results. Statistical analyses were performed using Prism software (GraphPad).

26. Cheng, J. *et al.* An extensive network of TET2-targeting microRNAs regulates malignant hematopoiesis. *Cell Rep.* **5,** 471–481 (2013).
27. Di Stefano, B. *et al.* A microRNA-based system for selecting and maintaining the pluripotent state in human induced pluripotent stem cells. *Stem Cells* **29,** 1684–1695 (2011).
28. Di Tullio, A. & Graf, T. C/EBPα bypasses cell cycle-dependency during immune cell transdifferentiation. *Cell Cycle* **11,** 2739–2746 (2012).
29. van Oevelen, C. *et al.* A role for mammalian Sin3 in permanent gene silencing. *Mol. Cell* **32,** 359–370 (2008).
30. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22,** 1616–1625 (2012).

©2014 Macmillan Publishers Limited. All rights reserved

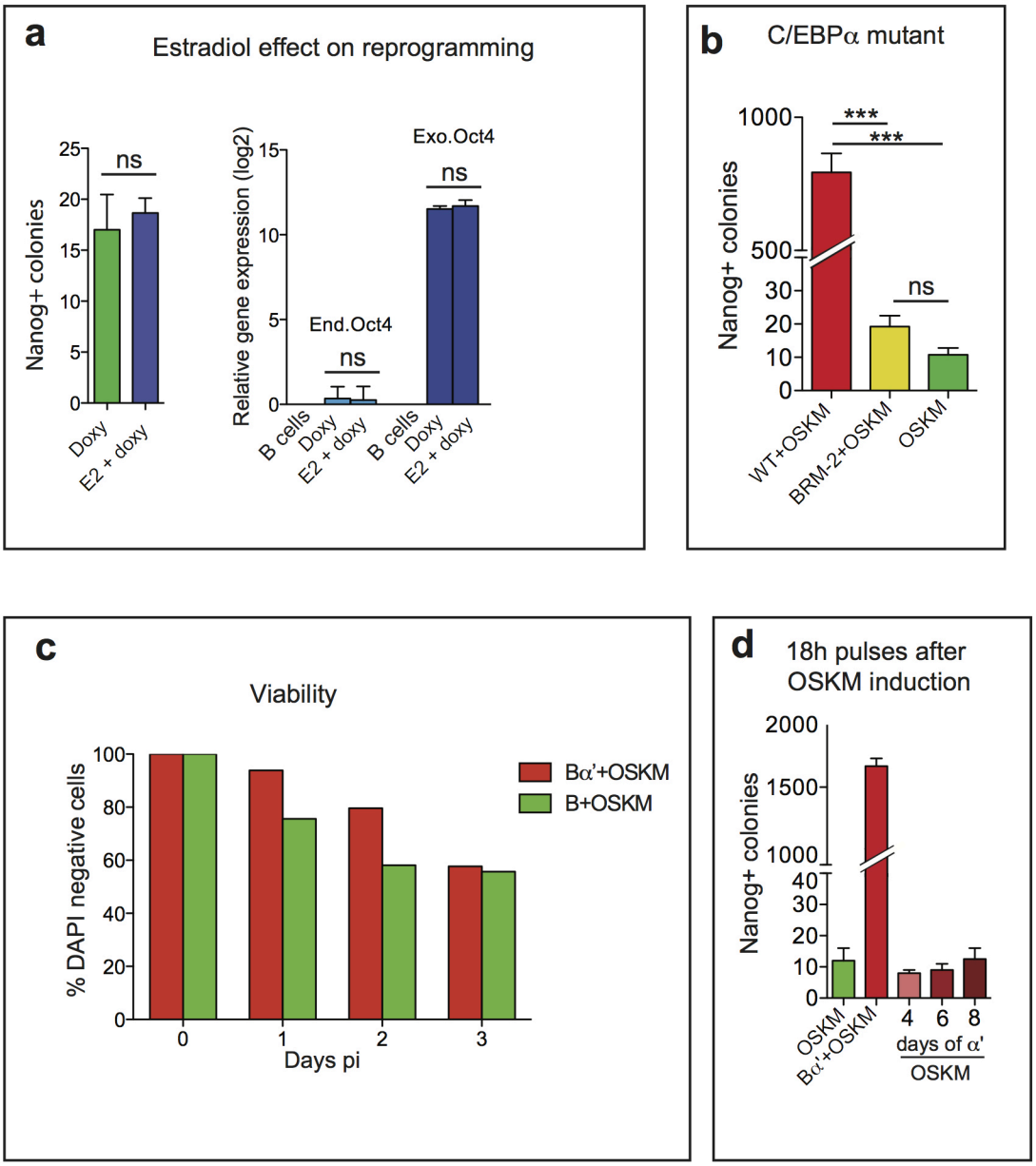


c

	# of wells with cell growth	# of wells with Nanog+ colonies
Exp 1	114	105
Exp 2	126	119

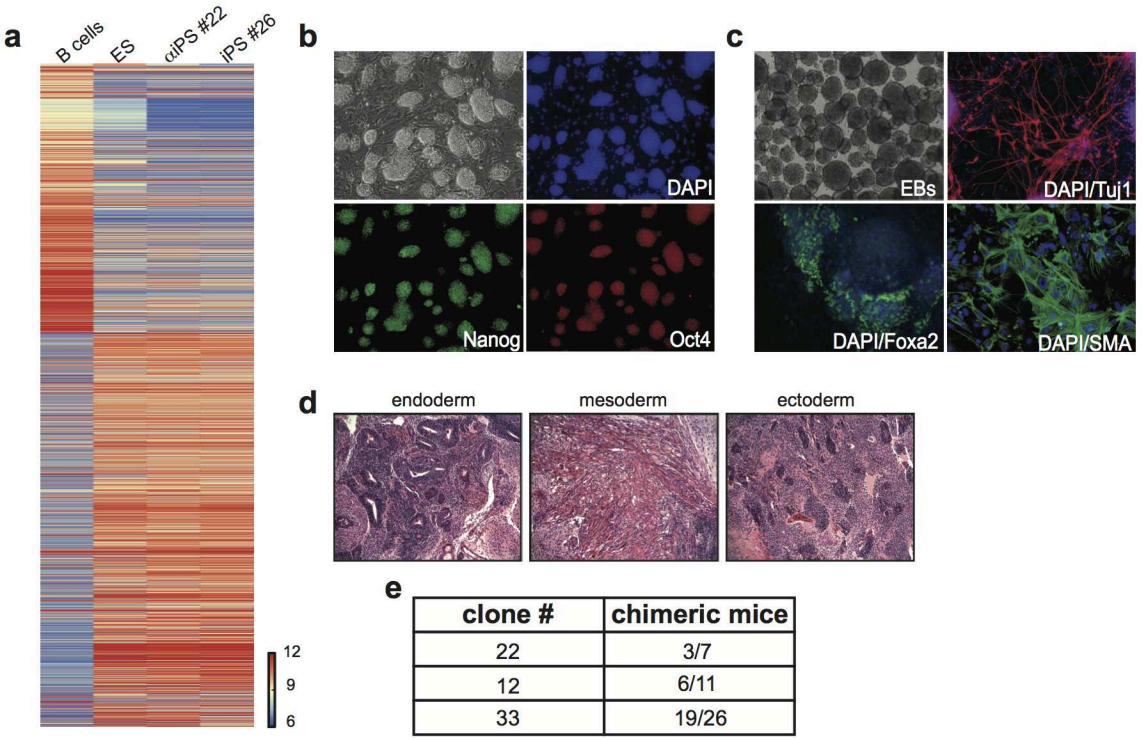
Extended Data Figure 1 | Single cell reprogramming. **a**, Bα′ cells were plated as single cells in 96-well plates on OP9 feeders in medium conditioned with IL-7, IL-15 and IL-4 cytokines. **b**, Representative Nanog-positive (upper panels) and Nanog-negative (lower panels) iPS cell clones in 96-well plates (scale bar, 100 μm). **c**, Efficiency of reprogramming obtained from two independent experiments.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 2 | C/EBPα mutant and cell viability during iPS cell reprogramming. **a**, E2 has no effect on reprogramming efficiency and ectopic Oct4 expression. Error bars indicate s.d. ($n = 3$). **b**, Effect of a C/EBPα mutant for the DNA binding domain on iPS cell reprogramming efficiency. Student's t-test $***P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$). **c**, Histogram showing percentages of live cells during the first 3 days of iPS cell reprogramming as analysed by FACS. Dead cells were scored as positive for DAPI staining. **d**, Effect of 18 h C/EBPα pulses after OSKM induction. Error bars indicate s.d. ($n = 3$).

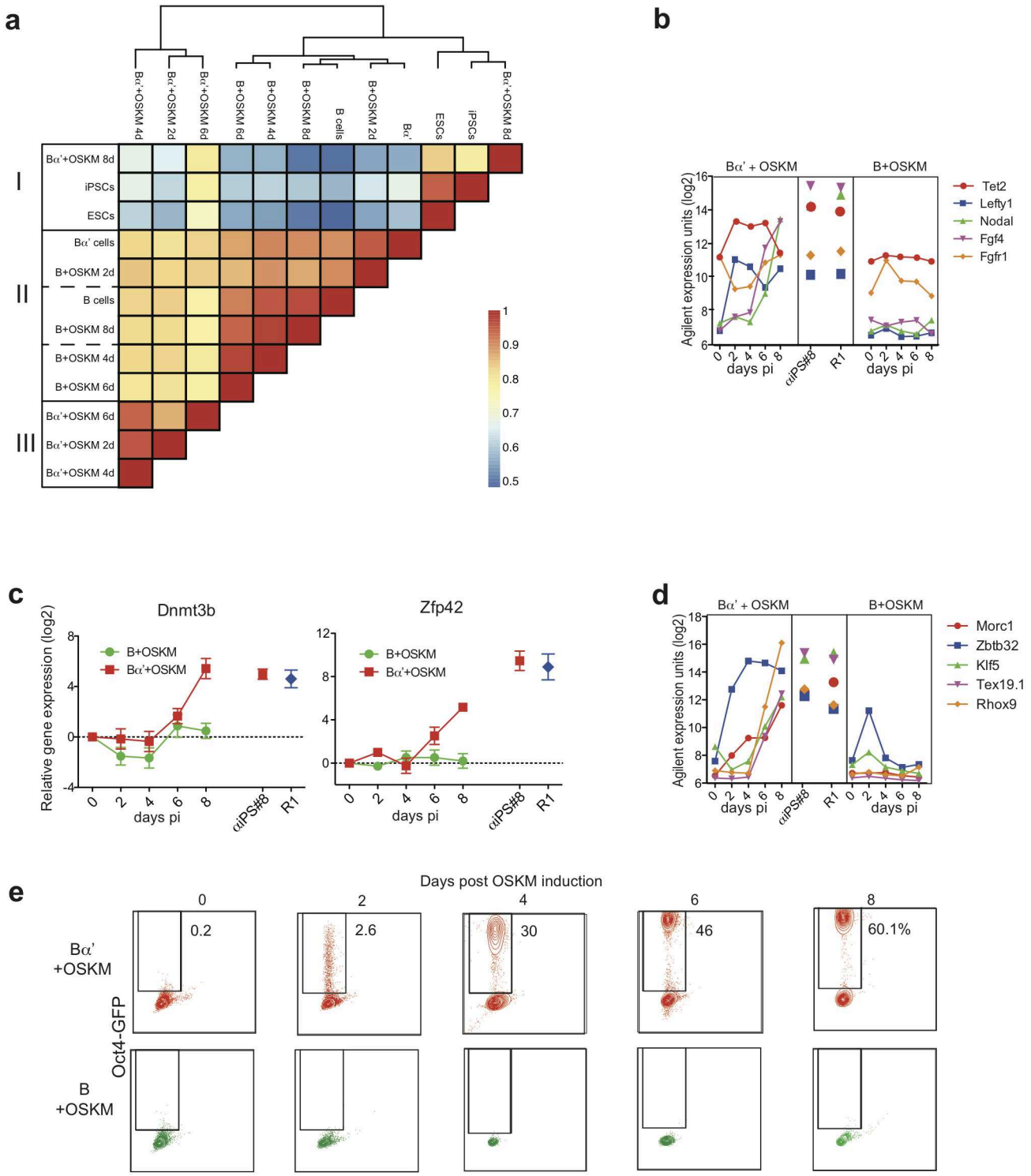
©2014 Macmillan Publishers Limited. All rights reserved



clone #	chimeric mice
22	3/7
12	6/11
33	19/26

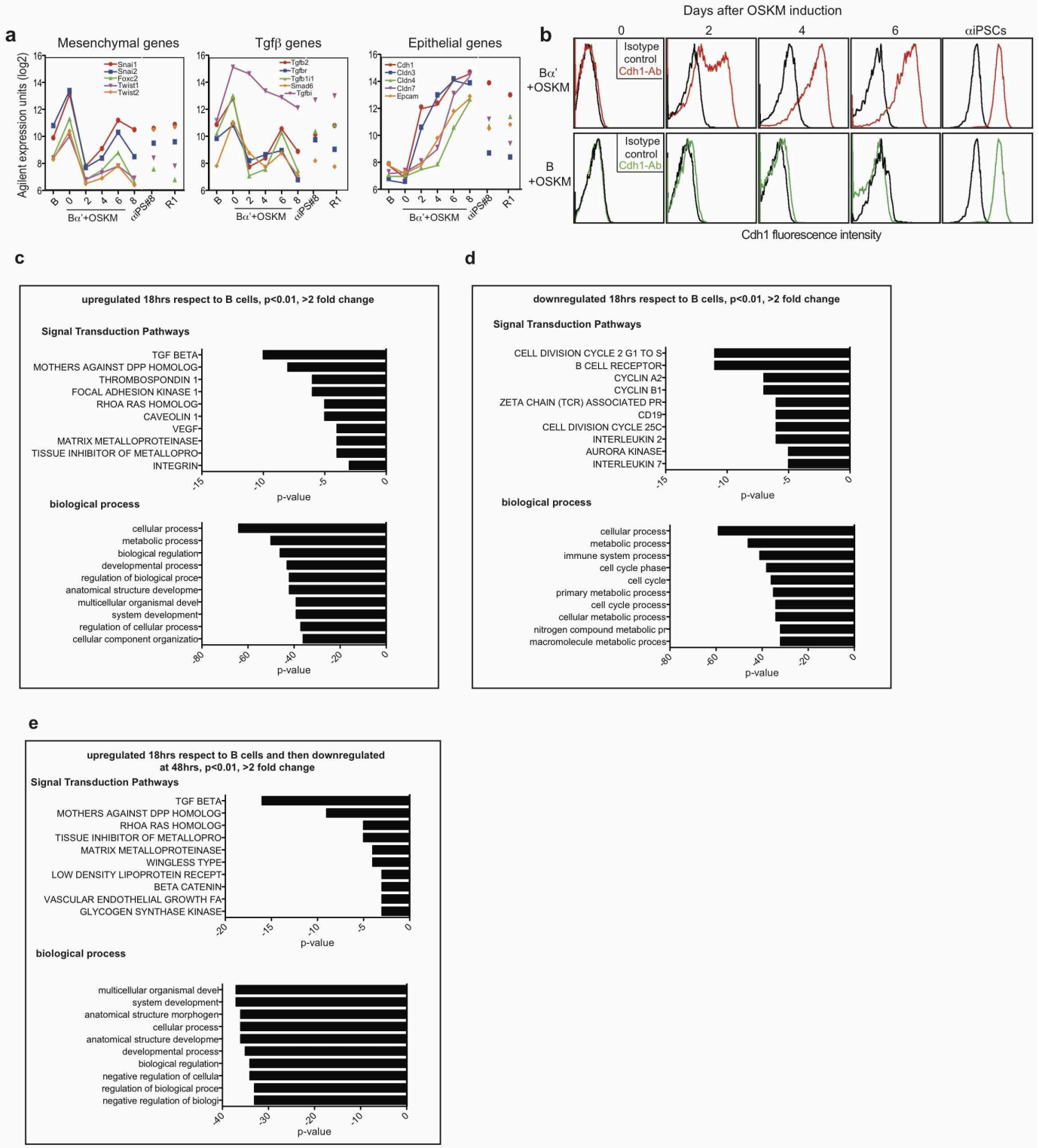
Extended Data Figure 3 | Properties of αiPS cells. **a**, Heat maps showing gene expression profiles of B cells, ES cells, αiPS cells (clone 22) and iPS cells (clone 26) derived from Bα′ and B cells, respectively. Genes were selected based on a >4-fold difference between B cells and ES cells. **b**, Micrographs (×10) of αiPS cell clone 22 growing on MEF feeders, showing colonies of ES-like morphology (bright field) and expression of Nanog (green) and Oct4 (red). **c**, Embryoid bodies generated from αiPS cell clone 22 (bright field) differentiated into Tuj1-positive cells (ectoderm), Foxa2-positive cells (endoderm) and SMA expressing cells (mesoderm). Original magnification, ×10. **d**, Section of a teratoma obtained from αiPS clone 22 stained with haematoxylin and eosin, showing differentiation into the major germ layers Original magnification, ×10. **e**, Proportion of chimaeric mice obtained after injection of CD1 blastocysts with three different αiPS cell clones.

©2014 Macmillan Publishers Limited. All rights reserved



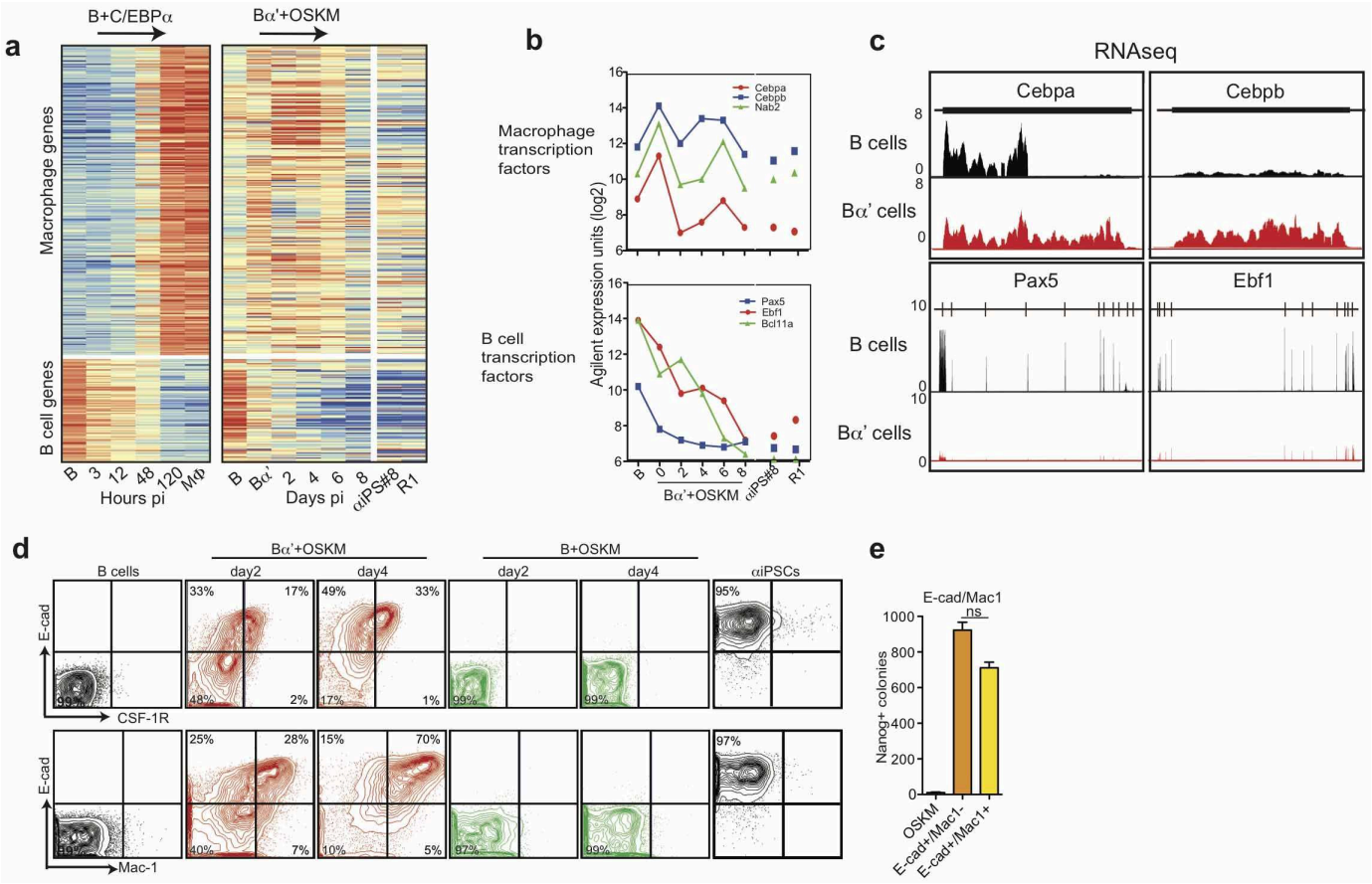
Extended Data Figure 4 | Pluripotency gene regulation during iPS cell reprogramming. **a**, Unsupervised hierarchical clustering analysis of gene expression array data. **b**, Agilent gene expression kinetics ($\log_2$ expression units) of pluripotency genes from Bα′+OSKM cells (red shaded panels), αiPS cells and ES cells (white panels) and B+OSKM cells (green shaded panels). **c**, qRT–PCR confirming array results for *Dnmt3b* and *Zfp42* (*Rex1*) genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, Agilent gene expression kinetics of late transition genes. **e**, Oct4–GFP expression during iPS cell reprogramming.

©2014 Macmillan Publishers Limited. All rights reserved



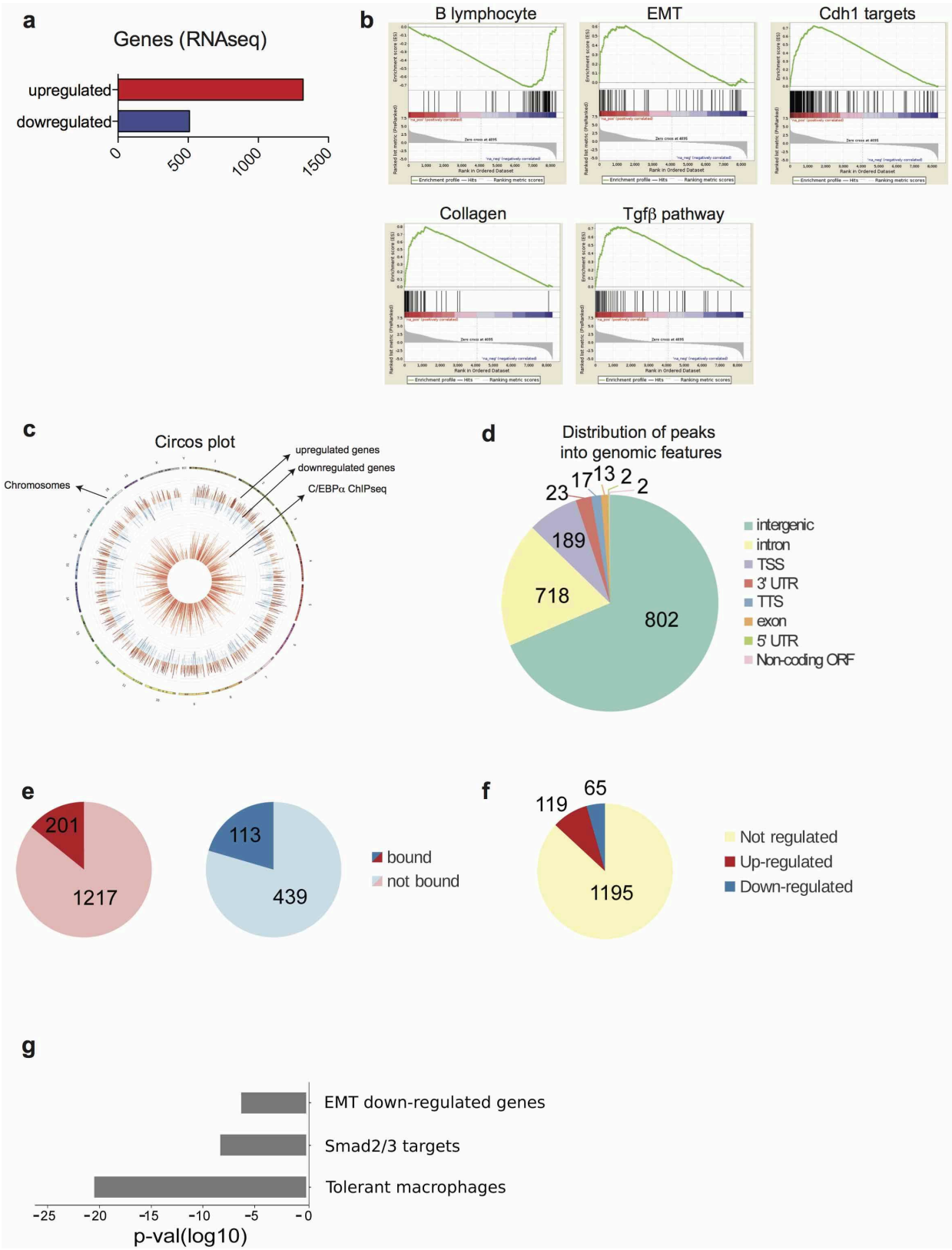
Extended Data Figure 5 | Changes in the expression of mesenchymal–epithelial transition genes. **a**, Agilent gene expression kinetics of mesenchymal–epithelial transition (MET) genes. **b**, E-cadherin expression by FACS at different times after induction. Isotype controls are shown in black; antibody-stained cells in red and green. **c**, GO analysis for genes upregulated >2-fold in Bα′ cells compared with B cells. **d**, GO analysis for genes downregulated >2-fold in Bα′ cells compared with B cells. **e**, GO analysis of genes upregulated >2-fold in Bα′ cells compared with B cells and subsequently downregulated >2-fold at 48 h.

©2014 Macmillan Publishers Limited. All rights reserved



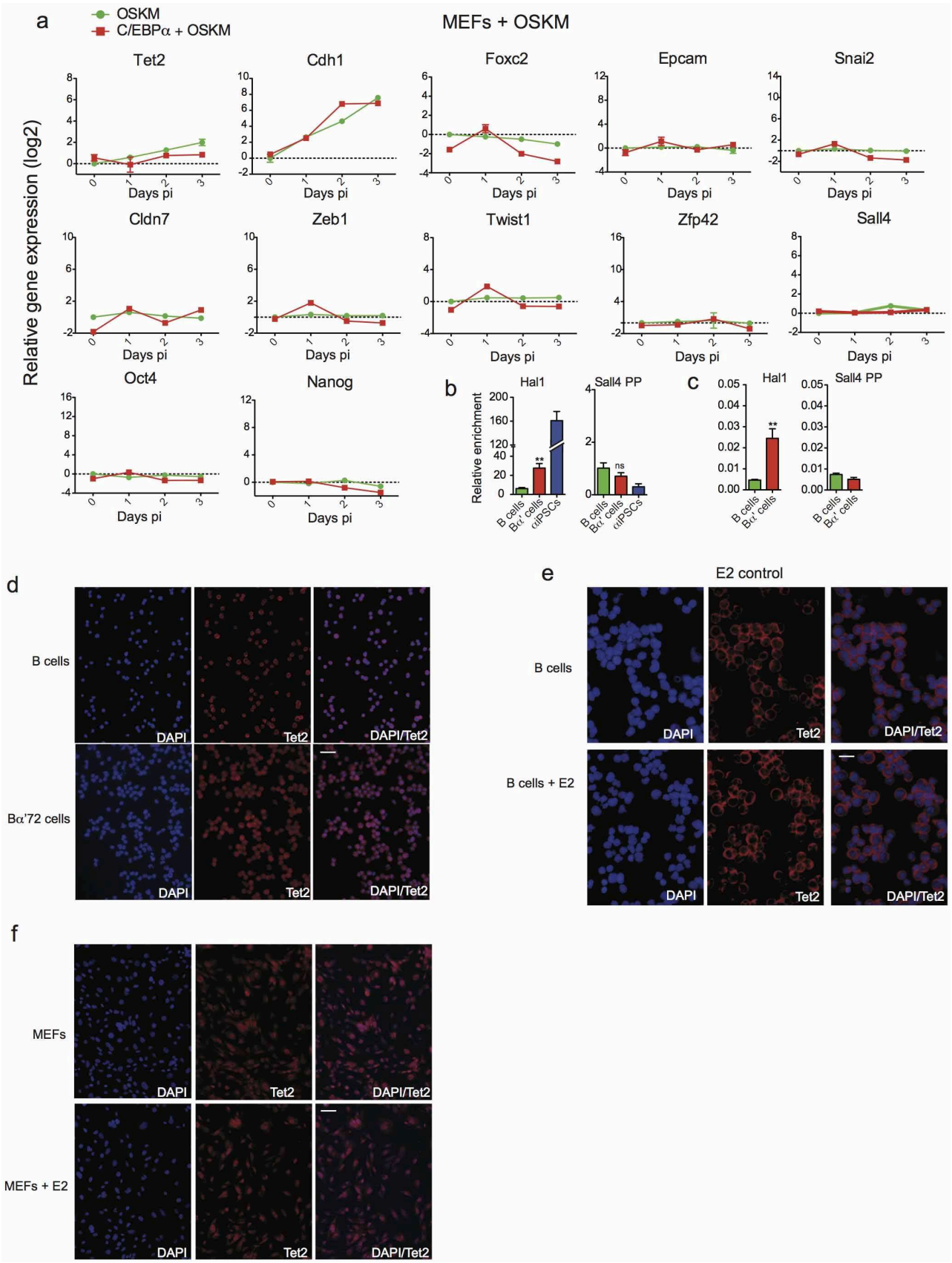
Extended Data Figure 6 | Changes in the expression of B cell and macrophage genes. **a**, Heat map of expression Affymetrix arrays with 263 macrophage and 83 B-cell-specific genes during C/EBPα-induced transdifferentiation of B cells and heat maps of the same genes using data from the Agilent expression array in Fig. 2. **b**, Agilent gene expression kinetics of selected macrophage and B-cell transcription factors. **c**, RNA-seq data for two representative macrophage and B-cell-specific genes each. **d**, FACS profiles of E-cadherin (Cdh1) expression, combined with either CSF-1R (upper panels) or Mac1 (lower panels) staining. **e**, Reprogramming potential of E-cadherin^{+}/Mac1^{-} and E-cadherin^{+}/Mac1^{+} cells. Error bars indicate s.d. ($n = 3$).

