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.

www.pnas.org/cgi/doi/10.1073/pnas.1610622114

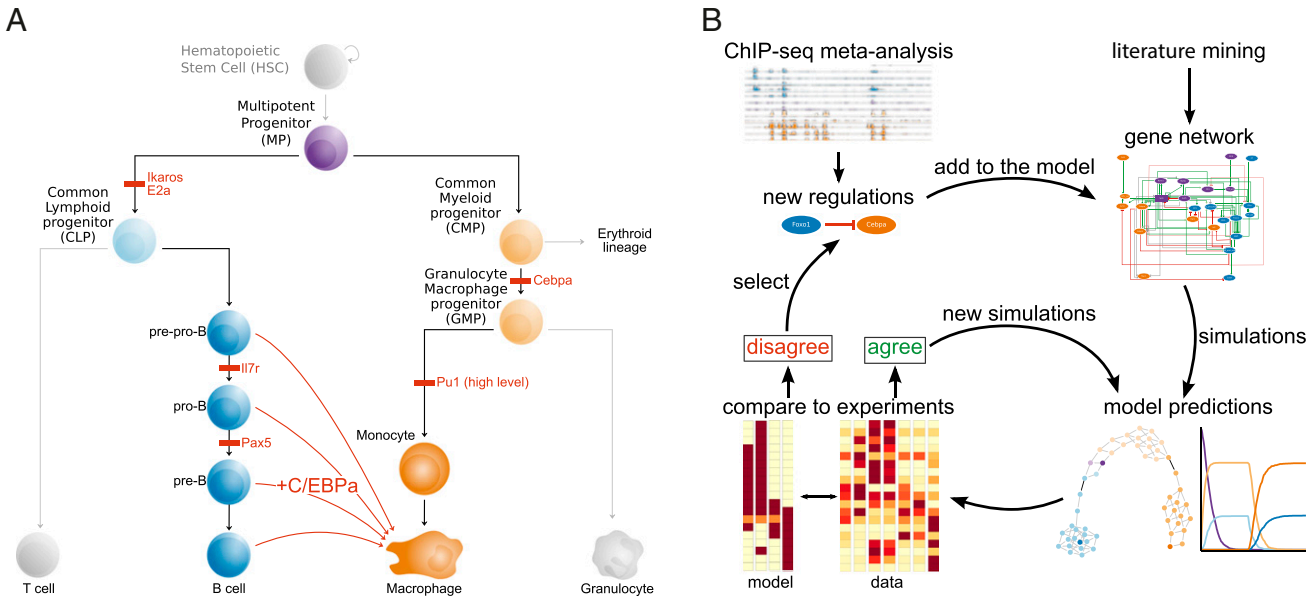

Fig. 1. (*A*) Schematic representation of hematopoietic cell specification. Genes in red are required for progression at the corresponding steps. C/EBPα-induced transdifferentiation is indicated by red arrows from B-lineage cells to macrophages. (*B*) Iterative modeling workflow. A model is first built 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

