



HAL
open science

Evidence for close molecular proximity between reverting and undifferentiated cells

Souad Zreika, Camille Fourneaux, Elodie Vallin, Laurent Modolo, Remi Seraphin, Alice Moussy, Elias Ventre, Matteo Bouvier, Anthony Ozier-Lafontaine, Arnaud Bonnaffoux, et al.

► To cite this version:

Souad Zreika, Camille Fourneaux, Elodie Vallin, Laurent Modolo, Remi Seraphin, et al.. Evidence for close molecular proximity between reverting and undifferentiated cells. *BMC Biology*, 2022, 20 (1), pp.155. 10.1186/s12915-022-01363-7 . hal-04134084

HAL Id: hal-04134084

<https://hal.science/hal-04134084>

Submitted on 5 Dec 2023

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


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

RESEARCH ARTICLE

Open Access



Evidence for close molecular proximity between reverting and undifferentiated cells

Souad Zreika^{1,2†}, Camille Fourneau^{1†}, Elodie Vallin¹, Laurent Modolo¹, Rémi Seraphin¹, Alice Moussy³, Elias Ventre^{1,4,5}, Matteo Bouvier^{1,6}, Anthony Ozier-Lafontaine⁷, Arnaud Bonnaffoux^{1,6}, Franck Picard¹, Olivier Gandrillon^{1,4†} and Sandrine Gonin-Giraud^{1*†} 

Abstract

Background: According to Waddington's epigenetic landscape concept, the differentiation process can be illustrated by a ball rolling down from the top of a hill (proliferation state) and crossing furrows before stopping in basins or "attractor states" to reach its stable differentiated state. However, it is now clear that some committed cells can retain a certain degree of plasticity and reacquire phenotypical characteristics of a more pluripotent cell state. In line with this dynamic model, we have previously shown that differentiating cells (chicken erythrocytic progenitors (T2EC)) retain for 24 h the ability to self-renew when transferred back in self-renewal conditions. Despite those intriguing and promising results, the underlying molecular state of those "reverting" cells remains unexplored. The aim of the present study was therefore to molecularly characterize the T2EC reversion process by combining advanced statistical tools to make the most of single-cell transcriptomic data. For this purpose, T2EC, initially maintained in a self-renewal medium (0H), were induced to differentiate for 24H (24H differentiating cells); then, a part of these cells was transferred back to the self-renewal medium (48H reverting cells) and the other part was maintained in the differentiation medium for another 24H (48H differentiating cells). For each time point, cell transcriptomes were generated using scRT-qPCR and scRNAseq.

Results: Our results showed a strong overlap between 0H and 48H reverting cells when applying dimensional reduction. Moreover, the statistical comparison of cell distributions and differential expression analysis indicated no significant differences between these two cell groups. Interestingly, gene pattern distributions highlighted that, while 48H reverting cells have gene expression pattern more similar to 0H cells, they are not completely identical, which suggest that for some genes a longer delay may be required for the cells to fully recover. Finally, sparse PLS (sparse partial least square) analysis showed that only the expression of 3 genes discriminates 48H reverting and 0H cells.

Conclusions: Altogether, we show that reverting cells return to an earlier molecular state almost identical to undifferentiated cells and demonstrate a previously undocumented physiological and molecular plasticity during the differentiation process, which most likely results from the dynamic behavior of the underlying molecular network.

Keywords: Cell-fate reversion, Cell differentiation, Erythroid progenitors, Single-cell RNA-seq, Single-cell RT-qPCR

[†]Souad Zreika, Camille Fourneau, Olivier Gandrillon and Sandrine Gonin-Giraud contributed equally to this work.

*Correspondence: sandrine.giraud@ens-lyon.fr

¹ Laboratory of Biology and Modelling of the Cell, Université de Lyon, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, Université Claude Bernard Lyon 1, Lyon, France

Full list of author information is available at the end of the article

Background

The integration and processing of endogenous and exogenous information constitute a fundamental requirement for cells to ensure functions and survival of unicellular or multicellular organisms. Cellular decision-making is then at the core of the physiological or pathological functioning of living organisms. Early



views of the mechanisms governing cell-fate decision-making, and in particular cell differentiation, were based on bulk population data, leading to an over-simplifying deterministic framework. In these first views, cell commitment to a predefined cell type was thought to be triggered through a stereotyped sequence of intermediate states under the influence of specific signals [1].

Single-cell approaches have allowed to change the scale of observation of many molecular processes and revealed that an important heterogeneity in gene expression lies at the heart of isogenic cell populations [2, 3]. Stochasticity in gene expression arises from different causes, such as the probabilistic nature of molecular interactions or transcriptional bursts [4]. Cell-to-cell variability is visible at all omics levels of gene expression, but is being widely studied at the transcriptomic level since various molecular biology tools are available for this scale of investigation [5]. Overall, this heterogeneity in gene expression has been shown to be critical for the process of differentiation, as it provides diversity without the cost of hard-wire-encoded fate programs [6, 7].

Furthermore, single-cell studies have also enabled the development of stochastic models to describe differentiation from single-cell transcriptomic data. One of the best-known models is Conrad Waddington's landscape, which also includes the non-genetic part of cell-to-cell heterogeneities [8]. According to Waddington's model, the shape of the landscape is determined by Gene Regulatory Networks (GRN) and state transitions are modelled as channeling events: a cell, presented as a ball, starts from a mountain top and crosses valleys before reaching a stable state by occupying basins or attractor states, shaped by an underlying GRN [9]. Once this stable state is reached, the state potential decreases and the associated cell fate is restricted or even irreversible [10].

However, it is now clearly accepted that some cells retain fate plasticity [11, 12]. Under the forced modification of transcription factors stoichiometry, a cell that has reached a differentiated state can return to a more pluripotent stage challenging the classical hierarchical view of differentiation [13, 14]. Quite interestingly, spontaneous fate reversion can be observed under a physiological or damaging condition where progenitors or even more committed cells return to an earlier stage, potentially more pluripotent, and reacquire progenitor or stem-cell-like phenotypes and characteristics [15–18]. In this view, our recent study has shown that chicken primary erythroid progenitor cells (T2EC) have retained the capacity to go back to a self-renewal state for up to 24H after the induction of differentiation before they irreversibly engaged in the differentiation process [19]. Despite intriguing and promising results, the molecular

determinants of this so-called fate reversion and the molecular characterization of the reverting cells remain unexplored.

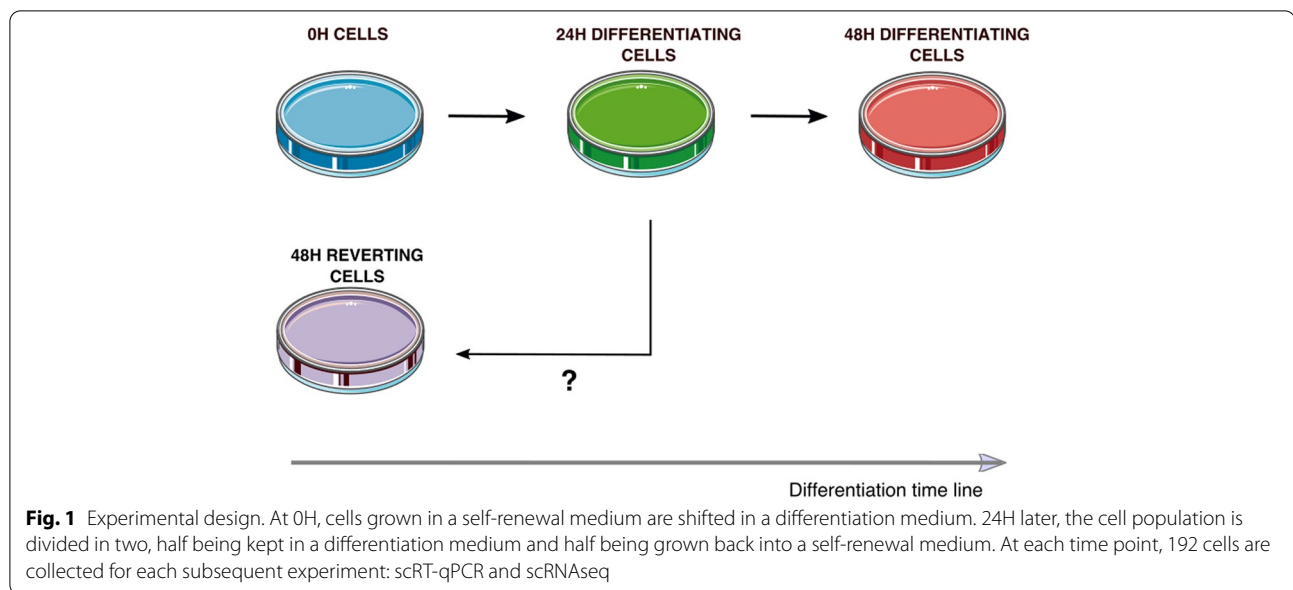
In this work, we go beyond the cellular and phenotypic characterization of the cell reversion process. We characterize the gene expression of primary erythroid progenitors and question if reverting cells undergo an actual fate reversion, i.e., in addition to regaining a comparable cellular state, reacquire a molecular state similar to undifferentiated cells.