©2014 Macmillan Publishers Limited. All rights reserved



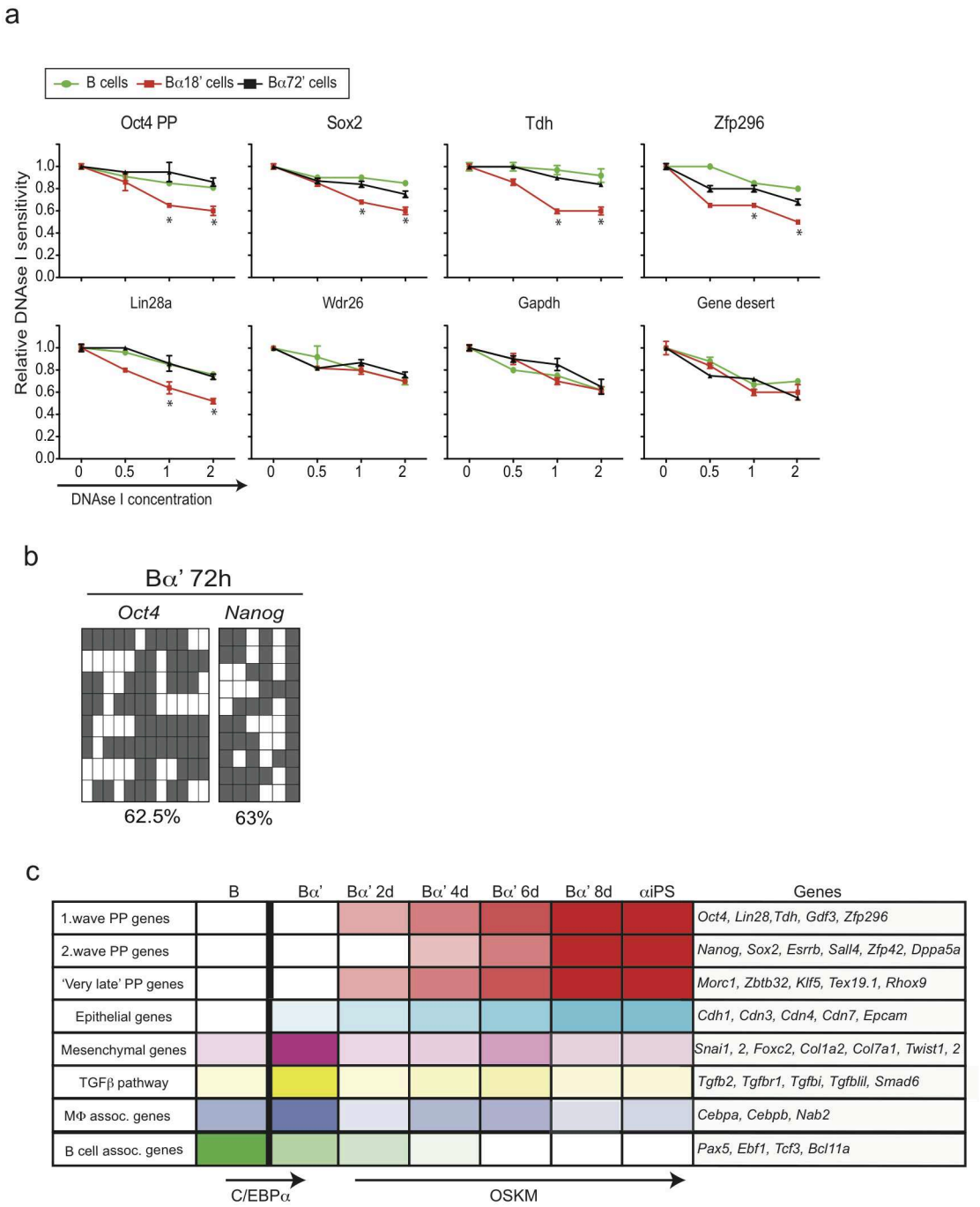
Extended Data Figure 7 | ChIP-seq and RNA-seq analysis. **a**, Number of upregulated and downregulated genes (>1.5-fold change) after a pulse of C/EBPα. **b**, Result of gene set enrichment analysis on RNA-seq data (ranked by fold change) for selected significantly enriched gene sets (from MSigDB cp.v4 database). **c**, Circular visualization of genomic distribution of differentially expressed genes (RNA-seq) and C/EBPα binding sites (ChIP-seq). **d**, Distribution of C/EBPα peaks into genomic features, on the basis of Refseq mm9 annotation. **e**, Fraction of upregulated (red chart) and downregulated genes (blue) after a pulse of C/EBPα showing a peak at 50 kb from TSS. **f**, Variation of expression after C/EBPα pulse of nearest genes to peaks. **g**, Selected gene sets significantly enriched in genes adjacent to C/EBPα peaks based on hypergeometric tests (from MSigDB cp.v4 database). Bonferoni procedure, P value <0.05.

©2014 Macmillan Publishers Limited. All rights reserved



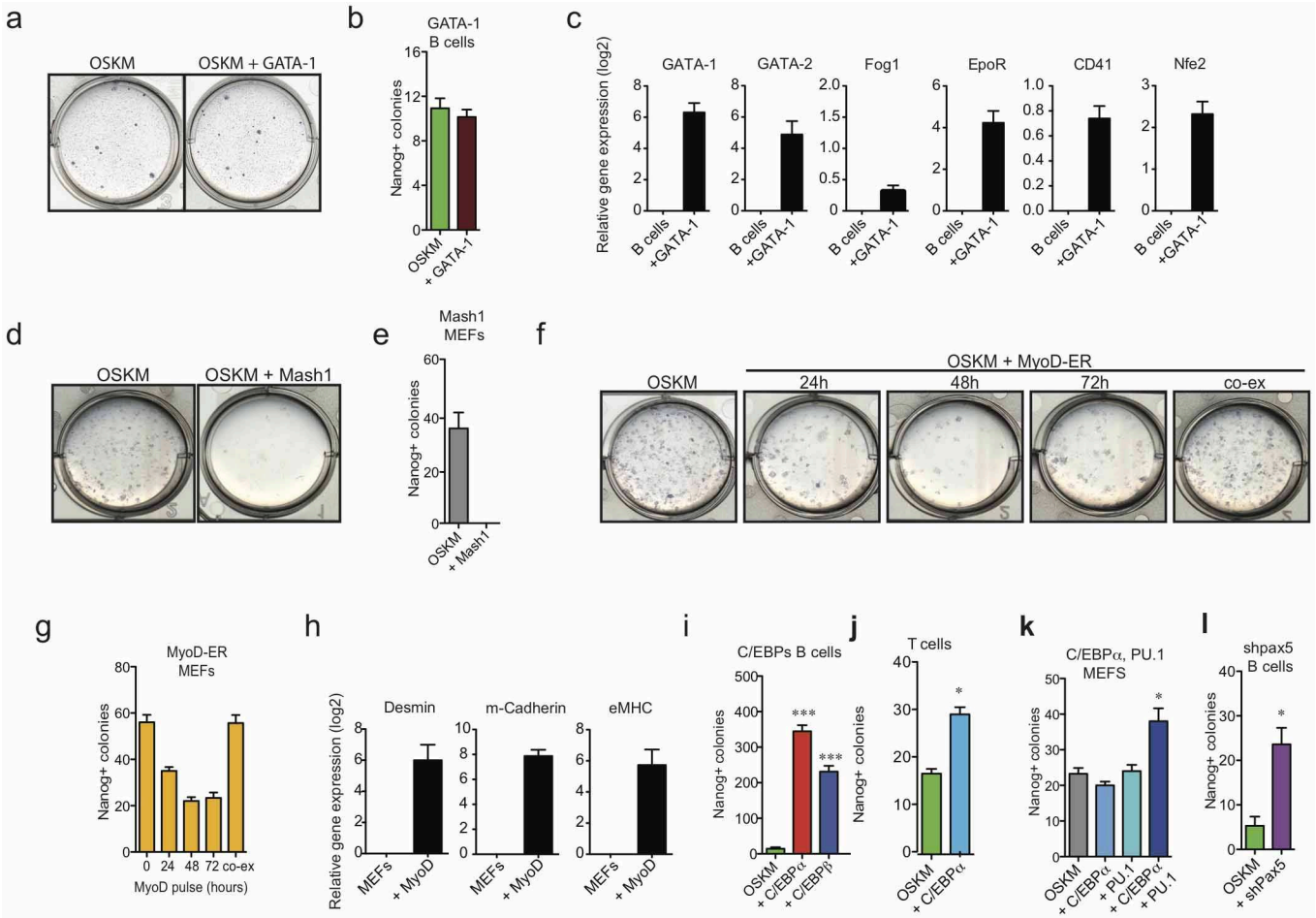
Extended Data Figure 8 | Tet2 expression in B cells and MEFs. **a**, RT–qPCR of Tet2, pluripotency and MET genes after induction of OSKM in MEFs pulsed or not with C/EBPα. Error bars indicate s.d. ($n = 3$). **b**, **c**, hMeDIP and Tet2 binding at regulatory regions of the Tet2 target gene *Hal1* (ref. 3) and *Sall4*. **d**, Tet2 localization in B cells pulsed with C/EBPα for 72 h. **e**, Oestradiol treatment has no effect on cellular localization of Tet2 protein in B cells (scale bar, 30 μm). **f**, Cellular localization of Tet2 protein in MEFs treated or not with E2 (scale bar, 100 μm).

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | DNase I sensitivity and gene expression summary. **a**, DNase I sensitivity at regulatory regions in B cells and Bα′ cells (data from Fig. 4a) and B cells pulsed for 72 h determined by enzyme titration. Error bars indicate s.d. ($n = 3$). Student's t-test $^*P < 0.05$ relative to control. **b**, CpGs analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters. Filled rectangles represent methylated CpGs, empty rectangles unmethylated residues. **c**, Summary of relevant gene expression changes during the transition from B cells to Bα′ cells and Bα′ cells to αiPS cells.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Transcription factor and cell type specificity. **a**, **b**, B+OSKM cells were infected with a retrovirus expressing GATA1 or left uninfected, induced with doxycycline and scored for 12 d.p.i. alkaline phosphatase$^+$ and Nanog$^+$ colonies. Error bars indicate s.d. ($n = 3$). **c**, qRT–PCR for megakaryocyte/erythroid-restricted genes in B cells infected with GATA1 retrovirus for 3 days. Error bars indicate s.d. ($n = 3$). **d**, **e**, Alkaline phosphatase$^+$ colonies at 15 d.p.i. in OSKM MEFs infected with a retrovirus expressing Mash1 or mock infected, and Nanog$^+$ colonies counted. Error bars indicate s.d. ($n = 3$). **f**, **g**, Effect of MyoD expression on iPS cell reprogramming of MEFs. Cells were infected with MyoD-ER and sequentially induced with E2 to activate MyoD and doxycycline to activate OSKM. Representative plates with alkaline phosphatase$^+$ iPS cell colonies 15 d.p.i. of MEFs pre-induced for 24, 48 or 72 h or continuously with E2. Error bars indicate s.d. ($n = 3$). **h**, Expression of muscle-restricted genes in OSKM MEFs 3 days after MyoD-ER induction. Error bars indicate s.d. ($n = 3$). **i**, Effects of C/EBPα and C/EBPβ in B+OSKM cells. Student's *t*-test ***$P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$). **j**, Effect of an 18-h C/EBPα pulse in pre-T+OSKM cells on the formation of 12 d.p.i. Nanog$^+$ colonies. Student's *t*-test *$P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **k**, Effects of C/EBPα, PU.1 and their combination in MEFs. Student's *t*-test *$P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **l**, Effect of Pax5 knock down in B-cell reprogramming. Student's *t*-test *$P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$).

©2014 Macmillan Publishers Limited. All rights reserved

5.2 Publication: Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells (Scientific Data, 2016)

Contribution:

I processed and analysed all the data.

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Reprogramming
» Induced pluripotent stem cells

Received: 21 February 2014
Accepted: 22 April 2014
Published: 27 May 2014

Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells

Bruno Di Stefano[1,2], Samuel Collombet[3,4,5] and Thomas Graf[1,2,6]

The reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) is lengthy and inefficient. The development of a reprogramming system that allows rapid and synchronous reprogramming to pluripotency is imperative for understanding the mechanism of iPSC formation and for future therapeutic applications. We have recently reported that a short expression in mouse primary B cells of the transcription factor C/EBPα before the induction of pluripotency factors increases the iPSC reprogramming efficiency >100-fold, involving 95% of the cells within a week. Here we present a dataset containing the time course of gene expression during this process as determined by microarray and RNA-seq techniques.

Design Type(s)	time series design • cell type comparison design
Measurement Type(s)	transcription profiling assay
Technology Type(s)	DNA microarray • next generation sequencing
Factor Type(s)	replicate • pulsed with C/EBPalpha • timepoint • technology type
Sample Characteristic(s)	Mus Musculus • pre-B-lymphocyte • embryonic stem cell • induced pluripotent stem cell

[1]Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003, Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003, Barcelona, Spain. [3]École Normale Supérieure, Institut de Biologie de l'ENS, 45 Rue d'Ulm, Paris, F-75005, France. [4]Inserm, U1024, Paris, F-75005, France. [5]CNRS, UMR 8197, Paris, F-75005, France. [6]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010, Barcelona, Spain.
Correspondence and requests for materials should be addressed to B.D.S. or T.G. (bruno.distefano@crg.eu or thomas.graf@crg.eu)

Background and Summary

Reprogramming somatic cells into induced pluripotent cells (iPSCs) by Oct4, Sox2, Klf4 and Myc (OSKM)[1] has revolutionized stem cell biology by opening unprecedented opportunities for cell therapy and disease modeling[2]. However, the finding that only a small proportion of the cells become reprogrammed, typically requiring >12 days, has hampered progress towards understanding the mechanism[3]. Recently, it has been proposed that reprogramming proceeds through a stochastic phase followed by a hierarchic phase in which the pluripotency network is established[4–6]. In addition, formation of iPSCs is preceded by the collapse of the somatic program and a mesenchymal to epithelial transition[7–10], followed by the upregulation of endogenous pluripotency genes. Our earlier work showed that C/EBPα very efficiently and rapidly induces the transdifferentiation of immature and mature B cells into macrophages by upregulating macrophage genes and silencing the B cell program, without involving retrodifferentiation[11–13]. C/EBPα has also been reported to enhance the reprogramming efficiency of B cells to pluripotency when co-expressed with OSKM[14,15], although the mechanism remains unknown and the final efficiency only reaches 1–3%.

We have recently found that an 18 h pulse of C/EBPα expression in B cells followed by OSKM activation induces a >100-fold increase in the iPSC reprogramming efficiency, involving up to 95% of the cells within a week[16]. Concomitantly, the cells undergo an epithelial to mesenchymal transition and pluripotency genes become upregulated to levels comparable to embryonic stem (ES) and iPS cells. In addition, B cells pulsed with C/EBPα (Bα′ cells) already initiate an epithelial to mesenchymal transition and downregulate the B cell program. In serum-free conditions, up to 70% Oct4-GFP positive cells are observed within 2 days, representing the fastest reprogramming to pluripotency reported so far. These results are consistent with the idea that the C/EBPα pulse helps to overcome the stochastic phase of iPSC reprogramming. The accelerated iPSC reprogramming approach described should help to fully elucidate the early events of reprogramming to pluripotency and, if applicable to human cells, could have potential clinical applications. Here, we describe gene expression profiling by two approaches, microarray and RNA-seq, during the reprogramming of mouse primary B cells, pulsed or not with C/EBPα, into iPSCs. Expression profiles from four time points across days 2 to 8 post-induction were analyzed in our previous work[16]. In this Data Descriptor we provide a detailed description of the full dataset, including the gene profiles from additional early time points not described earlier.

Methods

The following section is an expanded version of the methods description provided in Di Stefano *et al.*[16]

Overview

Mouse primary B cells were isolated from the Oct4-GFP reprogrammable mouse (containing a tetracycline-controlled transactivator (rtTa), a doxycycline-responsive OSKM cassette and an Oct4-GFP cassette)[17] and infected with a C/EBPα-hCD4 retrovirus[11]. hCD4-positive B cells were sorted and incubated for 18 h with β-estradiol (E2) to activate C/EBPα, followed by inducer washout and OSKM activation by doxycycline treatment. The RNA was then collected at different time points during reprogramming (Fig. 1), including samples at 3, 6, 12 and 24 h post-induction (hpi) with OSKM that were not incorporated into our previous study[16] but were actually part of the same experiment. In parallel, control B cells (not pulsed with C/EBPα) were exposed to doxycycline for OSKM induction and the RNA was collected following the scheme in Fig. 1. Embryonic stem cells and newly established iPSC lines (at passage 3) were used as a control for pluripotency gene expression.

Mice

The reprogrammable mouse line containing a tetracycline-inducible OSKM cassette as well as rtTA has been described previously[17]. Twelve mice (B6 strain, males and females, all homozygous for the OSKM and rtTA cassettes) were used to collect the RNA samples.

Cell cultures and reprogramming

Isolation of hematopoietic cells from bone marrow was performed as previously described[18]. $CD19^+$ cells (a mixture of pro-B and pre-B cells hereafter called B cells) were isolated using monoclonal antibody 1D3, purchased from BD Pharmingen, using MACS (Miltenyi Biotech). The purity of the sorted cell fractions (>98%) was confirmed by FACS using an LSR II flow cytometer (BD).

After isolation, B cells were grown in RPMI medium supplemented with 10% fetal bovine serum (FBS) and IL-7 (10 ng/ml) (B cell medium) for 1 week before inducing reprogramming on gelatinized plates seeded with a feeder layer of the OP9 stromal cell line.

Two different ESC lines (R1 (ATCC) and Bruce-4 (Millipore)) were used as positive controls for pluripotency gene expression. ESCs and B cell derived iPSCs were cultured on mitomycin C treated mouse embryo fibroblast (MEF) feeder cells in KO-DMEM medium (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, nonessential amino acids, β-mercaptoethanol, 1,000 U/ml LIF (ESC medium) and 15% FBS (Invitrogen) (ESC medium).

MEF cultures were established by trypsin digestion of mouse embryos (embryonic day 13.5) and the resulting cells cultured in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin.

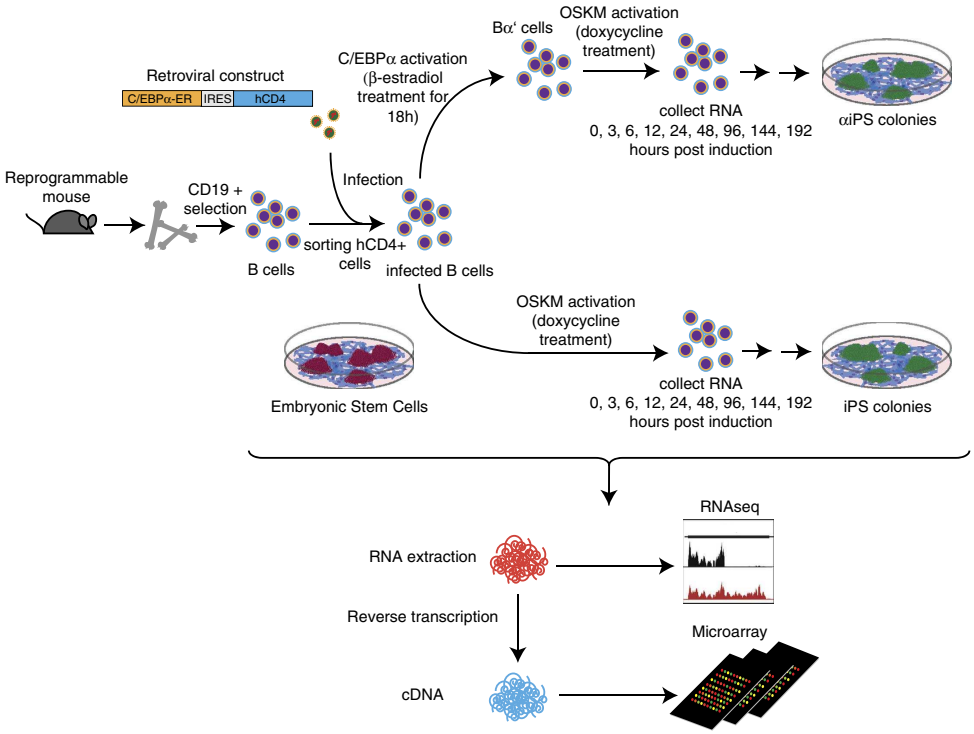


Figure 1. Experimental design.Mouse primary B cells were purified from bone marrow of reprogrammable mice and infected with a retrovirus carrying an estradiol (E2) inducible form of C/EBPα. Infected cells were sorted based on the expression of the cell surface marker hCD4 and plated in B cell medium on OP9 stromal cells. B cells were then exposed for 18 h to E2, followed by washout of the inducer and addition of doxycycline to activate OSKM factors. Control B cells were directly induced with doxycycline. Cells were collected at the indicated time points and FACS sorted for live cells before RNA extraction. Gene expression profiling was performed by microarray and RNA-seq.

For the reprogramming, B cell medium was changed to ESC medium supplemented with 2 μg/ml of doxycycline and 15% FBS, IL-4 (10 ng/ml), IL-7 (10 ng/ml) and IL-15 (2 ng/ml). B cells and Bα′ cells were seeded at a density of 500 cells/cm^2 in six-well plates. Bα′ cells were generated by exposing B cells for 18 h to 100 nM of E2 followed by inducer washout. To activate OSKM, the cells were subsequently treated with 2 μg/ml doxycycline. The reprogramming medium was changed every 2 days. For expansion of iPSC lines, colonies with ESC morphology were picked after doxycycline withdrawal at 12 days post induction and grown on MEF feeder layer in ESC medium.

Vectors and virus production and infection

The C/EBPαER-hCD4 retroviral vector has been described previously[11] and was used to overexpress C/EBPα in B cells. The lentiviral vector expressing a constitutive dsRED has been generated by substituting the GFP with the dsRED gene in the PTYF vector[19]. This virus was utilized to generate αiPS cell lines constitutively expressing dsRED to follow chimera contribution upon injection into host blastocysts. Viral production was performed as reported earlier[20]. Briefly, HEK293T cells were co-transfected with the vector plasmid and packaging plasmids (p8.9 and VSVG) using calcium phosphate transfection. Viral supernatants were harvested 48–72 h later and concentrated by ultracentrifugation at 20,000 *g* for 2 h at 20 °C. Viral concentrates were re-suspended in 1 × PBS and stored at −80 °C. Infection of B cells was carried out by centrifugation of the cells with concentrated virus for 2 h at 32 °C at 1000 *g* in B cell medium.

Chimeric mice

For the chimera formation assay, 10 to 15 iPSCs were injected into a 3.5 days post coitum (dpc) blastocyst of CD1 mice and transferred into pseudo-pregnant 2.5 dpc recipients for chimera generation. Chimerism was ascertained by the expression of the lentiviral PGK-dsRED.

RNA isolation and quantification

RNA isolation from MEFs and B cells was performed with the miRNeasy Mini Kit (Qiagen). RNA was collected from sorted live cells at 0, 3, 6, 12, 24, 48, 96, 144 and 192 hpi. To remove the feeders, ESCs and

iPSC clones at passage 2 or higher were seeded on gelatinized plates and processed with the above kits. RNA was eluted from the columns using RNase-free water or TE buffer and quantified by Nanodrop.

Gene expression arrays

RNA samples (with an RNA integrity number (RIN) >9) were subjected to gene expression analyses using Agilent expression arrays. Briefly: 100 ng of total RNA was labeled using LowInputQuick Amp Labeling Kit (Agilent 5190-2305) following the manufacturer's instructions. mRNA was reverse transcribed in the presence of T7-oligo-dT primer to produce cDNA. cDNA was then *in vitro* transcribed with T7 RNA polymerase in the presence of Cy3-CTP to produce labeled cRNA. The labeled cRNA was hybridized to the Agilent SurePrint G3 gene expression 8 × 60 K microarray according to the manufacturer's protocol. The arrays were washed and scanned on an Agilent G2565CA microarray scanner at 100% PMT and 3 μm resolution. Intensity data were extracted using the Feature Extraction software (Agilent). Raw data were taken from the Feature Extraction output files and corrected for background noise using the normexp method[21]. To assure comparability across samples, we used quantile normalization. Probe Level Quantile Normalization of High Density Oligonucleotide Array Data. http://bmbolstad.com/stuff/qnorm.pdf). (Bolstad, B. (2001)).

RNA-seq

Libraries were prepared using the TruSeq Stranded mRNA Sample Prep Kit v2 (ref. RS-122-2101/2, Illumina) according to the manufacturer's protocol. Briefly, 900 ng of total RNA was used for poly(A)-mRNA selection using streptavidin-coated magnetic beads, and subsequently fragmented to approximately 300 bp. cDNA was synthesized using reverse transcriptase (SuperScript II, ref. 18064–014, Invitrogen) and random primers. The second strand of the cDNA incorporated dUTP in place of dTTP. Double-stranded DNA was further used for library preparation. DNA was subjected to A-tailing and ligation of the barcoded Truseq adapters. All purification steps were performed using Qiagen PCR purification columns (refs. 50928106 and 50928006). Library size-selection was done with 2% low-range agarose gels. Fragments with insert sizes of 200 to 400 bp were cut out from the gel, and DNA was extracted using QIAquick Gel Extraction Kit (ref. 50928706, Qiagen) and eluted in 20 μl EB. Library amplification was performed by PCR on the size-selected fragments using the primer cocktail supplied in the kit.

Finally, the libraries were analyzed using Agilent DNA 1000 chips to determine the quantity and size distribution, and then quantified by qPCR using the KAPA Library Quantification Kit (ref. KK4835, KapaBiosystems) prior to amplification with Illumina's cBot. Libraries were loaded at a concentration of 10 pM onto the flowcell and sequenced on Illumina's HiSeq 2000. Base-calls were performed with CASAVA-1.8.2 software (Illumina). Reads were aligned to the mm9 genome with STAR v2.3.0.1, keeping only uniquely mapped reads with a maximum of two mismatches (option: outFilterMismatchNmax 2, outFilterMultimapNmax 1). Gene expression quantification was performed with the software HTSeq-count (option: mode union, stranded, features exons, attribute gene_id) on the RefSeq mm9 annotation (from the UCSC Genome Browser; the exact version used here is available on demand). Genes with uncounted reads in both conditions were removed, and normalization was performed with DESeq (option: method blind, sharingMode fit-only, fitType local).

Data Records

Gene expression profiling was performed, using both microarrays and RNA-seq approaches, on RNA samples collected from cells at 0, 3, 6, 12, 24, 48, 96, 144 and 192 hpi. Two biological replicates are available for each time point. Data from time points 0, 48, 96, 144 and 192 were analyzed in our previous work[16]. The full dataset is available in association with this Data Descriptor, including data from four earlier time points not analyzed in the previous work. A quality control report from Agilent Features Extraction software is also included. All samples are summarized in Table 1.

Data record 1

GEO accession number GSE46321 (Data Citation 1) contains microarray data for all samples listed in Table 1.

Data record 2

RNA-seq data for the B cell (T0_1) and Bα′ cell (aT0_1) samples are available under the GEO accession number GSE52396 (Data Citation 2).

Data record 3

The normalized microarray data for all the samples are available in Figshare (Data Citation 3).

Data record 4

A Quality Control report from the Feature Extraction software for the microarray data is available in Figshare (Data Citation 2).

Sample name	Organism	Tissue	Description	Sample name as appear in the manuscript
aT0_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, replicate 1	Bα′ cells
aT0_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, replicate 2	Bα′ cells
aT3_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 3 h with OSKM, replicate 1	Bα′+OSKM cells
aT3_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 3 h with OSKM, replicate 2	Bα′+OSKM cells
aT6_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 6 h with OSKM, replicate 1	Bα′+OSKM cells
aT6_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 6 h with OSKM, replicate 2	Bα′+OSKM cells
aT12_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 12 h with OSKM, replicate 1	Bα′+OSKM cells
aT12_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 12 h with OSKM, replicate 2	Bα′+OSKM cells
aT24_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 24 h with OSKM, replicate 1	Bα′+OSKM cells
aT24_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 24 h with OSKM, replicate 2	Bα′+OSKM cells
aT48_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 48 h with OSKM, replicate 1	Bα′+OSKM cells
aT48_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 48 h with OSKM, replicate 2	Bα′+OSKM cells
aT96_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 96 h with OSKM, replicate 1	Bα′+OSKM cells
aT96_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 96 h with OSKM, replicate 2	Bα′+OSKM cells
aT144_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 144 h with OSKM, replicate 1	Bα′+OSKM cells
aT144_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 144 h with OSKM, replicate 2	Bα′+OSKM cells
aT192_1	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 192 h with OSKM, replicate 1	Bα′+OSKM cells
aT192_2	Mus musculus	primary B cells	B cells pulsed for 18 h with C/EBPα, induced for 192 h with OSKM, replicate 2	Bα′+OSKM cells
T0_1	Mus musculus	primary B cells	B cells uninduced, replicate 1	B cells
T0_2	Mus musculus	primary B cells	B cells uninduced, replicate 2	B cells
T3_1	Mus musculus	primary B cells	B cells induced with OSKM for 3 h replicate 1	B+OSKM
T3_2	Mus musculus	primary B cells	B cells induced with OSKM for 3 h replicate 2	B+OSKM
T6_1	Mus musculus	primary B cells	B cells induced with OSKM for 6 h replicate 1	B+OSKM
T6_2	Mus musculus	primary B cells	B cells induced with OSKM for 6 h replicate 2	B+OSKM
T12_1	Mus musculus	primary B cells	B cells induced with OSKM for 12 h replicate 1	B+OSKM
T12_2	Mus musculus	primary B cells	B cells induced with OSKM for 12 h replicate 2	B+OSKM
T24_1	Mus musculus	primary B cells	B cells induced with OSKM for 24 h replicate 1	B+OSKM
T24_2	Mus musculus	primary B cells	B cells induced with OSKM for 24 h replicate 2	B+OSKM
T48_1	Mus musculus	primary B cells	B cells induced with OSKM for 48 h replicate 1	B+OSKM
T48_2	Mus musculus	primary B cells	B cells induced with OSKM for 48 h replicate 2	B+OSKM
T96_1	Mus musculus	primary B cells	B cells induced with OSKM for 96 h replicate 1	B+OSKM
T96_2	Mus musculus	primary B cells	B cells induced with OSKM for 96 h replicate 2	B+OSKM
T144_1	Mus musculus	primary B cells	B cells induced with OSKM for 144 h replicate 1	B+OSKM
T144_2	Mus musculus	primary B cells	B cells induced with OSKM for 144 h replicate 2	B+OSKM
T192_1	Mus musculus	primary B cells	B cells induced with OSKM for 192 h replicate 1	B+OSKM
T192_2	Mus musculus	primary B cells	B cells induced with OSKM for 192 h replicate 2	B+OSKM
R1	Mus musculus	Embryonic stem cells	Embryonic stem cells	ESCs
alpha8	Mus musculus	Induced pluripotent stem cells (iPS cells)	iPS cells derived from B cells pulsed with C/EBPα	αiPSCs
bruce_1	Mus musculus	Embryonic stem cells	Embryonic stem cells, replicate 1	ESCs
bruce_2	Mus musculus	Embryonic stem cells	Embryonic stem cells, replicate 2	ESCs
iPS26_1	Mus musculus	Induced pluripotent stem cells (iPS cells)	iPS cells derived from B cells, replicate 1	iPSCs
iPS26_2	Mus musculus	Induced pluripotent stem cells (iPS cells)	iPS cells derived from B cells, replicate 2	iPSCs
alpha22_1	Mus musculus	Induced pluripotent stem cells (iPS cells)	iPS cells derived from B cells pulsed with C/EBPα, replicate 1	αiPSCs
alpha22_2	Mus musculus	Induced pluripotent stem cells (iPS cells)	iPS cells derived from B cells pulsed with C/EBPα, replcate 2	αiPSCs

Table 1. Samples employed in the present study.

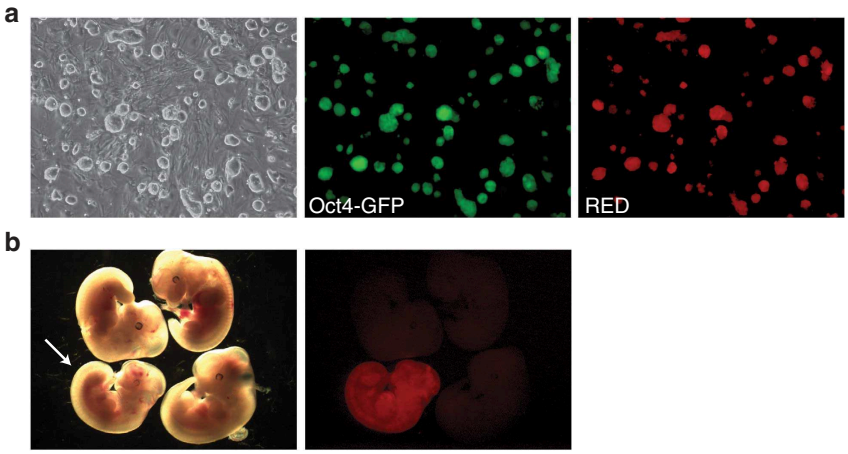


Figure 2. *In vivo* differentiation potential of αiPS cells. (**a**) Oct4-GFP^{+} αiPS cells generated from mouse primary B cells expressing dsRED under the constitutive PGK promoter. (**b**) Contribution of aiPS#3 dsRED-labeled cells to chimera formation after blastocyst injection. White arrow indicates a dsRED-expressing chimeric mouse.

