Logical modeling of dendritic cells *in vitro* differentiation from human monocytes unravels novel transcriptional regulatory interactions

Karen J. Nuñez-Reza1,
Aurélien Naldi2,
Arantza Sanchéz-Jiménez1,
Ana V. Leon-Apodaca1,
M. Angélica Santana3,
Morgane Thomas-Chollier2,
Denis Thieffry2*,
Alejandra Medina-Rivera1*.

1) Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, México.
2) Computational Systems Biology team, Institut de Biologie de l'Ecole normale supérieure, Inserm, CNRS, Université PSL, Paris, France.
3) Centro de Investigación en Dinámica Celular (IICBA), Universidad Autónoma del Estado de Morelos, Cuernavaca, México.
*Corresponding author

Abstract (Up to 200 words)

Dendritic cells are the major specialized antigen-presenting cells, thereby connecting innate and adaptive immunity. Because of their role in establishing adaptive immunity, they have been used as targets for immunotherapy. Monocytes can differentiate into dendritic cells *in vitro* in the presence of colony-stimulating factor 2 (CSF2) and interleukin 4 (IL4), activating four signalling pathways (MAPK, JAK/STAT, NFKB, and PI3K). However, the transcriptional regulation responsible for dendritic cell differentiation from monocytes (moDCs) remains unknown. By curating scientific literature on moDCs differentiation, we established a preliminary logical model that helped us identify missing information for the activation of genes responsible for this differentiation, including missing targets for key transcription factors (TFs). Using ChIP-seq and RNA-seq data from the Blueprint consortium, we defined active and inactive promoters, together with differentially expressed genes in monocytes, moDCs, and macrophages (which correspond to an alternative cell fate). We then used this functional genomic information to predict novel targets for the identified TFs. We established a second logical model integrating this information, which enabled us to recapitulate the main established facts regarding moDCs differentiation. Prospectively, the resulting model should be useful to develop novel immunotherapies based on moDCs regulatory network.

Keywords

Dendritic cells, differentiation, Logical modeling, Regulatory networks

Introduction

Dendritic cells (DCs) are the main antigen-presenting cells (1), whose role is to activate the innate immune response, by presenting antigens to the naïve lymphocytes in order to initiate the immune response (2). Dendritic cells have been used in immunotherapies for their capacity to activate the adaptive immune response, in particular, dendritic cells derived from monocytes (moDCs) (3), as monocytes circulate in peripheral blood, they are easily accessible. Furthermore, there is an established protocol for moDCs differentiation (3).

The protocol to differentiate monocytes to moDCs consists in cultivating monocytes with colony-stimulating factor 2 (CSF2) and interleukin 4 (IL4) (4). When only IL-4 is used, monocytes are activated, while treatment with CSF2 results in their differentiation into macrophages. Only the combined stimuli results in DC differentiation, pointing to the importance of signalling interplay for the differentiation of moDCs. CSF2 signalling leads to the activation of NFKB, MAPK, PI3K, JAK2, and STAT5 (5,6). IL4 signalling activates the JAK/STAT pathway, while JAK1 activates STAT3 and JAK3 activates STAT6 (7). There are some well-known transcription factors (TFs) ultimately activated by CSF2 and IL4 signalling pathways, but presumably, only a fraction of the target genes participating in moDCs differentiation have been reported (6,8,9).

A good way to integrate multiple signalling pathways into a comprehensive regulatory network and check its coherence consists of developing a dynamical model (10). As most of the available data are qualitative, it is natural to use a qualitative approach to build such a model. Logical models are well suited to represent this qualitative data and have been proposed for various similar processes (11–13). This qualitative formalism relies on the construction of a regulatory graph,

whose nodes denote molecular components, while arcs denote (positive, negative, or dual) regulatory interactions. In the simplest cases, nodes are associated with Boolean variables, which take the values 0 or 1, denoting absence/inactivity or presence/activation, respectively (14). Logical models are usually derived based on a careful manual curation of relevant scientific literature; but they can also be enriched using other sources of information, such as high-throughput sequencing data (15). Logical models can integrate different kinds of molecular entities (genes, proteins, lncRNA, etc.) (15).

GINsim is a computational tool dedicated to the building and analysis of logical models, enabling the delineation of logical regulatory graphs, together with various dynamical analyses, through model simulations, but also with the support of efficient algorithms to identify the attractors (stable states and/or oscillatory behavior) of the system, for wild-type or mutant conditions (14). The resulting model can be further analysed using the CoLoMoTo tool suite, an interactive toolbox integrating several logical modeling software tools, with a uniform interface to perform simulations and other analyses, which are easy to share, and reproduce through the use of notebooks (16).

The aim of our study was to integrate all the information gathered from scientific literature and high-throughput data (RNA-seq and ChIP-seq) into a logical model of the regulatory network underlying moDCs differentiation. After iterative enhancement, our final model is able to properly recapitulate cell commitment for each of the initial conditions considered: (i) IL4 alone fosters monocyte activation, (ii) CSF2 alone fosters macrophage commitments, while (iii) CSF2 and IL4 together foster moDC commitment.

Results

Information gathered from literature curation leads to a fragmentary model of monocytes to dendritic cells differentiation

To better understand the regulatory network controlling moDCs differentiation, we analysed the scientific literature and integrated relevant information into a regulatory graph. In this process, we focused on monocyte to moDCs differentiation studies carried on human cells, in particular, on studies where CSF2 and IL4 were used in similar culture conditions. The resulting regulatory graph is shown in Figure 1.

Figure 1

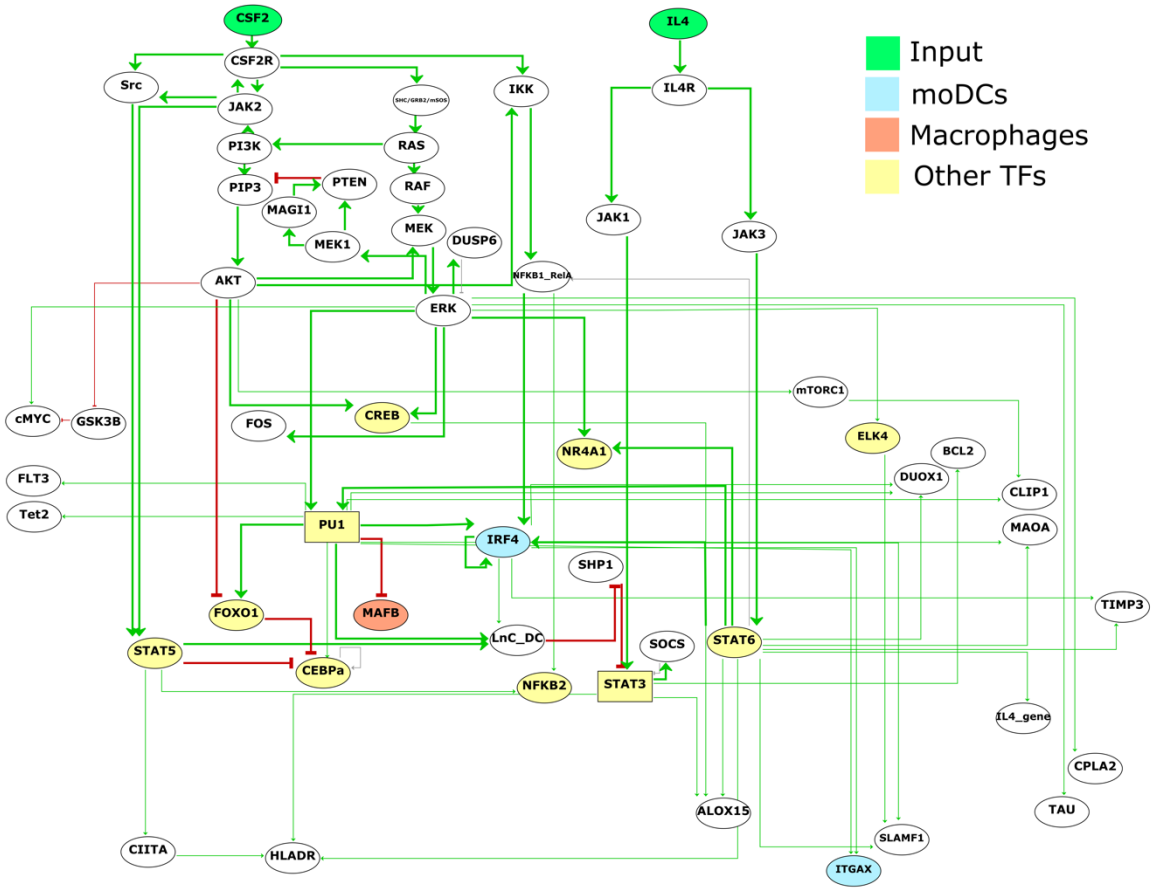


Figure 1. Regulatory graph controlling monocyte to moDCs differentiation, as derived from the scientific literature (last update "17april/2020"). The green nodes at the top represent the inputs (CSF2 and IL4), the yellow nodes denote transcription factors, the blue nodes denote moDCs specific genes, and orange nodes denote macrophage-specific genes. Nodes left in white correspond to components of generic signalling pathways. Green and red arcs denote positive and negative interactions, respectively.

Based on this first regulatory graph, we used GINsim to define logical rules (combining conditions on regulatory nodes with NOT, AND and OR Boolean operators), to compute the corresponding stable states and to perform simulations in order to determine the cellular phenotypes reached for each specific input condition. For this preliminary model, we found six stable states, but only one of them could be directly interpreted as a cellular phenotype (predenditic cells), while the other stable states did not recapitulate the typical signatures of activated monocytes or of macrophages.

Regarding the regulatory interactions between TFs and their target genes displayed in Figure 1, we observed that STAT6 has the highest number of interactions, while other TFs have only few interactions, such as STAT5, that only activates CIITA gene, or CREB that only activates ALOX15 gene. Furthermore, this regulatory graph contains very few specific moDCs markers. To complete this preliminary network, we decided to exploit epigenome and transcriptome data to infer novel regulatory interactions and integrate them into our logical model (a proof of concept of this approach can be found in Collombet *et. al.* 2016 (15)).

Epigenome annotations help to unravel relevant transcription factor regulatory interactions

In order to complete our model of the regulatory network controlling the differentiation of monocytes into moDCs, we included the TFs known to be activated by CSF2 and IL4 signals in moDCs, as well as established monocytes markers. Moreover, we included information regarding the differentiation of monocytes into macrophages, which occurs when monocytes are treated with CSF2 alone (17). In short, we (i) used monocytes, moDCs, and macrophage epigenome data to define chromatin states, (ii) defined genomic regions likely to be involved in the regulation of the genes of the model, and (iii) searched for putative TFs binding sites in these regions.

We analysed ChIP-seq data from the Blueprint consortium for six histone marks (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9me3 and H3K27me3) in monocytes, moDCs, and macrophages derived from monocytes. We then used ChromHMM (18) to annotate the epigenome in each cell type based on these data. The resulting states (segments) were classified as Quiescent/low signal, Polycomb repressed, Poised regulation, Active TSS, Active promoter, Primed enhancer, Active gene/enhancer, Low transcription, TSS repressed and Strong transcription (Figure 2a). As expected, it is possible to visualize clear differences in the epigenome of moDCs and monocytes when exploring genes with specific cell expression in a genome browser, for example, the gene IRF4, a TF that mediates the differentiation of moDCs, is only active in moDCs while it is poised on macrophages and monocytes (Figure 2b).