For this, differentiation of self-renewing cells was induced by medium change during 24H. Then, we split the differentiating population so that half could pursue differentiation, and the second half was shifted back in a self-renewal medium (Fig. 1). To provide robust quantitative measurements of gene expression variability, we combined a highly sensitive targeted quantification method (scRT-qPCR) with genome-wide scRNAseq data to characterize the transcriptome of each population at the single-cell level: undifferentiated (0H), differentiating (24H and 48H), and reverting (48H reverting) cells. Our statistical analyses show that 48H reverting cells and undifferentiated cells were much more similar, whereas a separation was clearly visible between cells maintained in differentiation (48H differentiating cells) and cells in reversion (48H reverting cells). Furthermore, a statistical comparison of cell distributions indicated no significant differences between 0H cells and 48H reverting cells. Moreover, gene expression pattern distribution of 48H reverting cells showed a shift towards expression pattern distribution of 0H cells. Finally, we identified genes that discriminate 48H reverting cells and 0H cells. Using sparse PLS [20], we were able to show that the expression of 3 genes, *HBBA*, *TBC1D7*, and *HSP90AA1*, was discriminant between 48H reverting cells and 0H cells showing that reverting cells kept transcriptional traces of their induction to differentiation. In conclusion, our results show that reverting cells display gene expression patterns that are very similar to undifferentiated cells while retaining traces of their response to differentiation induction, which suggests an almost complete molecular reversion after 24H of differentiation induction.

Results

Robustness of single-cell transcriptomics analysis

We sought to characterize at the molecular level the cells that were induced to differentiate for 24 h and that retained the ability to proliferate when placed back into a self-renewal medium. We used two different complementary single-cell transcriptomics technologies, scRT-qPCR and scRNAseq. scRT-qPCR allows for highly sensitive quantification but is knowledge-driven and offers information of a limited number of genes while



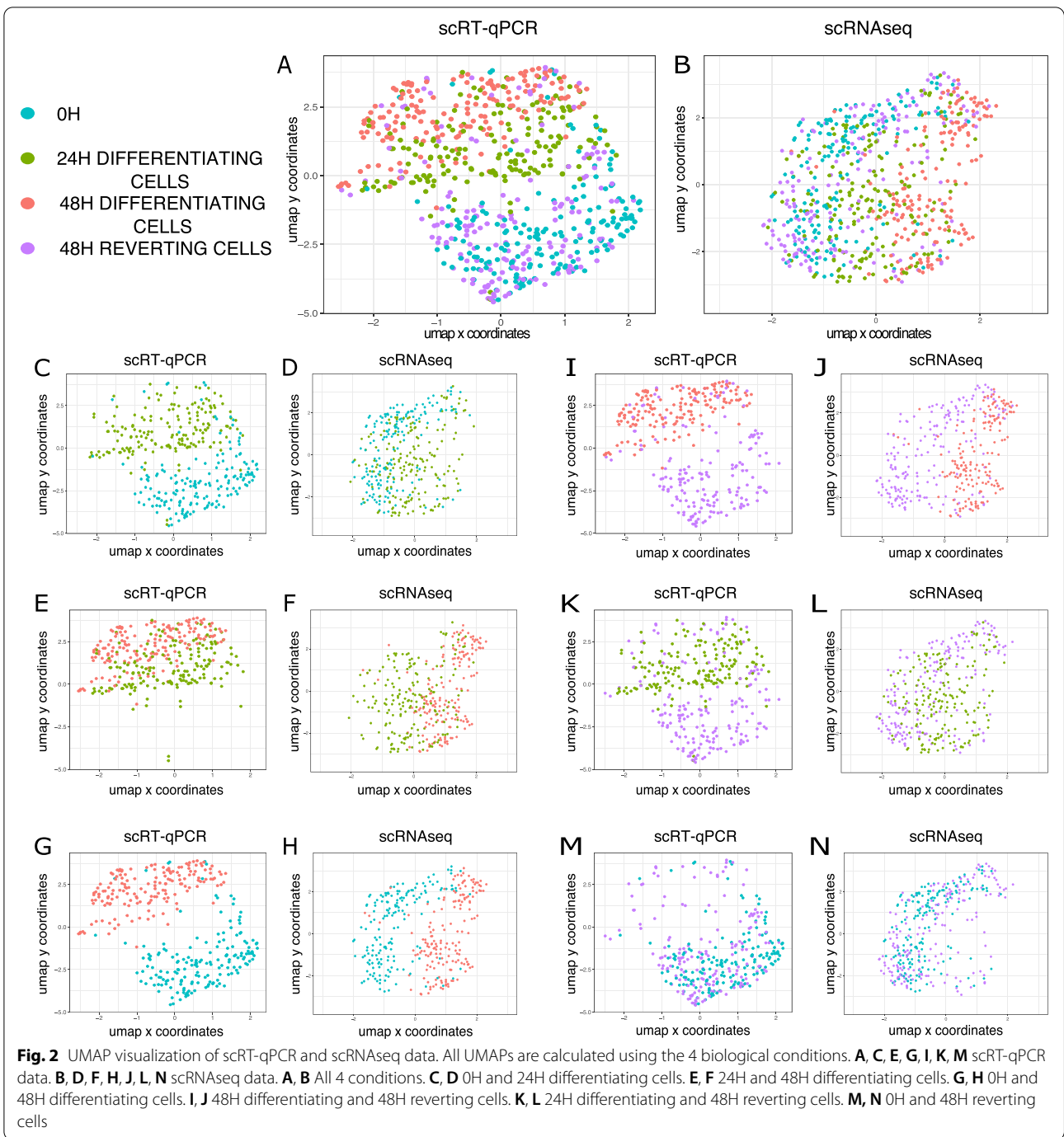
scRNAseq, although less precise for low expression level [21], enables genome-wide quantification without prior knowledge. Furthermore, using two different single-cell technologies allowed us to cross-validate our observations and point toward robust conclusions.

We first obtained by scRT-qPCR the expression level of 83 genes involved in T2EC differentiation in 173, 173, 168, and 171 cells for 0H, 24H, and 48H of differentiation and 48H reverting cells, respectively. Those genes are known to distinguish cells along the differentiation process and include sterol biosynthesis, metabolism, globin subunits, and transcription factors expressed by erythroid progenitors as published in [19]. The robustness of our measurements was confirmed by a Pearson's correlation of 0.85 (p -value = $2.2e-16$) between our experiments and the published data [19]. To investigate fate-reversion genome-wide by scRNAseq, we adapted the MARSseq (massively parallel single-cell RNA-Seq [22] — see the “Methods” section). Then, we obtained gene expression levels in 174, 181, 169, and 186 single cells for 0H, 24H, and 48H of differentiation and 48H reverting cells, respectively. The concordance between scRT-qPCR and scRNAseq data was confirmed by a Pearson's correlation of 0.73 (p -value = $1.34e-13$) between the 74 genes common to both datasets.

Similarity between reverting and undifferentiated cells revealed by dimension reduction

We used UMAP to uncover potential similarities between 48H reverting cells and subgroups of differentiating cells by projecting the 4 conditions (Fig. 2A,

scRT-qPCR data, and Fig. 2B, scRNAseq data). Then, we focused on the normal differentiation process using the 3 time points of differentiation (0H, 24H, and 48H differentiating cells) (Fig. 2C–H). For both experiments, pairwise representations show that 24H differentiating cells tend to overlap with both 0H cells (Fig. 2C, D) and 48H differentiating cells (Fig. 2E, F). On the contrary, the undifferentiated cells and 48H differentiating cells clearly differ (Fig. 2G, H). Interestingly, pairwise representations also reveal that 48H reverting cells separate well from the 48H differentiating cells (Fig. 2I, J) and from 24H cells (Fig. 2K, L), but are visually not distinguishable from the 0H cells (Fig. 2M, N). Almost identical results were observed when, instead of plotting cells on the UMAPs calculated from the mix of the 4 conditions, we recalculated the UMAPs for each pair of conditions (Additional file 1: Fig. S1). Principal Component Analysis (PCA) also captured this general separation of the data (Additional file 1: Fig. S2). Those analyses suggest that the transcriptomes of 48H reverting cells are more similar to the undifferentiated cells than to any other condition at both scales of observation. This was further confirmed by the pairwise statistical comparison of average scRNAseq distributions ([23] — see the “Methods” section). As shown in Table 1, the average transcriptomes of 48H reverting and 48H differentiating cells are significantly different, as well as of undifferentiated and 48H differentiating cells. In contrast, no significant difference in average transcriptomes was detected between 0H and 48H reverting conditions (p -value $\gg 0.05$), indicating a very close proximity of 48H reverting cells to undifferentiated cells.



48H reverting cells and undifferentiated cells have similar gene expression patterns

We then questioned if 48H reverting cells had gene expression patterns identical to 0H cells or retained, for some genes, an expression pattern more similar to 24H or 48H differentiating cells.