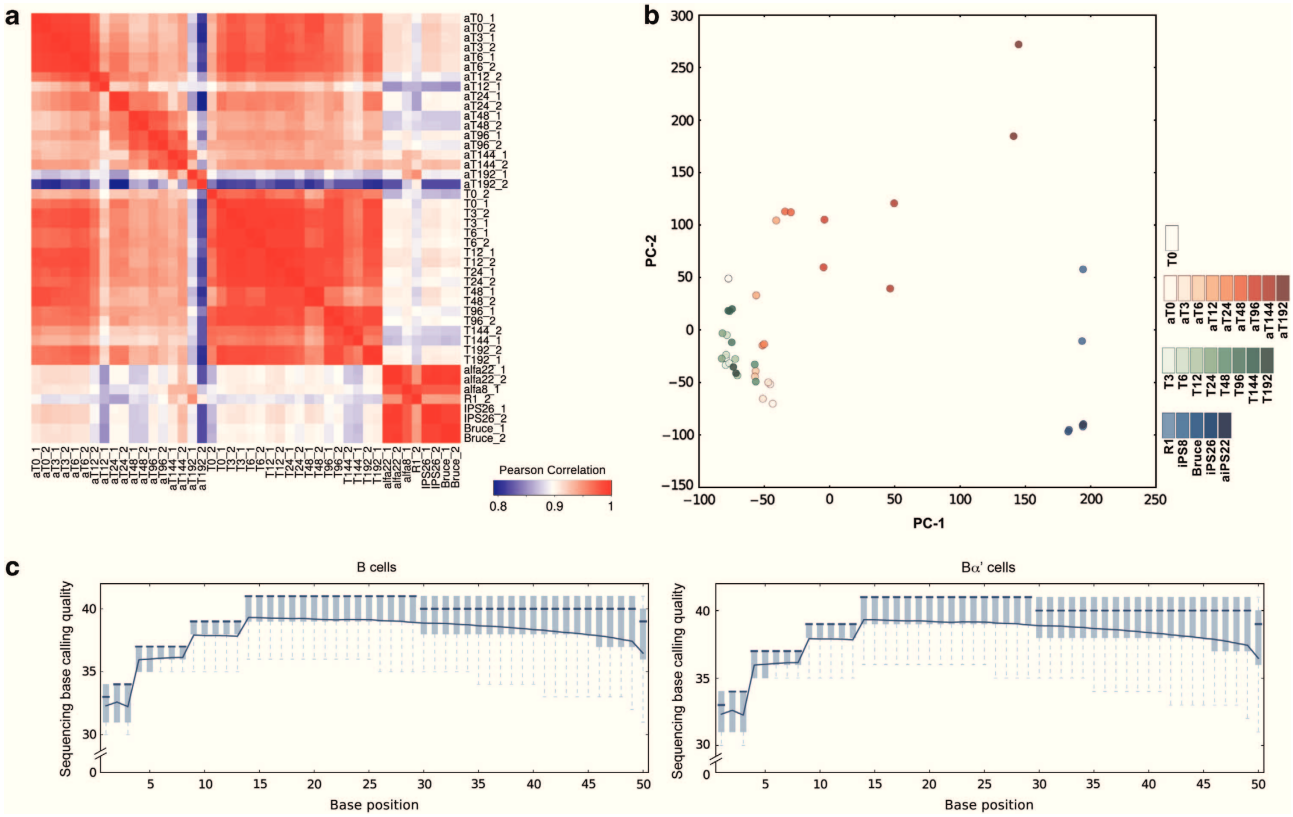


Figure 3. Quality control of high throughput data. (**a**) Heatmap showing Pearson correlation for duplicate microarray samples; the lower limit is 0.8 because correlation was calculated on all probes and most of them do not exhibit any change under any condition, leading to a high correlation between all samples. (**b**) Principal component analysis. Percentage of total variance explained by PC-1: 36%; by PC-2: 21%. (**c**) Distribution of RNA-seq sequencing quality along the reads, for B cells (left) and Bα′ cells (right). The X-axis corresponds to the position in the read and the Y-axis reflects the sequencing quality, with box plots representing 10th–25th–75th and 90th percentiles over all the reads. The lower part of the Y-axis scale has been omitted because all the observed values are above 30. The similarities between the B and Bα′ samples indicate uniformly high quality. For more explanations see technical validation part.

Technical Validation

Confirmation of pluripotency

Bone marrow isolated mouse primary B cells were pulsed with C/EBPα for 18 h and doxycycline treated to activate the OSKM cassette in ES medium on inactivated OP9 feeder cells. After doxycycline washout at day 12, transgene independent Oct4-GFP clones were picked and expanded in ES medium as stable iPSC lines (referred to as αiPSCs) (Fig. 2a). In order to assess the *in vivo* differentiation potential of the expanded cell lines, iPSCs expressing dsRED and Oct4-GFP were injected into host blastocysts. High contribution to chimerism was observed in developing embryos (Fig. 2b).

Quality control of high throughput data

For microarrays (Data Records 1, 3 and 4; Data Citations 1 and 3), after background correction and quantile normalization, we looked at sample correlation at the probe level (Fig. 3a). All pairs of replicates correlate very well (the diagonal of the heatmap), and all pluripotent stem cell samples show a very high correlation (lower right block). Principal component analysis (PCA) (Fig. 3b) confirmed the similarity between replicates and all pluripotent stem cells (ES/iPS), the absence of major changes during OSKM-only induction (green points) and the trajectory from B cells to pluripotent cells in Bα′+OSKM cells. The high correlation between all the OSKM-only samples (T0 to T192) was further confirmed by qRT-PCR analysis[16], showing that gene expression in these remains essentially unchanged.

For RNA-seq data (Data Record 2; Data Citation 2), we sequenced 87,105,048 and 83,487,348 reads for B and Bα′ cells, respectively. Sequencing quality was good and did not indicate read bias (Fig. 3c). Using the STAR aligner[22], we found 69,716,252 and 70,864,793 reads uniquely mappable onto the mm9 genome, and found 52,070,014 and 55,749,585 reads (60 and 67% of total sequenced reads) mapping to an annotated exon of the RefSeq mm9 annotation, indicating high quality mRNA sequencing.

Usage Notes

For gene expression microarray data, differentially expressed genes can be identified using various methods, in particular the popular LIMMA R package[23] or the EDGE software[24]. For RNA-seq, we recommend using TopHat as a read aligner. Recent protocols explain how to process and analyze RNA-seq data[25,26].

References

1. Takahashi, K., Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Robinton, D. A., Daley, G. Q. The promise of induced pluripotent stem cells in research and therapy. *Nature* **481,** 295–305 (2012).
3. Stadtfeld, M., Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24,** 2239–2263 (2010).
4. Hanna, J. H., Saha, K., Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143,** 508–525 (2010).
5. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150,** 1209–1222 (2012).
6. Buganim, Y., Faddah, D. A., Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* **14,** 427–439 (2013).
7. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
8. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
9. Hansson, J. *et al.* Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Reps* **2,** 1579–1592 (2012).
10. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
11. Xie, H., Ye, M., Feng, R., Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117,** 663–676 (2004).
12. Bussmann, L. H. *et al.* A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* **5,** 554–566 (2009).
13. Di Tullio, A. *et al.* CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108,** 17016–17021 (2011).
14. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
15. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat. Genet.* **41,** 968–976 (2009).
16. Di Stefano, B. *et al.* C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506,** 235–239 (2014).
17. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J., Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* **7,** 56–59 (2010).
18. Lo Celso, C., Scadden, D. Isolation and transplantation of hematopoietic stem cells (HSCs). *J. Vis. Exp* **2,** 157 (2007).
19. He, J., Kallin, E. M., Tsukada, Y., Zhang, Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nat. Struct. Mol. Biol.* **15,** 1169–1175 (2008).
20. Di Stefano, B. *et al.* A microRNA-based system for selecting and maintaining the pluripotent state in human induced pluripotent stem cells. *Stem Cells* **29,** 1684–1695 (2011).
21. Ritchie, M. E. *et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23,** 2700–2707 (2007).
22. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
23. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Molec. Biol.* **3,** 3 (2004).
24. Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G., Davis, R. W. Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. U.S.A.* **102,** 12837–12842 (2005).

25. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8,** 1765–1786 (2013).
26. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).

Data Citations

1. Di Stefano, B., Collombet, S., Graf, T. Gene Expression Omnibus GSE46321 (2014).
2. Di Stefano, B., Collombet, S., Graf, T. Gene Expression Omnibus GSE52396 (2014).
3. Di Stefano, B., Collombet, S., Graf, T. Figshare. http://dx.doi.org/10.6084/m9.figshare.939408 (2014).

Acknowledgements

We would like to thank Francesco Limone for his help with the figures, Laura Batlle for the generation of chimeric mice and the CRG genomic facility for high throughput sequencing and help with the bioinformatics analysis.

Author Contributions

T.G. and B.D.S. designed the project and the experiments and wrote the paper. B.D.S. performed the cell culture, the reprogramming experiments, the RNA collections and the molecular biology. S.C. was in charge of the bioinformatics analysis. This work was supported by the Ministerio de Educación y Ciencia, SAF.2007-63058 and AGAUR 2009 SGR768. B.D.S. holds a La Caixa International PhD Fellowship.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Di Stefano, B. *et al.* Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells. *Sci. Data* 1:140008 doi: 10.1038/sdata.2014.8 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/

Metadata associated with this Data Descriptor is available at http://www.nature.com/sdata/ and is released under the CC0 waiver to maximize reuse.

5.3 Publication: A Transcription Factor Pulse Can Prime Chromatin for Heritable Transcriptional Memory (Molecular and Cellular Biology, 2016)

Contribution:

I performed all the data processing and analysis, the identification of genes displaying long term memory, the integration of different ChIP-seq data, and I participated to the design of subsequent experiments.

AMERICAN SOCIETY FOR MICROBIOLOGY | Molecular and Cellular Biology®

RESEARCH ARTICLE

A Transcription Factor Pulse Can Prime Chromatin for Heritable Transcriptional Memory

Aimee Iberg-Badeaux,[a,b] Samuel Collombet,[c] Benoit Laurent,[a,b] Chris van Oevelen,[d*] Kuo-Kai Chin,[b] Denis Thieffry,[c] Thomas Graf,[d] Yang Shi[a,b]

Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA[a]; Department of Newborn Medicine, Boston Children's Hospital, Boston, Massachusetts, USA[b]; Computational Systems Biology Team, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, Ecole Normale Supérieure, PSL Research University, Paris, France[c]; Center for Genomic Regulation, Barcelona, Spain[d]

ABSTRACT Short-term and long-term transcriptional memory is the phenomenon whereby the kinetics or magnitude of gene induction is enhanced following a prior induction period. Short-term memory persists within one cell generation or in postmitotic cells, while long-term memory can survive multiple rounds of cell division. We have developed a tissue culture model to study the epigenetic basis for long-term transcriptional memory (LTTM) and subsequently used this model to better understand the epigenetic mechanisms that enable heritable memory of temporary stimuli. We find that a pulse of transcription factor CCAAT/enhancer-binding protein alpha (C/EBPα) induces LTTM on a subset of target genes that survives nine cell divisions. The chromatin landscape at genes that acquire LTTM is more repressed than at those genes that do not exhibit memory, akin to a latent state. We show through chromatin immunoprecipitation (ChIP) and chemical inhibitor studies that RNA polymerase II (Pol II) elongation is important for establishing memory in this model but that Pol II itself is not retained as part of the memory mechanism. More generally, our work reveals that a transcription factor involved in lineage specification can induce LTTM and that failure to rerepress chromatin is one epigenetic mechanism underlying transcriptional memory.

KEYWORDS trained immunity, transcriptional memory, epigenetic, priming, inflammation, chromatin

Received 29 June 2016 **Returned for modification** 29 July 2016 **Accepted** 29 November 2016

Accepted manuscript posted online 5 December 2016

Citation Iberg-Badeaux A, Collombet S, Laurent B, van Oevelen C, Chin K-K, Thieffry D, Graf T, Shi Y. 2017. A transcription factor pulse can prime chromatin for heritable transcriptional memory. Mol Cell Biol 37: e00372-16. https://doi.org/10.1128/MCB.00372-16.

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Thomas Graf, Thomas.Graf@crg.eu, or Yang Shi, yang_shi@hms.harvard.edu.

* Present address: Chris van Oevelen, HogeSchool Utrecht, ILC, Utrecht, the Netherlands.

A.I.-B. and S.C. contributed equally to this article.

Transcriptional memory in yeast and mammalian cells can be induced by signaling events culminating in transcription factor (TF) recruitment to target gene promoters and enhancers (1–4). Mechanistic studies of long-term transcriptional memory (LTTM) in mammalian cells are currently limited (5–11); however, short-term transcriptional memory (STTM) has been studied in greater detail. For example, primary macrophages display transcriptional memory of cytokine stimulation and pathogen exposure (1, 9, 10), yet the underlying molecular events are not established. Memory of this type in macrophages is thought to contribute to the "trained immunity" phenotype, whereby macrophages display a heightened and adapted response to repeat pathogen exposures (10). One epigenetic event thought to contribute to STTM in macrophages is the conversion of gene enhancers from a latent/off state to a poised state by transcription factor recruitment and decoration with active enhancer-associated histone modifications (1). Based on these studies, we reasoned that a short pulse of transcription factor activity could mimic a temporary signal transduction event and that such an approach could be used to study the mechanisms underlying long-term transcriptional memory in mammalian cells. Furthermore, we hypothesized that the use of a system that utilizes transcription factors with the ability to invade closed chromatin sites (i.e., pioneering

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

factors) could result in enhanced memory effects. Indeed, binding of these factors can make previously inaccessible chromatin sites transcriptionally competent and aid in the conversion of gene regulatory elements from a latent/off to a primed or active state (12). CCAAT/enhancer-binding protein alpha (C/EBPα) and AP-1 TF families have been reported to display pioneering factor activity in the myeloid lineage (13, 14), while C/EBPα overexpression in conjunction with endogenous PU.1 works as a module to mediate complete B cell-to-macrophage transdifferentiation (TD) (15, 16). Therefore, we reasoned that the inducible TD system developed previously (15) could be employed to decipher the epigenetic mechanisms contributing to the transcriptional memory phenotype in immune cells. We asked whether a short C/EBPα pulse, insufficient to commit B cells toward the macrophage lineage, might be sufficient to induce LTTM of target genes by initiating chromatin changes that are maintained through cell division. Here, we show that LTTM is indeed observed for a subset of C/EBPα target genes, many of which are important mediators of the pathogen- and interferon-induced monocyte/macrophage inflammatory response, e.g., *Cd14*, *Ifitm1*, and *Lrp1*. The TD model was then employed to define the molecular mechanisms responsible for memory and mitotic inheritance of chromatin state changes. Transcription factor recruitment and histone acetylation are not sufficient to induce memory; however, memory acquisition is blocked with RNA polymerase II (Pol II) elongation inhibitors and tends to occur on genes with a more repressed chromatin state upon the initial stimulus. We present data supporting a critical role for demethylation of histone H3 at K27 (H3K27) in establishing LTTM of *Ifitm1* through a mechanism of loss and failure to regain a repressive chromatin landscape. Taking our findings together, we demonstrate that epigenetic priming can indeed confer heritable cellular memory of temporary TF activity in mammalian cells.

RESULTS

A C/EBPα pulse induces long-term transcriptional memory at a subset of target genes. In the B cell-to-macrophage TD model (15), exogenous C/EBPα is induced and collaborates with endogenous PU.1 at enhancer elements to activate the macrophage gene expression program (14). Interestingly, a commitment point during the TD process is reached between 18 and 24 h whereby the induced, exogenous C/EBPα expression can be removed, but the cells continue to convert toward macrophage specification (15). Using this knowledge, we designed a pulse-chase-restimulation protocol to determine if any target genes display long-term memory after temporary activation by C/EBPα (Fig. 1A). The protocol includes an initial pulse period of 6 or 12 h, sufficient time for the induction of chromatin changes (14) but not for commitment, and a chase period of 6 days (doubling time of 15.4 h, resulting in 9.3 cell divisions in 144 h) (see Fig. S1A in the supplemental material). To validate the protocol design, we determined if the C/EBPα transgene or other known TFs induced in the pulse (17) were maintained at higher levels during the chase. To test this, we performed Western blot analyses on nuclear and cytoplasmic extracts to gauge C/EBPα levels (to track the shuttling of the transgene) and on nuclear extracts to gauge PU.1 and Runx1 levels. By day 3 of the chase period, total levels of tested TFs were comparable to levels in the control cells, suggesting that our memory protocol was designed with adequate time in the chase period for induced TFs to return to baseline and for C/EBPα-estrogen receptor (ER) to return to the cytoplasm (Fig. 1B). To generate a comprehensive list of possible memory events, transcriptome sequencing (RNA-seq) was performed to measure transcript levels before and after 6 h of C/EBPα stimulation from (i) naive cells (vehicle-treated control [CT] cells), (ii) cells previously pulsed for 6 h (6hP), and (iii) cells previously pulsed for 12 h (12hP) (average from 3 replicates). RNA-seq data were used to interrogate the following: (i) to determine if any transcripts induced by the initial pulse remain elevated after the chase period, as these would be genes displaying persistent memory of induction, and (ii) to determine if any of the genes that return to baseline levels display more robust induction upon restimulation. The latter group of genes would therefore display characteristics of LTTM and are the primary interest in this study. Working under the hypothesis that chromatin changes occurring during the

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

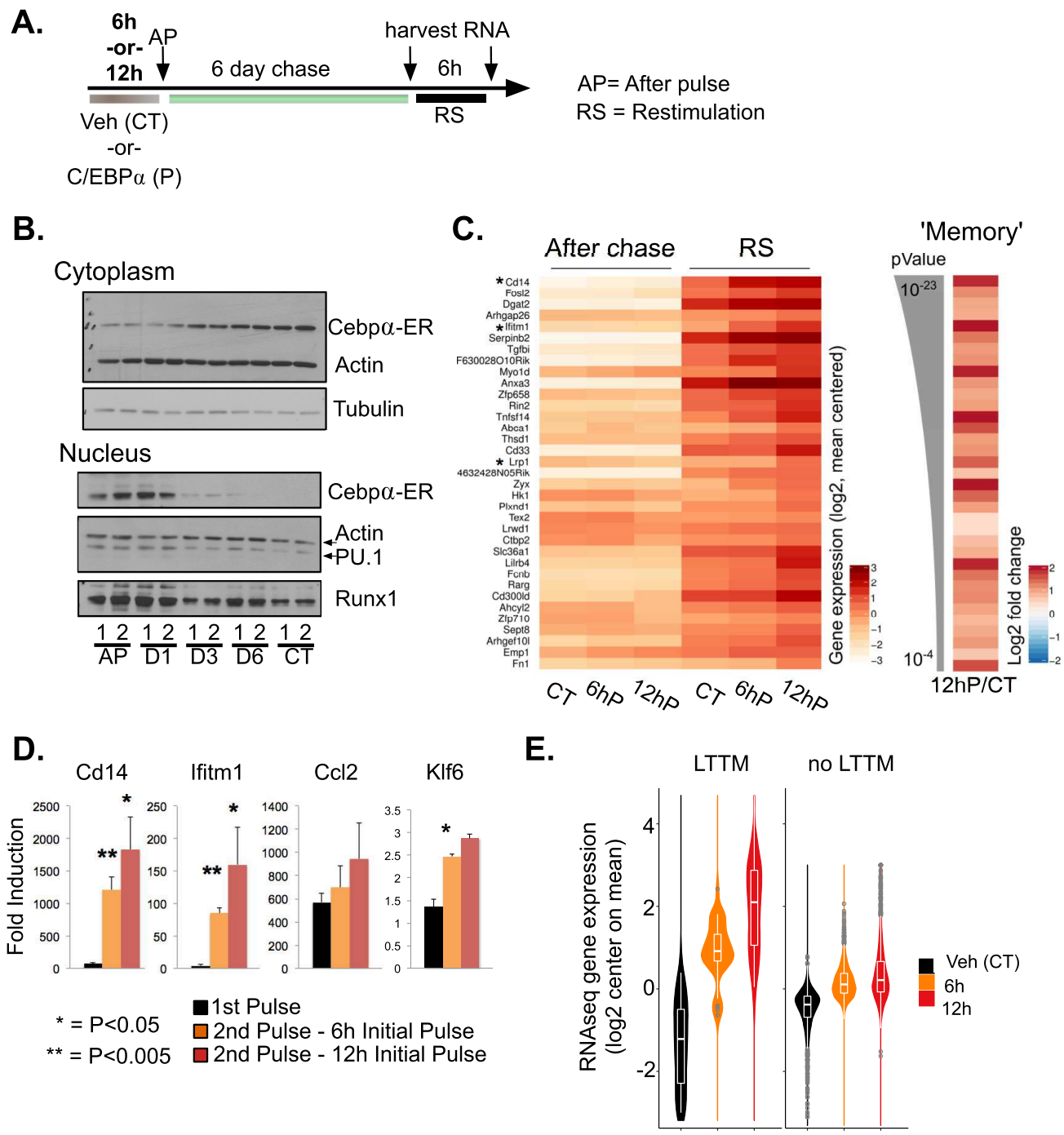


FIG 1 A C/EBPα pulse induces long-term transcriptional memory at a subset of target genes. (A) Experimental timeline showing the memory protocol, with time points analyzed in this figure indicated by arrows. For the initial pulse, C/EBPα activity was induced by the addition of β-estradiol (β-E2). Control cells (CT) were treated with vehicle (ethanol [EtOH]). After the pulse period, cells were washed three times in medium to remove the β-estradiol and relieve C/EBPα activity. (B) Cellular fractionation was performed, and lysates were analyzed by Western blotting to determine the relative levels of transgene C/EBPα-ER in the cytoplasm and nucleus after the pulse and throughout the chase period. Levels in the nucleoplasm of PU.1 and Runx1, two transcription factors induced in the pulse, were also analyzed. The numbers 1 and 2 indicate biological replicates. D1, day 1 of the chase, etc. (C) RNA-seq was performed in vehicle-treated (CT) and β-estradiol-treated pulsed cells (6 h and 12 h pulsed) after the 6-day chase period and 6 h into restimulation (RS). At left is a heat map of the 35 genes displaying highly significant LTTM at 6 h into restimulation with a 12-h initial pulse (FDR < 0.01). After chase and restimulation, transcript levels are shown for control, 6-h-pulsed (6hP), and 12-h-pulsed (12hP) samples. At right is a heat map displaying the fold difference, or memory effect, of 12-h-pulsed samples over control samples at 6 h into restimulation. Genes are listed in order from most to least significant, starting at the top. Locations in the heat map for the *Cd14*, *Lrp1*, and *Ifitm1* genes used as examples are highlighted with asterisks. Three biological replicates were sequenced for this analysis. (D) RT-qPCR confirmation of the RNA-seq results for LTTM genes *Cd14* and *Ifitm1*, as well as control genes *Ccl2* and *Klf6* that do not display significant memory. Data are displayed as fold induction during the 6-h restimulation period, with the first pulse being the initial induction of the CT cell sample during the restimulation period. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance. (E) Expression after the first pulse. Box plot diagrams show the average transcript levels in vehicle-treated CT cell samples and pulsed cells after the initial pulse (AP) for significant memory genes (LTTM, $n = 35$) and control genes that are induced by the pulse but do not acquire memory (no LTTM, $n = 1{,}562$).

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

pulse are responsible for establishing LTTM, we reasoned that a longer pulse might enhance the memory effect by allowing more time for chromatin changes to occur and accumulate. Therefore, we also used the RNA-seq data to determine if a longer pulse time (12 h versus 6 h) increases memory.

In total, only one gene (*Ccl9*) induced by the 12-h pulse remained significantly elevated after the chase period, while no gene reached this significance threshold with a 6-h pulse (log fold change, 2.2; false discovery rate [FDR], $\leq$0.01) (Table S1). Six hours into restimulation, 35 genes displayed significant LTTM after a prior 12-h pulse (termed memory genes; FDR, $\leq$0.01) (Table S1), representing approximately 3% of all genes induced by C/EBPα (Fig. 1C). Highlighted with asterisks in Fig. 1C are memory genes chosen here for further mechanistic study, namely, *Cd14*, *Lrp1*, and *Ifitm1*. *Ccl2* and *Klf6* were selected as control genes as both are induced by C/EBPα but do not display transcriptional memory as defined by the stringent criteria used to generate the list of 35 memory genes outlined in Fig. 1C.

To verify that transcripts induced at the pulse do indeed return to baseline throughout the chase, we performed reverse transcription-quantitative PCR (RT-qPCR) analysis of representative LTTM gene transcripts (*Cd14*, *Lrp1*, and *Ifitm1*) in a detailed time course through the chase period (Fig. S1B), which confirmed that transcripts returned to prestimulation levels by day 3 of the chase. RT-qPCR performed after an initial pulse in CT cells and during the secondary pulse in previously stimulated cells confirmed the existence of memory at representative genes and confirmed the increase in memory between a 6-h and 12-h initial pulse time (Fig. 1D and S1C). Interestingly, memory genes (LTTM) exhibit on average a greater fold induction at the initial pulse than control genes (no LTTM), as depicted by box plots of expression levels (Fig. 1E).

C/EBPα-pulsed cells display a more robust LPS response. The most significant memory gene in the TD model is *Cd14*. To confirm that *Cd14* displays memory at the protein level in addition to the transcript level, we performed flow cytometry analysis after the chase period and 24 h into restimulation with an antibody that recognizes membrane-bound Cd14. Following the 6-day chase, the signal intensities from CT and pulsed cells were indistinguishable. However, memory could be observed 24 h into restimulation as the percentage of cells in the population expressing high levels of Cd14 (Cd14^{high}) was significantly greater in the pulsed population (Fig. 2A). In the TD system, as cells convert from B cells to macrophages, they become increasingly responsive to lipopolysaccharide (LPS) stimulation, as measured by levels of cytokine production (15). Present on the host cell plasma membrane, Cd14 plays a major role in Toll-like receptor 4 (TLR4) sensing of bacterial wall molecule lipopolysaccharides and is essential for downstream signaling events and cytokine production (18). We therefore tested the ability of pulsed and control cells to respond to LPS during the early stages of TD (Fig. 2B shows the experimental timeline), expecting that the enhanced levels of Cd14 in the pulsed population could mediate more robust cytokine production. Confirming this hypothesis, pulsed cells transcribed significantly more of the LPS-responsive cytokine genes *Ccl2*, *Ccl3*, and *Il1b* than control cells in response to LPS treatment (Fig. 2C).

Transient transcription factor binding primes chromatin for enhancer activation. While LTTM can be induced with a transcription factor pulse in the TD model, the downstream events resulting in the acquisition and retention of memory are unknown and are of primary interest for the epigenetics community. The prevailing, albeit simplified, model of gene activation initiated by TF binding to enhancer sites consists of the following steps: (i) TF recruitment and possible nucleosome remodeling, (ii) histone acetyltransferase and mediator recruitment, (iii) BRD2/4 recruitment, and (iv) Pol II C-terminal domain (CTD) phosphorylation resulting in transcriptional elongation of target genes. Pol II elongation can be coupled to histone demethylation at H3K27 (19, 20) and methylation at H3K4 (enhancers and promoters) and H3K36 (gene body) (20–22) (Fig. 3A). Based on this scheme, we performed a stepwise analysis to determine which events are critical for memory formation.

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

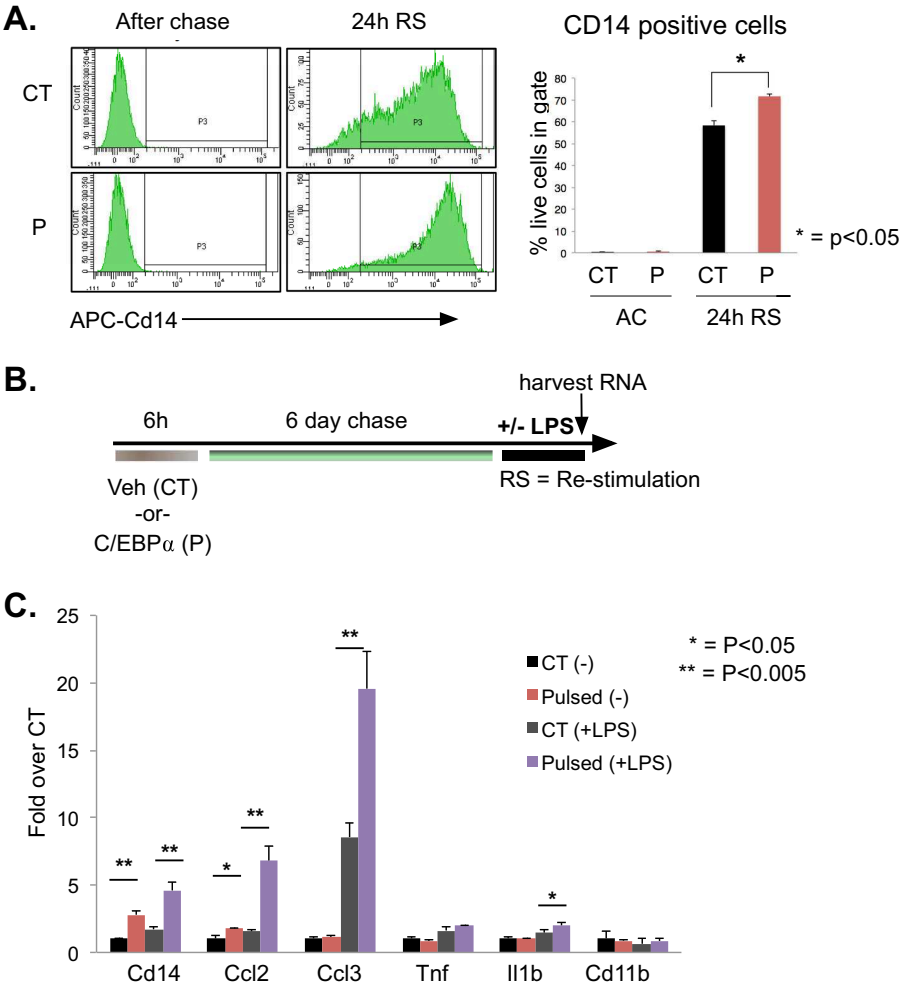


FIG 2 C/EBPα-pulsed cells display a more robust LPS response. (A) Flow cytometry analysis was performed after the chase period (AC) and 24 h into restimulation (RS) on control (CT) and pulsed (P) cells to measure Cd14 levels. The percentage of Cd14^{high} cells is graphed on the right. Allophycocyanin (APC)-conjugated Cd14 antibody was used. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance. (B) Experimental timeline for LPS stimulation test. After the chase, cells were simultaneously treated with β-estradiol and vehicle or LPS (1 μg/ml) to start the restimulation period. (C) Transcript levels at 6 h of restimulation. Pulsed and control cells were taken for transcript analysis after stimulation with vehicle or LPS for 6 h. Transcript levels of the cytokine genes *Cd14*, *Ccl2*, *Ccl3*, *Tnf*, and *Il1b* and the control gene *Cd11b* were measured. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance.