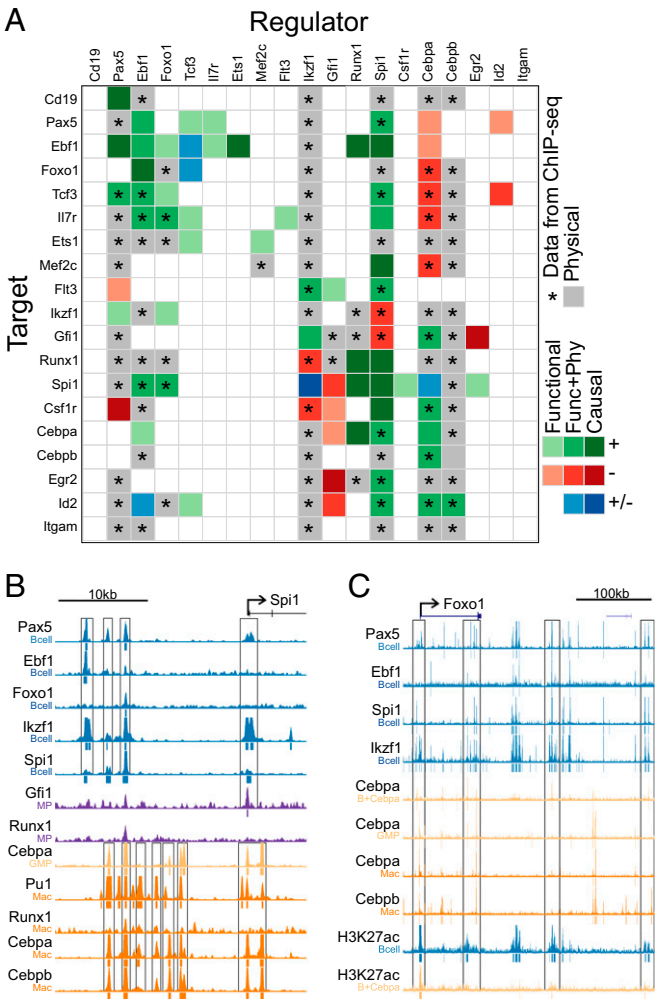

Fig. 2. (*A*) Heatmap showing the regulations inferred from the literature and from ChIP-seq meta-analysis. (*B*) ChIP-seq signals and peaks (under signal) at the *Spi1* locus. Black frames indicate known enhancers (24). The vertical axes represent reads per million (RPM) (maximum: 2 RPM for Ebf1 and 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

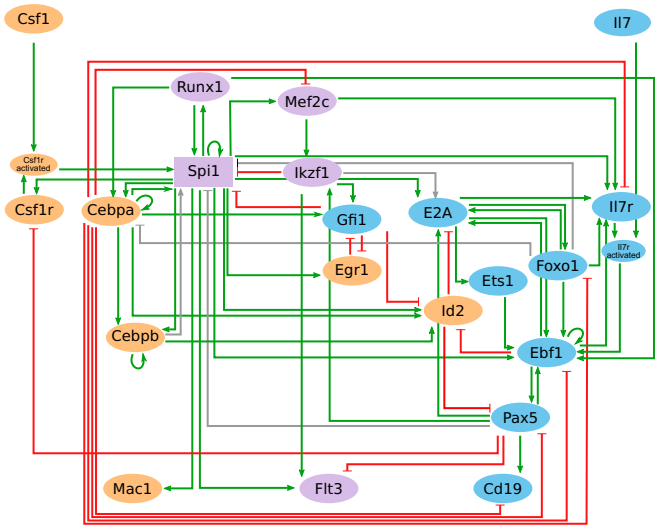

Fig. 3. A regulatory graph depicting the interactions inferred from the literature and ChIP-seq meta-analyses. Nodes represent genes (except for CSF1r_act and Il7r_act, which represent the activated forms of cytokine receptors), and arrows denote regulatory interactions. Orange nodes represent factors 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.

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

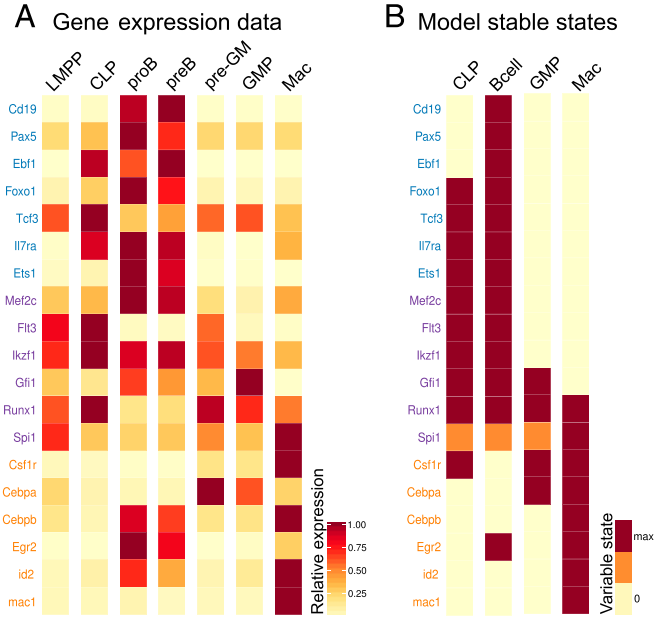

Fig. 4. (*A*) Gene-expression values (microarrays) in lymphoid/myeloid progenitors (LLPP), B cells, and macrophages (Mac) (29). These values are relative to the highest expression value. (*B*) Context-dependent stable states computed for the model. A yellow cell denotes the inactivation of the corresponding 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*.