Pairwise scRNAseq DE (differential expression) analysis revealed that the “normal” erythrocyte differentiation

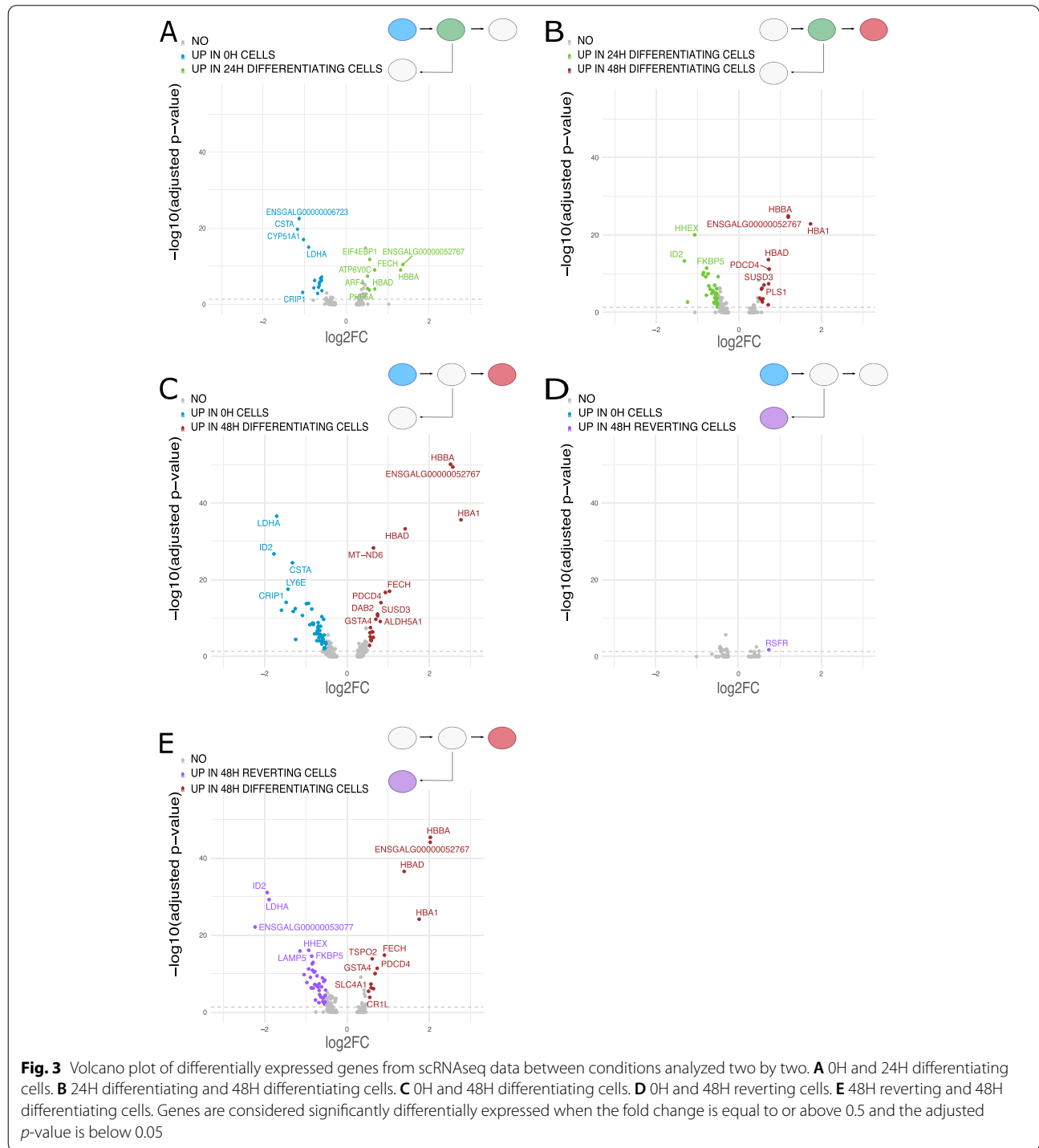
Table 1 *p*-value output of multivariate two tests between pair of conditions compared

	0H vs 48H reverting	0H vs 48H differentiating	48H reverting vs 48H differentiating
<i>p</i> -value	1.00	0	0.00000000369

process showed an increase in the expression of hemoglobin-related genes during the kinetics (hemoglobin subunit epsilon 1 (*HBBA*), hemoglobin alpha-locus 1 (*HBA1*), and hemoglobin alpha, subunit D (*HBAD*)) (Fig. 3A–C). On the other hand, 0H cells expressed a high level of *LDHA* (lactate dehydrogenase A), a marker for

glycolysis metabolism used by self-renewing cells [24], and *ID2* (inhibitor of DNA binding 2) coding for a transcription factor involved in differentiation inhibition [25].

Interestingly, when comparing 0H with 48H reverting cells, we saw only one gene that was significantly differentially expressed just above the threshold (Fig. 3D), the



RSFR (RNase super family related) gene, that is highly expressed in precursor cells from chicken bone marrow [26]. Furthermore, when comparing 48H reverting with 48H differentiating cells, we found hemoglobin-related genes up in the differentiating cells and *LDHA* and *ID2* up in reverting cells (Fig. 3E).

We more closely investigated gene expression distributions within the different conditions to see how gene expression patterns would evolve during the reversion process (Fig. 4). We selected 8 genes differentially expressed and which expression increases or decreases during the differentiation process. *HBA1*, *HBBA*, *HBAD* (different hemoglobin subunits), and *FECH* (Ferrochelatase) are involved in hemoglobin and heme pathways and are more expressed by differentiating cells while *LDHA*, *ID2*, *CSTA* (cystatin A1), and *CRIP1* (Cysteine-rich intestinal protein1) are more expressed by self-renewing undifferentiated cells. We plotted and compared their distribution between the 4 conditions. For the genes involved in differentiation, we see a gradual shift in the distributions towards a higher level of expression as cells get more differentiated (Fig. 4A–D) and we see the opposite shift for genes involved in proliferation (Fig. 4E–H). In all cases, the 48H reverting cell expression patterns for those genes shifted back to patterns closer to the 0H cells. At the time of observation and especially for genes up in differentiation, the 48H reverting cell expression patterns are not completely similar to those of 0H cells. This was further confirmed by using a dedicated statistical tool, sparse PLS (see below).

To go further on gene distribution comparisons, we computed Wasserstein distances, a geometric distance well suited for comparing multimodal distributions, for each 2000 genes of the scRNAseq dataset between each condition two by two. We then obtain 6 distributions of Wasserstein distance values. Finally, we computed the Gini index as a measure of statistical dispersion in each distribution (the higher the Gini index is, the higher inequality among the values). We performed 100 bootstraps and compared the Gini values obtained (Fig. 5A). Distribution of Wasserstein distances between 0H cells and 48H reverting cells had the smallest average Gini index among all distributions (Fig. 5B). This result points towards a closer global transcriptional state between 48H reverting cells and 0H cells.

48H reverting cells retain molecular traces of a commitment into differentiation

To further characterize the molecular changes that persisted after reversion, we sought to identify predictive genes that discriminate the most the 48H reverting cells and the undifferentiated cells. We performed logistic regression combined with dimension reduction (partial

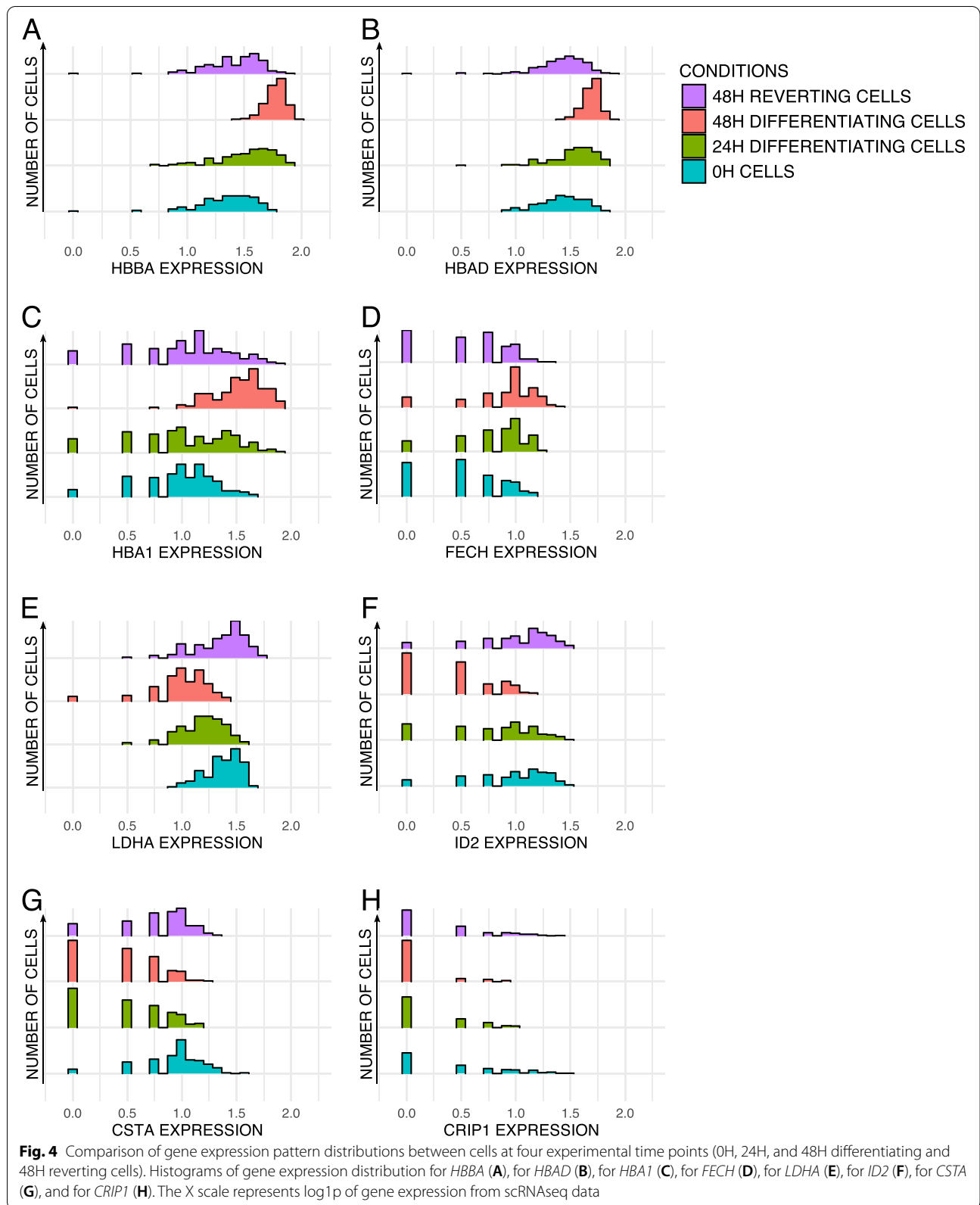
least square [20]) between 48H reverting cells and 0H cells and retained common most discriminating genes between scRT-qPCR and scRNAseq datasets. Interestingly, our results showed that only 3 common genes discriminate between the two cell groups: *HBBA*, *TBC1D7*, and *HSP90AA1*, the expression of which is shown in Fig. 6. *HBBA* is a subunit of the hemoglobin complex which carries oxygen, *TBC1D7* is presumed to have a role in regulating cell growth and differentiation [27], and *HSP90AA1* codes for an isoform of the HSP90 protein chaperone, which its specific transcription is known to be induced in response to insulin [28]. Looking closely, the 48H reverting cells have an intermediate expression level between differentiating cells and undifferentiated cells for the three predicted genes. The offset observed could be due to a longer duration of mRNA half-life at 24H of differentiation. We had previously performed a quantification of mRNA half-life during avian erythrocyte differentiation ([29] Additional file 1: Fig. S3). We focused on mRNA half-life at 24H for those three genes. *TBC1D7* and *HSP90AA1* have a relatively short half-life as opposed to *HBBA*. Other genes analyzed whose expression increases during differentiation, such as *DPP7*, *TPP1*, or *RPL22L1*, have also a long half-life duration mRNA, but only *HBBA* was identified in our statistical analysis as discriminating between undifferentiated and 48H reverting cells.