Past studies using experimental models with *Saccharomyces cerevisiae* have shown that short- and long-term transcriptional memory is observed at select galactose-responsive genes when yeast previously switched from glucose to galactose medium are reexposed to galactose. The mechanism attributed to LTTM in this system is debated, with some studies reporting histone variant H2AZ deposition and gene tethering to the nuclear pore as important events in memory formation (2, 23), while other studies demonstrate that cytoplasmic inheritance of transcription factors is the responsible mechanism (3, 4). We found that the global levels of predominant TFs C/EBPα and PU.1 returned to prestimulation levels in the chase (Fig. 1B); however, it is possible that some degree of TF induced by the pulse remained stably bound to chromatin throughout the chase period to mediate transcriptional memory. To test this, we performed chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) analysis for C/EBPα and PU.1 from control prestimulation, pulsed, and pulsed-chase samples. For both C/EBPα and PU.1, a small amount of retention was observed at target enhancer sites (see Materials and Methods for how enhancers were

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

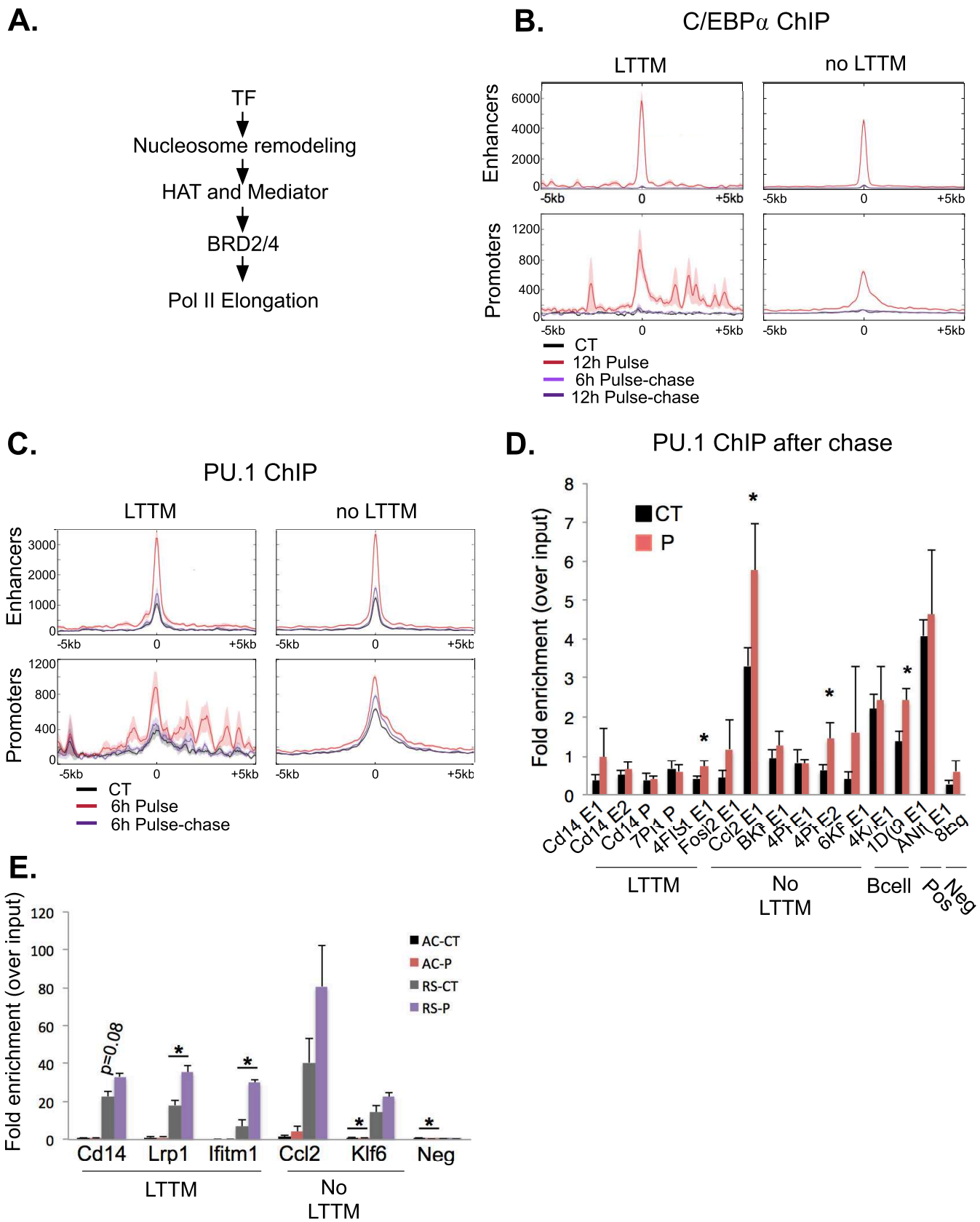


FIG 3 Transient transcription factor binding primes chromatin for establishment of active enhancer sites. (A) Stepwise model of chromatin events occurring downstream of TF recruitment used to dissect the molecular events responsible for memory acquisition. (B) ChIP-seq for C/EBPα at prestimulation levels (CT cells), after a 12-h pulse, and after the chase (pulse-chase) visualized as average peak trace files for LTTM and control gene enhancers and promoters. Three biological replicates were combined for a single sequencing run. (C) ChIP-seq for PU.1 at prestimulation levels (CT cells), after a 6-h pulse, and after the chase

(Continued on next page)

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

defined). However, this phenomenon was not unique to the memory genes (Fig. 3B to D). Interestingly, the gain of C/EBPα after the pulse at LTTM enhancers is greater than that at control gene enhancers, consistent with the observation that LTTM genes are induced to a higher degree (Fig. 1E). Previous work showed that during B cell-to-macrophage TD, the levels of C/EBPα on target enhancers become saturated 3 h after induction (14); however, we observed a time-dependent increase in memory between 6 and 12 h of the initial pulse. Together, these observations argue against TF retention as the key mechanism responsible for memory in this system.

Following our model (Fig. 3A), we next asked if histone acetylation induced by the pulse remained at elevated levels through the chase. We performed ChIP for histone H3 acetylated at K27 (H3K27ac) to gauge the level of enhancer activation after the chase period and during restimulation. We observed no difference in the levels of H3K27ac between pulsed and CT cells after the chase, suggesting that the H3K27ac deposited during the pulse (14) is lost. However, upon restimulation, we observed that the enhancers of the LTTM genes *Lrp1* and *Ifitm1* displayed higher levels of H3K27ac in pulsed cells (second stimulation) than in CT cells (initial stimulation), with a similar trend for the LTTM gene *Cd14* enhancer (Fig. 3E). Together, these data support the conclusion that memory genes are epigenetically primed for activation but not by augmented and retained levels of H3K27ac at enhancer sites. Therefore, in subsequent studies we focused our investigations on events occurring downstream of TF binding and H3K27 acetylation, including transcriptional elongation and histone methylation, as chromatin events potentially inherited to maintain LTTM.

Pol II elongation during the pulse is required for memory formation. Following histone acetylation, BRD4 is recruited to transition RNA polymerase II (Pol II) into the elongation phase by employing pTEFb to mediate CTD phosphorylation (24). Therefore, we speculated that LTTM establishment might require Pol II elongation during the pulse. To test this, during C/EBPα induction we treated cells with flavopiridol (Flavo), which is known to inhibit the release of poised Pol II into the elongating form capable of transcribing through the gene body. After the pulse, cells were washed and released into the memory protocol. Importantly, multiple groups have demonstrated that Flavo treatment during gene induction does not hinder recruitment of TFs, Pol II, and cofactors or histone acetylation (22, 25). Thus, while Flavo treatment impairs successful transcription of memory genes at the pulse (Fig. S2A), it does not prevent the engagement of chromatin at memory gene loci in response to TF induction. Strikingly, Flavo treatment at the pulse essentially blocked memory formation (Fig. 4A), suggesting that memory formation requires successful Pol II elongation.

Retention of low levels of elongating Pol II or retention of docked or poised polymerase at promoters may drive transcriptional memory (11, 26). Supporting this hypothesis, it has been demonstrated that the presence of poised Pol II at promoters enables rapid immediate early gene activation in neurons (27). It has been documented that within 1 h of Flavo treatment, total Pol II is lost from the gene body of actively transcribed genes (28). To first determine if low levels of elongating Pol II are retained and could thus bolster reactivation of LTTM genes, pulsed and control cells were treated for 4 h on day 4 of the chase to abrogate any elongating Pol II still present on memory genes. Posttreatment, cells were washed and released into fresh medium to recover and then on day 6 restimulated for 6 h to assess memory. While the treatment period was sufficient to temporarily impair Pol II transcription (Fig. S2B), Flavo treatment during the chase did not dampen the memory of tested LTTM genes (Fig. 4B). This

FIG 3 Legend (Continued)

(pulse-chase) visualized as average peak trace files for LTTM and control gene enhancers and promoters. Three biological replicates were combined and sequenced as one sample for analysis. (D) ChIP-qPCR for PU.1 after chase. ChIP-qPCR was performed at LTTM and no-LTTM gene enhancer sites in control (CT) cells and 6-h-pulsed (P) cells after the chase. Data are the averages from three biological replicates, and a two-sided *t* test was performed comparing the CT-chase to pulsed-chase values at each site. (E) H3K27ac ChIP-qPCR after the chase (AC) and after 6 h of restimulation (RS) in CT and 12-h-pulsed samples. Both memory and control gene enhancer sites were surveyed. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance.

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

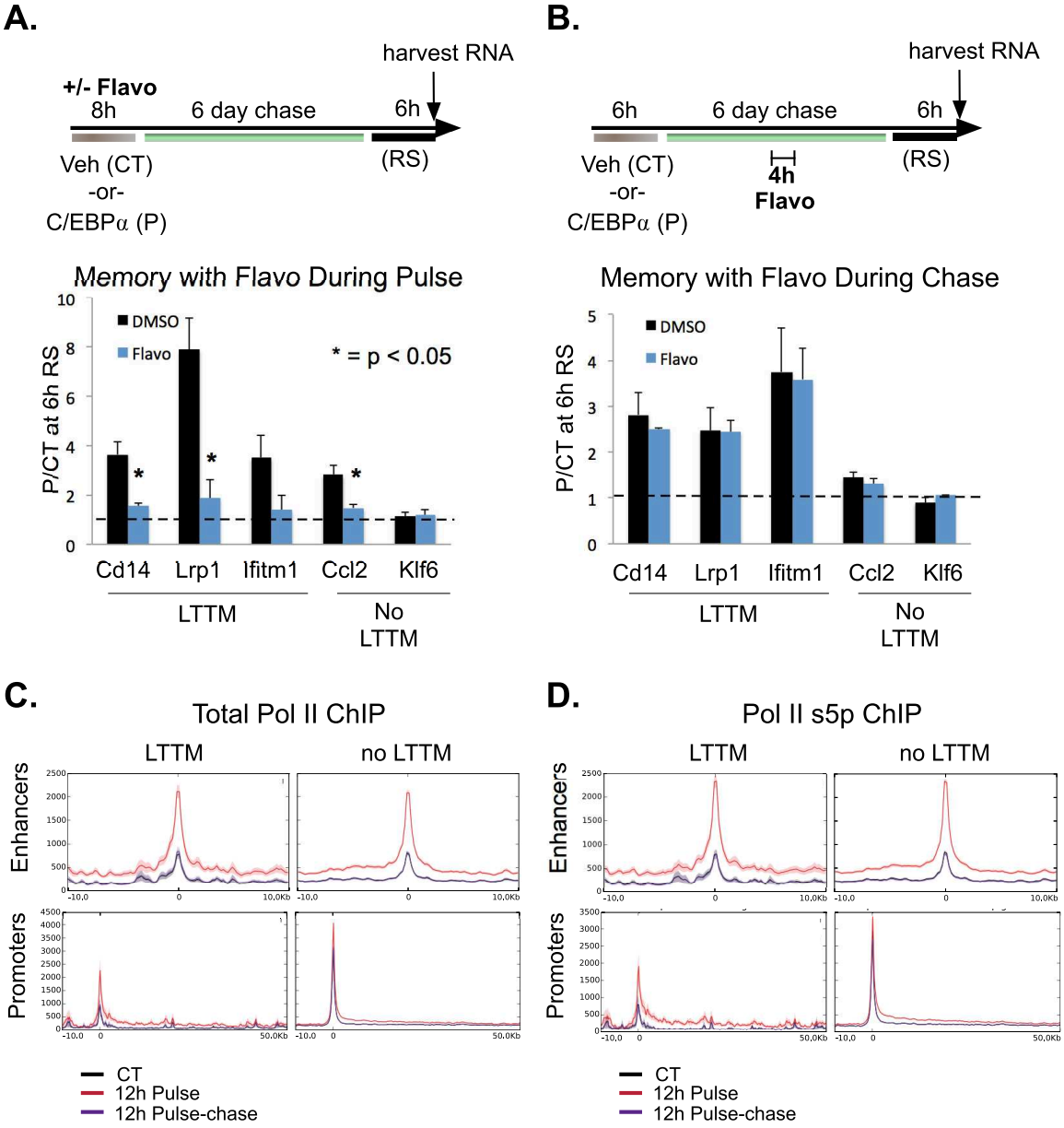


FIG 4 Pol II elongation but not retention is required for memory formation. (A) Flavopiridol (300 nM) was added to the cell culture simultaneously with C/EBPα induction and washed out after the pulse (P), and the cells were carried through the memory protocol. RNA was taken at 6 h of restimulation for qPCR analysis to determine the memory of selected LTTM genes. Data are represented as the fold induction of P/CT transcripts, where a value of 1 (dashed line) signifies no memory; data are the averages from three biological replicates. A two-sided *t* test was performed comparing the P/CT values for the dimethyl sulfoxide control to those for Flavo. (B) Flavo was added at 500 nM for 4 h during the chase, after which the cells were washed three times in medium and replated to continue the memory protocol. RT-qPCR analysis was performed in pulsed and control cells at 6 h into restimulation to measure transcript levels of the memory genes *Cd14*, *Lrp1*, and *Ifitm1* as well as of the control genes *Ccl2* and *Klf6*. Data are represented as fold induction of P/CT transcripts, where a value of 1 (dashed line) signifies no memory; data are the averages from three biological replicates. (C) ChIP-seq for total Pol II (unphosphorylated form using antibody 8WG16) at prestimulation levels (CT cells), after a 12-h pulse, and after the chase (12-h pulse-chase) visualized as average peak trace files for LTTM and control gene enhancers and promoters. Two biological replicates were combined and sequenced as one sample for analysis. (D) ChIP-seq for poised and elongating Pol II (phosphorylated at serine 5 [s5p] of the CTD using antibody 4H8) at prestimulation levels (CT cells), after a 12-h pulse, and after the chase (12-h pulse-chase) visualized as average peak trace files for LTTM and control gene enhancers and promoters. Two biological replicates were combined and sequenced as one sample for analysis.

led us to conclude that, for the genes tested, retained low levels of elongating Pol II are not responsible for LTTM. To determine if Pol II was retained in any form at LTTM gene enhancers or promoters, we performed ChIP assays for Pol II using antibodies that recognize (i) the unphosphorylated CTD to survey total Pol II (clone 8WG16) and (ii) the

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

CTD phosphorylated at serine 5 (clone 4H8) to survey Pol II in the poised and elongating forms (29). We found that at LTTM gene promoters and enhancers, Pol II levels increased with the pulse; however, after the chase the levels were indistinguishable from those of the control (prestimulation) (Fig. 4C and D). ChIP-qPCR analysis confirmed that after the chase period, the levels of poised Pol II at surveyed gene enhancers and promoters were comparable between pulsed and control cells (Fig. S2C), leading us to conclude that retained Pol II is not a contributing mechanism to LTTM in this model of transcriptional memory.

Memory gene regulatory regions are not marked by sustained H3K4 methylation. Impairing transcriptional elongation at gene bodies and enhancer RNAs (eRNAs) during cytokine stimulation in macrophages is known to also impair deposition of H3K4 methylation on *de novo* activation sites (22). As blocking transcriptional elongation abrogated memory in this study (Fig. 4A), we reasoned that deposition of H3K4 methylation at promoters and/or enhancer sites could be important for LTTM establishment and retention. Consistent with this hypothesis, high levels of histone H3 di- or trimethylated at K4 (H3K4me2/3) at gene promoters in yeast are known to persist after cessation of transcription (30), and short-term transcriptional memory in primary mouse macrophages is correlated with the deposition and retention of H3K4me1 at *de novo*-activated enhancers (1). Furthermore, both yeast and human studies of long-term transcriptional memory have demonstrated inheritance of elevated H3K4me2 deposited during the initial pulse (7, 26). To determine if H3K4 methylation acquired during the pulse persisted through the chase period, we performed ChIP-seq analysis to map H3K4me1, H3K4me2, and H3K4me3 at various points along the memory protocol. To identify chromatin events unique to memory genes, we performed analyses on the set of LTTM genes ($n = 35$) as well as on a set of control genes that are induced by C/EBPα but do not display LTTM ($n = 1{,}562$). At putative LTTM gene enhancers after the chase, H3K4me1 levels were slightly elevated in pulsed cells compared to levels in control cells. However, this same level of retention was seen at control gene enhancers as well (Fig. 5A, compare purple line to black line). ChIP-qPCR at model gene enhancer sites at the pulse and through the chase period revealed that while H3K4me1 was indeed deposited during the pulse, the levels fell gradually through the chase to prestimulation levels (Fig. S3A). No discernible difference was observed in the levels of H3K4me3 between control and pulsed cells at LTTM gene promoters (Fig. 5A, right, compare purple line to black line). Furthermore, ChIP-seq and ChIP-qPCR at selected LTTM gene regulatory regions confirmed that while H3K4me2 levels increase at the pulse, the increased levels return to prestimulation conditions after the chase (Fig. 5B and C and S3B). Hence, we conclude that LTTM in our model system does not involve significant retention of H3K4 methylation at promoters or enhancers.

LTTM genes display a more silent chromatin signature prior to stimulation. In our analysis of H3K4 methylation states on LTTM genes and control genes, we observed that LTTM genes generally displayed lower levels of H3K4 methylation at enhancers and especially at promoters (Fig. 5D). To ask if the lack of active histone marks indicated a higher level of repression, we analyzed the levels of the repressive mark H3K27me3 on LTTM versus control genes. This revealed that before the pulse, LTTM genes displayed higher levels of H3K27me3 across enhancers, promoters, and gene bodies (Fig. 5E). In addition, the prestimulation levels of total Pol II at LTTM gene promoters were lower than those of control no-LTTM genes (Fig. 4C and D). Taken together, these observations indicate that LTTM genes are in a more repressed chromatin state (low H3K4me, high H3K27me3, and low Pol II) than control genes that do not acquire memory with a pulse of C/EBPα activity.

Demethylation of H3K27me3 contributes to the mechanism of LTTM for *Ifitm1*. Our previous work with the TD model showed that upon C/EBPα induction, there is a gradual loss of H3K27me3 at C/EBPα binding sites, particularly at *de novo* enhancers that harbor high levels of H3K27me3 and low levels of H3K4me prior to induction (14). Furthermore, we along with others have shown that effective H3K27 demethylation is

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

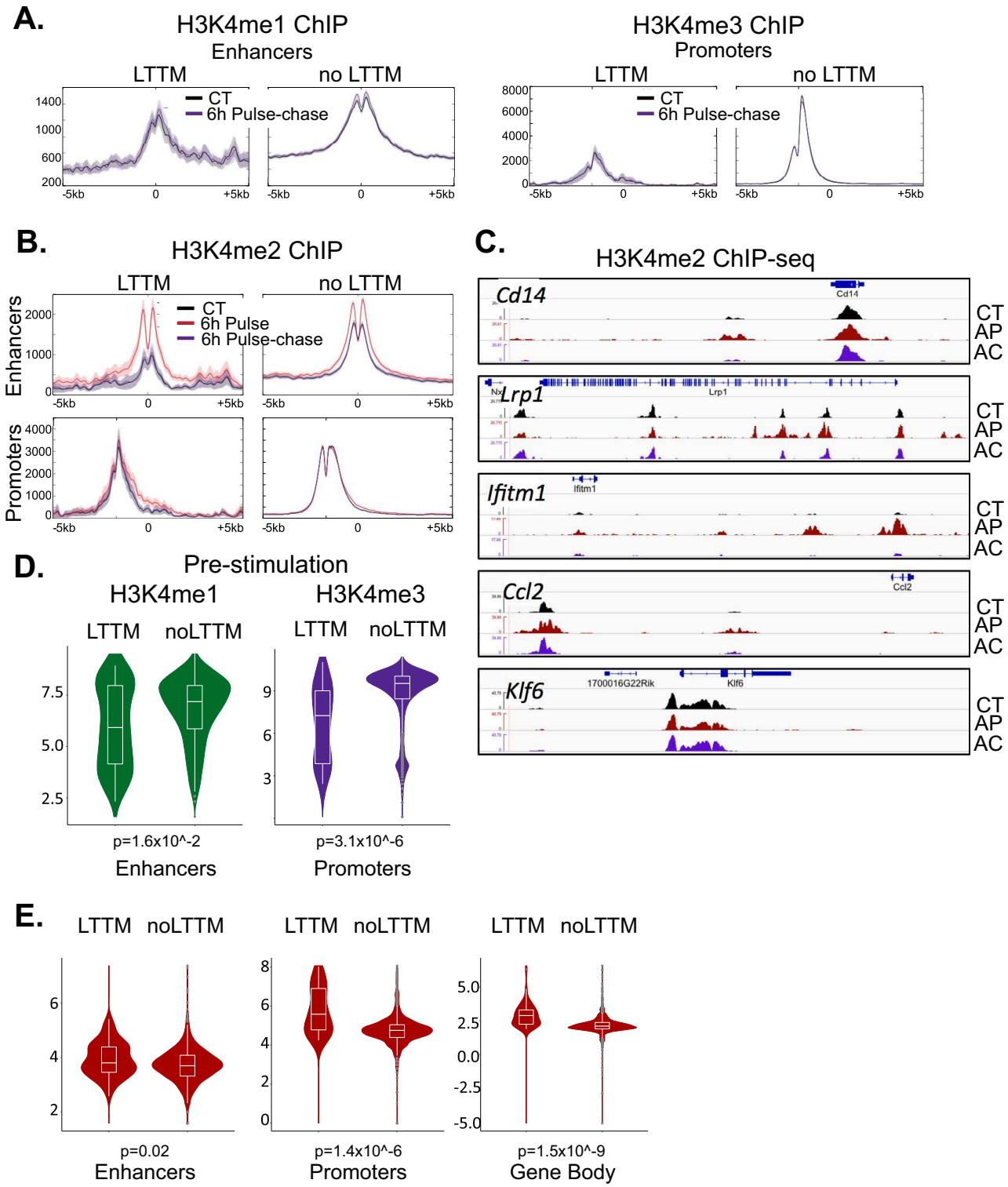


FIG 5 LTTM genes exist in a more repressed chromatin state prior to stimulation. (A) Average profiles for H3K4me1 levels at probable enhancers (centered on p300 binding sites) of LTTM and control genes are shown on the left. Two biological replicates were combined for a single sequencing run. Average profiles for H3K4me3 levels centered on LTTM and control gene promoters are shown on the right. Two biological replicates were combined for a single sequencing run. (B) Average profiles for H3K4me2 levels centered on LTTM and control gene enhancers or promoters. In this experiment, the signal in prestimulation control cells (CT), 6-h-pulsed cells, and pulsed cells after the chase (pulse-chase) are depicted. Two biological replicates were combined for a single sequencing run. (C) H3K4me2 ChIP-seq. Genome browser snapshots of H3K4me2 ChIP-seq in prestimulation control (CT), 6-h-pulsed (AP), and after-chase (AC) samples for representative memory (*Cd14*, *Lrp1*, and *Ifitm1*) and control (*Ccl2* and *Klf6*) genes. (D) Violin plots display the difference in H3K4 methylation levels before the pulse at LTTM and control gene enhancers and promoters. LTTM, $n = 35$; CT, $n = 1{,}562$. A two-sided Wilcoxon test was used to estimate P values. (E) H3K27me3 prestimulation. Violin plots display the difference in H3K27me3 levels before the pulse at LTTM and control genes at enhancers and promoters and in gene bodies. LTTM, $n = 35$; CT, $n = 1{,}562$. A two-sided Wilcoxon test was used to estimate P values.

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

linked to active Pol II elongation (20, 31, 32), an event critical for the establishment of LTTM (Fig. 4A). Therefore, we asked if LTTM gene enhancers and/or promoters were depleted of H3K27me3 in the pulse and if this depletion was sustained through the chase and could enable a quicker activation upon restimulation. When we looked across LTTM and control genes for retained loss of H3K27me3, we did not observe loss at gene promoters but did observe moderate loss of H3K27me3 at both control and LTTM gene enhancers (Fig. S4A). Looking specifically at the top 10 LTTM genes, here again we see that the enhancer regions show greater sustained loss of H3K27me3 than promoter regions (Fig. S4B). Although the difference is subtle, we reasoned that the removal of H3K27me3 at LTTM enhancers could be biologically significant as LTTM genes are decorated with higher baseline levels of H3K27me3 than those of the control gene group (Fig. 5E). Interestingly, when we studied the density of H3K27me3 signal at LTTM enhancers in control and pulsed cells after the chase, most LTTM genes displayed sustained loss of H3K27me3 at enhancer sites, with a particular enrichment for *Ifitm1* enhancer sites (Fig. 6A, asterisks mark enhancers for the genes *Cd14*, *Lrp1*, and *Ifitm1*). Sustained loss of H3K27me3 was confirmed by ChIP-qPCR at *Cd14* and *Ifitm1* enhancers, where H3K27me3 levels remained lower throughout the chase in pulsed cells (Fig. 6B and S4C).

The level of H3K27me3 at enhancers is generally low, making it difficult to draw conclusions based exclusively on ChIP studies. Therefore, to substantiate the importance of H3K27 demethylation on memory establishment and retention, we used small-molecule inhibitors that impair the enzymatic activity of the H3K27me3 demethylases and methylases, namely, Jmjd3/Utx (inhibited by glycogen synthase kinase [GSK]-J1/J4) (33) and Ezh1/2 (inhibited by UNC1999) (34). If H3K27 demethylation and retention contribute to memory, then blocking demethylation during the pulse should reduce the memory effect. Indeed, treatment of cells with GSK-J1/J4 during the pulse (added to cells at the same time as C/EBPα induction) significantly reduced memory observed for the LTTM genes *Cd14*, *Lrp1*, and *Ifitm1* but did not alter expression of *Ccl2* or *Klf6* (Fig. S4D). However, when we looked at the transcript levels during the pulse under GSK-J1/J4 treatment, the induction of LTTM genes was reduced by compound treatment (Fig. S4E). Therefore, it is not possible to conclude that the diminution of memory by compound is a result of changing H3K27me3 levels as reduced transcription could also be at play. We then asked if globally reducing the levels of H3K27me3 with the Ezh2 inhibitor UNC1999 (34) in the absence of a C/EBPα pulse could impart the same transcriptional memory phenomenon. Following the experimental design depicted in Fig. 6C, cells were treated for 48 h to allow for ~3 rounds of cell division under Ezh2 inhibition; the compound was washed out, and the cells were allowed to recover for a 6-day chase period. ChIP-seq analysis confirmed that H3K27me3 levels were reduced with compound treatment at C/EBPα target gene enhancers and, to a lesser extent, promoters (Fig. 6D, AT). After the 6-day chase, we observed a modest sustained loss of H3K27me3 (Fig. 6D, D6). Treatment with UNC1999 did indeed result in significant derepression of *Cd14* and *Ccl2* (Fig. 6E and S4F), but we observed a gradual return to pretreatment levels for all genes tested by day 6 of the chase (Fig. 6E and S4F). After the chase, cells were induced with C/EBPα to determine if prior Ezh2 inhibition resulted in a transcriptional memory phenotype. Upon C/EBPα induction, *Cd14* and *Ifitm1* retained significant memory of UNC1999 treatment (Fig. 6E). ChIP-qPCR studies at candidate gene regulatory regions confirmed that the *Ifitm1* enhancer site lost H3K27me3 after UNC1999 treatment and that this loss was retained over the 6-day chase period (Fig. 6F). These results indicate that prior Ezh1/2 inhibition can substitute for a C/EBPα pulse to impart memory at *Ifitm1* and that the resulting chromatin state is maintained for at least 6 days to mediate greater induction upon C/EBPα stimulation. Together, these data support the conclusion that for some genes under investigation in this study, H3K27 demethylation is critical for LTTM, either through maintained loss of H3K27me3 and/or through an unknown event linked to demethylation.

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

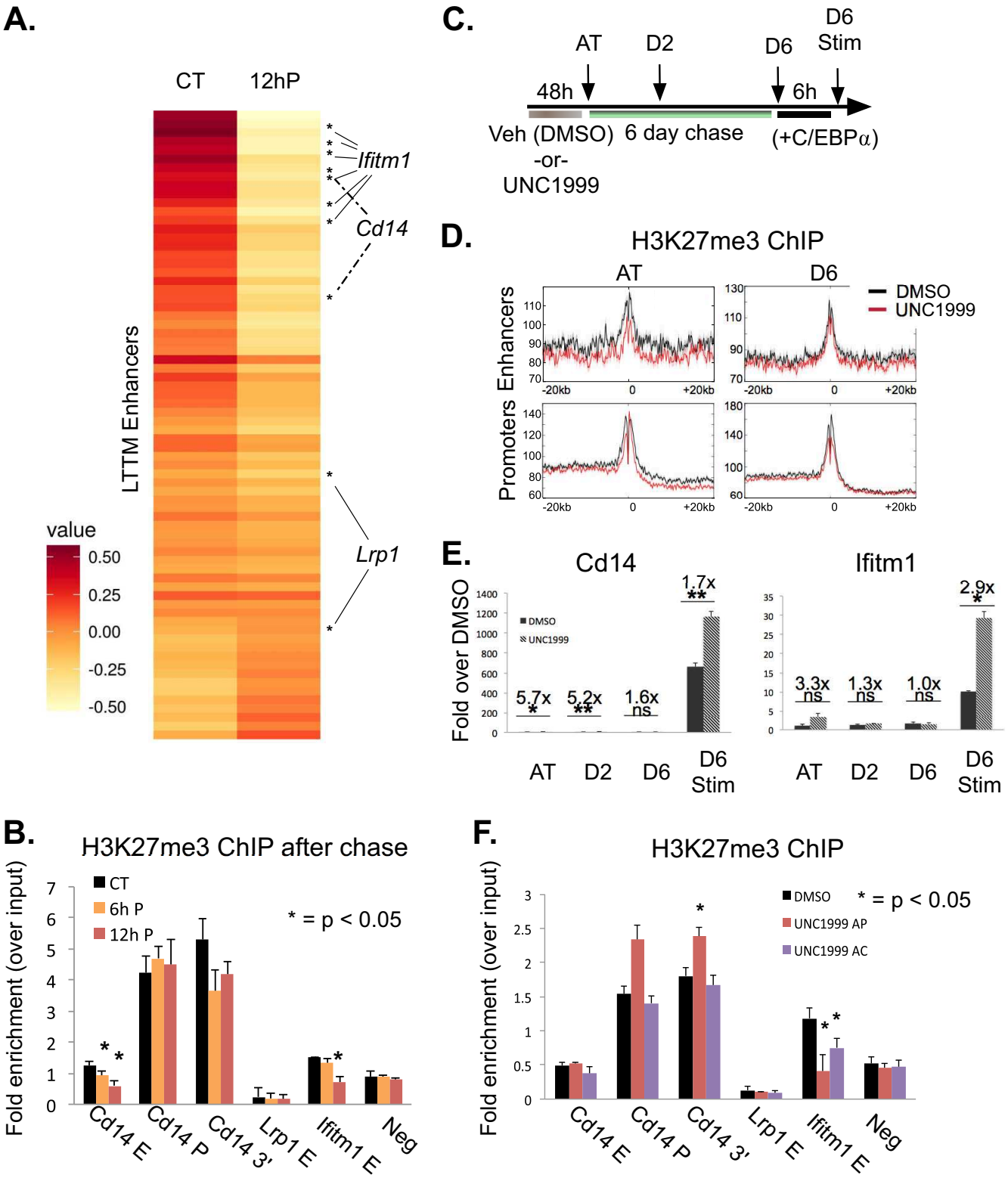


FIG 6 Loss of H3K27me3 is a heritable event that contributes to LTTM of *Ifitm1*. (A) Heat map showing the density of the H3K27me3 signal at LTTM gene enhancers in control and 12-h-pulsed cells after the chase derived from H3K27me3 ChIP-seq analysis. Three biological replicates were combined for a single sequencing run and used to calculate the intensity of signal over each enhancer region. Enhancer sites are ranked by the most to least amount of H3K27me3 lost with a 12-h pulse. Asterisks indicate enhancers associated with the memory genes chosen for study here (*Cd14*, *Lrp1*, and *Ifitm1*). (B) ChIP-qPCR of H3K27me3 at LTTM genes and associated regulatory regions from control cells (CT) and 6-h- and 12-h-pulsed cells after the chase. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance. (C) UNC1999 experiment depicted in the schematic. The Ezh1/2 inhibitor UNC1999 (or dimethyl sulfoxide [DMSO]) was added to cells at 3 μM for 48 h, and then the compound was washed out, and the cells were left to recover for 6 days (D6). Samples were collected after the treatment (AT), at day 2 of the chase period (D2), following the chase period (D6), and 6 h after C/EBPα

(Continued on next page)

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

DISCUSSION

Epigenetic changes, by definition, survive cell division and are therefore able to transmit to progeny information gained in parental cells. This mechanism can provide progeny with an adapted response to environmental cues. Here, we show that a short pulse of transcription factor C/EBPα activity can induce transcriptional memory on a subset of genes, and this memory survives multiple rounds of cell division. Many of the identified memory genes play important roles in the antimicrobial response of the innate immune system (e.g., *Cd14*, *Ifitm1*, *Abca1*, *Lrp1*, *Cd33*, and *Serpinb2*). We show that transcriptional memory of *Cd14* expression translates to the protein level, where pulsed cells display higher levels of Cd14 on the cell membrane. As Cd14 is a critical mediator of Toll-like receptor signaling in response to bacterial infection, pulsed cells in our model also transcribe larger amounts of *Ccl2*, *Ccl3*, and *Il1b* in response to LPS stimulation. These data suggest that memory of *Cd14* expression could enable a more robust response to bacterial infection.