Figure 2

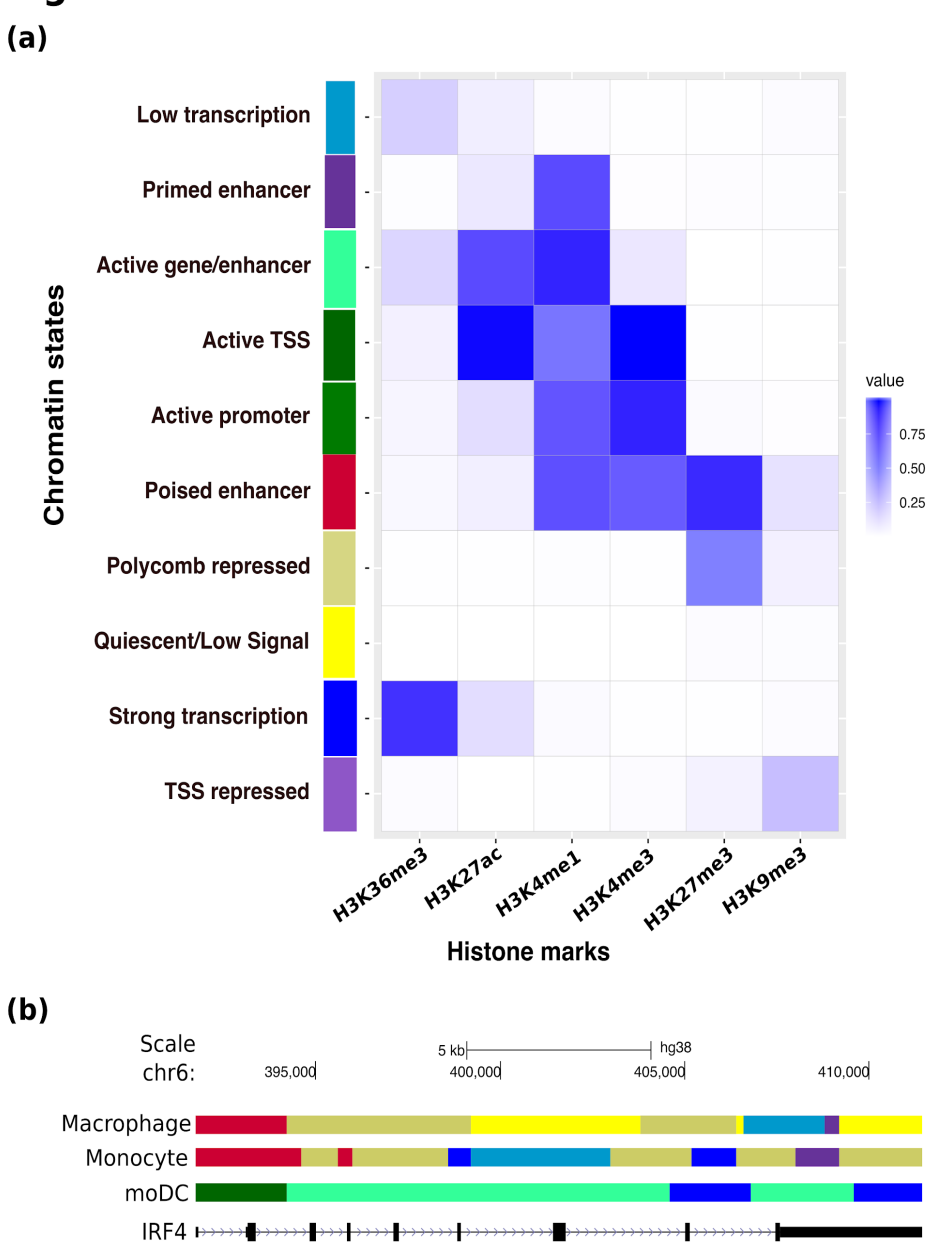


Figure 2. Epigenomic annotations of monocytes, moDCs, and macrophages. (a) Heatmap showing the histone mark enrichment in each of the states determined with ChromHMM. (b) IRF4 genomic region visualized in the UCSC browser, each row of segmentation corresponds to a specific cell type. Segmentation results from ChromHMM analysis, each color represents a state according to the color code used in the heatmap. State annotation was manually done based on biological knowledge. moDCs are the only cell type with the active gene marks (in green).

148 We selected the epigenome annotations regions with promoter-associated functions: Active TSS,
149 Repressed TSS, Active gene/enhancer, and Poised regulation. These regulatory regions were then
150 used to predict binding sites for the fourteen TFs activated by CSF2 and IL4 pathways (Figure 1)
151 using the position-weight matrices collected in the Jaspar database (19) with the pattern-matching
152 tool *matrix-scan* (20) from the RSAT suite (21). This led us to define novel regulatory interactions
153 targeting specific gene markers for moDCs, monocytes, and macrophages (Table 1), thereby
154 enabling us to complete the regulatory network controlling monocytes to moDCs differentiation.
155
156 **Table1. Cell type-specific gene markers selected to be added into the model.** Based on the
157 epigenome analysis we identified relevant regulatory interactions that helped select candidate
158 genes to be added to the model.

Cell-type	**Gene**
moDCs	TLR8
moDCs	TLR7
moDCs	TLR6
moDCs	TLR4
moDCs	TLR3
moDCs	NCOR2
moDCs	DEC205 (LY75)
moDCs	DCIR (CLEC4A)
moDCs	CD83
moDCs	CD48
moDCs	CD226
moDCs	CD209
moDCs	CD1C
moDCs	CD1B

moDCs	CD1A
moDCs	CD141 (THBD)
moDCs	ITGAX (CD11C)
moDCs	CCL22
moDCs	CCL2
Monocyte	CD14
Monocyte	SELL
Macrophage	CD163
Macrophage	CCDC151
Macrophage	MERTK
Macrophage	CD206

159

160 For thirteen out of 20 genes related to moDCs phenotype, we found putative binding sites for the
161 TF IRF4 (Figure 3a), corroborating a central role for IRF4 in the moDCs differentiation. In
162 particular, we predicted that IRF4 directly regulates TLR genes (TLR3, 4 and 7), which play a
163 crucial role in antigen recognition and are thus relevant for moDCs function. Furthermore, we
164 predicted that TLR6 and TLR8 are regulated by STAT6, another essential TF in moDCs (6). In
165 addition, we predicted that the genes encoding for the external proteins CD1A, CD1B, and CD1C
166 are regulated by IRF4, as well as by other TFs (PU.1, PRDM1, NR4A1, CEBPA) related to moDC
167 differentiation. Furthermore, we predicted that the gene coding for CD48, a costimulatory
168 molecule involved in T cell activation, is regulated by PU.1, which is known to participate in the
169 differentiation of STEM cell progenitors into leukocytes at different stages (22). We also looked
170 for regulatory interactions between the identified TFs. Additionally, we validated interactions of
171 PU.1 with CEBPA, IRF4, and IRF8. We also identified that AHR is regulating IRF4, MAFB, and
172 PRDM1, which represent interesting candidates to assess experimentally. Figure 3b summarizes

the regulatory interactions that compose our final logical model, where colored squares emphasize novel regulations, and asterisks denote interactions already described in the literature.

Figure 3

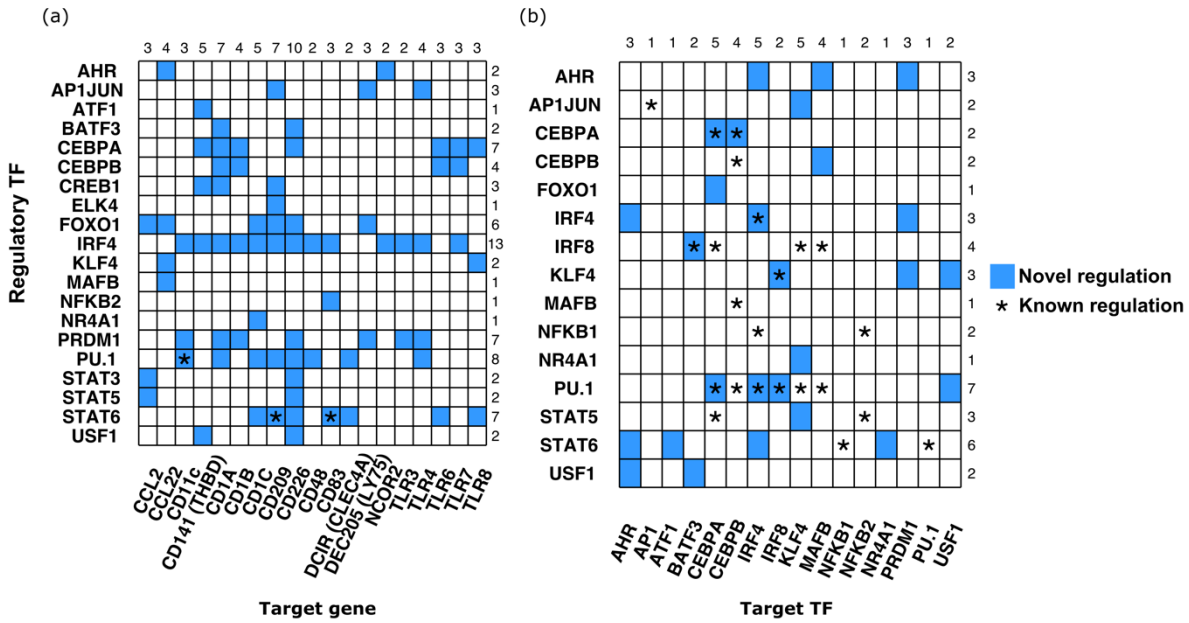


Figure 3. Predicted transcriptional regulatory interactions by TFs activated through CSF2 and IL4 signalling cascades. Each TF binding motif was used to search for putative binding sites in selected regulatory regions (based on chromatin state annotations) of specific gene or TF. (a) The y-axis shows the list of regulatory TFs and the x-axis shows the moDCs specific target genes. (b) The y-axis shows the list of regulatory TFs and the x-axis shows target TFs, in order to identify new regulatory interactions between TFs. Turquoise colored squares show predicted binding sites for the specified TF. Asterisks mark the binding sites for target genes or trans-regulation for target TF that have been already reported. Numbers at the end of each row correspond to the numbers of genes with regulatory interactions for each TF. Numbers at the top of every column correspond to the numbers of TF regulating each target gene.

Integration of new relevant regulatory interactions improves model accuracy

We integrated the selected gene markers for each cell type with the predicted regulatory TFs into our model by adding the discovered regulatory interactions summarized in Figure 3. Using the

new version of the model together with relevant Boolean rules, we set out to compute its stable states, which much better recapitulated the main cell fates compared to our first model (Figure 4a). Our revised model is characterised by four stable states. The first stable state corresponds to cell-death, which is the expected outcome for monocytes without cytokine stimulation. The second state, with IL4 ON, corresponds to monocyte signature (KLF4, SELL, and CD14 genes). The third stable state, with CSF2 ON, corresponds to monocytes that display a macrophage signature (MAFB, CEBPB, CD163, and CD206 genes). Finally, the last stable state, with CSF2 and IL4 ON, displays the moDCs signature (*i.e.* with IRF4, STAT6, CD1A, and CD209 all ON) (Figure 4b). Once we performed the analysis of stable states computation using our model, we validated that the PI3K signalling remained inhibited in order to reach moDCs commitment, this behavior was described by Van de Laar *et. al.* 2012 (23).