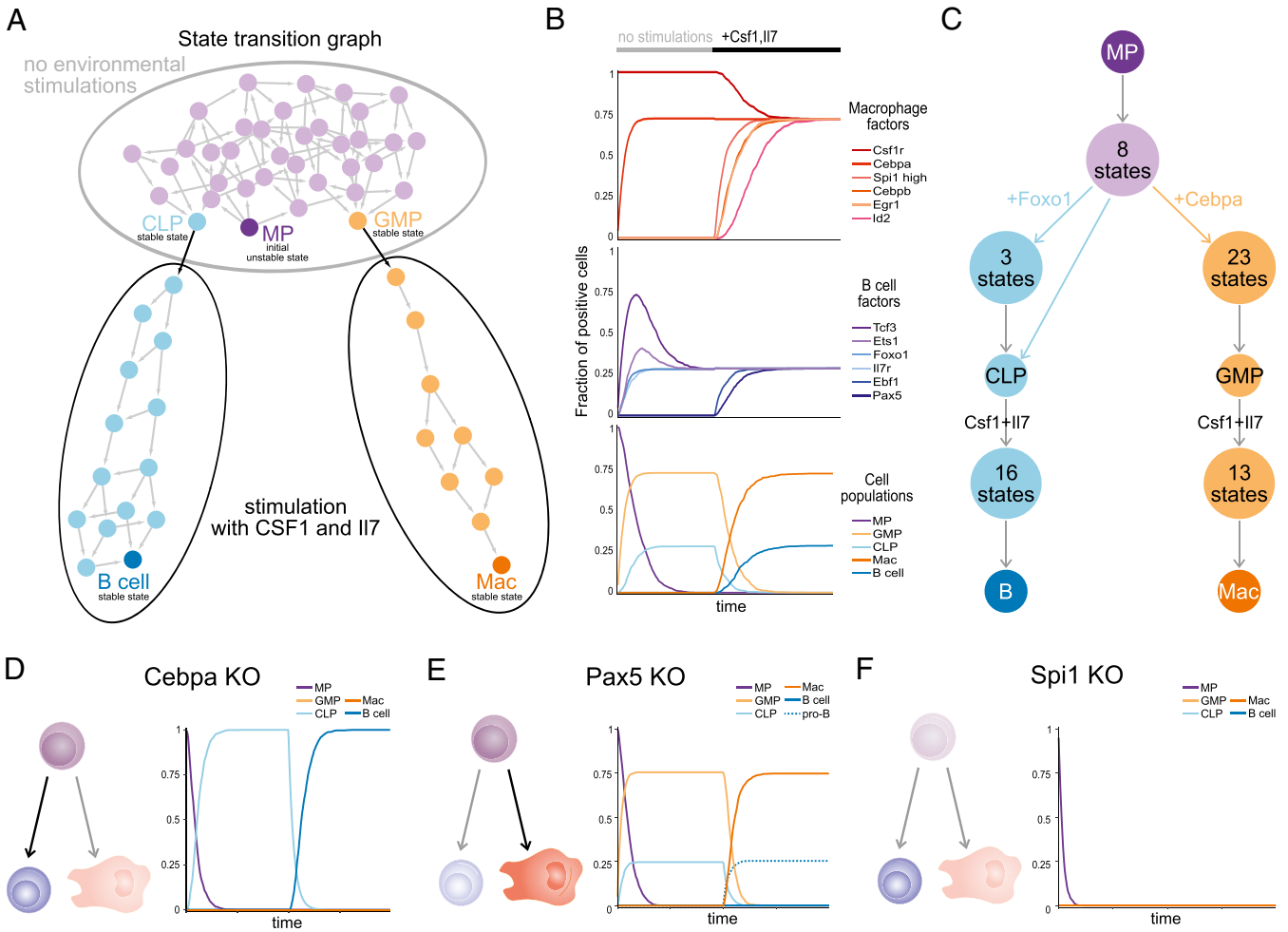

Fig. 5. (*A*) State transition graph generated by simulating the model starting from the unstable MP state in the absence of cytokine (*Upper*) and after the addition of CSF1 and IL7 (*Lower Left* and *Lower Right*). Nodes denote states, and arrows represent transitions between states. (*B*) Stochastic 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