These results confirmed that the 48H reverting cells display a gene expression pattern very close to those of 0H cells while still retaining traces of their engagement into the differentiation process independently of the mRNA half-life. The molecular process explaining such “lagging genes” will have to be explored.

Cells are distributed as a continuum along the differentiation path

At that stage, two hypotheses could be made: (1) Either all cells have engaged into a differentiation process and do molecularly revert to a self-renewal transcriptional state or (2) at 24H of differentiation two subpopulations coexist: one that is still undifferentiated and would give rise to the 48H reverting cells and a second more differentiated which would lead to the 48H differentiating population and die in the reversion experiment.

We hypothesized that the existence of two subpopulations at 24 h should lead to a higher number of modes in the distribution of some genes at that time point. To test this hypothesis, we therefore estimated for each condition the most-likely number of modes for the probability distribution of each gene, as assessed through a Gamma mixture on scRNAseq (see the “Methods” section). We found no significant difference in the number of modes observed between the 4 populations (Fig. 7), which



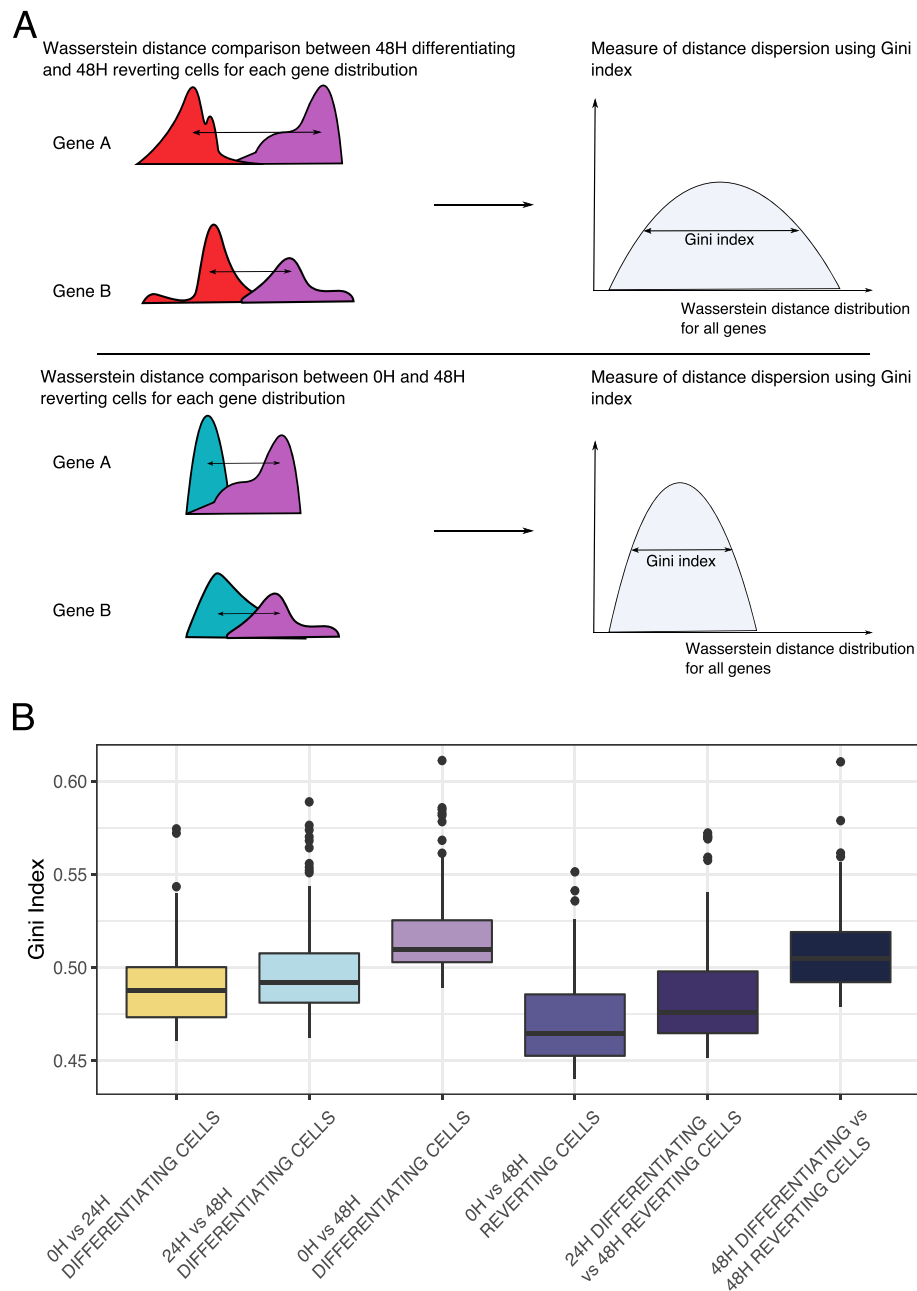
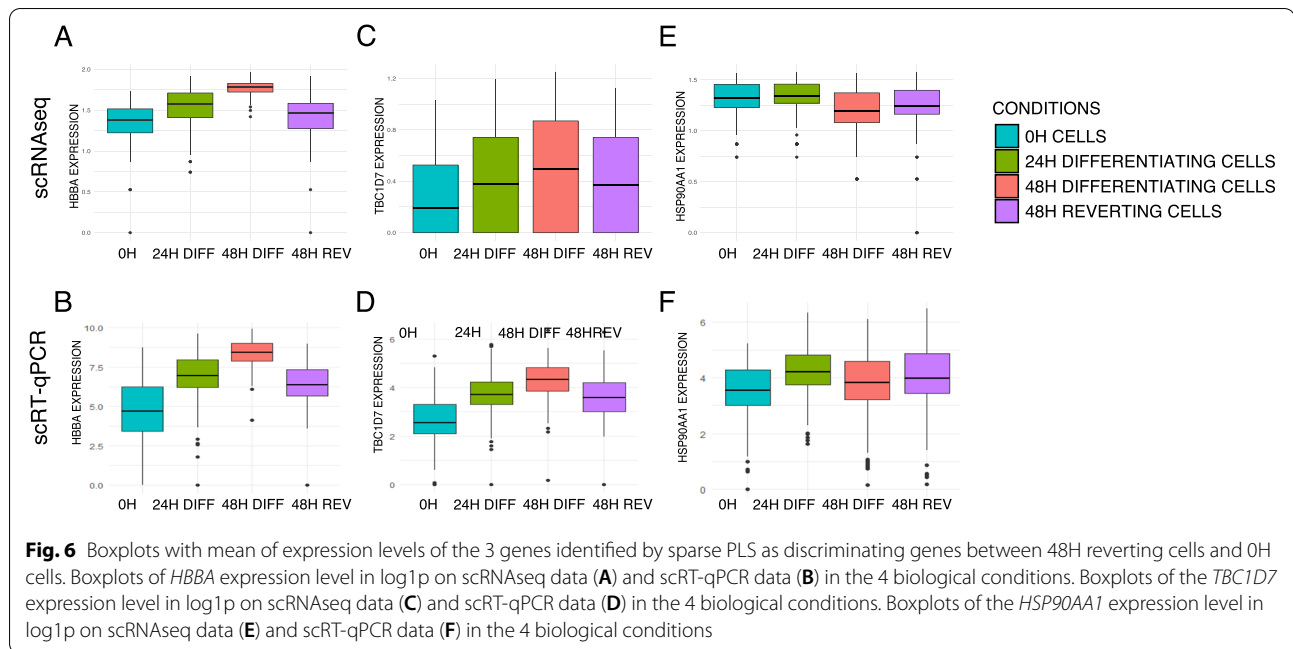


Fig. 5 Comparison of dispersion of gene distribution between cell populations. **A** Experiment design to compare gene distributions between the 4 biological conditions. Wasserstein distance is computed for each gene between pair of conditions, then dispersion of all gene distributions is calculated using Gini index. **B** Plot of Gini index values of Wasserstein distance distributions between conditions in pairs computed for each of the 2000 genes from scRNAseq data bootstrapped 100 times

confirms that the cells collected at 24H do not show more multi-stability than the other groups and are thus unlikely to be a mix of two populations.