Figure 4

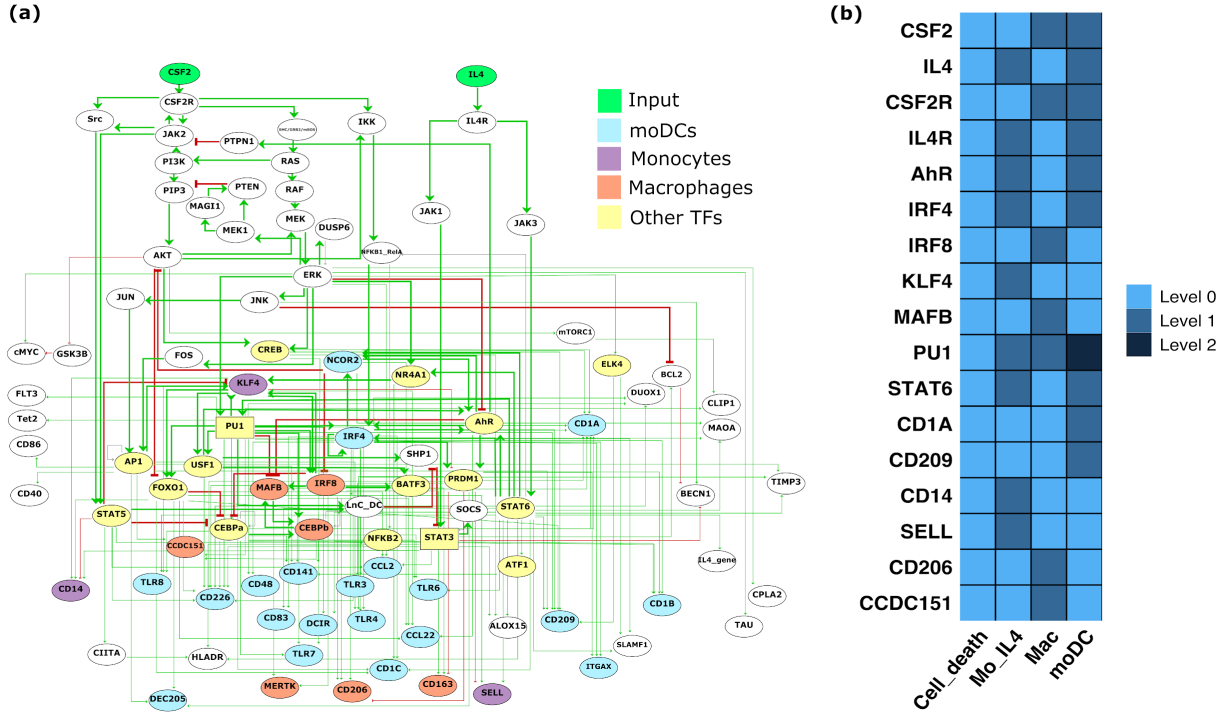
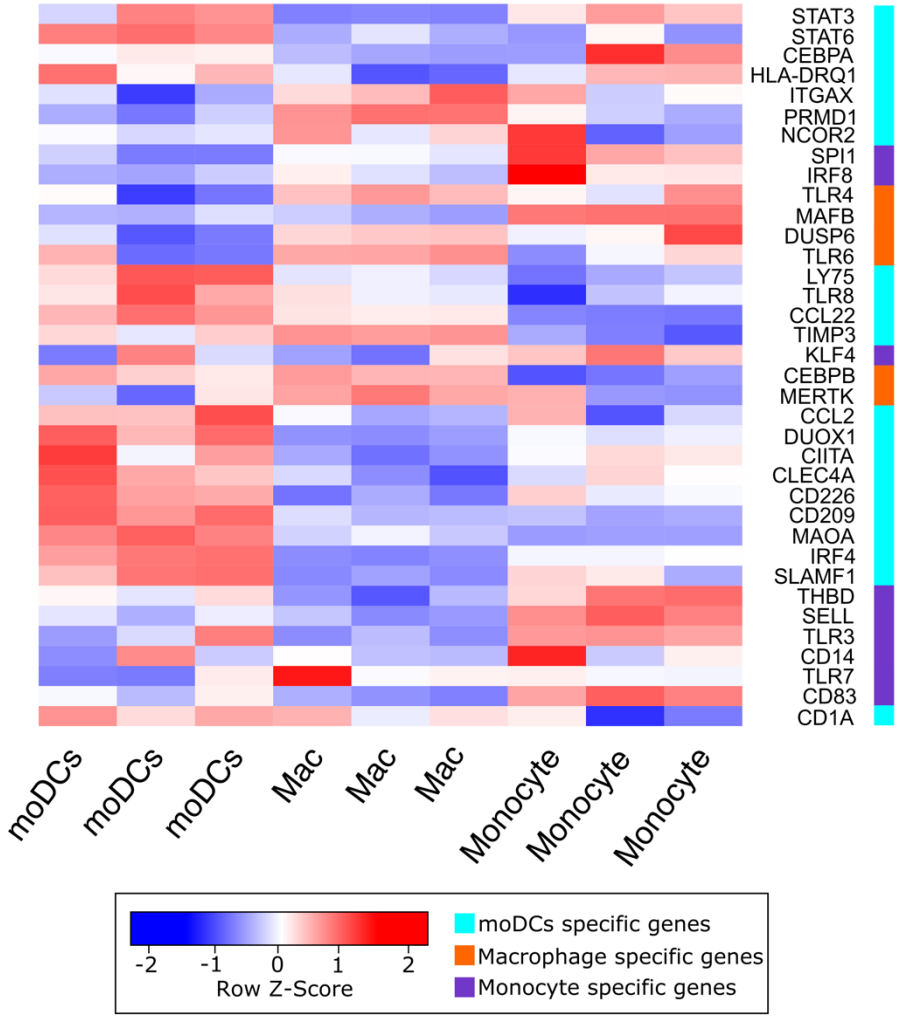


Figure 4. Logical model of monocytes to dendritic cells differentiation *in vitro*. **(a)** The green nodes at the top represent the inputs (CSF2 and IL4), the yellow nodes denote other TFs, blue nodes correspond to moDC specific genes, orange nodes to macrophage-specific genes, and purple nodes to monocyte specific genes. Green and red arcs denote positive and negative interactions, respectively. **(b)** Stables states of selected nodes (signature for each cell type), with the mention of the corresponding cell type. The first column corresponds to the final outcome in the absence of both IL4 and CSF2, *i.e.* cell-death of the monocytes. The second column corresponds to the stimulation of monocytes by IL4. The third column corresponds to the macrophage outcome, in the presence of the sole CSF2. Finally, the fourth column corresponds to moDCs commitment, in the presence of both IL4 and CSF2, where STAT3 reaches the level 2 in the presence of the long non-coding RNA LnC-DC, and PU.1 reaches the level 2, which is required to turn-off MAFB during moDCs commitment. SuppFig1 displays the complete set of nodes.

We used gene expression information to validate the different cell commitment expression signatures. To do that, we analysed RNA-seq data from monocytes, moDCs, and macrophages. Figure 5 displays the differential expression of the genes included in the model. Interestingly, we found two main clusters of genes highly expressed in moDCs, but down regulated in macrophages. These moDCs differentially express clusters of genes, including STAT3, STAT6, CEBPA, IRF4, TFs that participate in moDCs differentiation, also including CD206, MAOA, SLAMF1, that are specific markers for moDCs. Additionally, monocytes show highly expressed genes, like KLF4, IRF8, SELL, and CD14.

224 After integrating epigenome and transcriptome data into the model, we performed further
225 simulations using the CoLoMoTo toolbox, with the purpose of recapitulating documented cellular
226 commitment experiments.

Figure 5



227

228 **Figure 5.** Clustered heatmap showing differentially expressed genes between cell types. The first
229 three columns are moDCs, the next three columns are macrophages, and the last three columns are
230 monocytes (columns represent biological replicates). The Z-score indicates the level of differential
231 expression gene by gene bases. The colored column at the right represents specific gene markers

per cell type, in blue for moDCs, orange for macrophages, and purple for monocytes. The heatmap is clustered by differential expression.

Model simulations correctly estimate cellular commitment to differentiation

We imported our model into the CoLoMoTo Interactive Notebook, a digital notebook that enables integrated complementary analyses software (with PINT, BioLQM, and MaBOSS) and facilitates reproducibility (24). The notebook is available as supplementary material. We used the tool Pint (25) to assess nine single gene mutants (IRF4, STAT6, PU.1, IRF8, MAFB, NCOR2, AHR, JAK3, CEBPB) that have been reported in the literature to affect the differentiation process. We were able to replicate the behavior of each mutant (perturbations) with our model. Table 2 shows the summary of the results obtained for these perturbations, while Figure 6 shows the behavior of each node for each perturbation at the corresponding stable states.

245 **Table 2. Perturbations tested in the model of monocyte to moDCs differentiation.**

Protein	Function	Phenotype described	Perturbation simulated	Model phenotype
IRF4	Transcription factor	Monocytes were infected using lentiviral vectors containing shRNA against IRF4, silenced IRF4 induced a dramatic reduction of moDCs (9).	Loss of function	Lack of most of moDCs specific markers
STAT6	Transcription factor	The ectopic expression of STAT6 in monocytes, resulted in increased levels of the DC-specific marker DC-sign, following CSF2 stimulation and without IL4 (6).	Gain of function	STAT6 is almost sufficient to archive moDC differentiation
PU.1	Transcription factor	Inducible constructions of PU.1 and MAFB were used to infected monocytes. In cells with PU.1 induced DCs, MafB differentiated macrophages (26).	Loss of function	Abolish moDCs and macrophage phenotype commitment
IRF8	Transcription factor	Introduction of KLF4 into an Irf8-/- myeloid progenitor cell line induced a subset of IRF8 target genes and caused partial monocyte differentiation (27).	Loss of function	Abolish KLF4 expression, and the entire macrophage differentiation
MAFB	Transcription factor	Silencing of MAFB resulted in a strong decrease in mo-Macs and an increase in mo-DC differentiation (26).	Loss of function	moDCs differentiation is normal, Macrophage differentiation is abolished
NCOR2	Transcriptional regulator	NCOR2 silencing resulted in 1,834 variable genes that correspond with IL4 signature genes (28).	Loss of function	Lack of moDCs specific markers
AHR	Transcription factor	AHR silencing reduced mo-DC differentiation while slightly increasing mo-Mac (9).	Loss of function	Lack of every moDCs specific markers, macrophage differentiation is normal
JAK3	Tyrosine-protein kinase	STAT6 phosphorylation disappeared following JAK3 inhibition. In the case of MACs, we did not observe STAT6 phosphorylation, given the lack of stimulation of JAK3 (6).	Loss of function	Macrophage phenotype with CSF2 and IL4. Macrophage differentiation is not affected
CEBPB	Transcription factor	In the absence of CEBPb in monocytes CEBPb-KO, only a very low amount (5%) of this macrophage-like morphology was seen, and most of the cells stayed round (29).	Loss of function	Lack of some specific macrophage markers