The transcriptional memory we observed in this model increases in a manner dependent on the time of the initial pulse. This simple observation lends some critical insight into the mechanisms at play as it suggests that the epigenetic features that maintain memory are cumulative. For example, we know that peak binding of C/EBPα is achieved within the first 3 h of induction. If simply the binding of target sites by C/EBPα was sufficient to establish memory, then we would not observe an increase in memory with a 12-h pulse compared to the level with a 6-h pulse. Supporting these results, when we looked at C/EBPα binding by ChIP, we did not observe retention through the chase period.

Memory formation imparted by C/EBPα is blocked in the presence of the Pol II elongation inhibitor flavopiridol. Therefore, we focused our mechanistic studies on chromatin events dependent on Pol II elongation, including histone methylation and demethylation events. From previous work, we had insight into which events occur in the first 12 h of TD that correlate well with the kinetics of our model, and these include H3K4 methylation (increasing steadily in the first 12 h) and H3K27 demethylation (decreasing steadily in the first 12 h of TD). Therefore, we analyzed in depth the addition or removal of these marks and their maintenance through the memory protocol.

Memory formation has been associated with long-lived retention of H3K4me1/2 at enhancers or of H3K4me3 at gene promoters in previous models of transcriptional memory (14, 30). However, in our system, retained H3K4me1/2/3 histone modifications were not observed but were steadily lost throughout the chase. In regard to H3K4 methylation, however, we did observe a significant difference in the baseline levels of this mark between LTTM and control gene regions where LTTM gene enhancers and promoters had lower levels of H3K4 methylation before the pulse. This led us to interrogate the levels of repressive H3K27me3 to understand if LTTM gene regions are in a more repressed state in naive cells. Indeed, we observed higher levels of H3K27me3 at LTTM genes. Interestingly, prior to the first stimulus, we also observed less total Pol II at LTTM gene promoters than at control gene promoters. Together these findings indicate that the characteristic of LTTM genes compared to control genes (those induced by C/EBPα but show no memory) is germane to that of *de novo* versus primed enhancers. This observation supports the hypothesis that chromatin-priming events result in transcriptional memory, where LTTM genes are epigenetically primed during the pulse for greater induction upon restimulation.

FIG 6 Legend (Continued)

was induced (D6 stim). (D) Profile plots show H3K27me3 ChIP-seq signal after treatment (AT) and after the chase (day 6, D6) at all genes induced by C/EBPα, in both CT (dimethyl sulfoxide)- and UNC1999-treated cells. Three biological replicates were combined for a single sequencing run. (E) RT-qPCR was performed to determine transcript levels for LTTM genes *Cd14* and *Ifitm1* at time points throughout the UNC1999 experiment. The fold changes in transcript levels between UNC1999- and vehicle (DMSO)-treated samples are depicted over the graphs at each time point. Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance. (F) ChIP-qPCR of H3K27me3 at LTTM genes and associated regulatory regions from control cells (DMSO), UNC1999-treated cells after the 48-h treatment period (AP), and UNC1999-treated cells after the 6-day chase (AC). Data are the averages from three biological replicates; a two-sided *t* test was performed to determine statistical significance.

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

In our model, we postulate that the priming event conferring memory occurs in conjunction with Pol II elongation and at least in part involves H3K27 demethylation at certain gene regulatory regions (*Ifitm1*). We draw this conclusion from the results of different experimental approaches. First, small but detectible levels of H3K27me3 are lost with the pulse at enhancers of the most significant LTTM genes (including genes *Cd14* and *Ifitm1*), and this loss is retained through the chase. Second, treating cells with an Ezh2 inhibitor for 2 days in the absence of a C/EBPα pulse can induce a phenotype similar to transcriptional memory. Therefore, we conclude that LTTM seen in this model system is maintained at least in part by a passive mechanism, where loss of H3K27 methylation occurring in conjunction with gene activation is not reestablished. This could allow faster reactivation by removing one step in the transcriptional activation process and, for example, allow faster acetylation at H3K27 during active enhancer establishment. Indeed, we show that memory gene enhancer sites gain more H3K27ac upon restimulation. We propose that, in our model, removal of repressive histone marks is more critical for long-term memory than the addition of active histone marks. In line with our observation, a recent paper studying long-term memory in primary macrophages concluded that removal of repressive H3K9 methylation enabled transcriptional memory induced by LPS exposure (35). In future studies, it will be interesting to survey the nucleosome landscape to query whether nucleosome depletion occurring during Pol II elongation is another event that is inherited, as suggested in recent studies of transcriptional memory in mammalian cells (6, 7).

Polycomb-mediated repression is known to play a pivotal role in defining and securing cell fate by silencing lineage-specific genes in progenitor cells and then subsequently silencing developmental genes during terminal differentiation (36). Questions stemming from the data presented here concern which stimuli and what cell types *in vivo* have the ability to acquire transcriptional memory that can be transmitted to daughter cells. It is tempting to speculate that a biological event within innate immune cells with the ability to establish LTTM *in vivo* could be signaling in myeloid progenitor cells that induce transcription coupled to H3K27 demethylation at macrophage-specific genes prior to terminal macrophage differentiation. Interestingly, in response to type I interferon exposure, hematopoietic stem cells (HSCs) exit quiescence and produce progenitor cells to populate different blood lineages (37), and common myeloid progenitors are activated for emergency myelopoiesis to generate large numbers of inflammatory monocytes and neutrophils (38). In this scenario, myeloid progenitor cells could acquire memory of pathogen exposure through interferon-induced gene activation, pass transcriptional memory to daughter cells for many generations, and thereby display the phenotype of trained immunity.

The term epigenetic memory most commonly refers to chromatin changes in response to external cues that result in permanent cell fate commitment, usually in regard to cellular differentiation or reprogramming. For example, many models of differentiation, reprogramming, or transdifferentiation that involve the addition or removal of key lineage-specifying TFs harbor an intrinsic commitment point where the initiating event (TF induction) can be relieved, but the cells continue toward the newly specified fate (39). Often, a broad assumption is made that any chromatin changes occurring before the commitment point are lost as the overall cellular phenotype reverts to the original state. However, the data presented here support the idea that some chromatin changes occurring before the commitment point are indeed inherited and can be preserved as latent memory. Phenotypic consequences of this latent memory may not be revealed until cell descendants are presented with a similar stimulus or differentiation cue, upon which an adapted response is observed. Clinically, this finding could be relevant to understand the developmental origins of health and disease (DOHaD) (40), drug sensitization and addiction (41), and genomic priming for developmental competency (42). For example, fetal and neonatal exposures to abnormally high levels of hormone or endocrine disruptors could epigenetically prime target genes for altered responsiveness later in life, when similar hormonal signals increase during puberty. Long-lived epigenetic priming could also play a critical role in estab-

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

lishing immunological memory within innate immune cell progenitors that have the capacity to proliferate in response to pathogen challenge. Together, our data suggest that altering the repressive chromatin landscape has the capacity to impart long-lived priming effects and has implications in trained immunity and in DOHaD.

MATERIALS AND METHODS

Transdifferentiation and memory assay. C10 (HAFTL cells with a stably integrated C/EBPα transgene) cells were seeded at a concentration of 2e6 cells/ml in B cell medium (RPMI medium lacking phenol red, 10% charcoal-stripped fetal bovine serum [FBS], L-glutamine, 50 μM β-mercaptoethanol, penicillin-streptomycin). To initiate transdifferentiation (TD), a 2$\times$ TD mix was made by adding the following components to B cell medium: β-estradiol at 200 nM, interleukin-3 (IL-3; Peprotech) at 20 nM, and macrophage colony-stimulating factor (M-CSF; Peprotech) at 20 nM. The 2$\times$ mixture was added 1:1 to cell-containing medium to start the TD process. After the pulse period, cells were washed three times in warm B cell medium and replated in fresh B cell medium. Medium was refreshed every 2 days. TD was initiated as above at day 6 to start the restimulation period.

ChIP protocol. Cells at 10e6 cells/ml were fixed in 1% formaldehyde for 10 min at room temperature. See the experimental procedures in the supplemental material for the detailed protocol. Approximately 10 million cells were used to make chromatin for each immunoprecipitation. The antibodies and amounts used were as follows: 6 μl of PU.1 (sc-352x), 10 μl of C/EBPα p300 (sc-585x), 3 μl of H3K27me3 (07-449; Millipore), 5 μg of H3K4me1 (C15410037; Diagenode) 2 μl of H3K4me2 (04-790; Millipore), 5 μl of H3K4me3 (05-1339; Millipore), 2 μl of H3K27ac (39133; Active Motif), 5 μl of RNA polymerase II CTD repeat YSPTSPS antibody (8WG16) (ab817; Abcam), and 5 μl of RNA polymerase II phospho-S5 4H8 (39097; Active Motif).

For ChIP-seq library preparation, 10 ng of ChIP DNA was used for each library. An NEBNext DNA library prep reagent set for Illumina (E6000L) was used to generate the libraries, along with the standard Illumina barcoding primers according to the manufacturer's instructions.

RNA-seq. RNA was extracted using TRIzol (15596-026) according to the manufacturer's instructions. Five micrograms of RNA was used as starting material for rRNA depletion using an Ambion RiboMinus kit (A1083708). rRNA-depleted samples were analyzed on an Agilent Bioanalyzer to ensure that the majority of rRNA was depleted before the library was prepared. A total of 100 ng of rRNA-depleted RNA was then used in an NEBNext Ultra Directional RNA-seq kit (E7420L) to generate sequencing libraries according to the manufacturer's instructions.

Bioinformatic analysis. All parameters, software versions, and references can be found in Table S2 in the supplemental material. For RNA-seq, reads were mapped using STAR with the Ensembl gene annotation gtf (GRCm38.75). bigwig tracks were made using DeepTools BamCoverage. For quantitative analysis, reads were counted on gene exons using HTSeq-count (union mode). Sample scaling and statistical analysis were performed using the Rpackage DESeq2 (version 1.6.3) with default parameters. LTTM genes were selected for showing a significant difference between pulse-chase-restimulation and control-chase-restimulation values (P value of <0.01, negative binomial Wald test), no significant difference between pulse-chase and control-chase values (P value of >0.01), and strongly significant induction upon restimulation (pulse-chase-restimulation/pulse-chase of >1 and FDR of <0.01).

For ChIP-seq, reads were mapped using STAR. Duplicate reads were removed using picard, while bigwig tracks were made using DeepTools BamCoverage. Peak calling was performed using macs2.

For all quantitative analyses, reads were counted on merged regions for each time point using the R package csaw (function regionCounts). In order to estimate sequencing depth, we counted reads on 10-kb windows covering the whole genome (using the csaw package, function WindowCounts) and then used DEseq2 on these counts to compute the scaling factors. This ensures robustness of the scaling with respect to differences in peak numbers and to highly enriched outlier regions. Scaled counts were then transformed to a $\log_2$ scale (after adding a pseudocount of 1) and used in subsequent analyses.

For enhancer analysis, we merged peaks of p300 found at 0 h and 12 h after C/EBPα induction and selected those showing at least a 2-fold increase at 12 h. All p300 increasing peaks in $\pm$100 kb of gene transcription start sites (TSS) were selected as potential associated enhancers. We then measured the levels of the other ChIP-seq signals in these regions to make box plots. For promoter analysis, we used the region of $\pm$2 kb around the TSS and proceeded as for enhancer regions.

Average plots and heat maps were made using DeepTools (parameter, -binSize 10).

Accession number(s). Data are available under Gene Expression Omnibus (GEO) accession number GSE72488.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/MCB.00372-16.

TEXT S1, PDF file, 0.9 MB.

ACKNOWLEDGMENTS

We thank Yun Zhu and Wei Wang for undertaking some initial bioinformatic analysis for this project. We also thank Esteban Ballestar for reading the manuscript and providing scientific input. The chemical inhibitors used in this study were kindly

Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

provided by Cheryl Arrowsmith of the Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada.

A.I.-B. conceived of the project, conducted most of the experiments in the transdifferentiation system, and was the primary author of the manuscript. S.C. performed all bioinformatic analysis included in the paper and aided in writing the manuscript. B.L. performed key ChIP experiments. C.V.O. performed key ChIP experiments, provided ChIP-Seq data, and helped in revising the manuscript. K.-K.C. aided in RT-qPCR analysis. D.T. supervised the bioinformatic analysis. T.G. supervised the progress of this work and substantially aided in writing the manuscript. Y.S. supervised the project from inception.

REFERENCES

1. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent enhancers activated by stimulation in differentiated cells. Cell 152:157–171. https://doi.org/10.1016/j.cell.2012.12.018.
2. Brickner DG, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee P-CC, Widom J, Brickner JH. 2007. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. PLoS Biol 5:e81. https://doi.org/10.1371/journal.pbio.0050081.
3. Kundu S, Peterson CL. 2010. Dominant role for signal transduction in the transcriptional memory of yeast GAL genes. Mol Cell Biol 30:2330–2340. https://doi.org/10.1128/MCB.01675-09.
4. Zacharioudakis I, Gligoris T, Tzamarias D. 2007. A yeast catabolic enzyme controls transcriptional memory. Curr Biol 17:2041–2046. https://doi.org/10.1016/j.cub.2007.10.044.
5. Dunn J, McCuaig R, Tu WJ, Hardy K, Rao S. 2015. Multi-layered epigenetic mechanisms contribute to transcriptional memory in T lymphocytes. BMC Immunol 16:27. https://doi.org/10.1186/s12865-015-0089-9.
6. Bevington SL, Cauchy P, Piper J, Bertrand E, Lalli N, Jarvis RC, Gilding LN, Ott S, Bonifer C, Cockerill PN. 2016. Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J 35:515–535. https://doi.org/10.15252/embj.201592534.
7. Gialitakis M, Arampatzi P, Makatounakis T, Papamatheakis J. 2010. Gamma interferon-dependent transcriptional memory via relocalization of a gene locus to PML nuclear bodies. Mol Cell Biol 30:2046–2056. https://doi.org/10.1128/MCB.00906-09.
8. Wong MM, Byun JS, Sacta M, Jin Q, Baek S, Gardner K. 2014. Promoter-bound p300 complexes facilitate post-mitotic transmission of transcriptional memory. PLoS One 9:e99989. https://doi.org/10.1371/journal.pone.0099989.
9. Quintin J, Saeed S, Martens JH, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C, Jacobs L, Jansen T, Kullberg B-JJ, Wijmenga C, Joosten LA, Xavier RJ, van der Meer JW, Stunnenberg HG, Netea MG. 2012. *Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. Cell Host Microbe 12:223–232. https://doi.org/10.1016/j.chom.2012.06.006.
10. Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajanirefah A, Matarese F, Cheng S-CC, Ratter J, Berentsen K, van der Ent MA, Sharifi N, Janssen-Megens EM, Ter Huurne M, Mandoli A, van Schaik T, Ng A, Burden F, Downes K, Frontini M, Kumar V, Giamarellos-Bourboulis EJ, Ouwehand WH, van der Meer JW, Joosten LA, Wijmenga C, Martens JH, Xavier RJ, Logie C, Netea MG, Stunnenberg HG. 2014. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science 345:1251086. https://doi.org/10.1126/science.1251086.
11. Light WH, Freaney J, Sood V, Thompson A, D'Urso A, Horvath CM, Brickner JH. 2013. A conserved role for human Nup98 in altering chromatin structure and promoting epigenetic transcriptional memory. PLoS Biol 11:e1001524. https://doi.org/10.1371/journal.pbio.1001524.
12. Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25:2227–2241. https://doi.org/10.1101/gad.176826.111.
13. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576–589. https://doi.org/10.1016/j.molcel.2010.05.004.
14. van Oevelen C, Collombet S, Vicent G, Hoogenkamp M, Lepoivre C, Badeaux A, Bussmann L, Sardina JL, Thieffry D, Beato M, Shi Y, Bonifer C, Graf T. 2015. C/EBPα activates preexisting and de novo macrophage enhancers during induced pre-B cell. Stem Cell Rep 5:232–247. https://doi.org/10.1016/j.stemcr.2015.06.007.
15. Bussmann LH, Schubert A, Vu Manh TP, De Andres L, Desbordes SC, Parra M, Zimmermann T, Rapino F, Rodriguez-Ubreva J, Ballestar E, Graf T. 2009. A robust and highly efficient immune cell reprogramming system. Cell Stem Cell 5:554–566. https://doi.org/10.1016/j.stem.2009.10.004.
16. Xie H, Ye M, Feng R, Graf T. 2004. Stepwise reprogramming of B cells into macrophages. Cell 117:663–676. https://doi.org/10.1016/S0092-8674(04)00419-2.
17. Di Tullio A, Vu Manh TP, Schubert A, Castellano G, Månsson R, Graf T. 2011. CCAAT/enhancer binding protein alpha (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc Natl Acad Sci U S A 108:17016–17021. https://doi.org/10.1073/pnas.1112169108.
18. Jiang Z, Georgel P, Du X, Shamel L, Sovath S, Mudd S, Huber M, Kalis C, Keck S, Galanos C, Freudenberg M, Beutler B. 2005. CD14 is required for MyD88-independent LPS signaling. Nat Immunol 6:565–570. https://doi.org/10.1038/ni1207.
19. Wang AH, Zare H, Mousavi K, Wang C, Moravec CE, Sirotkin HI, Ge K, Gutierrez-Cruz G, Sartorelli V. 2013. The histone chaperone Spt6 coordinates histone H3K27 demethylation and myogenesis. EMBO J 32:1075–1086. https://doi.org/10.1038/emboj.2013.54.
20. Chen S, Ma J, Wu F, Xiong L-JJ, Ma H, Xu W, Lv R, Li X, Villen J, Gygi SP, Liu XS, Shi Y. 2012. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. Genes Dev 26:1364–1375. https://doi.org/10.1101/gad.186056.111.
21. Kanno T, Kanno Y, LeRoy G, Campos E, Sun H-WW, Brooks SR, Vahedi G, Heightman TD, Garcia BA, Reinberg D, Siebenlist U, O'Shea JJ, Ozato K. 2014. BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. Nat Struct Mol Biol 21:1047–1057. https://doi.org/10.1038/nsmb.2912.
22. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell 51:310–325. https://doi.org/10.1016/j.molcel.2013.07.010.
23. Light WH, Brickner JH. 2012. Nuclear pore proteins regulate chromatin structure and transcriptional memory by a conserved mechanism. Nucleus 4:357–360. https://doi.org/10.4161/nucl.26209.
24. Jang MK, Mochizuki K, Zhou M, Jeong H-SS, Brady JN, Ozato K. 2005. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell 19:523–534. https://doi.org/10.1016/j.molcel.2005.06.027.
25. Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. Genome Res 23:1210–1223. https://doi.org/10.1101/gr.152306.112.
26. D'Urso A, Takahashi YH, Xiong B, Marone J, Coukos R, Randise-Hinchliff C, Wang JP, Shilatifard A, Brickner JH. 2016. Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. eLife 5:e16691. https://doi.org/10.7554/eLife.16691.
27. Saha RN, Wissink EM, Bailey ER, Zhao M, Fargo DC, Hwang J-YY, Daigle KR, Fenn JD, Adelman K, Dudek SM. 2011. Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. Nat Neurosci 14:848–856. https://doi.org/10.1038/nn.2839.
28. Jonkers I, Kwak H, Lis JT. 2013. Genome-wide dynamics of Pol II elon-

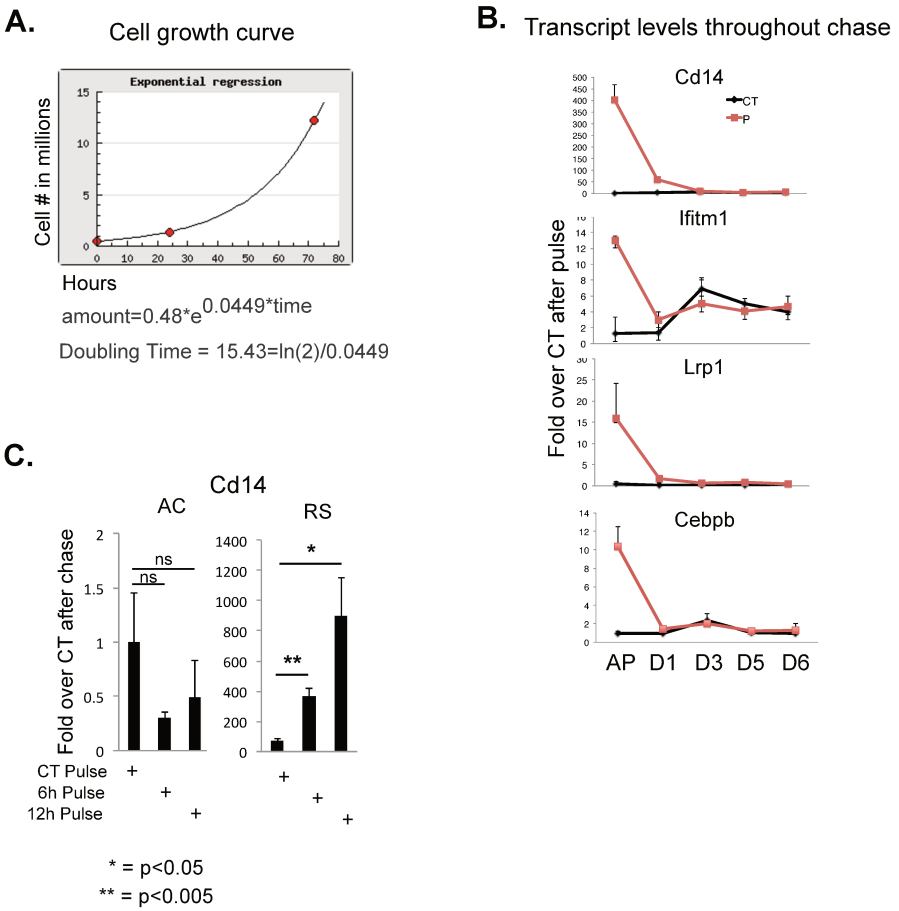
Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

gation and its interplay with promoter proximal pausing, chromatin, and exons. eLife 3:02407. https://doi.org/10.7554/eLife.02407.
29. Brookes E, Pombo A. 2009. Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO Rep 10:1213–1219. https://doi.org/10.1038/embor.2009.221.
30. Ng HH, Robert F, Young RA, Struhl K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell 11:709–719. https://doi.org/10.1016/S1097-2765(03)00092-3.
31. Estarás C, Fueyo R, Akizu N, Beltrán S, Martínez-Balbás MA. 2013. RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. Mol Biol Cell 24:351–360. https://doi.org/10.1091/mbc.E12-07-0561.
32. Smith ER, Lee MG, Winter B, Droz NM, Eissenberg JC, Shiekhattar R, Shilatifard A. 2008. Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Mol Cell Biol 28:1041–1046. https://doi.org/10.1128/MCB.01504-07.
33. Kruidenier L, Chung CW, Cheng Z, Liddle J, Che K, Joberty G, Bantscheff M, Bountra C, Bridges A, Diallo H, Eberhard D, Hutchinson S, Jones E, Katso R, Leveridge M, Mander PK, Mosley J, Ramirez-Molina C, Rowland P, Schofield CJ, Sheppard RJ, Smith JE, Swales C, Tanner R, Thomas P, Tumber A, Drewes G, Oppermann U, Patel DJ, Lee K, Wilson DM. 2012. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature 488:404–408. https://doi.org/10.1038/nature11262.
34. Xu B, On DM, Ma A, Parton T, Konze KD, Pattenden SG, Allison DF, Cai L, Rockowitz S, Liu S, Liu Y, Li F, Vedadi M, Frye SV, Garcia BA, Zheng D, Jin J, Wang GG. 2015. Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. Blood 125:346–357. https://doi.org/10.1182/blood-2014-06-581082.
35. Yoshida K, Maekawa T, Zhu Y, Renard-Guillet C, Chatton B, Inoue K, Uchiyama T, Ishibashi K-i, Yamada T, Ohno N, Shirahige K, Okada-Hatakeyama M, Ishii S. 2015. The transcription factor ATF7 mediates lipopolysaccharide-induced epigenetic changes in macrophages involved in innate immunological memory. Nat Immunol 16:1034–1043. https://doi.org/10.1038/ni.3257.
36. Laugesen A, Helin K. 2014. Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell 14:735–751. https://doi.org/10.1016/j.stem.2014.05.006.
37. Sato T, Onai N, Yoshihara H, Arai F, Suda T, Ohteki T. 2009. Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. Nat Med 15:696–700. https://doi.org/10.1038/nm.1973.
38. Buechler MB, Teal TH, Elkon KB, Hamerman JA. 2013. Cutting edge: type I IFN drives emergency myelopoiesis and peripheral myeloid expansion during chronic TLR7 signaling. J Immunol 190:886–891. https://doi.org/10.4049/jimmunol.1202739.
39. Graf T, Enver T. 2009. Forcing cells to change lineages. Nature 462:587–594. https://doi.org/10.1038/nature08533.
40. Hochberg Z, Feil R, Constancia M, Fraga M, Junien C, Carel JC, Boileau P, Le Bouc Y, Deal CL, Lillycrop K, Scharfmann R, Sheppard A, Skinner M, Szyf M, Waterland RA, Waxman DJ, Whitelaw E, Ong K, Albertsson-Wikland K. 2011. Child health, developmental plasticity, and epigenetic programming. Endocr Rev 32:159–224. https://doi.org/10.1210/er.2009-0039.
41. Nestler EJ. 2016. Transgenerational epigenetic contributions to stress responses: fact or fiction? PLoS Biol 14:e1002426. https://doi.org/10.1371/journal.pbio.1002426.
42. Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, Muth K, Palmer J, Qiu Y, Wang J, Lam DK, Raum JC, Stoffers DA, Ren B, Sander M. 2015. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell 16:386–399. https://doi.org/10.1016/j.stem.2015.02.013.

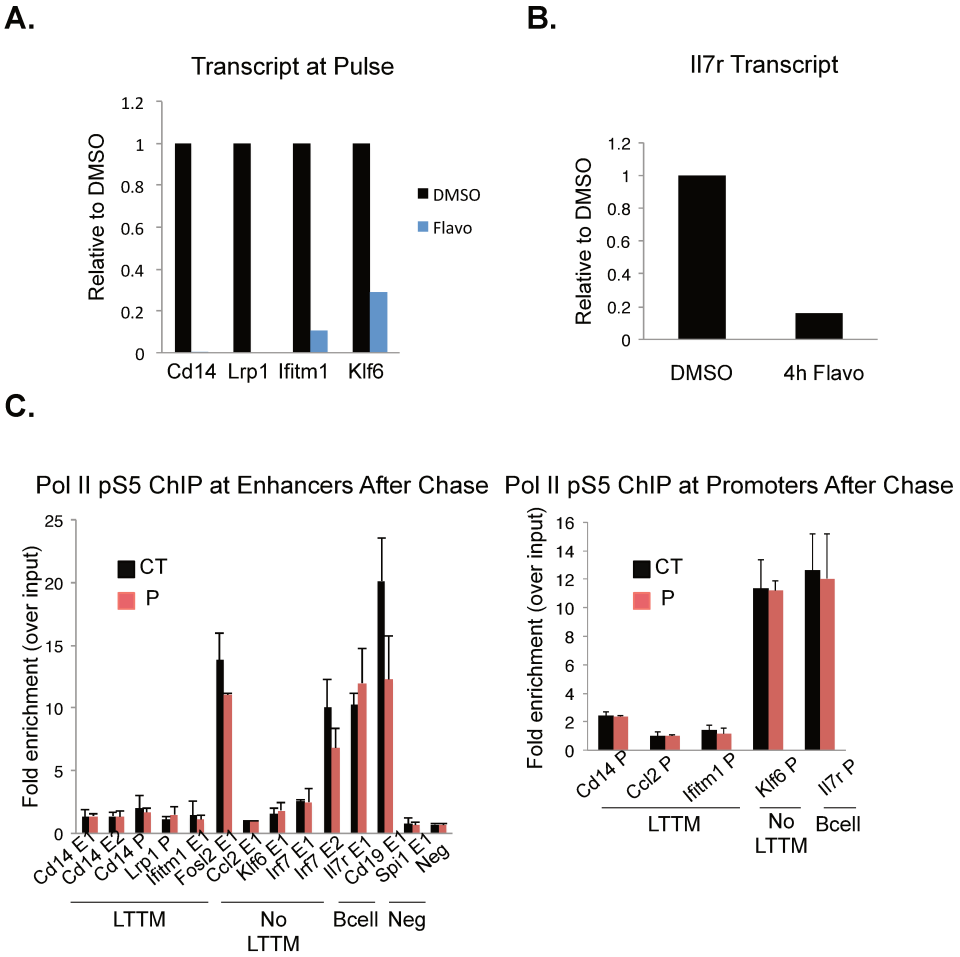
Downloaded from http://mcb.asm.org/ on August 22, 2017 by ENS BIOLOGIE

Supplemental Data

Figure Supplement Legends

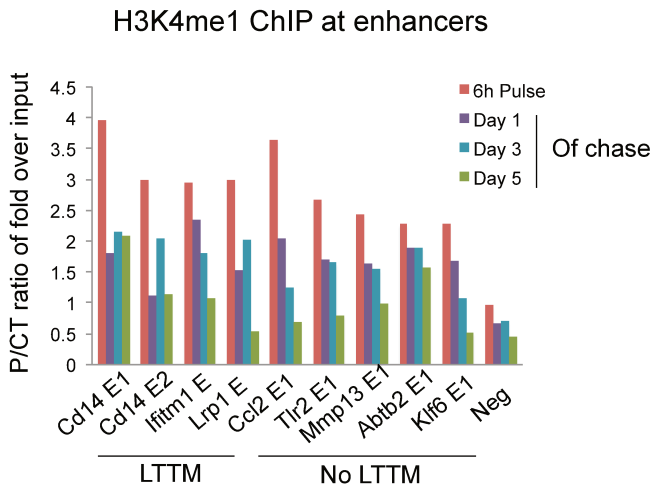


Supplemental Figure S1. (A) Counts of C10 cells were recorded for 3 days and used to calculate a growth curve to determine the number of cell divisions occurring over a period of 6 days. (B) Transcript levels of selected LTTM genes were measured by RT-qPCR during the chase period. Shown in the bar graphs is the average of 3 biological replicates, and a two-sided t-Test was performed to determine statistical significance. DX = Day x, AP = after pulse (C) RT-qPCR confirmation of RNA-seq results show memory for significant memory gene *Cd14*. Shown on one scale are the after chase (AC) samples, while the larger scale includes both the AC and after pulse (AP) samples. Average of 3 biological replicates, and a two-sided t-Test was performed to determine statistical significance.