The second hypothesis would also imply that in the 24H population, the cells engaged too far in the differentiation process would die a short time after media were

changed, while only the undifferentiated ones would survive. We then measured the viability rate during the kinetics and found no difference in viability between the conditions and especially between the 24H differentiation and the 48H reversion conditions (Additional file 1: Fig. S4).



Finally, the second hypothesis would also imply that the reverting cells are simply cells that have not yet entered the differentiation process. It would therefore be at odds with the evidence that the 48H reverting cells do retain traces of their engagement into the differentiation process (see upper).

Those results strongly suggest that the 24H cell population is not composed of two coexisting subpopulations of cells and that 48H reverting cells enter differentiation before going back to a transcriptomic state close to 0H cells.

Discussion

In the present study, we coupled two different single-cell transcriptomics techniques and state-of-the-art statistical approaches to demonstrate the fate reversibility of avian erythrocyte progenitors induced to differentiate for 24 h.

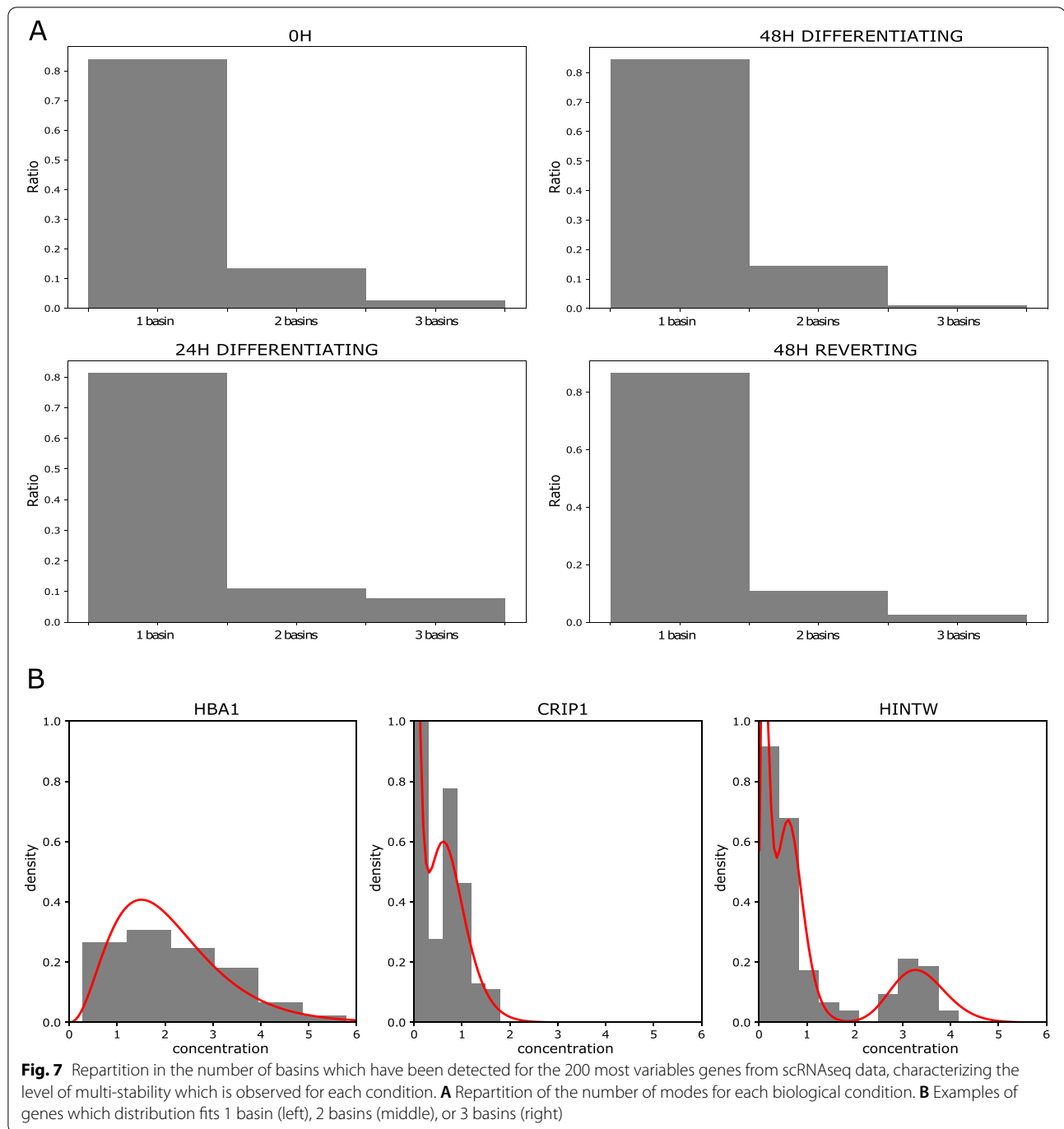
Our results revealed a very close proximity of reverting and undifferentiated cell transcriptomes. Indeed, statistical comparison of cell distributions showed no significant difference between 0H and 48H reverting cells while, as expected, significant changes in gene expression accompanied the differentiation sequence. The analysis of gene expression distribution patterns of the 48H reverting cells confirmed a switch toward the 0H cell gene expression profiles. First, DE analysis of scRNAseq data showed only one gene significantly differentially expressed between the two conditions. Second, Wasserstein distance analysis revealed closer distances between 48H reverting and 0H cells than to any other group of cells. Third, sparse

PLS analysis indicated that the expression level of only three genes, *HBBA*, *TBC1D7*, and *HSP90AA1*, was predictive of the 48H reverting and undifferentiated cells. Interestingly, the persistence of those three genes in 48H reverting cells could not be attributed solely to mRNA half-life duration. However, we cannot exclude that it could be a mere delay and thus a characterization of the reverting cells at a later time point may show a complete molecular reversion.

All of our results therefore favor the hypothesis that a vast majority of the 48H reverting cells responded to differentiation induction by modifying their gene expression profiles but then returned to the self-renewal transcriptional state.

One must note that this would not be the sole example of large-scale transcriptomic changes on (relatively) short time scales [18, 30]. The question as to whether such large-scale transcriptome changes are accompanied, or not, by (reversible) large-scale epigenetic changes remains an open question for future studies.

It has been described in the literature that during cellular decision-making, the cell state is maintained by dynamic interactions between positive and negative regulatory molecules [31] within the frame of a Gene Regulatory Network (GRN). These interactions can be repurposed by changing the stoichiometry of ubiquitous and specific regulatory molecules and factors [11, 13]. In our study, the analysis of gene expression patterns during the reversion process confirmed that the determination of the fate of erythrocyte progenitors is directed by the constraints of the dynamics of the GRN, influenced



by signals emitted by changing conditions of the environment surrounding the cells. In the absence of differentiation signals (or in the presence of self-renewal inducing signals), there is no ratchet in place that would prevent (at least at early stages in our case) the system to return back to its original quasi-steady state. This is in excellent agreement with the previous demonstration that there is a duration threshold for some GRN under which the

system can return back to its original state [32]. This was proposed to allow cells to discriminate between *bona fide* signals and random noise in their environment and could represent a physiological system for finely tuning the in vivo production of red blood cells while preserving the pool of progenitors. We recently proposed a methodology for inferring the GRN underlying T2EC differentiation [29]. For that, we kept in silico cells under constant