246

Figure 6

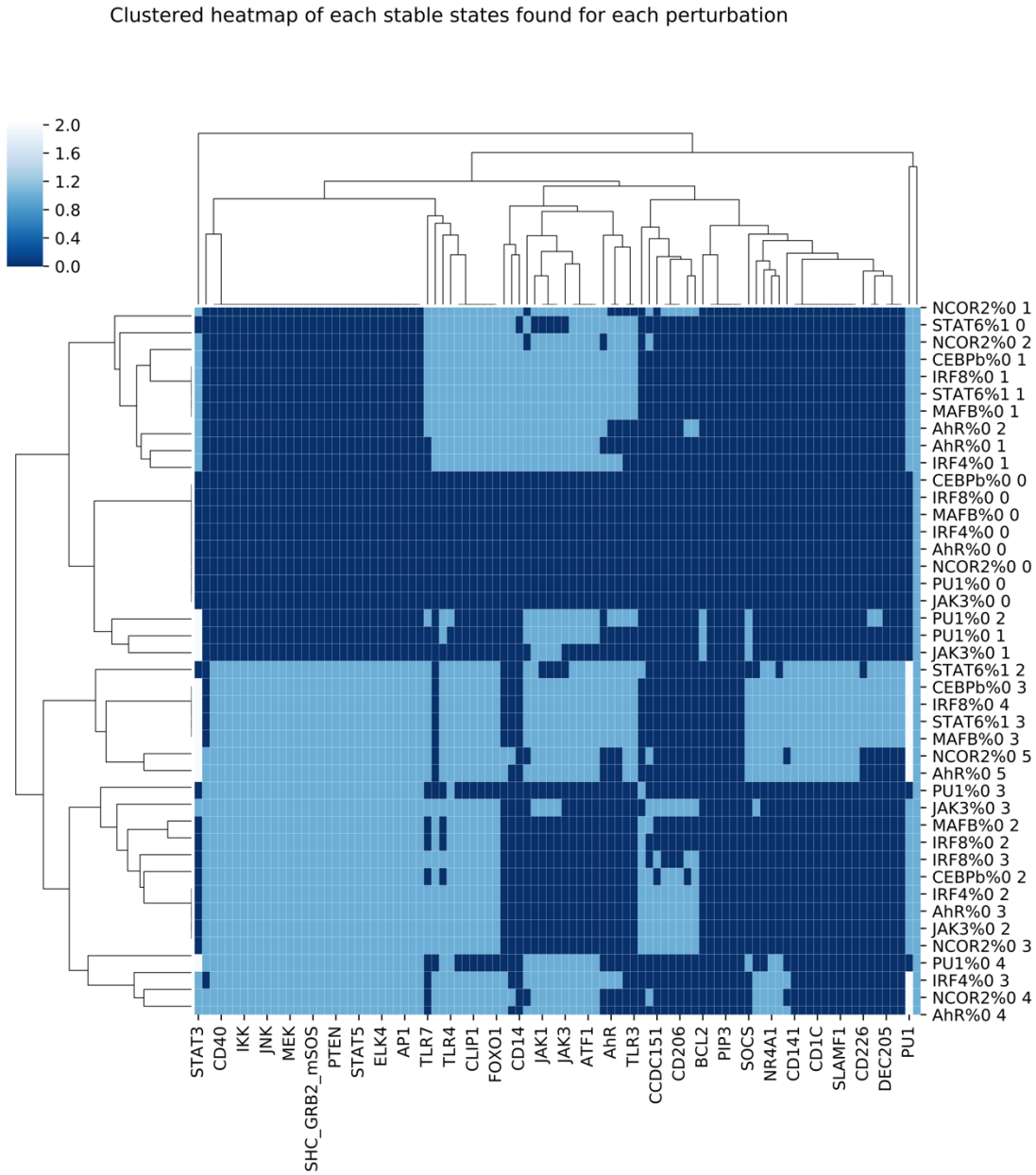
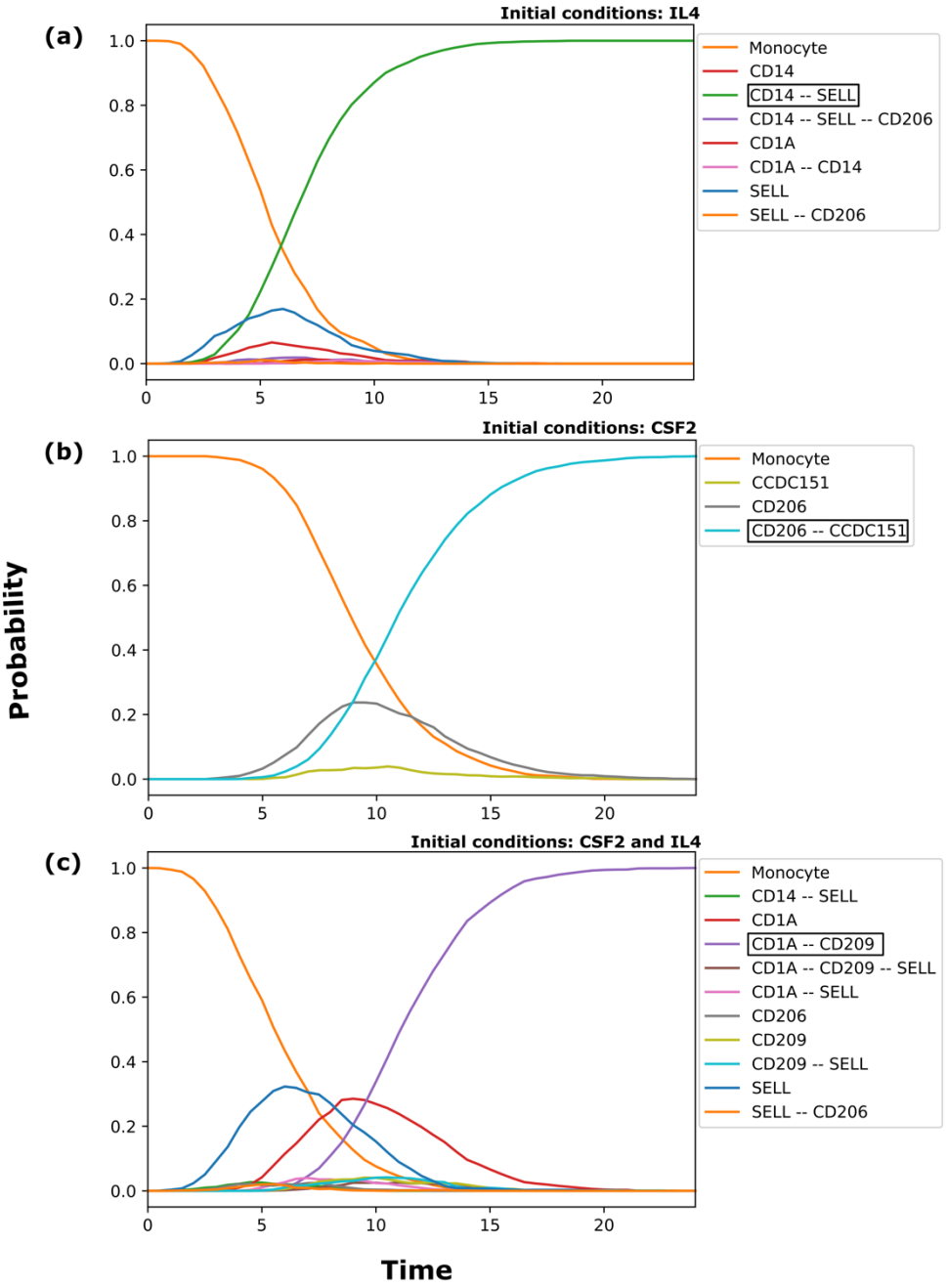


Figure 6. Clustered heatmap of the stable states obtained for all the perturbations considered. Each row represents one perturbation and one corresponding stable state. For example, NCOR2%0 1 denotes a knockout of NCOR2 and corresponds to the stable state 1 obtained for this condition. Every column represents one of the 95 nodes of the logical model.

We used the tool BioLQM to validate the reachability of each cell type commitment according to each stimulus combination. In order to verify the percentage of the final cell fate with the different initial stimulus, we used the stochastic Boolean simulation tool MaBoSS to estimate the probabilities to reach alternative states, where the final stable states represent alternative cell fate commitment (30).

From the literature, we know that CSF2 and ILF4 presence commits cells to differentiate to moDCs. Using MaBoSS, we tested cell commitment with the combination of CSF2 and IL4 ON at the initial state. This simulation showed that 100% of cells then commit to become moDCs (Figure 7c). For IL4 ON but not CSF2, 100% of the cells differentiate into stimulated monocytes, with the corresponding gene markers ON (figure 7a). For CSF2 ON but not IL4, 100% of the cells differentiated to macrophages, as expected (Figure 7b). In figure 7, we can clearly distinguish the three stable states corresponding to monocytes, moDCs, and macrophages, respectively.

Figure 7



264

265 **Figure 7. Stochastic simulations with** MaBoSS trajectories correctly recapitulate cellular
266 commitment. The x-axis shows the time, the y-axis shows the probability to reach every final
267 commitment. Highlighted in a rectangle is the final cellular commitment per stimulus. (a) IL4 ON
268 gives rise to 100% differentiated cells into stimulated monocytes. (b) CSF2 ON gives rise to100%

of the cells differentiated into macrophages. (c) With both CSF2 and IL4 ON 100% of cells commit to the moDC phenotype.

Discussion

The construction of logical models traditionally relies on manual curation of the literature of a biological system of interest. In this work, we further took advantage of public ChIP-seq and RNA-seq data from the Blueprint consortium (31) to delineate in more detail the network driving the differentiation of monocytes into moDCs. We were able to fill various gaps in this regulatory network, which allowed us to reach a better understanding of this particular differentiation process. In the process, we predicted a series of novel interactions, validated *in silico* through our simulations, and amenable to further experimental tests.

In particular, we delineated a series of target genes presumably important for the differentiation of monocytes into moDCs. Some TFs are already well known, such as IRF4, AHR, STAT6, and PU.1 (6,9,32). In our analysis, we were able to recapitulate key features regarding the expression of the corresponding genes, such as a high expression of IRF4 and STAT6 genes in moDCs. We further validated the results obtained by Vento et al 2016 (6), in which STAT6 is required for moDCs differentiation; according to our model, STAT6 is indeed required for moDCs differentiation, but not for macrophage commitment.

We were also able to unravel TFs not previously reported as relevant in this process, such as FOXO1, C/EBPα, AP1, and PRDM1, tentatively regulating specific moDCs genes. We predict that FOXO1 regulates at least six moDCs genes, while C/EBPα regulates at least seven of them.

Particularly, AP1 regulates TLR4, DEC205 (LY75), and CD209 (DC-SING), which are relevant for an antigen-presenting cell. CREB1 is also participating in the regulation of moDC genes, through the activation of CD141 y CD1A. We also predict for the first time that NR4A1 could regulate CD1C, a protein found at the surface of moDCs.

We further reviewed data recently published on the predicted TF-gene interaction considered in our model, and we found that some of these interactions have been recently experimentally confirmed. In particular, the regulation for the ITGAX gene was shown to be regulated by PU.1 and IRF4 (32), as predicted by our epigenomic analysis.

This study represents the first effort to integrate the current knowledge on monocytes to moDCs differentiation *in vitro* and should foster our understanding of this process. Additionally, we unraveled novel TFs and regulatory links potentially involved in this differentiation process.

Material and methods

GINsim implementation and simulations.

Using the software GINsim version 3.0 (14), we integrated the previously described signalling pathways that are activated when monocytes are cultured with CSF2 and IL4 (studies reviews are inside GINsim model as annotations, and in the SuppFile1). We also performed a review of the available literature related to the process of monocyte to moDCs differentiation. The logical model was built using GINsim (33), where nodes represent genes or proteins, and edges represent the interactions between them, these interactions can be negative (red arrow) or positive (green arrow).

313 In general, each node can take two values, zero or one, but in special cases, activation (ON)
314 requires to consider different qualitative levels of activation (*e.g.*: STAT3 expression is activated
315 by JAK1, but the presence of LncDC leads to a further increase of STAT3 expression). For these
316 special cases, it is possible to use multilevel nodes, *e.g.* ternary variables enabling an additional
317 level of activation (hence taking the values 0, 1, and 2). Logical rules are associated with each
318 component of the network, combining literals (*i.e.* regulatory variables with specific values) with
319 the classical Boolean operators AND (&), OR (|) and NOT(!), thereby defining in which conditions
320 each of these components can be activated or shut down.