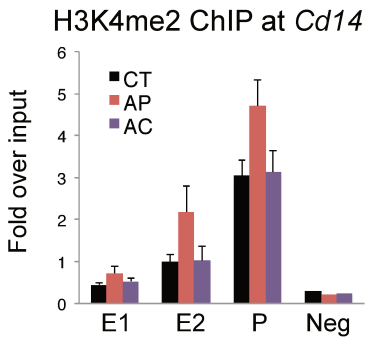


Supplemental Figure S2. (A) Relative transcript levels at the end of an 8h pulse for studied genes with and without flavopiridol treatment during the pulse period. (B) Relative transcript levels for highly expressed B cell gene *Il7r* after 4h flavopiridol treatment during the chase. (C) Pol II ChIP-qPCR with antibody 4H8 immunoprecipitates Pol II in the poised and elongating forms (phospho-serine 5 on the CTD). Pol II levels at selected LTTM and control gene enhancers (left) and promoters (right) were measured in pulsed and CT cells after the chase period. Values are displayed as fold over input. Average of 3 biological replicates, a two-sided t-Test performed to determine statistical significance.

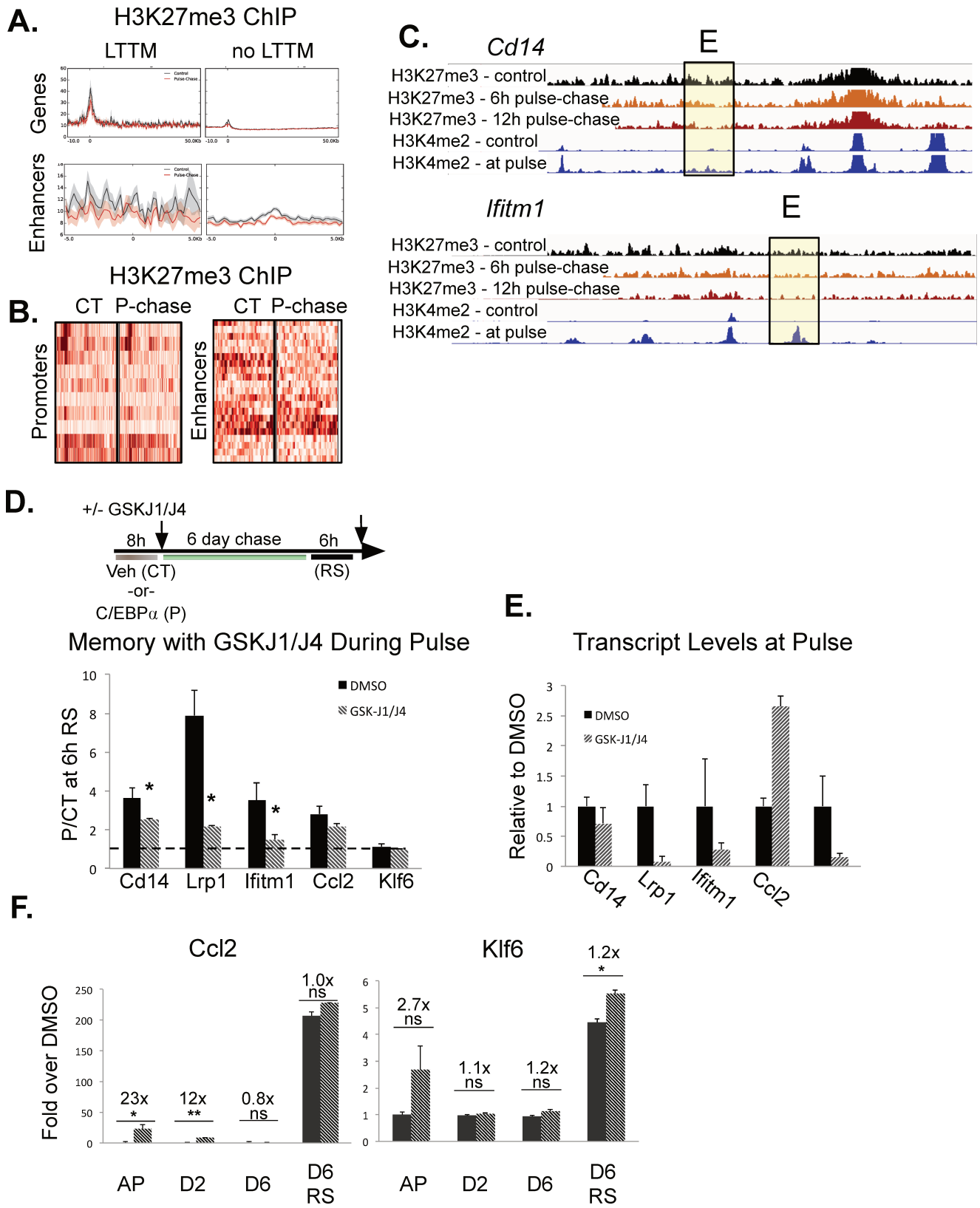
A.



B.



Supplemental Figure S3. (A) H3K4me1 ChIP-qPCR at selected LTTM and control gene enhancers at the pulse and throughout the chase period. Data are from a single ChIP experiment. (B) ChIP-qPCR of H3K4me2 at *Cd14* promoter and associated regulatory regions from control cells (CT), pulsed cells after the pulse (AP), and pulsed cells after the chase (AC). Average of 3 biological replicates, a two-sided t-Test was performed to determine statistical significance.


A.
H3K27me3 ChIP
LTTM
no LTTM
Genes
Enhancers
B.
H3K27me3 ChIP
CT P-chase
CT P-chase
Promoters
Enhancers
C.
Cd14
E
H3K27me3 - control
H3K27me3 - 6h pulse-chase
H3K27me3 - 12h pulse-chase
H3K4me2 - control
H3K4me2 - at pulse
Ifitm1
E
H3K27me3 - control
H3K27me3 - 6h pulse-chase
H3K27me3 - 12h pulse-chase
H3K4me2 - control
H3K4me2 - at pulse
D.
+/- GSKJ1/J4
8h
6 day chase
6h
Veh (CT)
-or-
C/EBPα (P)
(RS)
Memory with GSKJ1/J4 During Pulse
P/CT at 6h RS
DMSO
GSK-J1/J4
Cd14
Lrp1
Ifitm1
Ccl2
Klf6
E.
Transcript Levels at Pulse
Relative to DMSO
DMSO
GSK-J1/J4
Cd14
Lrp1
Ifitm1
Ccl2
F.
Ccl2
Fold over DMSO
23x
*
12x
**
0.8x
ns
1.0x
ns
AP
D2
D6
D6
RS
Klf6
2.7x
ns
1.1x
ns
1.2x
ns
1.2x
*
AP
D2
D6
D6
RS

Supplemental Figure S4. (A) H3K27me3 levels in control cells and 6h pulsed cells after the chase at LTTM and control gene enhancers and promoters (B) H3K27me3 signal across the top 10 most significant LTTM gene promoters and enhancers in control (CT) and pulsed cells after the chase (P-chase) (C) Genome browser window snapshots of the H3K27me3 ChIP-seq signals at enhancer sites used for qPCR in Figure 6B. ChIP tracks for H3K4me2 in control cells and after a 6h pulse are also shown to depict the sites of enhancer activation upon C/EBPα induction. (D) As depicted in the schematic, GSK-J1/J4 at 5 uM (or DMSO) was added to the cell culture simultaneously with C/EBPα induction and washed out after the pulse, and the cells carried through the memory protocol. RNA was taken at 6h re-stimulation for qPCR analysis to determine the memory of selected LTTM genes *Cd14, Lrp1,* and *Ifitm1.* Also included are control genes *Ccl2* and *Klf6.* Data is represented as fold P/CT transcript, where a value of 1 (dashed line) signifies no memory. Shown in the bar graphs is the average of 3 biological replicates, and a two-sided t-Test was performed to determine statistical significance. (E) Relative transcript levels after the pulse to determine the effect of GSK-J1/J4 on gene induction. (F) As related to Figure 6E, transcript levels for *Ccl2* and *Klf6* following UNC1999 treatment, in the chase period, and after 6h C/EBPα induction. The fold changes between UNC1999 and DMSO treated samples are depicted over the bar graph.

Table S1. LTTM genes and values used to generate Figure 1C, 'Memory' Column

ensemblID	geneName	pvalue	log2FoldChange
ENSMUSG00000051439	Cd14	7.58E-23	1.866462362
ENSMUSG00000029135	Fosl2	2.55E-12	1.175612747
ENSMUSG00000030747	Dgat2	1.10E-09	0.947403581
ENSMUSG00000036452	Arhgap26	3.11E-09	0.900953144
ENSMUSG00000025491	Ifitm1	7.67E-09	2.549707978
ENSMUSG00000062345	Serpinb2	2.53E-08	1.322551338
ENSMUSG00000035493	Tgfbi	2.99E-08	1.056033211
ENSMUSG00000078122	F630028O10Rik	3.48E-08	1.120390937
ENSMUSG00000035441	Myo1d	3.82E-08	1.939395143
ENSMUSG00000029484	Anxa3	1.11E-07	1.088718933
ENSMUSG00000056592	Zfp658	1.59E-07	0.95110332
ENSMUSG00000001768	Rin2	2.94E-07	1.031447199
ENSMUSG00000005824	Tnfsf14	3.43E-07	2.314269775
ENSMUSG00000015243	Abca1	1.21E-06	1.60516014
ENSMUSG00000031480	Thsd1	1.37E-06	1.060109498
ENSMUSG00000004609	Cd33	2.06E-06	1.118820622
ENSMUSG00000040249	Lrp1	2.35E-06	1.487331067
ENSMUSG00000020101	4632428N05Rik	2.99E-06	0.744159155
ENSMUSG00000029860	Zyx	3.51E-06	2.299076005
ENSMUSG00000037012	Hk1	4.98E-06	1.457404111
ENSMUSG00000097971	Gm26917	5.63E-06	2.983582952
ENSMUSG00000030123	Plxnd1	6.47E-06	1.034660457
ENSMUSG00000040548	Tex2	6.91E-06	0.470580968
ENSMUSG00000029703	Lrwd1	7.26E-06	0.475879091
ENSMUSG00000030970	Ctbp2	7.80E-06	0.761483555
ENSMUSG00000020261	Slc36a1	8.51E-06	0.902344531
ENSMUSG00000062593	Lilrb4	9.25E-06	1.905936847
ENSMUSG00000026835	Fcnb	1.07E-05	1.371642224
ENSMUSG00000001288	Rarg	1.18E-05	1.170737058
ENSMUSG00000034641	Cd300ld	1.56E-05	1.195032404
ENSMUSG00000029772	Ahcyl2	1.70E-05	0.995956793
ENSMUSG00000048897	Zfp710	2.59E-05	0.830513527
ENSMUSG00000018398	Sept8	3.26E-05	0.945059127
ENSMUSG00000040964	Arhgef10l	5.21E-05	1.075904334
ENSMUSG00000001227	Sema6b	6.21E-05	0.709012242
ENSMUSG00000030208	Emp1	6.80E-05	0.689175856
ENSMUSG00000026193	Fn1	7.05E-05	1.608665313

Supplemental Materials and Methods

Extended ChIP protocol

Cells at 10e6 cells/mL were fixed in 1% formaldehyde for 10 minutes at room temperature. Fixation was quenched with 0.125M glycine for 5 minutes. Cells were then spun down at 1200 rpm for 4 minutes, and the pellet washed 3 times in cold PBS. The cells were then re-suspended in swelling buffer (25 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.1% NP-40) containing protease and phosphatase inhibitors, and incubated for 10 minutes in ice. At this point, the cell suspension was transferred to a glass Dounce homogenizer with a 'tight' pestle and the cell membrane broken with 25 strokes of the pestle. Nuclei were spun down at 3000xg for 5 minutes at 4 degrees. Nuclei were re-suspended in sonication buffer (50 nM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS) at a concentration of ~10e6 cell nuclei per mL sonication buffer. Nuclei were sonicated with a Branson probe sonicator for 3 minutes total, 10 seconds on and 45 seconds off, at 30 mA. Following sonication, the lysate was spun down at 16,000xg for 15 minutes to remove the insoluble material, and this step repeated and the supernatant retained (the 'chromatin') for ChIP. 5% of the chromatin lysate was frozen to serve as input material. 1 mL (about 10e6 cells of starting material) of chromatin was used for each IP. Antibodies and amount used were as follows: PU.1 (sc-352x, 6 ul), p300 (sc-585x, 10 ul), H3K27me3 (07-449, 3 ul), H3K4me1 (Diagenode C15410037, 5 ug) H3K4me2 (04-790, 2 ul), H3K4me3 (05-1339, 5ul). 50 ul of protein-G Dynabeads were added with the chromatin and antibody mixture, and tubes were placed on a 360 rotor at 4 degrees overnight. The following morning, the IP was washed with the following buffers (all at 4 degrees, except for TE which is at room temperature): 1X in sonication buffer, 1X in wash A (50 mM HEPES pH 7.9, 500 mM NaCl, 1mM EDTA, 1%

Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), 1X in wash B (20mM Tris pH 8.0, 1mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), 2X in TE (10mM Tris pH 8.0, 1mM EDTA). Chromatin was then eluted from the washed beads by incubating with 250ul elution buffer (1% SDS, 0.1M NaHCO3) for 15 minutes at room temperature. This step was repeated, and the eluted fractions combined for a total of 500 ul. To this, 20 ul 5M NaCl was added, and the samples incubated at 65 degrees for 4 hours to reverse the crosslinks. Afterwards, to remove the protein in the sample, 10 ul of 0.5M EDTA, 20 ul of 1M Tris pH 6.5, and 2 ul of 10mg/ml Proteinase K was added and incubated for 1 hr at 45 degrees. DNA was recovered by Phenol/chloroform extraction.

For ChIP-seq library preparation, 10 ng of ChIP DNA was used for each library. The NEBNext DNA library Prep Reagent set for Illumina (E6000L) was used to generate the libraries, along with the standard Illumina barcoding primers, following the manufacturers instructions.

Inhibitors

The following compounds and concentrations were used to treat C10 cells either during the C/EBPa pulse, or during the chase (for flavopiridol): Flavopiridol at 300 nM, GSK-J1/J4 at 5uM. GSK-J1/J4 and UNC1999 were obtained in collaboration with the SGC Consortium in Toronto. Flavopiridol was purchased from VWR Catalog #89151-284. UNC1999 was used at 3uM for 48 hours in naïve C10 cells to induce a memory phenotype in the absence of a C/EBPα pulse in Figure 5.

FACS analysis

C10 cells after the chase period, and 24 hours into re-stimulation were taken for analysis. Cells were spun down to remove culture media and resuspended in PBS + 5% FBS, and incubated for 10 minutes. Then, cells were spun down, and resuspended at 1e7/mL in PBS + 5% FBS, and 100 ul taken for labeling with antibody. 5 ul of APC-CD14 antibody (Biolegend, #123312) was added to the 100 ul of cell suspension, and this mixed and incubated for 30 minutes at 4 degrees. Following this, the cells were washed once in PBS to remove unbound antibody, then resuspended in 0.5 mL PBS and transferred to a FACS tube for analysis. FACS analysis was performed on a BD LSRII machine.

Primers for RT-qPCR

Gene	Forward	Reverse
Klf6	TGCTCCCAATGTGTAGCATC	TTTGGTCCACAGGTCTTCCT
Cd11b	CTTCTGGTCACAGCCCTAGC	AGAGGGCACCTGTCTGGTTA
Fosl2	TATCCCGGGAACTTTGACAC	GTTGATTGTGGGGATGAAGG
Cd19	CATTGCAAGGTCAGCAGTGT	AAAAGCCACCAGAGAAACCA
Il7r	TGGCTCTGGGTAGAGCTTTC	GTGGCACCAGAAGGAGTGAT
ActB	CAGCTTCTTTGCAGCTCCTT	CACGATGGAGGGGAATACAG
Ccl2	AGCTCTCTCTTCCTCCACCA	GCTGCTGGTGATCCTCTTGT
Lrp1	GCTGCTTTCAGCTCTGGTCT	GGCAATCTCTTTCACCGTCA
Tlr2	TAGGGGCTTCACTTCTCTGC	CAGAACAGCGTTTGCTGAAG
Cebpb	ATCGACTTCAGCCCCTACCT	TAGTCGTCGGCGAAGAGG
Ifitm1	ATGTGGTCTGGTCCCTGTTC	AGACAACGATGACGACGATG
miR223	TTGCAGCCAGAATAGCAGAA	AGTGCCAAGAGGAATTGGAA
Cd14	CAGAGAACACCACCGCTGTA	TGGCTTCGGATCTGAGAAGT
Abca1	TGTCCATGTTGTAGCGCAGT	ATTCAGCTTGGTGATGCGGA
Ddx58	ACAACAAGGGCCCAATGGAA	TAACCTGCATGGTACAGGGC
Ccl3	ATATGGAGCTGACACCCCGA	GACAGCCCAGGTCAAAGGTT
Irf7	CTCTGCCCACACAGGTTCTG	CATCCAGGAACCCAAGGCTG
Fosl2	TATCCCGGGAACTTTGACAC	GTTGATTGTGGGGATGAAGG
Tnfa	CCACCACGCTCTTCTGTCTA	GGTTGTCTTTGAGATCCATGC

Primers used for ChIP-qPCR

Gene	location name	Forward	Reverse
Cd14	E1	TGGAGTTCAACCCACCAAATGA	TGGTTGTTATTTGCAACGGCT
Cd14	E2	CAGGTCCTTGTGCTTGTTCA	AGCAAAGACCAGCAAAGCAT
Cd14	P	ACGTTGCGGAGGTTCAAGAT	CTTAAAGCGGCTTACGGTGC
Cd14	gene body 1	CAGCATCCCGCAGTGAATTG	CCTCCAAGTTTTAGCGCTGC
Cd14	gene body 2	AGCGCTAAAACTTGGAGGGT	TGAATTGGGCGAGAGAGGAC
Irf7	E1	AGCATTGCTGAGGCTCACTT	GAGGTACCTGGTGGTTTGGG
Irf7	E2	CTCCCAAATGAACGTGCGTG	CCAGAGTCTTAAGGCGGGTG
Ifitm1	E	TGCCGGGAATGGTTCACTTT	TCCAACGAATGCTGCTGTCA
Ifitm1	P	CAAAGCAAACTCGGAGCGAC	GTTTGAAGCTGTGAGGCAGC
Ifitm1	gene body	CATGTGGTCTGGTCCCTGTT	GGAGATACCTTGACCCACGG
Ccl2	E1	ACACCTGAAAACTGCACCCA	GGAAAGCCTTGCCCAATGTG
Ccl2	P	GCTCCCTTCCCAGCATATCAC	TTGGGTTGGCTCTTTGTCCTT
Ccl2	gene body 1	GGCTGGAGAGCTACAAGAGG	AGGGCCGGGGTATGTAACT
Ccl2	gene body 2	TCTACCCAAGACTGTGAGCCT	GAGAAGAGAAGGACCACGGG
Lrp1	E	TAGCTTGGGGAAGGTGCCTA	CGAGGTAAGCGGTCAACAGG
Tlr2	E	GCCTATGGTTGCTCACGGAA	TCCCTGTCGTCATCTCCCAA
Klf6	E	ACCAAGCAAGCAGGTCACTT	TGGGTTTTGTCAGTTGGGCT
Klf6	P	TGAGCCCTTGGCACAAGTTT	GCTCGTTCGTATGACAGGGT
Il7r	E	GAGCTGTTGGACAGCATCAG	TTTGACAGGGAGGAGAATGG
Il7r	P	CCATCCTGGGCATTTCCACT	GGCTCTGGGTAGAGCTTTCG
Il7r	gene body	TCACTCGTAAAAGAGCCCCAC	TCCTCCTCACCATTAACTTGCT
Cd19	E	CCTGGTGACTCGGACTGTTT	GGCAACATTCCAGCTTTTGT
Cd19	gene body	AGCGAGACAGAGATACGGAGT	CTGGACAGTGAACGTGGAGG
SpiB	E	CATGGGCTAGGGGCTAGAGA	AAGCCTGGATGCTTCTGACC
Negative	N/A	CTTTTCCCGCTTTCTCCTCT	CGACAAGGGAGCTAAAACCA

List of LTTM gene designated enhancers (with locations relative to the TSS) used to generate the heatmap in Figure 6A

TOP

Rin2_74622
Ifitm1_49297
Lilrb4_-26748
Ifitm1_45613
Ifitm1_30616
Rin2_37151
Cd14_18161
Ifitm1_32221
Cd33_-12838
Slc36a1_5264
Ifitm1_55008
Arhgef10I_99253
Ifitm1_23419
Cd33_-11334
Cd300Id_82270
Zfp710_-9427
Thsd1_21858
4632428N05Rik_-94357
4632428N05Rik_2228
Cd33_63
Plxnd1_-9766
Cd14_6764
Cd33_2893
Cd300Id_72700
Fcnb_-7165
Arhgef10L_-33516
F630028O10Rik_-332
Cd300Id_58951
Fosl2_59408
Dgat2_-7199
Tnfsf14_-98042
Fosl2_55525
Cd300Id_55697
4632428N05Rik_-66860
4632428N05Rik_-2724
Fosl2_57819
Abca1_-67873
Zfp710_49963
Dgat2_-5859
Lrp1_-46907
Abca1_39567
Plxnd1_-18702
Thsd1_34995
Tnfsf14_-95987
Myo1d_68625
Tnfsf14_-90853
Plxnd1_93861
Plxnd1_93442
Arhgap26_2906
4632428N05Rik_-84427
Rarg_14159
Serpinb2_-9300
4632428N05Rik_-86218
Zfp710_19521
Fn1_-77803
Fn1_-75879
Ctbp2_71377
Lrp1_16715
Zyx_3452
Sept8_-11295
Arhgef10I_3472
Ifitm1_14575
Tgfbi_4001
Cd33_51843
Emp1_1414
Thsd1_39495
Tgfbi_1577
Lilrb4_38423
Tgfbi_2139
Ahcyl2_46556

5.4 Publication: Krox20 hindbrain regulation incorporates multiple modes of cooperation between cis-acting elements (PLoS genetic, 2017)

Contribution:

I participated to the design of the ATAC-seq experiment, I processed and analysed the ATAC-seq data, and participated to the interpretation of the results compared to the other experiments.

PLOS | GENETICS

RESEARCH ARTICLE

Krox20 hindbrain regulation incorporates multiple modes of cooperation between *cis*-acting elements

Elodie Thierion[1,2¤], Johan Le Men[1], Samuel Collombet[1], Céline Hernandez[1], Fanny Coulpier[1], Patrick Torbey[1], Morgane Thomas-Chollier[1], Daan Noordermeer[3], Patrick Charnay[1]*, Pascale Gilardi-Hebenstreit[1]*

1 Ecole normale supérieure, PSL Research University, CNRS, Inserm, Institut de Biologie de l'Ecole normale supérieure (IBENS), Paris, France, **2** Sorbonne Universités, UPMC Univ Paris 06, IFD, Paris, France, **3** Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, University Paris-Saclay, Gif-sur-Yvette, France

¤ Current address: University of Cambridge, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, United Kingdom
* gilardi@biologie.ens.fr (PGH); charnay@biologie.ens.fr(PC)

Check for updates

OPEN ACCESS

Citation: Thierion E, Le Men J, Collombet S, Hernandez C, Coulpier F, Torbey P, et al. (2017) *Krox20* hindbrain regulation incorporates multiple modes of cooperation between *cis*-acting elements. PLoS Genet 13(7): e1006903. https://doi.org/10.1371/journal.pgen.1006903

Editor: Jose Luis Gomez-Skarmeta, Centro Andaluz de Biología del Desarrollo, SPAIN

Received: March 13, 2017

Accepted: July 3, 2017

Published: July 27, 2017

Copyright: © 2017 Thierion et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE94716.

Funding: The P.C. laboratory was financed by the Institut National de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Ministère de la Recherche et Technologie and the Fondation pour la Recherche Médicale (FRM). It has received support under the PIA launched by the French Government and implemented by the ANR, with the references: ANR-10-LABX-54

Abstract

Developmental genes can harbour multiple transcriptional enhancers that act simultaneously or in succession to achieve robust and precise spatiotemporal expression. However, the mechanisms underlying cooperation between *cis*-acting elements are poorly documented, notably in vertebrates. The mouse gene *Krox20* encodes a transcription factor required for the specification of two segments (rhombomeres) of the developing hindbrain. In rhombomere 3, *Krox20* is subject to direct positive feedback governed by an autoregulatory enhancer, element A. In contrast, a second enhancer, element C, distant by 70 kb, is active from the initiation of transcription independent of the presence of the KROX20 protein. Here, using both enhancer knock-outs and investigations of chromatin organisation, we show that element C possesses a dual activity: besides its classical enhancer function, it is also permanently required in *cis* to potentiate the autoregulatory activity of element A, by increasing its chromatin accessibility. This work uncovers a novel, asymmetrical, long-range mode of cooperation between *cis*-acting elements that might be essential to avoid promiscuous activation of positive autoregulatory elements.

Author summary

The formation of multicellular organisms from the egg to the adult stage is largely under genetic control. The activation of specific genes is governed by regulatory DNA sequences present nearby on the chromosome. Most of these sequences promote activation and are called enhancers. In this paper, we study two enhancers governing the expression of a gene involved in the formation of the posterior brain in vertebrates. One of these enhancers is involved in a positive feedback loop: it is itself activated by the protein product of the gene that it regulates. The other enhancer was thought to be only involved in the initial

MEMOLIFE and ANR-11-IDEX-0001-02 PSL* Research University. The genomic platform of the IBENS is supported by France Génomique national infrastructure and funded as part of the Programme "Investissements d'Avenir" (PIA) managed by the Agence Nationale de la Recherche (ANR, contract ANR-10-INBS-0009). ET and JLM were supported by doctoral grants from Sorbonne Universités and by the FRM. PT is supported by doctoral grant from PSL Research University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

accumulation of the protein, necessary for the subsequent activation of the feedback loop. Here we show that the second enhancer directly cooperates with the autoregulatory enhancer to increase its accessibility and its activity. Our work uncovers a novel, long-range mode of cooperation between enhancers that restricts the domain of action of auto-regulatory enhancers within embryos and might be essential to avoid their inappropriate activation.

Introduction

DNA *cis*-acting elements play key roles in the regulation and evolution of gene expression by controlling spatiotemporal transcription patterns. A major class of *cis*-regulatory elements are transcriptional enhancers, which can recruit combinations of transcription factors (TFs) to modulate transcription initiation from (a) cognate gene promoter(s), in general independently of their relative distance and orientation [1–3]. So far, most enhancers have been functionally characterized by assay of their transcriptional activity using transgenic constructs carrying the enhancer and a reporter gene driven by a minimal promoter [4]. Another strategy consists in the random insertion of a transposon that senses enhancer activity within the surrounding genomic region. It is particularly useful to detect multiple *cis*-regulatory elements with similar activities and long-distance modulation of gene expression [5,6]. Transgenesis using BACs allows the introduction of large DNA fragments containing enhancers in their native context. This approach is helpful in the analysis of multiple enhancers controlling the same gene [7], but can be challenging for the study of mammalian enhancer that are located far away from the promoter that they control. These different approaches provide useful information on spatial and temporal activity of the putative enhancer element, but they usually do not establish whether and how the enhancer actually participates in the control of the expression of its suspected cognate gene in its full normal genomic context. Answer to this latter question requires *in vivo* analyses involving deletion or mutation of the endogenous enhancer. This issue is particularly important in situations where multiple, overlapping enhancers operate within the same *cis*-regulatory landscape. In such cases, various types of regulatory crosstalk can occur between the enhancers, resulting in additive, synergistic, competitive or repressive effects [3]. In vertebrates, very few studies have addressed such situations.

Enhancer activity is intimately linked to chromatin organization. Hence, association of pioneer TFs to an enhancer can lead to chromatin decompaction and facilitate the binding of additional TFs and/or recruitment of various epigenetic machineries [8]. In return, chromatin configuration can affect gene expression by modulating long-range interactions between enhancers and promoters [9], that are usually constrained within regions called topologically associated domains (TADs) [10,11]. TADs, which are approximately Mb-sized in mammals, form constitutive "regulatory neighbourhoods" that provide specificity to enhancer-promoter interactions by reducing aberrant contacts between *cis*-regulatory elements located in distinct TADs [6,10].

To provide insights into the mechanisms involved in the regulation of a vertebrate gene by multiple enhancers during development, we investigated the case of the mouse *Krox20/Egr2* gene [12] for which several hindbrain-specific enhancers have been identified [13]. The hindbrain is an attractive model to investigate the genetic control of morphogenesis in vertebrates, as it is subject to a transient segmentation process leading to the formation of 7–8 segments called rhombomeres (r) [14,15]. The formation and specification of segments r3 and r5 are governed by the transcription factor KROX20/EGR2 [15–17]. So far three evolutionarily

conserved sequences exhibiting enhancer activity in the hindbrain have been identified within the *Krox20* locus and are termed A, B and C [13]. Element A, located 217 kb upstream of the promoter in the mouse, is active in both r3 and r5. This element carries several KROX20 binding sites and requires direct binding of the protein for its activity, suggesting that it acts as an autoregulatory element [13]. Indeed, upon deletion of element A, *Krox20* expression is normally initiated in r3 and r5, but is not amplified nor maintained at later stages [18]. Additional studies have indicated that element A underlies a positive feedback loop that acts as a binary switch for specification of odd- versus even–numbered rhombomere identity [18]. Element B, located 164 kb upstream of the promoter, drives the expression of reporter constructs specifically in r5 [13,19,20]. Finally, element C, located 144 kb upstream of the promoter, is active in the r3-r5 region [13,19–21]. Several observations suggest that elements B and C, in contrast to the autoregulatory element A, are involved in the initial steps of *Krox20* expression in r3 or r5 (initiator elements): i) they are transcriptionally active at the early stages of *Krox20* hindbrain expression [13]; ii) they are activated by transcription factors known to act upstream of *Krox20* [19–21]; iii) they do not require the presence of the KROX20 protein for their activity [13].

In the present study, we have investigated the contribution of element C to *Krox20* expression, as it was the only characterized initiator element with an activity in r3. Using a conditional knock-out mutation of element C, we show that, unexpectedly, this element is not necessary for *Krox20* initial expression in r3. In contrast, it appears absolutely required for the maintenance of *Krox20* expression in this rhombomere. This activity involves a cooperation in *cis* with element A, element C potentiating its activity and increasing its accessibility. These observations reveal that a *cis*-acting element can cooperate with other enhancers within the same locus according to different modalities and suggest a scheme for protecting autoregulatory elements from inappropriate activation.

Results

In r3, element C is required for late, but not initial *Krox20* expression

To assess the contribution of element C to the regulation of *Krox20* expression, we generated a mouse line carrying a deletion of this element. The details of the knock-out strategy are presented in Fig 1A. Two alleles were generated: $Krox20^{Cflox}$, in which element C is present, but flanked by two loxP sites, and $Krox20^{\Delta C}$, in which element C is deleted. The impact of element C deletion on *Krox20* transcription was analysed by mRNA *in situ* hybridization on 4 to 14 somite stage (s) embryos, comparing homozygous ($Krox20^{\Delta C/\Delta C}$) with heterozygous mutants ($Krox20^{+/\Delta C}$), the knock-out of one allele of *Krox20* having no phenotype [15,16,22]. Unexpectedly, up until 6s the expression of *Krox20* in r3 and r5 is similar in $Krox20^{\Delta C/\Delta C}$ embryos and control littermates (Fig 1B). However, at 8s, *Krox20* expression is severely reduced in r3 from $Krox20^{\Delta C/\Delta C}$ embryos as compared to controls and, at later stages, it is completely lost (Fig 1B). During the considered period, although *Krox20* expression does not appear to be dramatically affected in r5, in contrast to r3, the width of the corresponding domain of expression appears to be slightly reduced after 10s (Fig 1B).