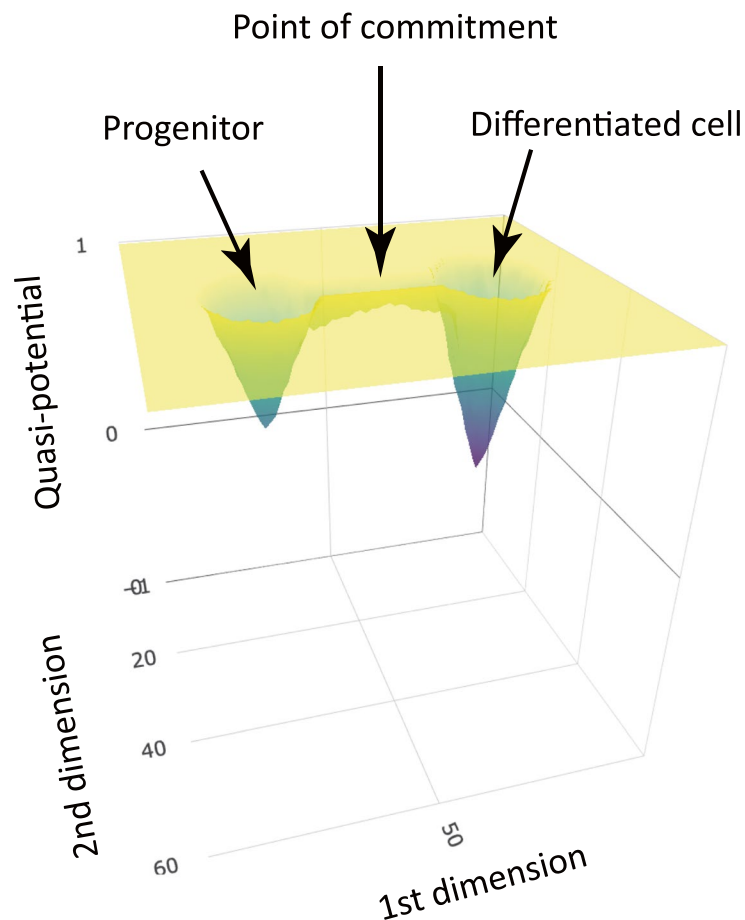


Fig. 8 A quasi-potential well depiction of the erythroid differentiation process. While the cells have not escaped the zone of influence of the progenitor attractor (i.e., when they have not passed the point of commitment, aka the point of no return [19]), the removal of the environmental influences results in their relaxing back to their original attractor state

differentiation stimulus. It would be of interest to see if the inferred GRN would be able to revert, up to a certain point where no “spontaneous” return is possible [19], to its original state. This would be a very strong constraint to impose and should severely limit the number of putative GRN able to reproduce experimental data and thus approaching the most accurate network.

Taken together, our results point towards a physiological plasticity and reversibility with respect to erythrocyte decision-making. It is also reminiscent of the plasticity observed in cancer stem cells that might not be specific to tumor cells [33]. In terms of the epigenetic landscape, our work implies that instead of a continuous gradient that the cells will roll down as in the classical Waddington’s depiction [8], they may go through an unstable state and may, sometimes, roll upwards over a bump in the landscape [34]. Thus, differentiation should be more appropriately described as cells moving from well to well, that is, from one metastable state [35–37] to another one

(Fig. 8). This view abides by the multi-stability framework where a complex quasi-potential landscape aims at describing both normal and pathological differentiation processes [37, 38], and exemplifies the fact that “commitment (is) a dynamical property of the landscape” [39]. It is important at this stage to remember that Waddington himself was aware that his drawing was but a simplification. Adapting and refining this landscape should not be considered as departing from his views. Such a non-monotonous landscape has been proposed to account for the depiction of regeneration in adult tissues [12] and is consistent with previously proposed dynamical principles of cell fate restriction [10]. It is in excellent accordance with the recent depiction that cells can “climb uphill on Waddington’s epigenetic landscape” during cranial neural crest cell development [15] and would also be more relevant to account for the “hesitant” behavior of human CD34+ stem cells in vitro [40] than a straight slope. It is beyond the scope of this discussion to go into more

details, but a cell “climbing uphill” should be seen as equivalent to “the landscape bending into a new valley.”

Conclusions

Our work has provided a detailed molecular characterization of the probabilistic nature of erythrocyte cell fate determination, influenced by the constraints of the underlying Gene Regulatory Network dynamics, and driven by environmental influences.

In conclusion, our results clearly depart from a deterministic view of the differentiation process and fully support the importance of gene expression stochasticity in all systems examined to date [4, 41–44], both in vitro [19, 21, 40, 45–48] and in vivo [49–51].

These new insights into the process of cell reversion could also lead to significant improvements of the executable GRN inference scheme [29].

Methods

Cellular biology

T2EC were extracted from the bone marrow of 19-day-old SPAFAS white leghorn chicken’s embryos (INRA, Tours, France). Cells were grown in self-renewal in a LM1 medium (α -MEM, 10% fetal bovine serum (FBS), 1 mM HEPES, 100 nM β -mercaptoethanol, 100 U/mL penicillin and streptomycin, 5 ng/mL TGF- α , 1 ng/mL TGF- β , and 1 mM dexamethasone) as previously described [52].

Differentiation was induced by removing the LM1 medium and placing the cells into a DM17 medium (α -MEM, 10% fetal bovine serum (FBS), 1 mM Hepes, 100 nM β -mercaptoethanol, 100 U/mL penicillin and streptomycin, 10 ng/mL insulin, and 5% anemic chicken serum (ACS [53])).

Differentiation kinetics were achieved by collecting a sub-fraction of the cells at different times after induction of differentiation (0H and 24H). After 24H, the DM17 medium was removed and half of the cells were placed back into the LM1 medium while the other half was kept in the DM17 medium to achieve 48H reversion and 48H differentiation time points respectively (Fig. 1).

Cell population mortality was assessed by counting dead and living cells from the different time points and conditions after Trypan blue staining and using a Malassez cell.

Single-cell sorting

For both single-cell transcriptomics methods, cells were sorted in 96-well plates using FACS Aria II μ , BD: 8 plates were produced for scRNAseq (2 plates per time point) and 8 plates were produced for scRT-qPCR (2 plates per time point). Since the first steps of library construction are performed per plate, we refer as “batch” the different plates.

Single-cell RT-qPCR analysis

All the manipulations related to the high-throughput scRT-qPCR experiments in microfluidics were performed according to the protocol recommended by the Fluidigm company (PN 68000088 K1, p.157-172). All steps from single-cell isolation to scRT-qPCR, gene selection, data generation, and cleaning are described in detail in [19]. The expression matrix was log_{1p} transformed before subsequent analysis.

Single-cell RNAseq

scRNAseq was performed using an adapted version of the MARSseq protocol [22]. Unless specified, all indicated concentrations correspond to final concentrations.

Individual cells were sorted into 96-well plates containing 4 μ L of lysis buffer and index RT primers (0.2% Triton (Sigma Aldrich), 0.4 U/ μ L RNaseOUT (ThermoFisher Scientific), 400nM RT_primers (Sigma Aldrich)). Index RT_primers (Table 2) contain oligo-dT chain to capture mRNA, a T7 RNA polymerase promoter for further in vitro transcription (IVT), unique cell barcodes for subsequent de-multiplexing, and unique molecular identifiers (UMIs) for PCR bias deduplication. After cell sorting, plates were immediately centrifuged and frozen on dry ice before storage at -80°C until reverse transcription (RT) was performed. The plates were put at 72°C for 3 min for denaturation. A total of 4 μ L of RT mix was added in each well (2mM dNTP (Thermo scientific), 20mM DTT, 2X First stranded buffer, 5 U/ μ L Superscript III RT enzyme (Superscript III RT enzyme kit Thermo scientific), 10% (W/V) PEG 8000 (Sigma Aldrich)). ERCC RNA spike-in (Thermo Scientific) was diluted into the RT mix (dilution 5×10^{-7}). The plates were then transferred into a thermocycler (program: 42°C -2min, 50°C -50min, 85°C -5min, 4°C hold).

Table 2 List and sequences of primers used for scRNAseq library construction

Primer name	5' to 3'
Index_RT_primers (cell BC and UMI)	5'-CGATTGAGGCCGGTAATACGACTCACTATAGGG GCGACG TGTGCTCTCCGATCTXXXXXXXXNNNNNNNTTT TTTTTTTTTTTTTTTTTV-3'
P5N6_XXXX (Plate BC)	5'-CTACACGACGCTCTTCCG ATCTXXXXXXXXNNNN-3'
P5.rd1	5'-AATGATACGGCGACCACCGAGATCTACACTCTT TCC CTACACGACGCTCTTCCGATCT-3'
P7.rd2	5'-CAAGCAGAAGACGGCATACGAGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

After reverse transcription, samples were pooled by plate and Exonuclease I (NEB) digestion was performed, followed by 1.2X AMPure beads purification (Beckman Coulter). Samples were eluted in 10mM Tris-HCl, pH=7.5. Second strand cDNA synthesis was performed with 1X SSS buffer and SSS enzyme (NebNext mRNA second strand synthesis kit NEB; thermocycler program: 16°C-150min, 65°C-20min, 4°C hold). Resulting double-strand cDNA were linearly amplified by IVT overnight (10mM ATP, 10mM GTP, 10mM UTP, 10mM GTP, 1X reaction buffer, 1/10 T7 RNA polymerase mix (HighScribe T7 High Yield RNA synthesis NEB)) at 37°C. IVT products were purified with 1.3X Ampure beads and eluted with 10mM Tris-HCl, 0.1mM EDTA. Amplified RNAs were fragmented (1X RNA fragmentation buffer (RNA fragmentation reagents Invitrogen)) at 70°C for 3 min. The fragmentation reaction was stopped with 34μL of STOP mix (0.3X Stop solution (RNA fragmentation reagents Invitrogen), TE buffer 1X (10mM Tris, 1mM EDTA, pH 8 - Invitrogen), and 0.7X AMPure beads to proceed with sample purification). Differing from the original MARSseq protocol, instead of ligation, a second RT was done to incorporate P5N6 primers (Table 2) containing random hexamers and specific barcodes to distinguish the different plates (5mM DTT, 500μM dNTP, 10μM P5N6_XXXX, 1X First stranded buffer, 10U/μL Superscript III RT enzyme, 2U/μL RNaseOUT; thermocycler program: 25°C 5min, 55°C 20min, 70°C 15min, 4°C hold). The cDNAs were then purified with 1.2x AMPure beads. Illumina primers (Table 2) were added by PCR (0.5 μM Mix primer P5.rd1/P7.Rd2, 1X KAPA Hifi HotStart PCR Mix (Kapa Biosystem); thermocycler program: 95°C 3min, 12 times [98°C 20s, 57°C 30s, 72°C 40s], 72°C 5min, 4°C hold), and PCR products were purified with 0.7x AMPure beads and eluted in 15μL.