321

322 **ChIP-seq data analysis.**

323 Raw fastq files from ChIP-seq experiments were retrieved from the Blueprint Consortium (31)
324 data access portal (http://dcc.blueprint-epigenome.eu/#/datasets) with dataset identifiers
325 EGAD00001001552, EGAD00001002484, EGAD00001002485, EGAD00001001576,
326 EGAD00001002504. We processed six histone marks data (H3K4me1, H3K4me3, H3K27ac,
327 H3K36me3, H3K9me3, and H3K27me3) with two biological replicates from human monocytes,
328 macrophages, and moDCs. We performed quality control of read sequences with FastQC/0.11.3
329 tool (34), then we used Trimmomatic/0.33 (35) to improve quality reads before mapping them
330 with bowtie2-2.2.6 (36) to the human hg38 reference genome. Second quality control is required
331 after alignment, for which we used ENCODE QC, which consists of three major tests: NRF (non-
332 redundant fraction), PBC1 (PCR Bottleneck coefficient 1), and PBC2 (PCR Bottleneck coefficient
333 2) (37). IDR analysis (37) was performed to replicate control.

334

335 **Chromatin states definition.**

We used one set of the six histone marks (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9me3, and H3K27me3) ChIP-seq data for each cell type (monocytes, macrophages, and moDCs) and their respective input control. Chromatin states were defined using ChromHMM (18) version 1.12 (38) with the recommended parameters (BinarizeBed -b 200, assembly hg38), and specifying 10 states. In order to define the description for the states, we used the probability of appearance of different marks in every state (*e.g.* H3K27ac-Enhancers, H3Kme1-Enhancers, H3K4me3-Promoters, H3K27me3-Repressive, H3K9me3-Repressive, H3K36me3-Transcribed (39), and then we looked into the enrichment of the states for several genome annotations (CpGIsland, RefSeqExon, RefSeqGene, RefSeqTES, RefSeqTSS, and RefSeq2kb). Integrating this information, we were able to manually assign a functional description to each state. Once we described every state, we focused on Active TSS, Repressed TSS, Active gene/enhancer, and Poised regulation regions to look for regulatory interactions between TFs and target genes. For poised regulation, it is well known that regions go from poised to active regions when cells are under differentiation. Additionally, we took the whole segment for each state selected result from ChromHMM.

Search for TFBS using matrix-scan

From manual curation of literature, we identified 22 TFs that participate after monocyte stimulation leading to the differentiation of macrophages or moDCs. We retrieved one PSSM (Position-Specific Scoring Matrix) for each of the 22 TFs (SuppTable1) from the JASPAR2018 database human collection (19). We performed pattern-matching searches for TF motif instances using the 22 PSSMs in the selected chromatin regions (Active TSS, Repressed TSS, Active gene/enhancer, and Poised regulation) from ChromHMM results. For this task we used the tool

359 *matrix-scan* (20) from the RSAT suite (21) with the following main parameters: background model
360 of Markov order 1 and stringent thresholds of p-value $\leq 10^{-5}$ and score 1 (-markov 1 -lth score 1
361 -uth pval 1e-5).
362
363 **RNA-seq analyses**
364 Raw fastq files from RNA-seq experiments were retrieved from the Blueprint Consortium (31)
365 data access portal (http://dcc.blueprint-epigenome.eu/#/datasets) with dataset identifiers:
366 EGAD00001002308, EGAD00001001506, EGAD00001002526, EGAD00001002507, and
367 EGAD00001001582. For this analysis, we used the methods described in Law et al 2016 (40). In
368 brief, that is quality control with FastQC/0.11.3 (41), pseudo-alignment and count determination
369 with Kallisto 0.43.1 (42) using the release-90 from Ensembl (ftp://ftp.ensembl.org/pub/release-
370 90/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh38.cdna.all.fa.gz) to create our index with the
371 following command: kallisto index -i index_kallisto_hsap_90_cdna --make-unique
372 Homo_sapiens.GRCh38.cdna.all.fa.gz. Counts were assigned to genes using Tximport 1.14.0 (43),
373 and were processed from raw-scale to counts per million (CPM), then they were transformed to
374 log-CPM. Genes below 1 of expression were removed. Then we normalized raw library sizes using
375 the *calcNormFactors* function from edgeR library in R. Afterwards, we performed a differential
376 gene expression analysis with edgeR 3.28.0 (44). Finally, we used heatmap.2 from the gplots
377 library to plot the genes found in our model (Figure 5).
378
379 **CoLoMoTo analysis**
380 In order to assure reproducibility, we used the CoLoMoTo toolbox(16) that integrates several
381 logical modeling tools, including GINsim, bioLQM, Pint, and MaBoSS. We used GINsim to

compute the stable states, and bioLQM to identify trap spaces approximating cyclic attractors. The computation of mean stochastic trajectories was performed using MaBoSS (30). The GINsim model and the CoLoMoTo notebook are available at https://github.com/karenunez/moDC_model_differentiation.

Figures generation

Figure 1, and 4A were generated with the GINsim software. The plots in Figures 2A, 3A, 3B, and 4B were done using the ggplot2 library from R. Figures 6, and 7 are from the CoLoMoTo notebook constructed in this study.

Supplemental material

Supplementary files are available at https://github.com/karenunez/moDC_model_differentiation.

SuppFile1. Model_annotation.doc

SuppFile2. Mo_Mac_moDCs_ChromHMM_summary.html

SuppFile3. moDC_E7_ActiveGeneEnhancer.bed

SuppTable1. TFs_JASPARID_matrixes.xlsx

SuppFigure1.StableStates_95nodes.png

Acknowledgments

Karen Julia Nuñez Reza is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship

CVU/Becario: 634764/331823 from CONACYT, México. This work received support from Luis Aguilar, Alejandro Leon and Jair García of the Laboratorio Nacional de Visualización Científica Avanzada. We thank Carina Uribe Díaz and Alejandra Castillo Carbajal for their technical support. We want to thank Marc Dalod and Thien Phong Vu Manh for your valuable comments on our work.

Funding

This work was supported by CONACYT grants 269449 and 1690; Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica – Universidad Nacional Autónoma de México (PAPIIT-UNAM) grant [IA206517-IA201119]; M.T.-C., A.M.-R, A.S. and D.T. further acknowledge SEP-CONACYT-ECOS-ANUIES (291235) support. MTC is supported by Institut Universitaire de France.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' contributions

K.J.N.-R. carried out the manual literature curation, performed the ChIP-seq and RNA-seq data analysis, participated in the construction of the two model versions, notebook implementation,

study design, and drafted the manuscript; A.N. participated in the construction of the two model versions, notebook implementation, and drafted the manuscript; A.S.-J. participated in the manual literature curation, and drafted the manuscript; A.V.L.-A. participated in the manual literature curation, and drafted the manuscript; A.S. participated in the design of the study, and drafted the manuscript; M.T.C. mentored the ChIP-seq, ChromHMM, and matrix-scan analysis, participated in the design of the study and drafted the manuscript; D.T participated in the construction of the two model versions, notebook implementation, study design, and drafted the manuscript; A.M.-R. mentored the ChIP-seq, ChromHMM, and matrix-scan analysis, participated in the design of the study, and drafted the manuscript.

References

1. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. Nature [Internet]. 1998 Mar 19;392(6673):245–52. Available from: http://dx.doi.org/10.1038/32588

2. Segura E. Review of Mouse and Human Dendritic Cell Subsets. Methods Mol Biol [Internet]. 2016;1423:3–15. Available from: http://dx.doi.org/10.1007/978-1-4939-3606-9_1

3. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. Nat Rev Immunol [Internet]. 2019 Aug 29; Available from: http://dx.doi.org/10.1038/s41577-019-0210-z

4. León B, López-Bravo M, Ardavín C. Monocyte-derived dendritic cells. Semin Immunol [Internet]. 2005 Aug;17(4):313–8. Available from: http://dx.doi.org/10.1016/j.smim.2005.05.013

5. Laar LVD, Coffer PJ, Woltman AM. Review article Regulation of dendritic cell development by GM-CSF : molecular control and implications for immune homeostasis and therapy. 2017;119(15):3383–94. Available from: http://dx.doi.org/10.1182/blood-2011-11-370130.The

6. Vento-tormo R, Company C, Rodríguez-ubreva J, Rica LD, Urquiza JM, Javierre BM, et al. IL-4 orchestrates STAT6-mediated DNA demethylation leading to dendritic cell differentiation. 2016;1–18. Available from: http://dx.doi.org/10.1186/s13059-015-0863-2

7. Ranasinghe C, Trivedi S, Wijesundara DK, Jackson RJ. IL-4 and IL-13 receptors: Roles in immunity and powerful vaccine adjuvants. Cytokine Growth Factor Rev [Internet]. 2014 Aug;25(4):437–42. Available from: http://dx.doi.org/10.1016/j.cytogfr.2014.07.010

8. Wang C, Ye Z, Kijlstra A, Zhou Y, Yang P. Activation of the aryl hydrocarbon receptor affects

activation and function of human monocyte-derived dendritic cells. Clin Exp Immunol [Internet]. 2014;177(2):521–30. Available from: http://dx.doi.org/10.1111/cei.12352

9. Goudot C, Coillard A, Villani AC, Gueguen P, Cros A, Sarkizova S, et al. Aryl Hydrocarbon Receptor Controls Monocyte Differentiation into Dendritic Cells versus Macrophages. Immunity [Internet]. 2017 Sep 19;47(3):582–96.e6. Available from: http://dx.doi.org/10.1016/j.immuni.2017.08.016

10. Le Novère N. Quantitative and logic modelling of molecular and gene networks. Nat Rev Genet [Internet]. 2015 Mar;16(3):146–58. Available from: http://dx.doi.org/10.1038/nrg3885

11. Martínez-Méndez D, Villarreal C, Mendoza L, Huerta L. An Integrative Network Modeling Approach to T CD4 Cell Activation. Front Physiol [Internet]. 2020 Apr 23;11:380. Available from: http://dx.doi.org/10.3389/fphys.2020.00380

12. Rodríguez-Jorge O, Kempis-Calanis LA, Abou-Jaoudé W, Gutiérrez-Reyna DY, Hernandez C, Ramirez-Pliego O, et al. Cooperation between T cell receptor and Toll-like receptor 5 signaling for CD4+ T cell activation. Sci Signal [Internet]. 2019 Apr 16;12(577). Available from: http://dx.doi.org/10.1126/scisignal.aar3641

13. Remy E, Rebouissou S, Chaouiya C, Zinovyev A, Radvanyi F, Calzone L. A Modeling Approach to Explain Mutually Exclusive and Co-Occurring Genetic Alterations in Bladder Tumorigenesis. Cancer Res [Internet]. 2015 Oct 1;75(19):4042–52. Available from: http://dx.doi.org/10.1158/0008-5472.CAN-15-0602

14. Naldi A, Hernandez C, Abou-Jaoudé W, Monteiro PT, Chaouiya C, Thieffry D. Logical Modeling and Analysis of Cellular Regulatory Networks With GINsim 3.0. Front Physiol [Internet]. 2018;9(June):1–16. Available from: http://dx.doi.org/10.3389/fphys.2018.00646

15. Collombet S, Van Oevelen C, Luis J, Ortega S, Abou-jaoudé W, Di B. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. 2016; Available from: http://dx.doi.org/10.1073/pnas.1610622114

16. Naldi A, Hernandez C, Levy N, Stoll G, Monteiro PT, Chaouiya C, et al. The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks. Front Physiol [Internet]. 2018 Jun 19;9:680. Available from: http://dx.doi.org/10.3389/fphys.2018.00680

17. Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells. Science [Internet]. 2010 Feb 5;327(5966):656–61. Available from: http://dx.doi.org/10.1126/science.1178331

18. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods [Internet]. 2012 Feb 28;9(3):215–6. Available from: http://dx.doi.org/10.1038/nmeth.1906

19. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res [Internet]. 2018 Jan 4;46(D1):D1284. Available from: http://dx.doi.org/10.1093/nar/gkx1188

20. Turatsinze J-V, Thomas-Chollier M, Defrance M, van Helden J. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat Protoc [Internet].