To investigate longer-term consequences of element C deletion on cell specification, we analysed the expression of a KROX20 target gene, *EphA4* [22], which is known to persist beyond the period of *Krox20* expression in r3 and r5 [18]. In control embryos ($Krox20^{+/\Delta C}$), at 10s and 25s, *EphA4* is expressed at high levels in r3 and r5 and at a lower level in r2 (S1 Fig). At both stages, the r3 domain, as demarcated by *EphA4* expression, is reduced in $Krox20^{\Delta C/\Delta C}$ embryos as compared to controls, whereas the r5 domain is similar in both genotypes (S1 Fig). This is consistent with the premature loss of *Krox20* expression in r3, known to reduce the extension of this rhombomere [18,22]. The limited variation of *Krox*20 expression in r5 in

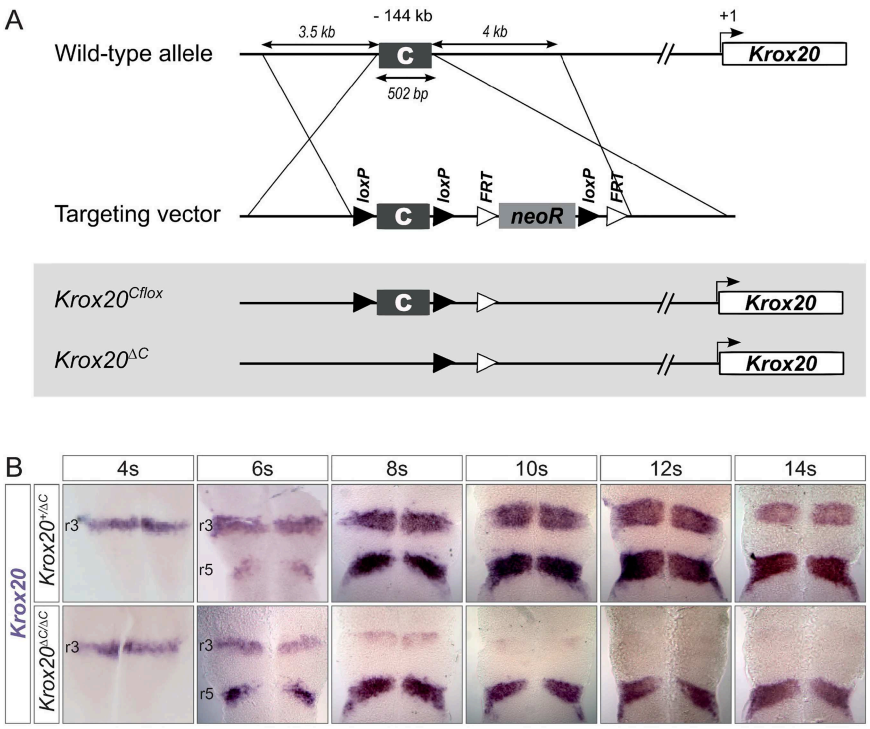


Fig 1. Genetic analysis of element C function. (A) Strategy for the construction of conditional and null alleles of element C. The targeting vector was introduced into the locus in ES cells by homologous recombination and one of the ES clones subsequently allowed germ line transmission in the mouse. The floxed allele, *Krox20*^{Cflox}, was obtained by crossing the founder mouse line with a *Flp* (targeting FRT sites) deletor line. The null allele, *Krox20*$^{\Delta C}$, was obtained by crossing the *Krox20*^{Cflox} line with a *Cre* (targeting loxP sites) deletor line, PGK-Cre. **(B)** In situ hybridization for *Krox20* mRNA performed on *Krox20*$^{+/\Delta C}$ and *Krox20*$^{\Delta C/\Delta C}$ embryos at the indicated somite stages. Embryos were flat-mounted with anterior toward the top. Rhombomere positions are indicated on the left.

https://doi.org/10.1371/journal.pgen.1006903.g001

Krox20$^{\Delta C/\Delta C}$ embryos after 10s does not appear to perturb the size of this rhombomere at later stages, consistent with the fact that *Krox20* expression is not required for the maintenance of *EphA4* expression in r3 and r5 [18].

In conclusion, these data indicate that i) element C is dispensable for the initiation of *Krox20* expression in r3 or r5, suggesting the existence of other elements in charge of these functions; ii) in r3, in contrast, element C is absolutely required for expression beyond 6s, leading to a reduction in size of this rhombomere at later stages. Notably, the phenotype observed in r3 in *Krox20*$^{\Delta C/\Delta C}$ embryos is very similar to what was previously described in *Krox20*$^{\Delta A/\Delta A}$ embryos (Figs 1B and 2A, and S1 Fig) [18]; iii) in r5, the contribution of element C to *Krox20* expression is rather limited, without significant effect on the size of this rhombomere.

Elements A and C cooperate in *cis* for the establishment of the autoregulatory loop

The similarity of the phenotypes observed in r3 upon deletion of elements A or C led us to investigate the possibility of an involvement of element C in *Krox20* autoregulation, together with element A. For this purpose, we first analysed the expression of *Krox20* in composite heterozygous embryos, *Krox20*$^{\Delta A/\Delta C}$, carrying deletions of element A on one allele and of element C on the other (Fig 2A). Although this combination does not affect *Krox20* expression in r3 at early stages, at 8s *Krox20* mRNA level is severely reduced and, at 12s, it is completely lost,

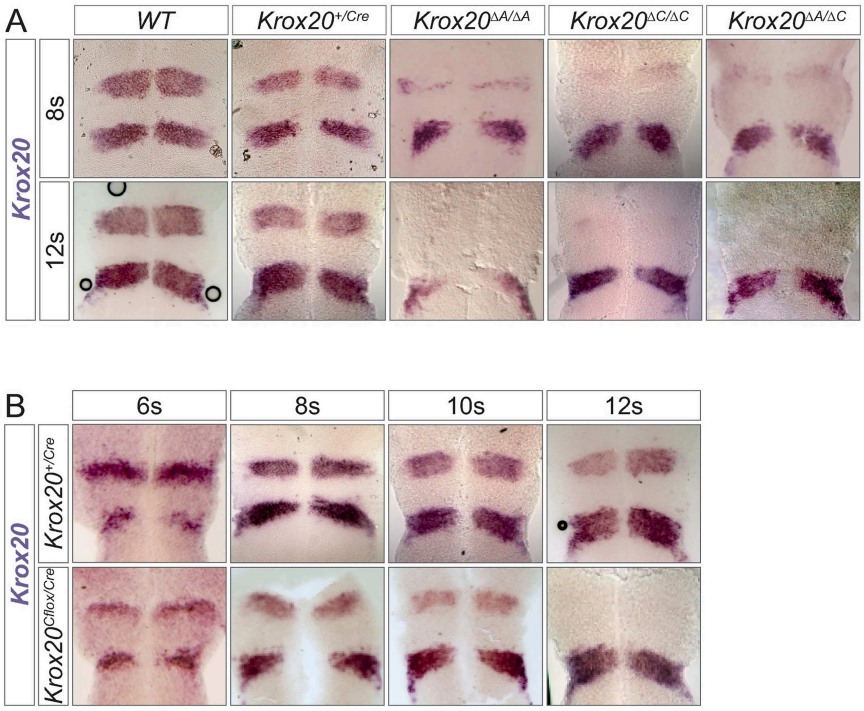


Fig 2. Cooperation in *cis* between elements A and C. (A) In situ hybridization for *Krox20* mRNA was performed on wild type (WT), *Krox20*$^{+/Cre}$, *Krox20*$^{\Delta A/\Delta A}$, *Krox20*$^{\Delta C/\Delta C}$ and composite heterozygous *Krox20*$^{\Delta A/\Delta C}$ embryos at the indicated somite stages. **(B)** In situ hybridization for *Krox20* mRNA was performed on *Krox20*$^{+/Cre}$ and *Krox20*$^{Cflox/Cre}$ embryos at the indicated somite stages. In (A) and (B) embryos were flat-mounted with anterior toward the top.

https://doi.org/10.1371/journal.pgen.1006903.g002

mimicking the phenotype observed in *Krox20*$^{\Delta A/\Delta A}$ or *Krox20*$^{\Delta C/\Delta C}$ embryos at both stages (Fig 2A). In r5, *Krox20* expression is only slightly affected in *Krox20*$^{\Delta A/\Delta C}$ embryos, similarly to *Krox20*$^{\Delta C/\Delta C}$ embryos (Fig 2A). This defect in the maintenance of *Krox20* expression in r3, combined with apparently normal expression at early stages, contrasts with the fact that a single wild type *Krox20* allele is sufficient to activate and maintain the autoregulatory loop (Fig 2A, *Krox20*$^{+/Cre}$) [12]. This suggests that in *Krox20*$^{\Delta A/\Delta C}$ embryos the level of expression of *Krox20* is not a limiting factor for the activation of the only wild type allele of element A. Therefore, the most likely explanation for the defect in *Krox20* maintenance is that, in r3, the deletion of element C impairs the activity of element A located on the same chromosome and that the two elements synergistically cooperate, in *cis*, for the establishment and/or maintenance of the autoregulatory loop. A more conventional, partial redundancy between elements A and C appears much less likely.

This cooperation does not preclude an early involvement of element C, for instance to poise element A for the subsequent autoregulation phase. To investigate whether element C has a function only at the early phase of *Krox20* activation, or whether it is required during the autoregulation phase as well, we generated a genetic condition in which element C is initially active, but is deleted at a later stage. This was achieved by combining the *Krox20*^{Cflox} allele (Fig 1A) with a knock-in allele, *Krox20*^{Cre}, in which the *Krox20* coding sequence has been replaced by the coding sequence of the Cre recombinase [23]. In such embryos, *Krox20* and *Cre* are expected to be synthetized at early somitic stages. Subsequently, the recombinase leads to deletion of element C in r3 and r5. In *Krox20*$^{Cflox/Cre}$ embryos, *Krox20* expression is progressively reduced in r3 from 6s to 10s, as compared to *Krox20*$^{+/Cre}$ controls (Fig 2B), although less

abruptly than in *Krox20*$^{\Delta C/\Delta C}$ mutants (Figs 1B and 2A). At 12s, *Krox20* expression is completely abolished in r3 in *Krox20*$^{Cflox/Cre}$ embryos (Fig 2B). These data indicate that the presence of element C only during the early phase of *Krox20* expression is not sufficient to establish and/or maintain the autoregulatory loop. The higher level of *Krox20* in r3 in *Krox20*$^{Cflox/Cre}$ as compared to *Krox20*$^{\Delta C/\Delta C}$ embryos is likely to originate from transient activation of the loop, followed by termination of its activity, due to Cre excision of element C.

In conclusion, these results indicate that elements A and C synergistically cooperate in *cis* for establishing and/or maintaining this loop in r3. More precisely, they show that element C is permanently required for activity of the *Krox20* feedback loop.

Physical interactions between *Krox20 cis*-elements

The existence of a cooperation in *cis* between elements A and C led us to investigate the existence of possible physical 3D interactions between the different *Krox20 cis*-elements, which are separated by large distances on the mouse chromosome. A previous Hi-C analysis [11] in embryonic stem cells identified a TAD that includes the gene and its *cis*-regulatory elements (Fig 3A). The left boundary of the TAD spreads out over a relatively large and undefined transition zone (S2A and S2B Fig). To better characterize the *Krox20* regulatory neighborhood, we used circular chromosome conformation capture (4C-seq) on multiple viewpoints in the locus [24]. In samples prepared from total embryos at embryonic day (E) 9.5, when *Krox20* is no more transcribed [25], the *Krox20* gene and its distant regulatory element A (separated by over 200 kb) show highly similar distributions of 4C-seq signal (Fig 3A and S2B and S2C Fig) preferentially located in the *Krox20* TAD. In contrast, the distribution of the *Nrbf2* gene, which is located in the TAD transition zone and is separated from element A by a 35 kb genomic interval (including a cluster of CTCF binding sites) spreads its interactions about equally over the two neighboring TADs (S2C Fig). Repositioning of the TAD boundary to the cluster of CTCF binding sites results in strongly increased separation of signal between the *Nrbf2* gene on one hand and the *Krox20* gene and its regulatory elements on the other hand, indicating they are located in different regulatory neighborhood (Fig 3A and S2C Fig).

To determine if 3D chromatin interactions in the *Krox20* regulatory neighborhood were dynamic at these early stages of embryogenesis, and possibly linked to the autoregulatory loop, we compared our E9.5 samples to micro-dissected embryonic heads at E8.5, when the autoregulatory loop is active in a subset of cells [18]. For all viewpoints, very similar patterns were obtained between E8.5 heads and E9.5 (Fig 3B). At both stages, the *Krox20* promoter forms long-range interactions within the *Krox20* TAD that cover elements A and B (Fig 3B). In addition, bi-directional interactions are formed between elements A and B and, to a lesser extent, between elements A and C (Fig 3B).

In conclusion, these data reveal that the *Krox20* regulatory neighbourhood adopts a higher-order configuration that incorporates long-range interactions between the various *cis*-regulatory elements and is mostly invariant at different positions in the early embryo.

Chromatin accessibility of *Krox20* enhancers correlates with transcriptional activity

To investigate the correlation between the activity of the *Krox20 cis*-regulatory elements and their chromatin modifications and conformation, we first performed ChIP-seq experiments [26] to analyse two histone modifications: H3K4me1 (broad peaks covering active enhancers) and H3K27ac (punctuated peaks covering active enhancers and promoters) [27]. In E8.5 wild type embryo heads, a number of H3K4me1 peaks were observed, including those that expectedly correspond to the A, B and C elements and to a previously known neural crest element

(NCE; Fig 4A) [28]. The signals observed for the H3K27ac mark were low across the *Krox20* locus except for the promoter (S3A Fig). We can observe the same pattern of H3K27ac at the *EphA4* locus with low signal at the enhancer driving its expression in r3 and r5 [29] and higher signal at the promoter (S3B Fig). In contrast, a gene widely expressed at E8.5 in the whole embryo head, like *Sox2* [30], displays a high H3K27ac enrichment (S3C Fig). The low signals observed for the *Krox20* and *EphA4* genes are most likely due to the limited number of *Krox20*-expressing cells in the sample.

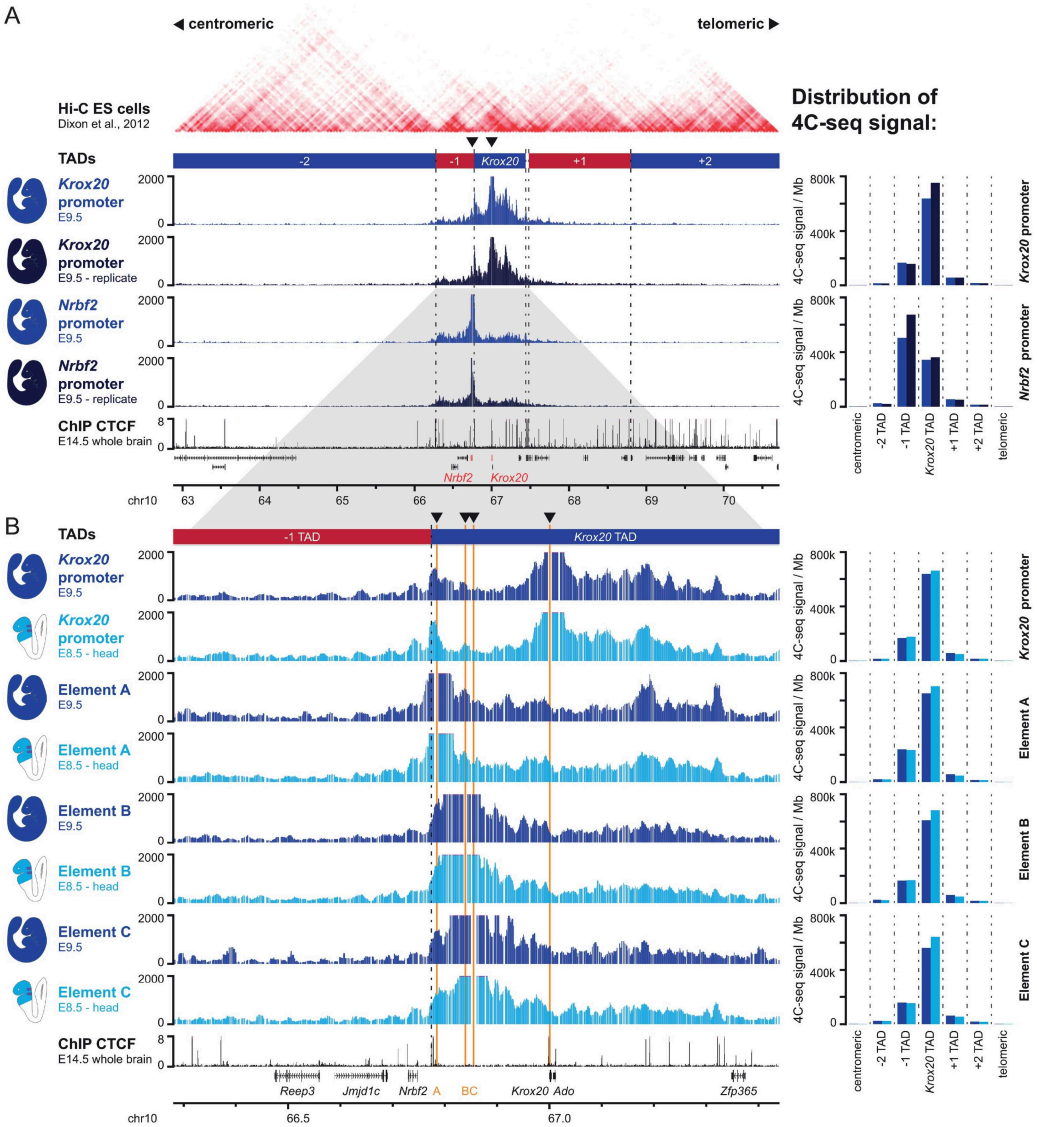


Fig 3. Physical interactions within the *Krox20* locus. (A) Alignment of data in the *Krox20* and adjacent loci from Hi-C in ES cells [11], 4C-seq in E9.5 whole mouse embryos, using the *Krox20* and *Nrbf2* promoters as viewpoints (this work, 2 biological replicates) and CTCF ChIP-seq in E14.5 mouse brain (ENCODE, [58]). **(B)** Zoom in on the *Krox20* locus, showing 4C-seq data from the *Krox20* promoter, element A, element B and element C as viewpoints. CTCF ChIP-seq data in E14.5 mouse brain (ENCODE) are indicated below. Signals from simultaneously processed E9.5 whole embryo (dark blue) and E8.5 embryo head (light blue) samples are shown. On the right, normalized distributions of the 4C-seq signals in different genomic regions are indicated. TADs as defined in [7] or by our additional analysis (S2 Fig) are indicated above, with dashed lines in the graphs demarcating TAD boundaries. Genes (black/red), *cis*-regulatory elements (orange) and genomic coordinates are indicated below each set of data. Arrowheads above each 4C track pinpoint viewpoints.

https://doi.org/10.1371/journal.pgen.1006903.g003

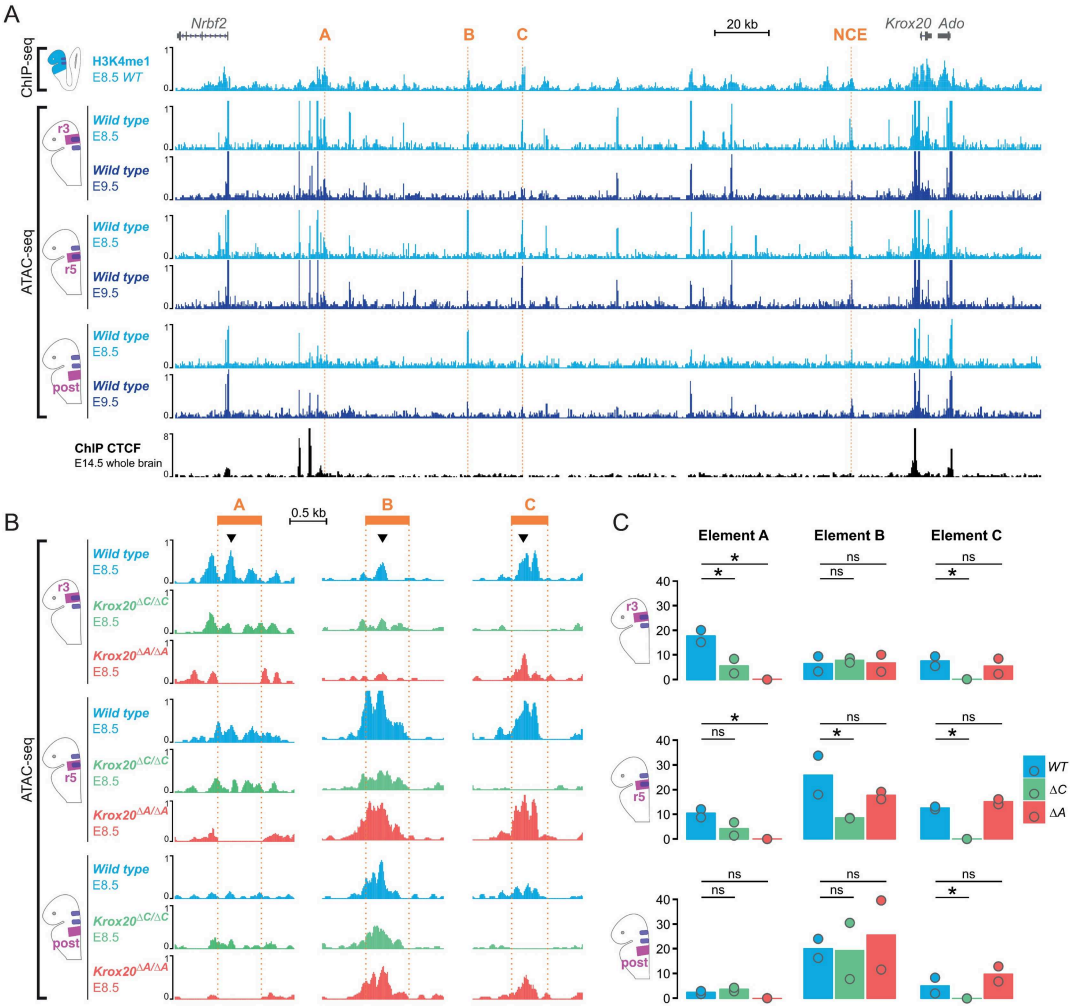


Fig 4. Chromatin modifications and accessibility within the *Krox20* locus. (A) ChIP-seq was performed for the H3K4me1 mark on wild type E8.5 embryo heads (light blue) using biological duplicates. Only one set of data is shown. ATAC-seq was performed on dissected regions (r3, r5 and a more posterior region ("post"; see text) from wild type embryos at E8.5 (light blue) and E9.5 (dark blue) using biological duplicates. Only one set of data is shown. CTCF ChIP-seq data from E14.5 mouse brain (ENCODE) are indicated below (see Fig 3A). Genes, *cis*-regulatory elements (orange) and a genomic scale are indicated at the top. **(B)** ATAC-seq was performed on dissected parts from wild type (light blue), and *Krox20*$^{\Delta C/\Delta C}$ (green) and *Krox20*$^{\Delta A/\Delta A}$ (red) E8.5 embryos using biological duplicates. Only one set of data is shown. *Cis*-regulatory elements (orange) and a genomic scale are indicated at the top. Arrowheads indicate the summits (defined by macs2 after peak calling, see Material and methods) used for quantifications in (C). **(C)** Barplots showing signal intensity of ATAC-seq (normalized fragment counts) at the summit of each element (arrowheads in panel B) for wild type (WT, blue) and *Krox20*$^{\Delta C/\Delta C}$ (ΔC, green) embryos at E8.5 for each dissected part. The statistical significance was calculated using a negative binomial Wald Test (R package DESeq2) on the 2 replicates, which are represented by dots. Star indicates p-value < 0.05. ns: non-significant.

https://doi.org/10.1371/journal.pgen.1006903.g004

To overcome this limitation, we performed micro-dissections and assessed chromatin structure by ATAC-seq [31], a technique that requires much lower cell numbers (a few thousand). Enhancer activity is usually associated with increased local chromatin accessibility. E8.5 (8s-10s stage) or E9.5 embryos were dissected to isolate samples derived from three regions: an anterior region, obtained by cutting within r2 and r4 (r3 sample); an intermediate region, for which cutting was performed within r4 and r6 (r5 sample), and a posterior region, for which cutting was performed within r6 and r7-8 (posterior sample). We observed peaks of accessibility at the level of the promoter at both stages and in all of the regional samples (Fig 4A). At

both stages, element A was specifically accessible in the r3 and r5 samples, but not in the posterior sample (Fig 4A and 4B), in accordance with its activity restricted to r3 and r5 [13]. At E8.5, element B was compacted in the r3 sample, but highly accessible in the r5 and posterior samples (Fig 4A and 4B). This accessibility largely decreased at E9.5 (Fig 4A). The limited accessibility of element B in r3 is in agreement with its lack of activity in this rhombomere [13]. Finally, element C was particularly accessible in the r3 and r5 samples at E8.5 (Fig 4A and 4B), consistent with its activity that spans the r3-r5 region [13]. This accessibility was only maintained in the r5 sample at E9.5 (Fig 4A).

The pattern of chromatin accessibility observed in our ATAC-seq experiments revealed additional potential enhancers involved in the regulation of the *Krox20* gene in the hindbrain. Indeed, we have identified an element located 107 kb downstream to *Krox20* with high accessibility at E8.5 (S3A Fig). We have tested the transcriptional activity of this new element (NE) by transgenesis in the zebrafish by cloning it upstream of a *GFP* reporter gene. In a transgenic line, this element drives specific expression in r3 at the time of the initiation of *Krox20* expression in this rhombomere (S4 Fig). These data raise the possibility that element NE might be the missing element involved in the initiation of *Krox20* expression in r3, although its activity still needs to be verified in the mouse.

In conclusion, this analysis reveals that the patterns of accessibility of the different known elements largely correlate with their enhancer activities and helped us to identify a novel candidate element for the regulation of *Krox20* expression in r3.

Element C modulates the accessibility of enhancer A

A final step was to investigate the effects of enhancer deletions on the accessibility of the other elements. Deletion of element A did not significantly affect the accessibility of elements B or C in any samples (Fig 4B and 4C and S3A Fig). In contrast, deletion of element C significantly reduced the accessibilities of element A in r3 and of element B in r5 (Fig 4B and 4C and S3A Fig). These data establish that element C has the capacity to specifically modulate the accessibility of elements A and B and therefore probably their activities. They may provide a mechanism for the involvement of the late activity of element C in the control of *Krox20* autoregulation governed by element A. Furthermore, this analysis establishes the existence of an asymmetry in the relationship between elements A and C: whereas element C affects A accessibility and presumably potentiate its activity, the reverse is not true.

Discussion

In the present study, we have made a further step in the understanding of the molecular mechanisms governing the expression of a master developmental regulator, using both enhancer knock-outs and investigations of chromatin structure. This analysis reveals that *Krox20* regulation relies on a complex crosstalk between several *cis*-acting elements that interact simultaneously according to multiple modes (redundant/additive/synergistic, symmetric/asymmetric) to shape the pattern of expression of the gene. Among these enhancers, element C performs a dual function, as a classical enhancer and as a potentiator in *cis* of element A. We propose that this latter role may constitute a general means to prevent promiscuous activities of autoregulatory elements.

Control of the early phase of *Krox20* expression in r3

Previous analyses had suggested a rather straightforward mode of regulation of the *Krox20* gene. Element C was responsible for the initiation of its expression in r3, whereas element B, possibly together with element C, was in charge of the initiation in r5. Subsequently, element

A governed the maintenance of the expression through a positive feedback loop towards a definitive engagement into an odd-numbered rhombomere fate [13,18,19]. Knock-out analysis of element C now leads to major revisions of this scenario. Despite the early r3 enhancer activity of element C, as demonstrated by transgenic experiments in mouse and fish [13,32], the deletion of mouse element C does not affect early *Krox20* expression in r3 (Fig 1B). This suggests that another *cis*-acting element contributes to this expression. This is not element B, which is only active in r5, as revealed in transgenic experiments [13], nor element A, which is absolutely dependent on the presence of the KROX20 protein [13,18]. Therefore, another enhancer, active in r3 and not dependent on KROX20, must be involved. Indeed, the identification of the NE element, fully accessible at early time in the hindbrain and specifically active in r3, makes it an attractive candidate for being involved in the initiation of *Krox20* expression in this rhombomere (S3A Fig and S4 Fig).

The absence of phenotype in *Krox20*$^{\Delta C/\Delta C}$ embryos during the early phase of *Krox20* expression does not preclude a role for element C during this phase. In support of this idea, enhancer activity of element C in r3 is dependent on the binding of Meis and Hox/Pbx factors [21], as well as of the Sp5 factor mediating FGF signalling [19,20], factors that are precisely known to act upstream of *Krox20* in r3 [33–39]. It is therefore possible that elements C and NE cooperate in a redundant manner (S4 Fig) and further analyses will be required to determine whether this is indeed the case. Several examples of redundancy have been reported for enhancers governing the expression of developmental genes [3,40–42]. Redundant enhancers, or shadow enhancers, often share the same regulatory logic, since their activities have to be, at least in part, concomitant [43]. The analysis of the characteristics of the NE enhancer should reveal whether it depends on the same TFs as element C. In a few cases of redundant *cis*-acting elements that have been investigated in detail so far, it has been shown that redundancy provides robustness to the system and that, in specific genetic or environmental conditions, phenotypes can be revealed in absence of one of the elements [44].

Dual function of element C

Our study also revealed an unexpected function of element C: it is required for autoregulation, which was thought to be only dependent on element A. Using genetic approaches, we showed that an interaction must occur in *cis* between the two elements and that it is permanently required during the autoregulatory phase. ATAC-seq experiments indicated that element C is likely to act by modulating the accessibility of element A. Therefore, simultaneous to its classical enhancer function, element C performs another type of activity, which we propose to name enhancer potentiator. Potentiator characteristics (asymmetrical interaction, permanent requirement, long-range) clearly distinguishes this function from classical enhancer cooperative activities (additive, synergistic) and possibly from other hierarchical logic modes of interactions [3,45].

At this point, it is not known whether this additional enhancer potentiator function of element C, which is functionally distinct from its classical enhancer activity, is dependent on enhancer activity. Several recent studies have shown that enhancers can be transcribed and that the products of this transcription can act locally in *cis* to promote the expression of the target gene [46]. It is possible that such a mechanism could be involved in the potentiator activity of element C. It is interesting to note that Nlz factors, which are likely to repress *Krox20* expression by reducing the accessibility of the KROX20 protein to element A [18], are also involved in repressing element C [32]. It is therefore possible that Nlz factors only indirectly affect accessibility of KROX20 on element A, by modulating the potentiator activity of element C.

In r5, *Krox20* is almost normally expressed in the absence of element C, suggesting that cooperation between elements A and C is not essential in this rhombomere to support element

A activity. It is possible that element B, which is likely to constitute the major initiator element in r5 and physically interacts with element A (Fig 3B) performs a dual function similar to element C and potentiates the activity of element A in this rhombomere, in addition to its classical enhancer activity.

A security lock on autoregulatory elements

Analyses by Hi-C in embryonic stem cells [11] and by 4C-seq in various embryonic samples (this report) revealed the existence of a regulatory neighbourhood that contains interactions between the *Krox20* promoter and element A, irrespective of the considered stages or regions of the embryo (Fig 3). This chromatin configuration might therefore constitute a permissive environment for the activation of the autoregulatory loop. Such a situation, in which an autoregulatory element might be only dependent on the presence of its cognate TF for its activity, would be rather dangerous for an organism, as any transcription of the TF gene, even illegitimate, could end up activating the feedback loop and lead to high-level and long-term expression of the gene. This danger would be increased by the stochastic nature of the activation of the autoregulatory loop [19]. Furthermore, developmental genes may possess several positive autoregulatory enhancers that have to function at specific stages or in different tissues. This is exemplified by the *Krox20* gene, that has at least three of such elements and that are differentially active in r3/r5, the neural crest and developing bones [13,28]. Therefore, mechanisms must exist as well to prevent the inappropriate activation of these elements in the other embryonic tissues. Indeed, we have shown that ectopic expression of exogenous *Krox20* in the entire zebrafish embryo only leads to activation of the autoregulatory loop in the r2-r6 region of the hindbrain [18].