Libraries were sequenced on a Next500 sequencer (Illumina) with a custom paired-end protocol to avoid a decrease of sequencing quality on read1 due to the high number of T added during polyA reading (130pb on read1 and 20pb on read2). We aimed for a depth of 200,000 raw reads per cell.

Bio-informatic pipeline

Fastq files were pre-processed through a bio-informatic pipeline developed in the team on the Nextflow platform [54]. Briefly, the first step removed Illumina adaptors. The second step de-multiplexed the sequences according to their plate barcodes. Then, all sequences containing at least 4T following cell barcode and UMI were kept. Using UMIttools whitelist, the cell barcodes and UMI were extracted from the reads. The sequences were then mapped on the reference transcriptome

(Gallus GallusGRCG6A.95 from Ensembl) and UMI were counted. Finally, a count matrix was generated for each plate.

Data filtering, normalization, and analysis

All analyses were carried out using R software (version 4.0.5; [55]). Matrixes from the eight plates were pooled together. Cells were filtered based on several criteria: reads number, gene number, count number, and ERCC content. For each criterion, the cutoff values were determined based on SCONE [56] pipeline and were calculated as follows:

$$\text{mean} - 3 * \text{SD}$$

We selected genes present in at least two cells. The filtered matrix was then normalized using SCTransform from the Seurat package [57] and we corrected for batch effect, time effect, and sequencing depth effect. The expression matrix was finally log1p transformed.

Variable genes were identified using FindVariableFeatures from Seurat, vst method [58]. Based on visualization of gene variance, we retained the 2000 most variable features. Differentially expressed genes were identified using the FindMarkerGenes function from Seurat [58]. Analysis was done by pairwise comparisons between conditions; genes with log fold change ≥ 0.5 and adjusted p -value < 0.05 were kept as significant. More information on QC filtering is given in Additional file 1: Fig. S5.

Statistical analysis

All statistical analyses were performed using the R software (version 4.0.5; [55]). Dimensionality reduction and visualization were performed using UMAP [59]. UMAP was performed directly on the 2000 most variable genes (from the scRNAseq dataset) or 83 genes (from the scRT-qPCR dataset) using default parameters. PCA was performed using prcomp function from the stats R package (version 3.6.2). Adaptive sparse PLS for logistic regression was performed using the plsgenomics package [20]. For this analysis, scRT-qPCR data were scaled. Sparse PLS is a supervised statistical analysis that allows to predict the most discriminant variables between two groups.

Wasserstein distance computation was done using the Transport R package [60] and was accomplished for each gene of the scRNAseq dataset.

Gini indexes were calculated using the Ineq R package on Wasserstein distance distributions [61].

Bootstrapping was done using the sample_frac function from the Dplyr R package [62].

Estimation of multi-stability levels

For estimating the level of multi-stability in the data, we considered that the probability distribution of each

gene can be approximated by a Gamma distribution, or a mixture of Gamma distributions, since they are known to describe continuous single-cell data accurately [63]. More precisely, we parameterized the distribution of a gene i by:

$$p_i^m(x) = \sum_{j=1}^m \mu(j) \frac{x^{a_i^j-1} b_i^{a_i^j} e^{-b_i x}}{\Gamma(a_i^j)},$$

where Γ denotes the Gamma function: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Note that only the parameters $(a_i^j)_{j=1, \dots, m}$ depend on the mixture component j : this is related to the distribution arising from the well-established two-state model of gene expression [64], when only the frequency of mRNA bursts is regulated, as described in [65].

For every condition, we constructed 10 training sets consisting of 80% of the cells in the population (randomly chosen), and we estimated the parameters $[(a_i^j)_{j=1, \dots, m}, b_i]$ with a MCMC algorithm for the numbers of mixture components $m = 1, 2, 3$ successively. We then considered that the optimal number of components for gene i was the one which minimized the average BIC score estimated on the 10 corresponding test sets.

Multivariate two-sample test

Samples were compared using a multivariate two-sample test based on the 2000 most variable genes. We suppose that the normalized gene expression X_1 and X_2 of two conditions (0H vs 48H reversion, 0H vs 48H differentiation, 48H reversion vs 48H differentiation) follow a multivariate Gaussian distribution $\mathcal{N}(\mu_1, \Sigma)$ and $\mathcal{N}(\mu_2, \Sigma)$ respectively, and we denote by $n = n_1 + n_2$ the total number of cells. Then, we test the null hypothesis $H_0: \mu_1 = \mu_2$ using the generalized Hotelling's T^2 test [23]. The data being high dimensional ($p > n$), the between-gene pooled covariance matrix is not invertible and is replaced by its Moore-Penrose inverse. In this setting, the asymptotic distribution of the generalized Hotelling statistics is $\chi^2(n-2)$. The p -values were adjusted according to the Benjamini-Hochberg correction [66]. Analysis was performed using the `fdahotell` R package [67].

Abbreviations

ACS: Anemic chicken serum; DE: Differential expression; GRN: Gene Regulatory Networks; MARSseq: Massively parallel single-cell RNA-Seq; PCA: Principal Component Analysis; scRNAseq: Single-cell RNA sequencing; scRT-qPCR: Single-cell reverse transcription-quantitative polymerase chain reaction; SD: Standard deviation; sparse PLS: Sparse partial least square.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01363-7>.

Additional file 1: Figure S1. [UMAP visualization of scRT-qPCR and scRNAseq data]. **Figure S2.** [PCA of scRT-qPCR and scRNAseq data]. **Figure S3.** [Half-life of mRNA table]. **Figure S4.** [Histograms of viability rate during reversion and differentiation processes]. **Figure S5.** [Summary table of scRNAseq data filtering steps and threshold values for each step].

Acknowledgements

We gratefully thank all members of the SBDM team and particularly Gerard Benoit for very fruitful discussions, suggestions, and commentaries on our project. We also thank Ghislain Durif for his great technical support during PLS computation and G. Yvert for helpful comments about the manuscript. We thank the computational center of IN2P3 (Villeurbanne/France) and Pôle Scientifique de Modélisation Numérique (PSMN, Ecole Normale Supérieure de Lyon) where computations were performed. We acknowledge the contribution of the AniRA-Cytométrie core facility of SFR BioSciences (UAR3444/US8). We thank the BioSyl Federation and the LabEx Ecofect (ANR-11-LABX-0048) of the University of Lyon for inspiring scientific events.

Authors' contributions

SZ and CF contributed equally to the conceptualization of the study, data generation, data analysis, statistical analysis, and interpretation, as well as writing the manuscript. CF participated in the pipeline development. EV1 provided technical support for scRT-qPCR and scRNAseq experiments. LM provided support for data analysis and participated to the pipeline development. RS participated to the pipeline development. AM provided support for the scRNAseq protocol. EV2, MB, AO, and FP performed the statistical analysis. AB participated to the conceptualization of the study. OG performed data analysis and obtained the funding. OG and SG participated to the conceptualization of the study, the project administration, the project supervision, and writing of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by funding from the French agency ANR (SinCity; ANR-17-CE12-0031).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information file and publicly available repositories. Pipelines and analysis scripts are available at https://gitbio.ens-lyon.fr/LBMC/sbdm/mars_seq. Previous scRT-qPCR data are available in the SRA repository <http://www.ncbi.nlm.nih.gov/sra/SRP076011> [19]. Previous data mRNA half-life are available in the OSF repository <https://osf.io/gkedt/> [29]. The datasets supporting the conclusions of this article are available in the NIH repository, accession number PRJNA802343 [68], and in the OSF repository <https://osf.io/upw8d/>.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratory of Biology and Modelling of the Cell, Université de Lyon, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, Université Claude Bernard Lyon 1, Lyon, France. ²Azm Center for Research in Biotechnology and its Applications, LBA3B, EDST, Lebanese University, Tripoli 1300, Lebanon. ³Ecole Pratique des Hautes Etudes, PSL Research University, UMR5951, INSERM, Univ-Evry, Paris, France. ⁴Inria Team Dracula, Inria Center Grenoble Rhone-Alpes,

Grenoble, France. ⁵Institut Camille Jordan, CNRS UMR 5208, Université Claude Bernard Lyon 1, Villeurbanne, France. ⁶Vidium solutions, Lyon, France. ⁷Nantes Université, Centrale Nantes, Laboratoire de mathématiques Jean Leray, LMJL, F-44000 Nantes, France.