2008 Sep 18;3:1578. Available from: https://doi.org/10.1038/nprot.2008.97

21. Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, et al. RSAT 2018: Regulatory sequence analysis tools 20th anniversary. Nucleic Acids Res [Internet]. 2018;46(W1):W209–14. Available from: http://dx.doi.org/10.1093/nar/gky317

22. McArdel SL, Terhorst C, Sharpe AH. Roles of CD48 in regulating immunity and tolerance. Clin Immunol [Internet]. 2016 Mar;164:10–20. Available from: http://dx.doi.org/10.1016/j.clim.2016.01.008

23. van de Laar L, Coffer PJ, Woltman AM. Regulation of dendritic cell development by GM-CSF: molecular control and implications for immune homeostasis and therapy. Blood [Internet]. 2012 Apr 12;119(15):3383–93. Available from: http://dx.doi.org/10.1182/blood-2011-11-370130

24. Naldi A, Monteiro PT, Müssel C, Consortium for Logical Models and Tools, Kestler HA, Thieffry D, et al. Cooperative development of logical modelling standards and tools with CoLoMoTo. Bioinformatics [Internet]. 2015 Apr 1;31(7):1154–9. Available from: http://dx.doi.org/10.1093/bioinformatics/btv013

25. Paulevé L. Pint: A Static Analyzer for Transient Dynamics of Qualitative Networks with IPython Interface. In: Computational Methods in Systems Biology [Internet]. Springer International Publishing; 2017. p. 309–16. Available from: http://dx.doi.org/10.1007/978-3-319-67471-1_20

26. Bakri Y, Sarrazin S, Mayer UP, Tillmanns S, Nerlov C, Boned A, et al. Balance of MafB and PU.1 specifies alternative macrophage or dendritic cell fate. Blood [Internet]. 2005 Apr 1;105(7):2707–16. Available from: http://dx.doi.org/10.1182/blood-2004-04-1448

27. Aguilera-Montilla N, Chamorro S, Nieto C, Sánchez-Cabo F, Dopazo A, Fernández-Salguero PM, et al. Aryl hydrocarbon receptor contributes to the MEK/ERK-dependent maintenance of the immature state of human dendritic cells. Blood [Internet]. 2013;121(15):108–18. Available from: http://dx.doi.org/10.1182/blood-2012-07-445106.N.A.-M.

28. Sander J, Schmidt SV, Cirovic B, McGovern N, Papantonopoulou O, Hardt A-L, et al. Cellular Differentiation of Human Monocytes Is Regulated by Time-Dependent Interleukin-4 Signaling and the Transcriptional Regulator NCOR2. Cell Immunity [Internet]. 2017;47:1051–66. Available from: http://dx.doi.org/10.1016/j.immuni.2017.11.024

29. Gutsch R, Kandemir JD, Pietsch D, Cappello C, Meyer J, Simanowski K, et al. CCAAT/enhancer-binding protein beta inhibits proliferation in monocytic cells by affecting the retinoblastoma protein/E2F/cyclin E pathway but is not directly required for macrophage morphology. J Biol Chem [Internet]. 2011 Jul 1;286(26):22716–29. Available from: http://dx.doi.org/10.1074/jbc.M110.152538

30. Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics [Internet]. 2017 Jul 15;33(14):2226–8. Available from: http://dx.doi.org/10.1093/bioinformatics/btx123

31. Stunnenberg HG, Abrignani S, Adams D, de Almeida M, Altucci L, Amin V, et al. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell [Internet]. 2016;167(5):1145–9. Available from: http://dx.doi.org/10.1016/j.cell.2016.11.007

32. Yashiro T, Kasakura K, Oda Y, Kitamura N, Inoue A, Nakamura S, et al. The hematopoietic cell-

specific transcription factor PU.1 is critical for expression of CD11c. Int Immunol [Internet]. 2017;29(2):87–94. Available from: http://dx.doi.org/10.1093/intimm/dxx009

33. Gonzalez AG, Naldi A, Thieffry D, Chaouiya C. GINsim : A software suite for the qualitative modelling , simulation and analysis of regulatory networks. 2006;84:91–100. Available from: http://dx.doi.org/10.1016/j.biosystems.2005.10.003

34. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res [Internet]. 2018 Aug 24;7:1338. Available from: http://dx.doi.org/10.12688/f1000research.15931.2

35. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics [Internet]. 2014;30(15):2114–20. Available from: http://dx.doi.org/10.1093/bioinformatics/btu170

36. Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min [Internet]. 2015 Jan 8;8(1):1. Available from: http://dx.doi.org/10.1186/s13040-014-0034-0

37. Landt S, Marinov G, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res [Internet]. 2012;22(Park 2009):1813–31. Available from: http://genome.cshlp.org/content/22/9/1813.short

38. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc [Internet]. 2017 Dec;12(12):2478–92. Available from: http://dx.doi.org/10.1038/nprot.2017.124

39. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol [Internet]. 2010 Aug;28(8):817–25. Available from: http://dx.doi.org/10.1038/nbt.1662

40. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res [Internet]. 2016 Jun 17 [cited 2019 Dec 11];5(1408):1408. Available from: https://f1000research.com/articles/5-1408/v1/pdf

41. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Front Genet [Internet]. 2013 Dec 17;4:288. Available from: http://dx.doi.org/10.3389/fgene.2013.00288

42. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol [Internet]. 2016;34(5):525–7. Available from: http://dx.doi.org/10.1038/nbt.3519

43. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res [Internet]. 2015 Dec 30;4:1521. Available from: http://dx.doi.org/10.12688/f1000research.7563.2

44. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics [Internet]. 2010 Jan 1;26(1):139–40. Available from: http://dx.doi.org/10.1093/bioinformatics/btp616

Figure 1

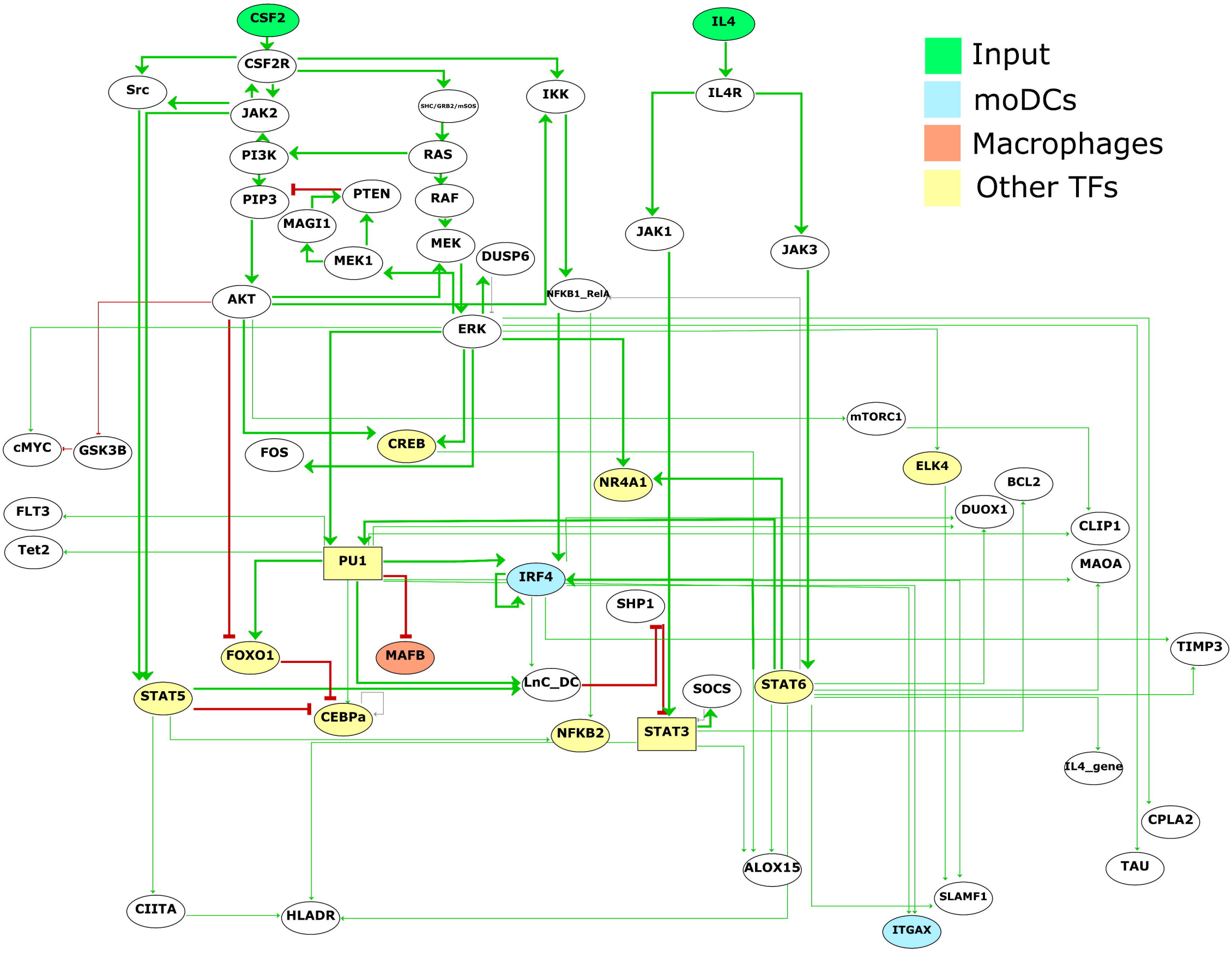
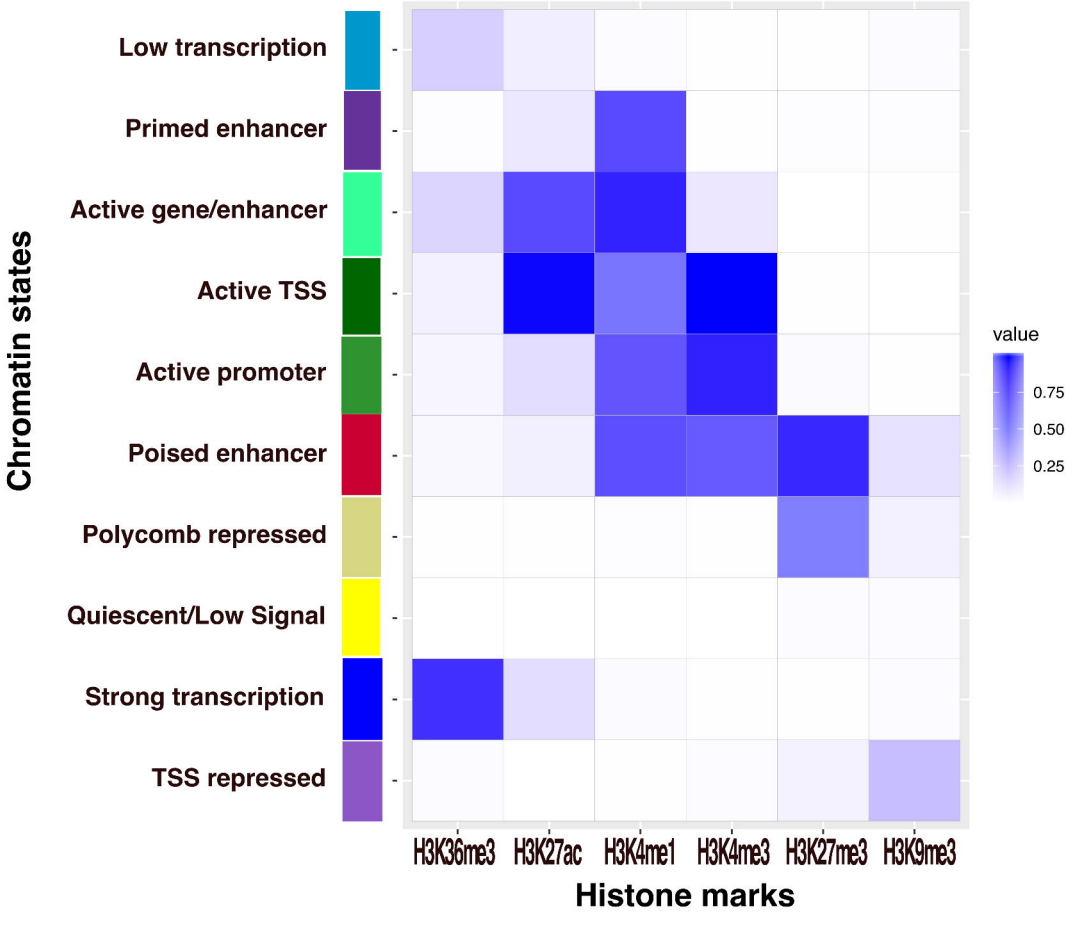