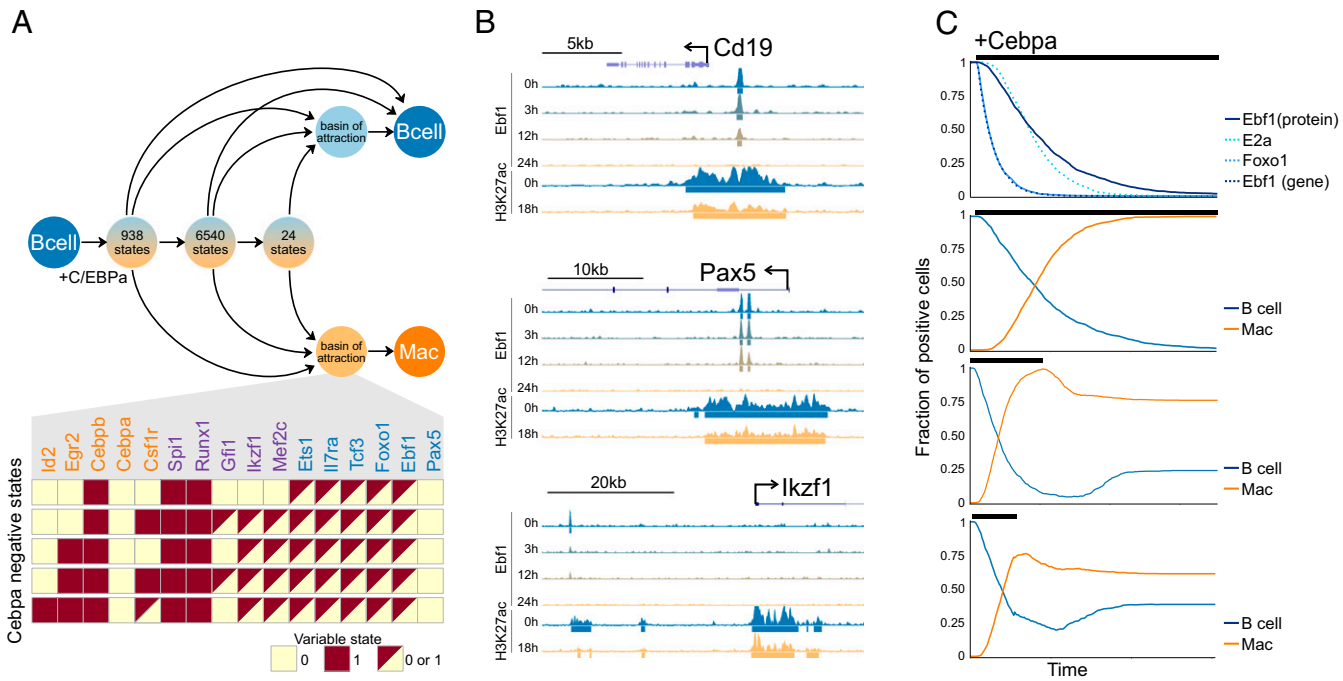

Fig. 6. (*A*, *Upper*) Hierarchical transition graph of the simulation of B-cell transdifferentiation upon transient C/EBPα expression, taking into account all possible C/EBPα pulse durations. Nodes represent clusters of states, and arcs correspond to transitions between these clusters. (*Lower*) Cebpa$^-$ 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.

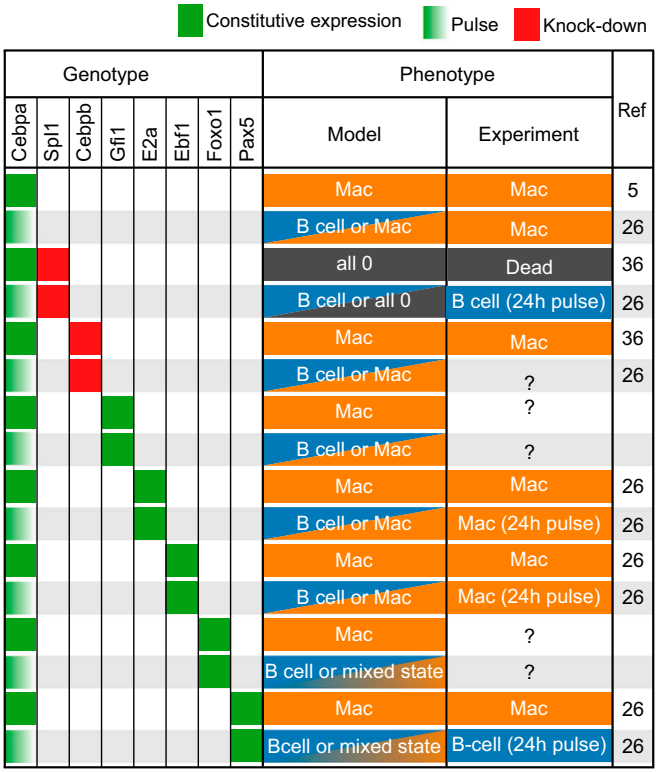

Fig. 7. Table summarizing the impact of selected perturbations (knockin or knockout) on B-cell transdifferentiation into macrophages (Mac) upon either a permanent or a transient induction of C/EBPα. Orange boxes represent macrophages, blue boxes B cells, gray boxes all 0 stable states or cell death.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