Received: 7 March 2022 Accepted: 27 June 2022

Published online: 06 July 2022

References

- Wolpert L. Do we understand development? *Science*. 1994;266:571–2.
- Elowitz MB. Stochastic gene expression in a single cell. *Science*. 2002;297:1183–6.
- Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet*. 2002;31:69–73.
- Symmons O, Raj A. What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol Cell*. 2016;62:788–802.
- Kolodziejczyk AA, Lönnberg T. Global and targeted approaches to single-cell transcriptome characterization. *Brief Funct Genomics*. 2018;17:209–19.
- Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol*. 2009;7:e1000149.
- Blake WJ, Balázs G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell*. 2006;24:853–65.
- Waddington CH. *The strategy of the genes*. 1st ed: Routledge; 1957.
- Shi J, Teschendorff AE, Chen W, Chen L, Li T. Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. *Brief Bioinform*. 2018;21:248–61.
- Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet*. 2016;17:693–703.
- Baron MH. Reversibility of the differentiated state in somatic cells. *Curr Opin Cell Biol*. 1993;5:1050–6.
- Rajagopal J, Stanger BZ. Plasticity in the adult: how should the Waddington diagram be applied to regenerating tissues? *Dev Cell*. 2016;36:133–7.
- Johnson NC, Dillard ME, Baluk P, McDonald DM, Harvey NL, Frase SL, et al. Lymphatic endothelial cell identity is reversible and its maintenance requires Prox1 activity. *Genes Dev*. 2008;22:3282–91.
- Ladewig J, Koch P, Brüstle O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol*. 2013;14:225–36.
- Zalc A, Sinha R, Gulati GS, Wesche DJ, Daszczuk P, Swigut T, et al. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science*. 2021;371:eabb4776.
- Buczacki SJA, Zecchini HI, Nicholson AM, Russell R, Vermeulen L, Kemp R, et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature*. 2013;495:65–9.
- Tata PR, Mou H, Pardo-Saganta A, Zhao R, Prabhu M, Law BM, et al. Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature*. 2013;503:218–23.
- Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*. 2008;453:544–7.
- Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol*. 2016;14:e1002585 <http://www.ncbi.nlm.nih.gov/sra/SRP076011>.
- Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P. A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol*. 2008;7(1).
- Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, et al. Cell fate decision as high-dimensional critical state transition. *PLoS Biol*. 2016;14:e2000640.
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science*. 2014;343:776–9.
- Secchi P, Stamm A, Vantini S. Inference for the mean of large p small n data: a finite-sample high-dimensional generalization of Hotelling's theorem. *Electron J Stat*. 2013;7:2005–31.
- Richard A, Vallin E, Romestaing C, Roussel D, Gandrillon O, Gonin-Giraud S. Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements. *PLoS One*. 2019;14:e0221472.
- Yokota Y, Mori S, Narumi O, Kitajima K. In vivo function of a differentiation inhibitor, Id2. *IUBMB Life*. 2001;51:207–14.
- Klenova EM, Botezato I, Laudet V, Goodwin GH, Wallace JC, Lobanenkov VV. Isolation of a cDNA clone encoding the RNASE-superfamily-related gene highly expressed in chicken bone marrow cells. *Biochem Biophys Res Commun*. 1992;185:231–9.
- Dibble CC, Elis W, Menon S, Qin W, Klekota J, Asara JM, et al. TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1. *Mol Cell*. 2012;47:535–46.
- Zuehlke AD, Beebe K, Neckers L, Prince T. Regulation and function of the human HSP90AA1 gene. *Gene*. 2015;570:8–16.
- Bonnafox A, Herbach U, Richard A, Guillemin A, Gonin-Giraud S, Gros P-A, et al. WASABI: a dynamic iterative framework for gene regulatory network inference. *BMC Bioinformatics*. 2019;20:220 <https://osf.io/gkedt/>.
- Nichols JM, Antolović V, Reich JD, Brameyer S, Paschke P, Chubb JR. Cell and molecular transitions during efficient dedifferentiation. *eLife*. 2020;9:e55435.
- Blau HM. Differentiation requires continuous active control. *Annu Rev Biochem*. 1992;61:1213–30.
- Sokolik C, Liu Y, Bauer D, McPherson J, Broecker M, Heimberg G, et al. Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst*. 2015;1:117–29.
- Thankamony AP, Saxena K, Murali R, Jolly MK, Nair R. Cancer stem cell plasticity – a deadly deal. *Front Mol Biosci*. 2020;7:79.
- Kimmel JC, Yi N, Roy M, Hendrickson DG, Kelley DR. Differentiation reveals latent features of aging and an energy barrier in murine myogenesis. *Cell Rep*. 2021;35:109046.
- Guillemin A, Stumpf MPH. Noise and the molecular processes underlying cell fate decision-making. *Phys Biol*. 2021;18:011002.
- Pisco AO, Fouquier d'Hérouël A, Huang S. Conceptual confusion: the case of epigenetics. preprint. bioRxiv. 2016:053009. <https://doi.org/10.1101/053009>.
- Ventre E, Espinasse T, Bréhier C-E, Calvez V, Lepoutre T, Gandrillon O. Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastability. *J Math Biol*. 2021;83:59.
- Huang S. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*. 2009;31:546–60.
- Sáez M, Blassberg R, Camacho-Aguilar E, Siggia ED, Rand DA, Briscoe J. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Syst*. 2022;13:12–28.e3.
- Moussy A, Cossette J, Parmentier R, da Silva C, Corre G, Richard A, et al. Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLoS Biol*. 2017;15:e2001867.
- Smith S, Grima R. Single-cell variability in multicellular organisms. *Nat Commun*. 2018;9:345.
- Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. *Nat Rev Genet*. 2019;20:536–48.
- Losick R, Desplan C. Stochasticity and cell fate. *Science*. 2008;320:65–8.
- Paldi A. Stochastic or deterministic? That is the Question. *Org J Biol Sci*. 2020;4:77–9.
- Guillemin A, Duchesne R, Crauste F, Gonin-Giraud S, Gandrillon O. Drugs modulating stochastic gene expression affect the erythroid differentiation process. *PLoS One*. 2019;14:e0225166.
- Moris N, Edri S, Seyres D, Kulkarni R, Domingues AF, Balayo T, et al. Histone acetyltransferase KAT2A stabilizes pluripotency with control of transcriptional heterogeneity. *Stem Cells Dayt Ohio*. 2018;36:1828–38.
- Stumpf PS, Smith RCG, Lenz M, Schuppert A, Müller F-J, Babbie A, et al. Stem cell differentiation as a non-Markov stochastic process. *Cell Syst*. 2017;5:268–282.e7.

48. Stockholm D, Edom-Vovard F, Coutant S, Sanatine P, Yamagata Y, Corre G, et al. Bistable cell fate specification as a result of stochastic fluctuations and collective spatial cell behaviour. *PLoS One*. 2010;5:e14441.
49. Wernet MF, Mazzoni EO, Çelik A, Duncan DM, Duncan I, Desplan C. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*. 2006;440:174–80.
50. Dussiau C, Boussaroque A, Gaillard M, Bravetti C, Zaroili L, Knosp C, et al. Hematopoietic differentiation is characterized by a transient peak of entropy at a single-cell level. *BMC Biol*. 2022;20:60.
51. Toh K, Saunders D, Verd B, Steventon B. Zebrafish neuromesodermal progenitors undergo a critical state transition *in vivo*. *bioRxiv*. 2022.02.25.481986. <https://doi.org/10.1101/2022.02.25.481986>.
52. Gandrillon O, Schmidt U, Beug H, Samarut J. TGF- β cooperates with TGF- α to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *EMBO J*. 1999;18:2764–81.
53. Gandrillon O, Samarut J. Role of the different RAR isoforms in controlling the erythrocytic differentiation sequence. Interference with the v-erbA and p135gag-myb-ets nuclear oncogenes. *Oncogene*. 1998;16:563–74.
54. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35:316–9.
55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2021.
56. Cole MB, Risso D, Wagner A, DeTomaso D, Ngai J, Purdom E, et al. Performance assessment and selection of normalization procedures for single-cell RNA-Seq. *Cell Syst*. 2019;8:315–328.e8.
57. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296.
58. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single cell data. *Cell*. 2019;177(7):1888–1902.e21.
59. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018;37:38–44.
60. Schuhmacher D, Bähre B, Gottschlich C, Hartmann V, Heinemann F, Schmitzer B. transport: computation of optimal transport plans and Wasserstein distances. R package version 0.12-2. 2020. <https://cran.r-project.org/package=transport>.
61. Zeileis A. ineq: measuring inequality, concentration, and poverty; 2014.
62. Hadley Wickham, Romain François, Lionel Henry, Kirill Müller. dplyr: a grammar of data manipulation. 2021.
63. Albayrak C, Jordi CA, Zechner C, Lin J, Bichsel CA, Khammash M, et al. Digital quantification of proteins and mRNA in single mammalian cells. *Mol Cell*. 2016;61:914–24.
64. Peccoud J, Ycart B. Markovian modeling of gene-product synthesis. *Theor Popul Biol*. 1995;48(2):222–34.
65. Ventre E. Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics. *In Silico Biol*. 2021;14(3–4):89–113.
66. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57:289–300.
67. Stamm A, Pini A, Vantini S. fdahotelling: inference for functional data analysis in R. R (>= 3.1.3). <https://github.com/astamm/fdahotelling>.
68. Molecular characterization of pre-commitment cell reversion during erythroid differentiation. NCBI BioProject accession: PRJNA802343. 2022. <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA802343>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