Figure 2

(a)


Chromatin states
Low transcription
Primed enhancer
Active gene/enhancer
Active TSS
Active promoter
Poised enhancer
Polycomb repressed
Quiescent/Low Signal
Strong transcription
TSS repressed
H3K36me3
H3K27ac
H3K4me1
H3K4me3
H3K27me3
H3K9me3
Histone marks
value
0.75
0.50
0.25

(b)

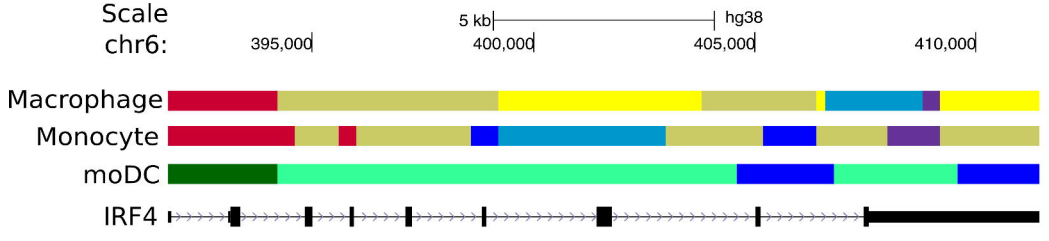

Scale
chr6:
5 kb
hg38
395,000
400,000
405,000
410,000
Macrophage
Monocyte
moDC
IRF4

Figure 3

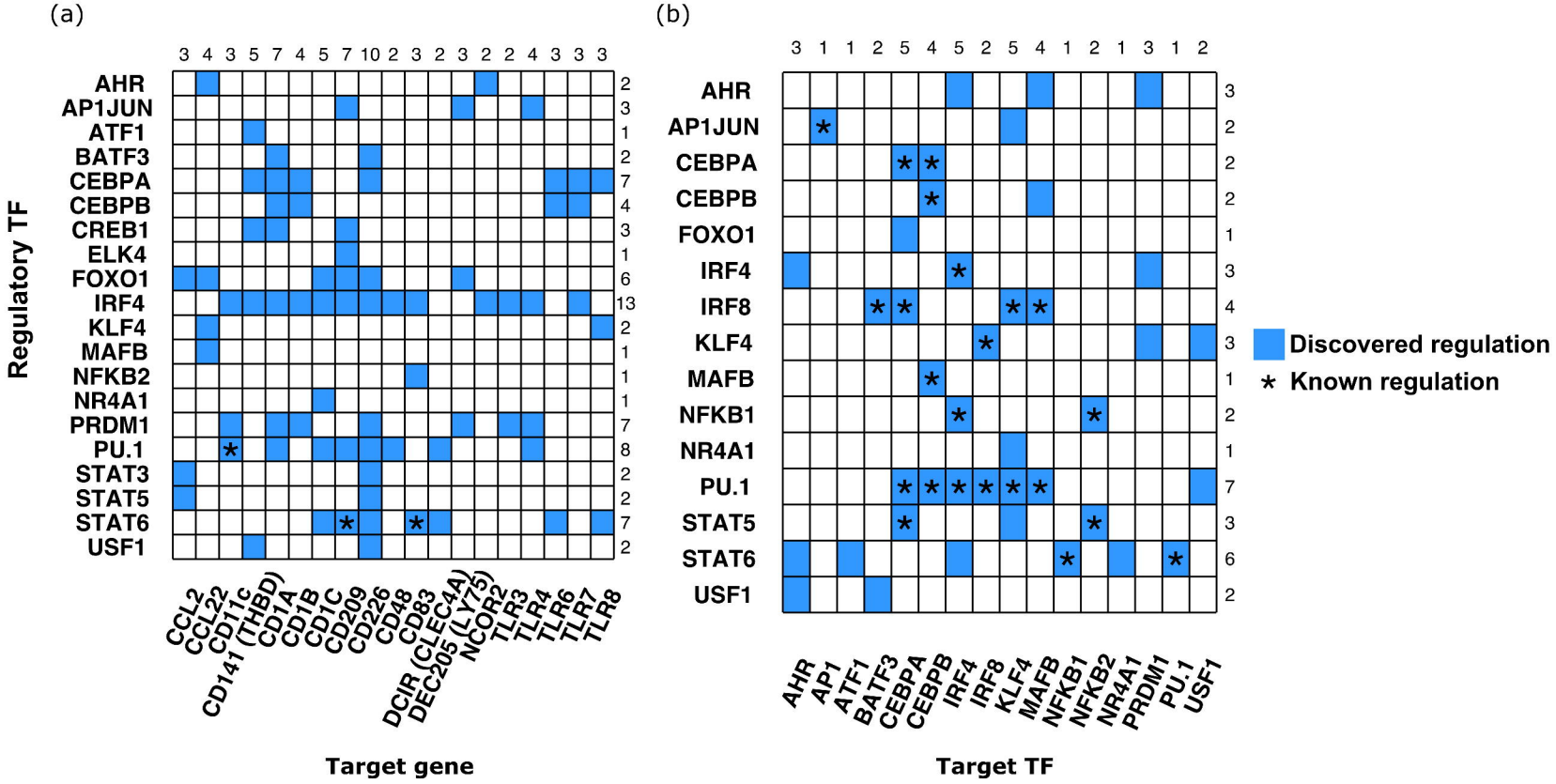
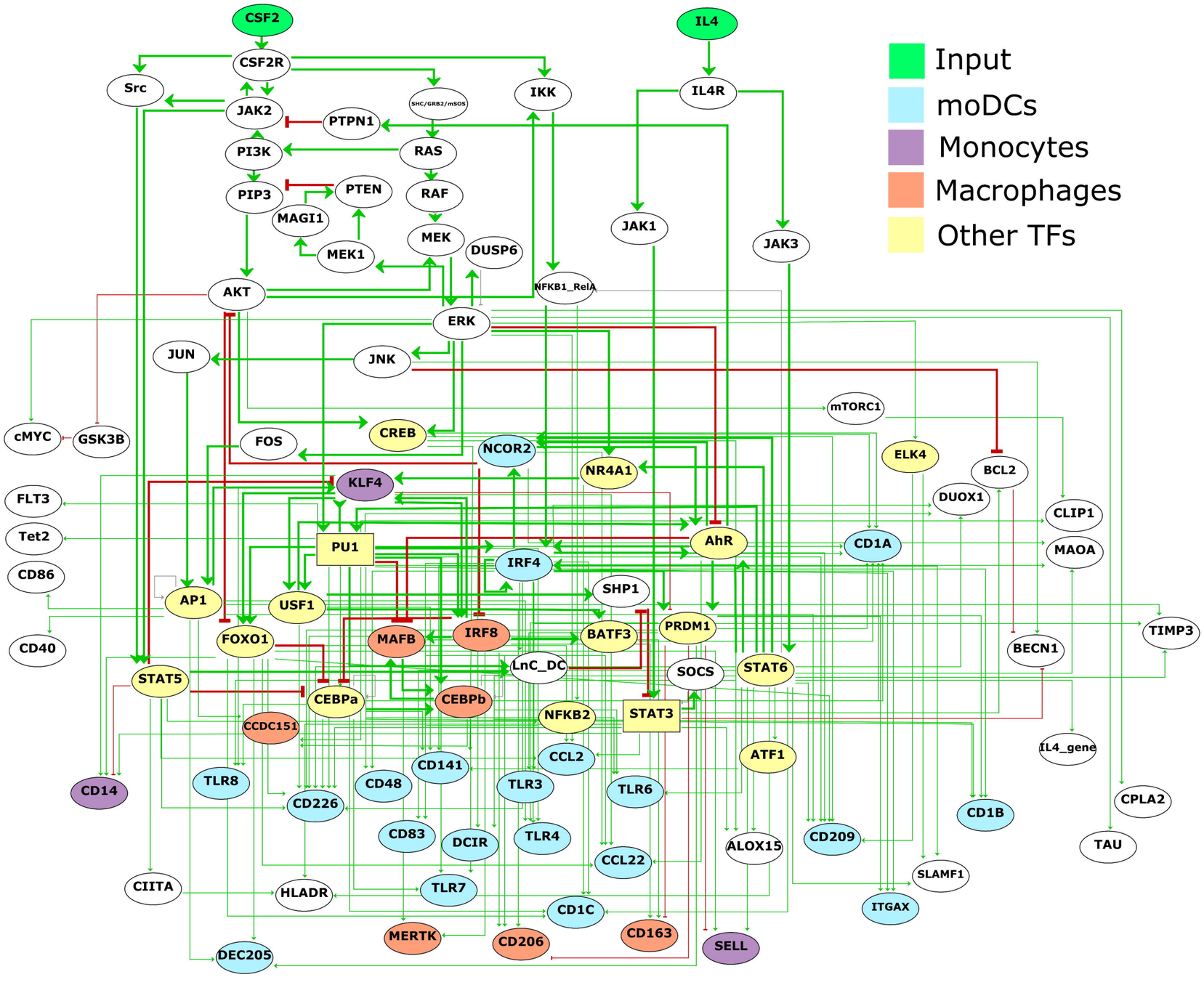

(a)
3 4 3 5 7 4 5 7 10 2 3 2 3 2 2 4 3 3 3
AHR 2
AP1JUN 3
ATF1 1
BATF3 2
CEBPA 7
CEBPB 4
CREB1 3
ELK4 1
FOXO1 6
IRF4 13
KLF4 2
MAFB 1
NFKB2 1
NR4A1 1
PRDM1 7
PU.1 8
STAT3 2
STAT5 2
STAT6 7
USF1 2
Regulatory TF
CCL2
CCL22
CD11c
CD141 (THBD)
CD1A
CD1B
CD1C
CD209
CD226
CD48
CD83
DCIR (CLEC4A)
DEC205 (LY75)
NCOR2
TLR3
TLR4
TLR6
TLR7
TLR8
Target gene
(b)
3 1 1 2 5 4 5 2 5 4 1 2 1 3 1 2
AHR 3
AP1JUN 2
CEBPA 2
CEBPB 2
FOXO1 1
IRF4 3
IRF8 4
KLF4 3
MAFB 1
NFKB1 2
NR4A1 1
PU.1 7
STAT5 3
STAT6 6
USF1 2
AHR
AP1
ATF1
BATF3
CEBPA
CEBPB
IRF4
IRF8
KLF4
MAFB
NFKB1
NFKB2
NR4A1
PRDM1
PU.1
USF1
Target TF
Discovered regulation
Known regulation

Figure 4

(a)



(b)