The introduction of an enhancer potentiator within a positive feedback loop may constitute an efficient prevention (safety lock) against inappropriate activation of autoregulatory elements. According to our model (Fig 5A), in the ground state, the autoregulatory element (element A in the case of *Krox20*) is locked in a configuration that is not accessible to its cognate TF and therefore unable to activate transcription, despite its possible interaction with the promoter. This lock will be released when the potentiator function is provided by a second *cis*-acting element (element C). It is interesting to note that in transgenic constructs, element A is able to activate a promoter in the absence of element C *in cis*. This difference in behaviour might be explained by two types of reasons: in transgenic constructs, element A is very close to the promoter, in contrast to the endogenous enhancer, located far upstream to the promoter; the chromatin context of a transgene is likely to be different, possibly more permissive than that of a highly regulated endogenous locus.

In the endogenous locus, if the unlocker element is also responsible for the early accumulation of the cognate TF, through its classical enhancer activity, the autoregulatory element will be placed under the control of the upstream factors regulating the initial expression of the TF. In this way, the asymmetrical cooperation between the two *cis*-acting elements becomes essential for establishing the appropriate specificity of the positive feedback loop. As indicated in the model, such a feedback loop can be broken by mutation of either the TF or of any of the two *cis*-acting elements (Fig 5B).

Materials and methods

Ethics statement

All animal experiments were performed in accordance with the guidelines of the council of European Union directive n°2010/63/UE and were approved by the "Comité d'éthique pour l'expérimentation animale Charles Darwin" (Project Number: CE5/2012/120).

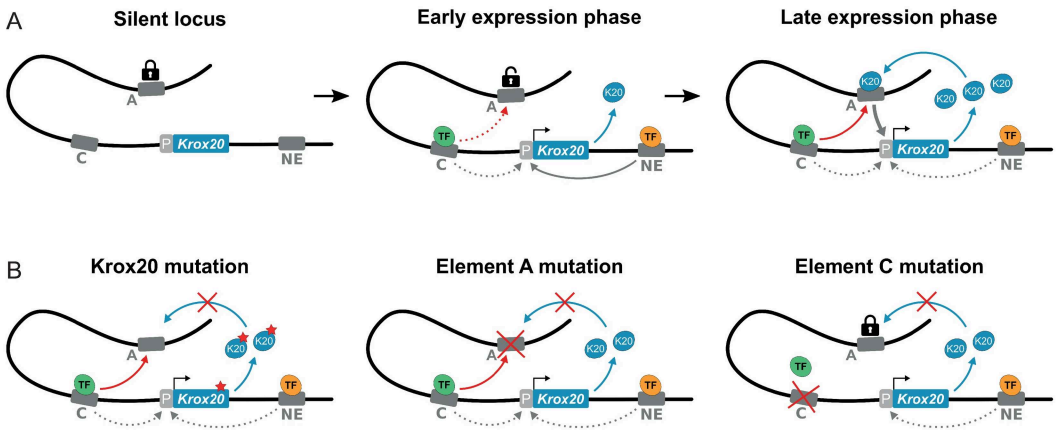


Fig 5. A model for *Krox20* regulation and the dual function of element C. (**A**) Schematic representation of the regulation of *Krox20* in r3. Three situations are envisaged in wild type embryos. Left: silent locus. If both element C and the new enhancer (NE) are inactive, no expression occurs. Middle: early expression phase. At this stage, elements C and NE have been bound by their respective transcription factors and have initiated the expression of *Krox20* via their classical enhancer functions. Nevertheless, element C has not yet been unlocked (decompacted) element A and/or the concentration of the KROX20 protein has not reached high enough levels to allow the establishment of a stable feedback loop with a significant probability. Right: late expression phase. Via its potentiator function, element C has unlocked element A, which can bind the KROX20 protein, which has now accumulated at a high enough concentration. Activation of enhancer A establishes the autoregulatory loop. **(B)** Three mutations that disrupt the positive feedback loop are presented at late expression phase. Left: mutation of the KROX20 protein preventing binding to element A. Middle: mutation of element A, preventing the binding of the KROX20 protein. Right: mutation of element C, preventing unlocking of element A.

https://doi.org/10.1371/journal.pgen.1006903.g005

Mouse lines and situ hybridization

All mouse lines were maintained in a mixed C57BL6/DBA2 background. We used the following alleles: $Krox20^{Cre}$ [23] and $Krox20^{\Delta A}$ [18]; the mouse $Krox20^{Cflox}$ line was generated at the Institut Clinique de la Souris (Illkirch, France) by homologous recombination in ES cells; the $Krox20^{\Delta C}$ allele was obtained as described in Fig 1, using the maternally expressed PGK-Cre transgene as deletor [47]. In situ hybridizations were performed on whole embryos as previously described [48], with the following digoxigenin-labelled riboprobe: *Krox20* [49] and *EphA4* [50].

ChIP-seq

ChIP-seq experiments were performed as previously described [51]. Briefly, 10 embryos at E9.5 or 20 embryos at E8.5 were dissected in cold PBS. Cell suspensions were obtained by passing them through a 21G needle fitted onto a 5ml syringe. The cells were cross-linked with 1% formaldehyde for 10 min and washed twice in PBS, 1 mM PMSF, 1 X PIC (Protease Inhibitor Cocktail). Sonication was performed on a Covaris S220 using the following programme: duty factor = 10/5, peak incident power = 140, cycles per burst = 200 during 600/480 seconds. 5–10 μg of chromatin was used for each IP using 3 μg of the following antibodies: anti H3K4me1 (C15410037, Diagenode) and anti H3K27ac (ab4729, Abcam) in RIPA buffer. The libraries were prepared with the MicroPlex Library Preparation kit (Diagenode, E8.5 embryos) and with the NEXTflex ChIP-Seq Kit (Bioo Scientific, E9.5 embryos). ChIP Seq experiments involved biological duplicates.

Sequencing was performed on multiplexed samples using 50 bp single-end reads on an Illumina HiSeq system (E9.5 embryos) or using 42 bp paired-end reads on an Illumina NextSeq (E8.5 embryos) according to the manufacturer's specifications. Chip-seq data were analysed as follows, using Eoulsan [52] with extended support for ChIP-seq workflows (https://github.

com/GenomicParisCentre/eoulsan/tree/branch-chip-seq). First, reads were filtered out when they would not pass Illumina filters (module filterreads with option illuminaid). Files corresponding to technical replicates were merged (module technicalreplicatemerger, with option format = fastq), followed by trimming of the reads using Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; version 0.4.1 in module trimadapt, with cutadapt v1.8.1 and options: length = 41, quality = 20, error = 0.1, stringency = 8, and is.paired = yes for E8.5 data and is.paired = no for E9.5 data). Mapping was performed using STAR [53] (version 2.4.0k in module mapreads with mapper.arguments = "—outFilterMultimapNmax 1—outFilterMismatchNmax 999—outFilterMismatchNoverLmax 0.06—alignIntronMax 1—alignEndsType EndToEnd—alignMatesGapMax 2000—outSAMunmapped Within"). Further filters were applied to the mapped reads before conversion into BAM (module filtersam with removeunmapped = true; module sortsam; module rmdupgalax with is_sort = true; module sam2bam with compression.level = 5). BIGWIG files were created from the resulting BAM files using deepTools' bamCoverage [54] (version 1.6.0, with options:—binSize 1—normalizeTo1x 200000000—fragmentLength 200—outFileFormat bigwig).

4C-seq

4C-seq libraries were constructed as previously described [55] with small adjustments. 35 E9.5 embryos (for each biological duplicate) or 250 E8.5 embryos were dissected in cold PBS, followed by dissociation in collagenase type I (Gibco). DpnII (New England Biolabs, Ipswich, MA) was used as the primary restriction enzyme and NlaIII (New England Biolabs) was used as the secondary restriction enzyme. For each viewpoint, up to 600 ng of each 4C-seq library were amplified using 16 individual PCR reactions with inverse primers including Illumina adapter sequences (S1 Table). Illumina sequencing was performed on multiplexed samples, containing PCR amplified material of up to 7 viewpoints, using 100 bp single-end reads on the Illumina HiSeq system according to the manufacturer's specifications at the iGE3 Genomics Platform of the University of Geneva (Switzerland). Reads were sorted, aligned, translated to restriction fragments and smoothed (11 fragments running mean) using the 4C-seq pipeline of the BBCF HTSstation [56] according to ENSEMBL Mouse assembly NCBIM37 (mm9). For the calculation of the 4C-seq signal distribution, reads were normalized to the entire chromosome 10, based on an approach adapted from [http://www.ncbi.nlm.nih.gov/pubmed/25959774]. For visualizations, smoothed 4C-seq reads were normalized to the 5 TADs surrounding the *Krox20* locus (chr10:62,880,000–70,720,000).

Position of mouse TADs in ES cells and associated 40 kb normalized Hi-C matrices [11] were obtained from http://promoter.bx.psu.edu/hi-c/download.html. Directionality indexes were calculated as described using tools described previously [7]. Interaction matrices are visualized using standard cut-offs.

ATAC-seq

ATAC experiments were performed according to Buenrostros and colleagues [31], using a homemade transposome [57]. 7–8 embryos at 10-12s were dissected in cold PBS for each genotype and cells were mechanically dissociated. Biological duplicates were performed for ATAC experiments. Cells were lysed before transposition using 1 μl of transposome and purified using a Qiagen MinElute Kit with 10 μl of Elution Buffer. Transposed DNA was amplified by PCR as previously described [57] and quantified by qPCR using 5 μl of PCR products. The number of additional cycles was determined by plotting linear Rn versus cycle and corresponded to one third of the maximum fluorescence intensity. The remaining PCR products (45 μl) were treated with the additional number of cycles. The final product was purified with

Qiagen PCR Cleanup Kit and eluted in 20 μl Elution Buffer. Sequencing was performed on multiplexed samples using 42 bp paired-end reads on an Illumina NextSeq according to the manufacturer's specifications. For computational analysis, paired-end reads were mapped onto the mouse genome assembly mm9, using STAR (outFilterMultimapNmax 1; outFilterMismatchNmax 999; outFilterMismatchNoverLmax 0.06; alignIntronMax 1; alignEndsType EndToEnd; alignMatesGapMax 2000). Duplicates reads were removed using Picard (http://picard.sourceforge.net) (MarkDuplicates, REMOVE_DUPLICATES = true). To consider only fragments coming from transcription factors protected DNA (and not from nucleosomes), only fragment with size lower than 100 bp were kept. Bigwig tracks were obtained using DeepTools BamCoverage (1.5.9.1). Peak calling was performed using MACS2 (2.1.0.20140616), using a q-value< = 0.01 threshold (other parameters as default). For quantification, we first defined a set of non-redundant enriched regions for all samples by taking the union of all peak summits from both replicates of all samples, grouped together all summits distant from less than 50 bp, and for each group kept only the summit with the lowest q-value (calculated by MACS2). We then quantified the signal at all summits in each sample by counting the number of fragments (using the R bioconductor package csaw, v. 1.0.7). Normalisation and statistical analysis were performed using the bioconductor DESeq2 package (1.6.3). Library size factors were calculated on fragment counts in genomic bins of 10 kb. Comparison between wild type, *Krox20*$^{\Delta C/\Delta C}$ and *Krox20*$^{\Delta A/\Delta A}$ embryos was performed using negative binomial Wald Test (DESeq2).

Accession codes

The data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE94716 and is available at the following address: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94716

Supporting information

S1 Fig. Element C is required for normal development of the r3 territory. (A) Sequence of the deleted element C and associated genomic coordinates (mouse genome assembly mm9). This sequence corresponds to the cloned sequence of the mouse element C described in [13]. **(B)** *In situ* hybridization for *EphA4* mRNA performed on control (*Krox20*$^{+/\Delta C}$) and homozygous (*Krox20*$^{\Delta C/\Delta C}$) mouse embryos at the indicated somite stages. Note the reduction of the size of r3 in the homozygous mutant. Rhombomeres positions are indicated on the left of each embryo.
(TIF)

S2 Fig. The Krox20 regulatory neighborhood includes its regulatory elements but not the *Nrbf2* gene. (A) The left boundary of the *Krox20* TAD, as previously determined from Hi-C in ES cells [11], spreads out over a transition zone with a low Directionality Index of Hi-C interactions in both ES cells and adult cortex ([11]). Hi-C in ES cells and TADs as defined in [11] or by our additional analysis ('with CTCF', this figure) are indicated above. Dashed lines in the graphs demarcate the boundaries between the -2, -1, *Krox20* and +1 TADs. The transition zone between the -1 and *Krox20* TADs is highlighted in grey. CTCF ChIP-seq data in E14.5 mouse brain (ENCODE, [58]) and genes are indicated below. Orange bars and gene names above pinpoint 4C-seq viewpoints. **(B)** Coordinates of the different regions used for the analysis of the distribution of 4C-seq signal. **(C)** Normalized distribution of 4C-seq signal for the *Nrbf2*, Element A and *Krox20* viewpoints using previously determined TAD boundaries ([11], left) or after the repositioning of the TAD boundary to the cluster of CTCF binding sites

between the *Nrbf2* gene and Element A (right). The ratio of signal between the -1 and *Krox20* TADs (average of 2 replicates) is shown above for each viewpoint. When the previously determined TAD boundary is used (left), the 4C-seq signal of the *Nrbf2* viewpoint is almost equally distributed over the -1 and *Krox20* TADs, whereas a much more discrete distribution is observed when the cluster of CTCF sites is used (right). In contrast, the nearby Element A viewpoint always restricts its strongest signal to the *Krox 20* TAD, similar to the associated but much more distant *Krox20* gene.
(TIF)

S3 Fig. Chromatin state and accessibility within the *Krox20* locus. (A-C) ChIP-seq was performed for the H3K27ac mark on wild type E8.5 embryo heads (light blue) in duplicates and only one replicate is shown. ATAC-seq was performed on dissected regions (r3, r5 and a more posterior region ("post"; see text) from wild type (light blue), *Krox20*$^{\Delta C/\Delta C}$ (green) and *Krox20*$^{\Delta A/\Delta A}$ (red) E8.5 embryos. Genes, *cis*-regulatory elements (orange) and a genomic scale are indicated at the top. CTCF ChIP-seq data in E14.5 mouse brain (ENCODE) are indicated below each panel.
(TIF)

S4 Fig. Dynamics of enhancer activity of the new mouse element (NE) in zebrafish. A zebrafish transgenic line *Tg(NE:gfp)*, carrying a construct in which NE is linked to the gfp gene driven by a minimal promoter, was analysed by double ISH with *krox20* (orange) and *gfp* (purple) probes at 3s and 10s stages.
(TIF)

S1 Table. Sequences of 4C-seq primers including Illumina adaptors.
(PDF)

S1 File. Supplementary methods.
(PDF)

Acknowledgments

We are grateful to Benoît Noël for help with the UCSC trackhubs, to Melody Matelot for help with the 4C-experiments and to Michel Wassef for critical comments on the manuscript and we thank the genomic platform of the IBENS. 4C-seq data analysis was performed at the Vital-IT (http://www.vital-it.ch) computing center of the Swiss Institute of Bioinformatics using tools developed by the BBCF (http://bbcf.epfl.ch).

Author Contributions

Conceptualization: Elodie Thierion, Johan Le Men, Patrick Charnay, Pascale Gilardi-Hebenstreit.

Data curation: Samuel Collombet, Céline Hernandez, Morgane Thomas-Chollier.

Formal analysis: Elodie Thierion, Samuel Collombet, Céline Hernandez, Morgane Thomas-Chollier, Pascale Gilardi-Hebenstreit.

Funding acquisition: Patrick Charnay, Pascale Gilardi-Hebenstreit.

Investigation: Elodie Thierion, Johan Le Men, Fanny Coulpier, Patrick Torbey.

Methodology: Elodie Thierion.

Project administration: Daan Noordermeer, Patrick Charnay, Pascale Gilardi-Hebenstreit.

Resources: Morgane Thomas-Chollier, Daan Noordermeer, Patrick Charnay.

Software: Samuel Collombet, Céline Hernandez, Morgane Thomas-Chollier.

Supervision: Morgane Thomas-Chollier, Patrick Charnay, Pascale Gilardi-Hebenstreit.

Validation: Elodie Thierion, Patrick Torbey, Morgane Thomas-Chollier, Daan Noordermeer.

Visualization: Elodie Thierion, Johan Le Men, Patrick Torbey, Daan Noordermeer.

Writing – original draft: Elodie Thierion, Patrick Charnay, Pascale Gilardi-Hebenstreit.

Writing – review & editing: Elodie Thierion, Daan Noordermeer, Patrick Charnay, Pascale Gilardi-Hebenstreit.

References

1. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell. 1981; 27(2 Pt 1):299–308.
2. Maniatis T, Goodbourn S, Fischer J a. Regulation of inducible and tissue-specific gene expression. Science. 1987 Jun 5; 236(4806):1237–45. PMID: 3296191
3. Long HK, Prescott SL, Wysocka J. Review Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016; 167(5):1170–87. https://doi.org/10.1016/j.cell.2016.09.018 PMID: 27863239
4. Kvon EZ. Using transgenic reporter assays to functionally characterize enhancers in animals. Genomics. 2015; 106(3):185–92. https://doi.org/10.1016/j.ygeno.2015.06.007 PMID: 26072435
5. Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet. 2011 Apr; 43 (4):379–86. https://doi.org/10.1038/ng.790 PMID: 21423180
6. Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014; 24(3):390–400. https://doi.org/10.1101/gr.163519.113 PMID: 24398455
7. Dunipace L, Saunders A, Ashe H, Stathopoulos A. Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. Dev Cell. 2013; 26(5):536–43. https://doi.org/10.1016/j.devcel.2013.08.010 PMID: 24044892
8. Zaret KS, Carroll JS. Pioneer transcription factors: Establishing competence for gene expression. Genes Dev. 2011; 25(21):2227–41. https://doi.org/10.1101/gad.176826.111 PMID: 22056668
9. Gibcus JH, Dekker J. The hierarchy of the 3D genome. Mol Cell. 2013 Mar 7; 49(5):773–82. https://doi.org/10.1016/j.molcel.2013.02.011 PMID: 23473598
10. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 May 17; 485(7398):381–5. https://doi.org/10.1038/nature11049 PMID: 22495304
11. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 May 17; 485(7398):376–80. https://doi.org/10.1038/nature11082 PMID: 22495300
12. Chavrier P, Zerial M, Lemaire P, Almendral J, Bravo R, Charnay P. A gene encoding a protein with zinc fingers is activated during G0/G1 transition in cultured cells. EMBO J. 1988; 7(1):29–35. PMID: 3129290
13. Chomette D, Frain M, Cereghini S, Charnay P, Ghislain J. Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. Development. 2006 Apr; 133(7):1253–62. https://doi.org/10.1242/dev.02289 PMID: 16495311
14. Lumsden A, Krumlauf R. Patterning the vertebrate neuraxis. Science. 1996 Nov 15; 274(5290):1109–15. PMID: 8895453
15. Schneider-Maunoury S, Topilko P, Seitandou T, Levi G, Cohen-Tannoudji M, Pournin S, et al. Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. Cell. 1993 Dec 17; 75(6):1199–214. PMID: 7903221
16. Schneider-Maunoury S, Seitanidou T, Charnay P, Lumsden a. Segmental and neuronal architecture of the hindbrain of Krox-20 mouse mutants. Development. 1997 Mar; 124(6):1215–26. PMID: 9102308
17. Voiculescu O, Taillebourg E, Pujades C, Kress C, Buart S, Charnay P, et al. Hindbrain patterning: Krox20 couples segmentation and specification of regional identity. Development. 2001 Dec; 128 (24):4967–78. PMID: 11748134

18. Bouchoucha YX, Reingruber J, Labalette C, Wassef M a, Thierion E, Desmarquet-Trin Dinh C, et al. Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning. Mol Syst Biol. 2013 Sep 24; 9(690).
19. Labalette C, Bouchoucha YX, Wassef MA, Gongal PA, Le Men J, Becker T, et al. Hindbrain patterning requires fine-tuning of early krox20 transcription by Sprouty 4. Development. 2011 Jan; 138(2):317–26. https://doi.org/10.1242/dev.057299 PMID: 21177344
20. Labalette C, Wassef MA, Desmarquet-Trin Dinh C, Bouchoucha YX, Le Men J, Charnay P, et al. Molecular dissection of segment formation in the developing hindbrain. Development. 2015 Jan 1; 142 (1):185–95. https://doi.org/10.1242/dev.109652 PMID: 25516974
21. Wassef M a, Chomette D, Pouilhe M, Stedman A, Havis E, Desmarquet-Trin Dinh C, et al. Rostral hindbrain patterning involves the direct activation of a Krox20 transcriptional enhancer by Hox/Pbx and Meis factors. Development. 2008 Oct; 135(20):3369–78. https://doi.org/10.1242/dev.023614 PMID: 18787068
22. Seitanidou T, Schneider-Maunoury S, Desmarquet C, Wilkinson DG, Charnay P. Krox-20 is a key regulator of rhombomere-specific gene expression in the developing hindbrain. Mech Dev. 1997 Jul; 65(1–2):31–42. PMID: 9256343
23. Voiculescu O, Charnay P, Schneider-Maunoury S. Expression pattern of a Krox-20/Cre knock-in allele in the developing hindbrain, bones, and peripheral nervous system. Genesis. 2000 Feb; 26(2):123–6. PMID: 10686605
24. van de Werken HJG, Landan G, Holwerda SJB, Hoichman M, Klous P, Chachik R, et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods. 2012; 9(10):969–72. https://doi.org/10.1038/nmeth.2173 PMID: 22961246
25. Irving C, Nieto M a, DasGupta R, Charnay P, Wilkinson DG. Progressive spatial restriction of Sek-1 and Krox-20 gene expression during hindbrain segmentation. Dev Biol. 1996; 173(1):26–38. https://doi.org/10.1006/dbio.1996.0004 PMID: 8575627
26. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-Resolution Profiling of Histone Methylations in the Human Genome. Cell. 2007; 129(4):823–37. https://doi.org/10.1016/j.cell.2007.05.009 PMID: 17512414
27. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet. 2011 Jan; 12(1):7–18. https://doi.org/10.1038/nrg2905 PMID: 21116306
28. Ghislain J, Desmarquet-Trin-Dinh C, Gilardi-Hebenstreit P, Charnay P, Frain M. Neural crest patterning: autoregulatory and crest-specific elements co-operate for Krox20 transcriptional control. Development. 2003; 130(5):941–53. PMID: 12538520
29. Theil T, Frain M, Gilardi-Hebenstreit P, Flenniken A, Charnay P, Wilkinson DG. Segmental expression of the EphA4 (Sek-1) receptor tyrosine kinase in the hindbrain is under direct transcriptional control of Krox-20. Development. 1998 Feb; 125(3):443–52. PMID: 9425139
30. Wood HB, Episkopou V. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech Dev. 1999; 86(1–2):197–201. PMID: 10446282
31. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10(12):1213–8. https://doi.org/10.1038/nmeth.2688 PMID: 24097267
32. Labalette C, Wassef MA, Desmarquet-Trin Dinh C, Bouchoucha YX, Le Men J, Charnay P, et al. Molecular dissection of segment formation in the developing hindbrain. Development. 2015; 142(1):185–95. https://doi.org/10.1242/dev.109652 PMID: 25516974
33. Wiellette EL, Sive H. vhnf1 and Fgf signals synergize to specify rhombomere identity in the zebrafish hindbrain. Development. 2003; 130(16):3821–9. PMID: 12835397
34. Cordes SP, Barsh GS. The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. Cell. 1994 Dec 16; 79(6):1025–34. PMID: 8001130
35. Waskiewicz AJ, Rikhof H a, Moens CB. Reveals a Hindbrain Ground State. Dev Cell. 2002; 3:723–33. PMID: 12431378
36. Choe S-K, Vlachakis N, Sagerström CG. Meis family proteins are required for hindbrain development in the zebrafish. Development. 2002; 129(3):585–95. PMID: 11830560
37. McNulty CL, Peres JN, Bardine N, van den Akker WMR, Durston AJ. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. Development. 2005; 132 (12):2861–71. https://doi.org/10.1242/dev.01872 PMID: 15930115
38. Walshe J, Maroon H, McGonnell IM, Dickson C, Mason I. Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. Curr Biol. 2002; 12(13):1117–23. PMID: 12121619

39. Aragon F, Pujades C. FGF signaling controls caudal hindbrain specification through Ras-ERK1/2 pathway. BMC Dev Biol. 2009; 9:61. https://doi.org/10.1186/1471-213X-9-61 PMID: 19958530

40. Hong J-W, Hendrix D a, Levine MS. Shadow enhancers as a source of evolutionary novelty. Science. 2008; 321(5894):1314. https://doi.org/10.1126/science.1160631 PMID: 18772429

41. Kurokawa D, Kiyonari H, Nakayama R, Kimura-Yoshida C, Matsuo I, Aizawa S. Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. Development. 2004 Jul; 131(14):3319–31. https://doi.org/10.1242/dev.01220 PMID: 15201224

42. Lam DD, de Souza FSJ, Nasif S, Yamashita M, López-Leal R, Otero-Corchon V, et al. Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. PLOS Genet. 2015; 11(2):e1004935. https://doi.org/10.1371/journal.pgen.1004935 PMID: 25671638

43. Lagha M, Bothma JP, Levine M. Mechanisms of transcriptional precision in animal development. Trends Genet. 2012 Aug; 28(8):409–16. https://doi.org/10.1016/j.tig.2012.03.006 PMID: 22513408

44. Ludwig MZ, Manu, Kittler R, White KP, Kreitman M. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS Genet. 2011; 7(11).

45. Leddin M, Perrod C, Hoogenkamp M, Ghani S, Assi S, Heinz S, et al. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood. 2011; 117(10):2827–38. https://doi.org/10.1182/blood-2010-08-302976 PMID: 21239694

46. Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet. 2016; 17(4):207–23. https://doi.org/10.1038/nrg.2016.4 PMID: 26948815

47. Lallemand Y, Luria V, Haffner-Krausz R, Lonai P. Maternally expressed PGK-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. Transgenic Res. 1998; 7(2):105–12. PMID: 9608738

48. Giudicelli F, Taillebourg E, Charnay P, Gilardi-Hebenstreit P. Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. Genes Dev. 2001 Mar 1; 15(5):567–80. https://doi.org/10.1101/gad.189801 PMID: 11238377

49. Wilkinson DG, Bhatt S, Chavrier P, Bravo R, Charnay P. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature. 1989 Feb 2; 337(6206):461–4. https://doi.org/10.1038/337461a0 PMID: 2915691

50. Gilardi-Hebenstreit P, Nieto MA, Frain M, Mattéi MG, Chestier A, Wilkinson DG, et al. An Eph-related receptor protein tyrosine kinase gene segmentally expressed in the developing mouse hindbrain. Oncogene. 1992 Dec; 7(12):2499–506. PMID: 1281307

51. Vitobello A, Ferretti E, Lampe X, Vilain N, Ducret S, Ori M, et al. Hox and Pbx Factors Control Retinoic Acid Synthesis during Hindbrain Segmentation. Dev Cell. 2011; 20(4):469–82. https://doi.org/10.1016/j.devcel.2011.03.011 PMID: 21497760

52. Jourdren L, Bernard M, Dillies MA, Le Crom S. Eoulsan: A cloud computing-based framework facilitating high throughput sequencing analyses. Bioinformatics. 2012; 28(11):1542–3. https://doi.org/10.1093/bioinformatics/bts165 PMID: 22492314

53. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

54. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42(W1):187–91.

55. Matelot M, Noordermeer D. Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq). 2016;223–41.

56. David FPA, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y, et al. HTSstation: A web application and open-access libraries for high-throughput sequencing data analysis. PLoS One. 2014; 9(1).

57. Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014; 24(12):2033–40. https://doi.org/10.1101/gr.177881.114 PMID: 25079858

58. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012; 488(7409):116–20. https://doi.org/10.1038/nature11243 PMID: 22763441

Résumé

Les cellules immunitaires proviennent d'un ensemble commun de cellules souches hématopoïétiques qui se différencient hiérarchiquement en lignées myéloïdes et lymphoïdes. Ce processus est étroitement régulé par un réseau entrelacé de facteurs de transcription et de régulateurs épigénétiques, qui contrôlent l'activation et la répression des gènes impliqués.

Les travaux récents sur la reprogrammation cellulaire ont montré que certaines protéines peuvent reprogrammer des cellules différenciées, comme le facteur de transcription C/EBPa qui peut induire la trans-differenciation de cellules B en macrophages. De plus, une courte induction de Cebpa suivie de l'expression des quatre facteurs de transcription Oct4-Sox2-Klf4-cMyc permet une reprogrammation extrêmement rapide en cellules pluripotentes.

Afin de déchiffrer le réseau de régulation moléculaire contrôlant la spécification et la reprogrammation des cellules immunitaires, J'ai combiné différentes méthodes à haut débit pour analyser l'expression des gènes et leur régulation épigénétique, et ce au court de la reprogrammation des cellules B.

J'ai découvert des interactions entre différents facteurs de transcription, au niveau des régions régulatrices de gènes des différents programmes génétiques impliqués (lymphoide, myeloide et pluripotence), et j'ai identifié des facteurs régulant l'état de la chromatine également impliqués dans la reprogrammation (notamment Lsd1, Hdac1, Brd4 et Tet2).

Enfin, J'ai intégré ces données dans un modèle dynamique du réseau moléculaire régulant la spécification des cellules B et des macrophages à partir de progéniteurs multipotents. J'ai utilisé à la fois des méthodes analytiques (analyse des états stables) et des simulations (simulations logiques asynchrones, chaînes de Markov à temps continu) pour étudier in silico la différenciation et la reprogrammation cellulaire.

Ces analyses ont révélés des régulations transcriptionelles encore inconnues, que nous avons pu confirmer expérimentalement. Nous avons ainsi obtenu une meilleure compréhension des circuits de régulation contrôlant le destin cellulaire.

Mots Clés

Cellules souches, réseau de régulation, reprogrammation cellulaire, modélisation dynamique.

Abstract

Immune cells arise from a common set of hematopoietic stem cells, which differentiate hierarchically into the myeloid and lymphoid lineages. This process is tightly regulated by an intertwined network of transcription and epigenetic factors, which control both the activation and repression of gene programs, to ensure cell commitment.

However, recent work on cellular reprogramming has shown that the ectopic expression of some specific factors can enforce the trans-differentiation of committed cells. The transcription factor C/EBPa can induce the reprogramming of B-cells into macrophages. Furthermore, a pulse of Cebpa expression in B cells followed by the expression of the four transcription factors Oct4-Sox2-Klf4-cMyc leads to an extremely fast and efficient reprogramming into induced pluripotent stem cells.

Despite the many data we have on the molecular mechanisms by which specific genes are regulated, we are still lacking a global understanding of the interplay between these factors and how they control cell fate.

In order to decipher the molecular regulatory network controlling immune cell specification and their reprogramming, I have combined a variety of high-throughput methods to measure changes in gene expression and epigenetic regulation during B cells reprogramming.

I have revealed the interplay between different transcription factors at enhancers regulating genes of the different programs (B cells, macrophages and pluripotent cells) and identified epigenetic regulators forming complexes and controlling enhancers activities (such as Lsd1, Hdac1, Brd4 and Tet2) and consequently regulating cell fate.

Finally, I integrated these data together with published data, in a computational model of the regulatory network controlling the specification of B-cells and macrophages from multipotent progenitors. I used both analytic tools (stable states analysis) and simulations (logical asynchronous simulations, continuous time Markov chains) to study in silico differentiation and reprogramming.

These analyses have revealed previously unknown transcriptional regulations, which we confirmed experimentally, and allowed us to get a better understanding of the regulatory circuits controlling cell fate commitment.

Keywords

Stem cells, regulatory network, cellular reprogramming, dynamical modeling.