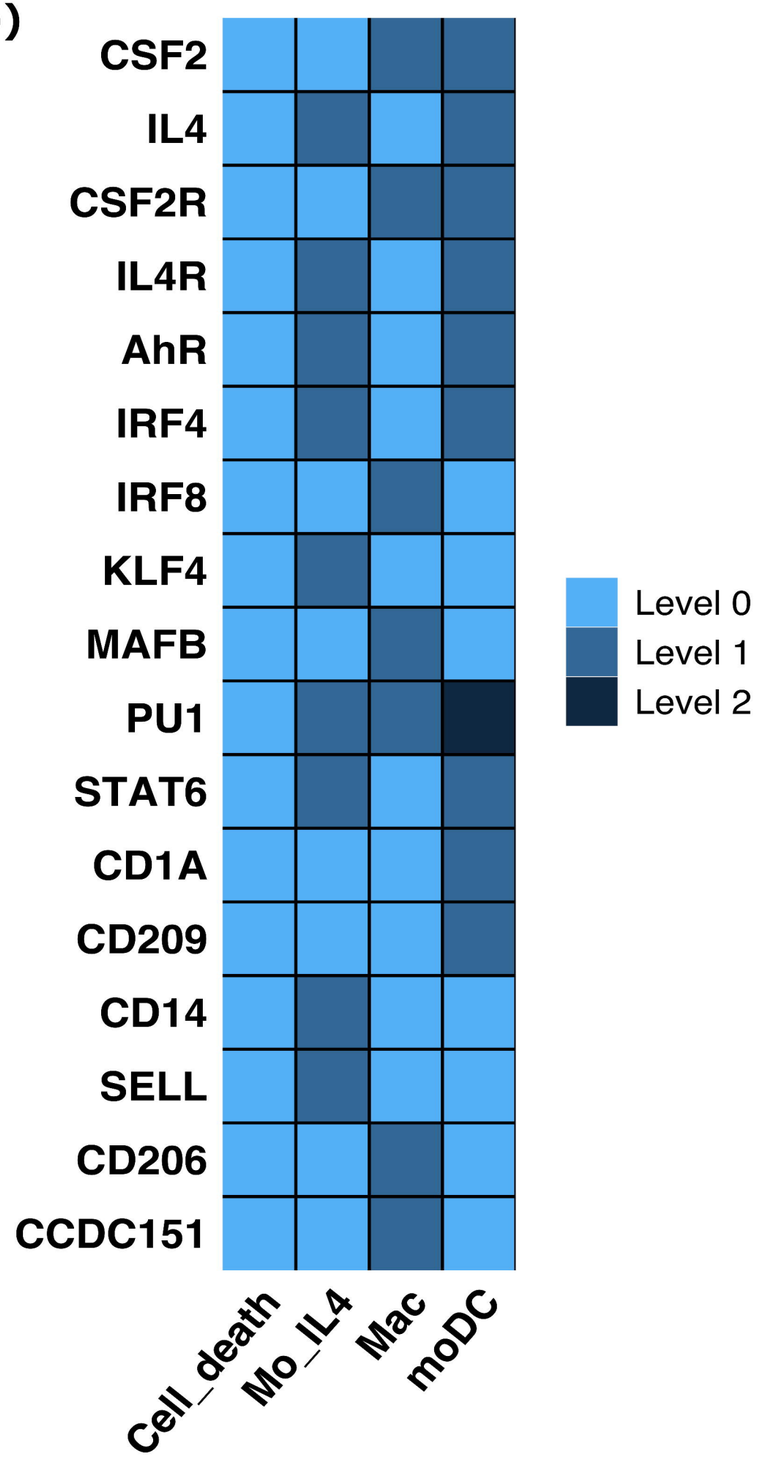


Figure 5

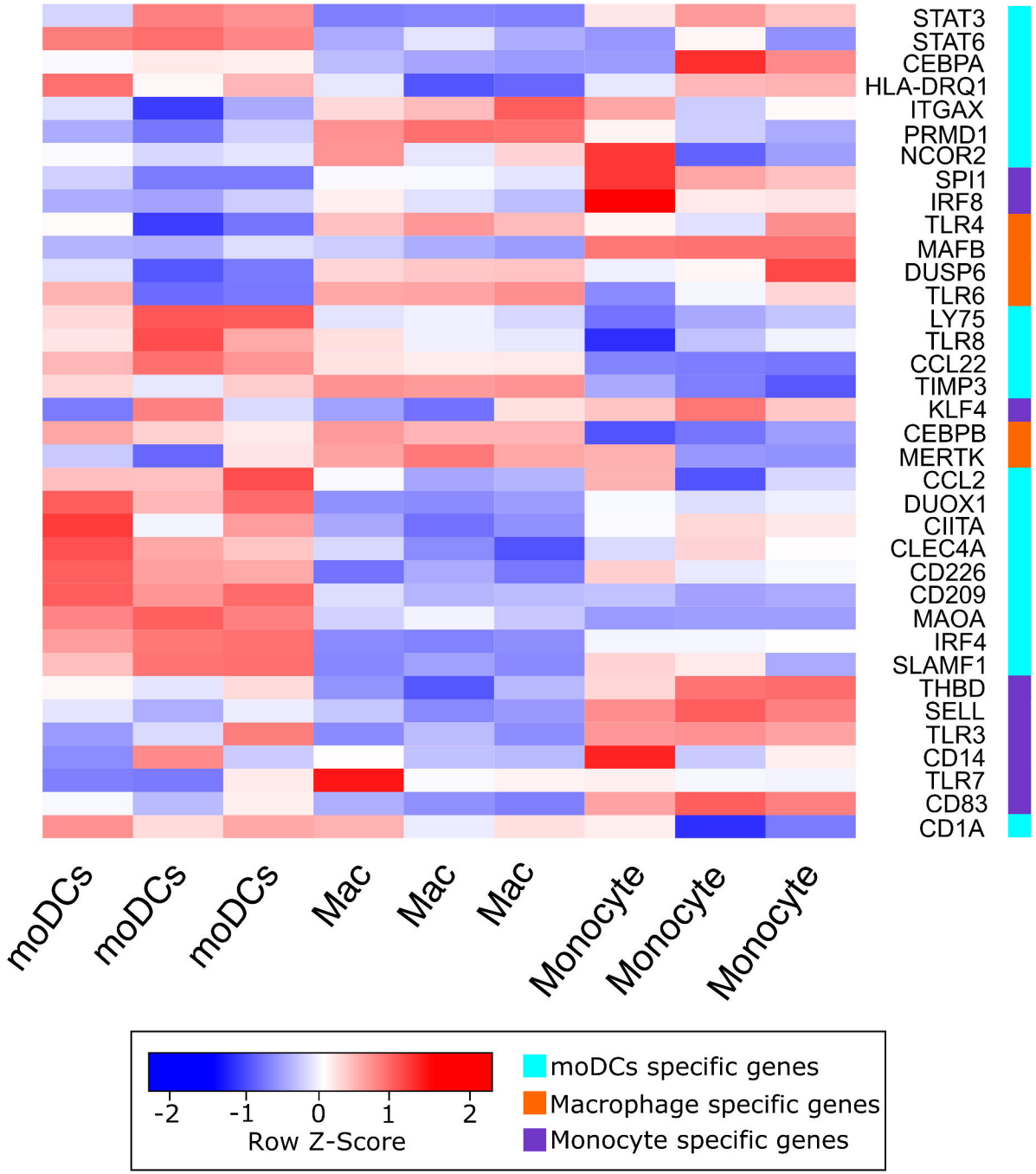

STAT3
STAT6
CEBPA
HLA-DRQ1
ITGAX
PRMD1
NCOR2
SPI1
IRF8
TLR4
MAFB
DUSP6
TLR6
LY75
TLR8
CCL22
TIMP3
KLF4
CEBPB
MERTK
CCL2
DUOX1
CIITA
CLEC4A
CD226
CD209
MAOA
IRF4
SLAMF1
THBD
SELL
TLR3
CD14
TLR7
CD83
CD1A
moDCs
moDCs
moDCs
Mac
Mac
Mac
Monocyte
Monocyte
Monocyte
-2
-1
0
1
2
Row Z-Score
moDCs specific genes
Macrophage specific genes
Monocyte specific genes

Figure 6

Clustered heatmap of each stable states found for each perturbation

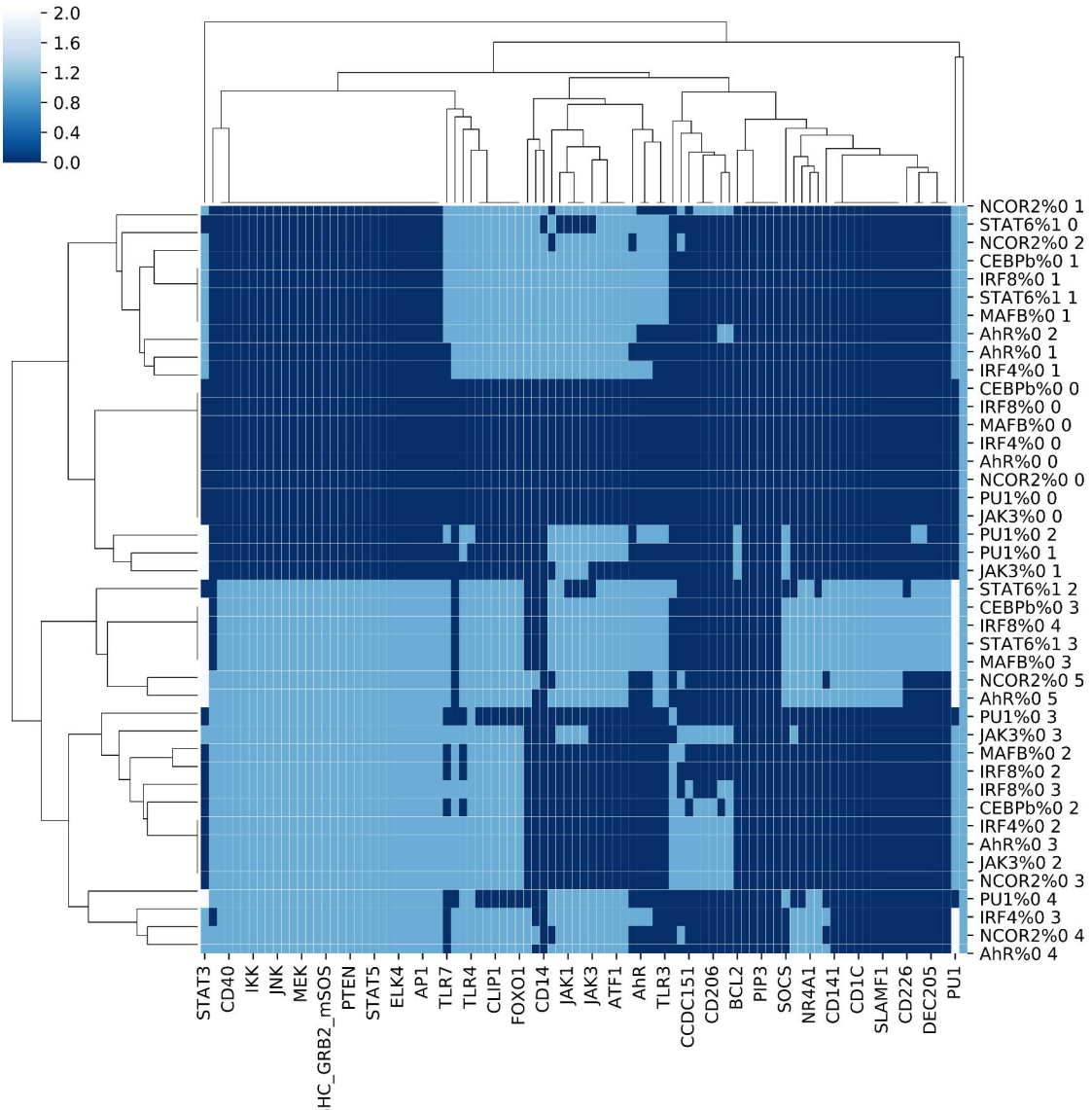


Figure 7