The basal level of gene expression associated with chromatin loosening shapes Waddington landscapes and controls cell differentiation

Gilles Flouriot[1], Charly Jehanno[2], Yann le Page[1], Pascale le Goff[1], Benjamin Boutin[3], Denis Michel[1]*

[1] Univ Rennes, Inserm, EHESP, Irset UMR 1085, Rennes, France

[2] University of Basel, Department of Biomedicin, Basel, Switzerland

[3] Univ Rennes, Institut de Recherches Mathématiques de Rennes, France

* Corresponding author: denis.michel@live.fr.

Abstract

The baseline level of transcription is variable and seriously complicates the normalization of comparative transcriptomic data, but its biological importance remains unappreciated. We show that this ingredient is in fact crucial for the interpretation of molecular biology results. It is correlated to the degree of chromatin loosening measured by DNA accessibility, and systematically leads to cellular dedifferentiation as assessed by transcriptomic signatures, irrespective of the molecular and cellular tools used. A theoretical analysis of gene circuits formally involved in differentiation, reveals that the epigenetic landscapes of Waddington are restructured by the level of non-specific expression, such that the attractors of progenitor and differentiated cells can be mutually exclusive. Together, these results unveil a generic principle of epigenetic landscape remodeling in which the basal gene expression level, notoriously important in pluripotent cells, allows the maintenance of stemness by generating a specific landscape and in turn, its reduction favors multistability and thereby differentiation. This study highlights how heterochromatin maintenance is essential for preventing pathological cellular reprogramming, age-related diseases and cancer.

Introduction

Data from the litterature show that basal expression, chromatin loosening and stemness, are intimately connected phenomena: **(i)** Stem cell chromatin is loosened compared to that of differentiated cells [1, 2, 3] and the differentiation of stem cells is accompanied by the progressive condensation of their chromatin [4]. **(ii)** A high level of basal expression is a hallmark of stem cells, distinguishing them from their terminally differentiated counterparts [5, 6]. **(iii)** The histone mark H3K9me3 associated to closed chromatin, prevents reprogramming [7, 8]. Its inhibition forbids differentiation [4] whereas its forced demethylation facilitates reprogramming [9, 10]. **(iv)** H3K9 acetylation characterizes puripotency and reprogramming capacity [11]. **(v)** More generally, opening chromatin by inhibition of DNA methyltransferases and histone deacetylases, improves the induction of pluripotent stem cells [12]. Such observations have also been reported for specialized cases of terminal differentiation. For instance, a defect of the H3K9 trimethylase Suv39h1 maintains the re-programming capacity of CD8$^+$ lymphocytes [13] and chromatin acetylation induces the developmental plasticity of oligodendrocyte precursors [14]. Long before these studies, it had already been shown that cellular differentiation is associated to an overall loss of DNA accessibility, measured experimentally with DNAseI [15]. This impressive list of convergent observations with stem and progenitor cells can be further extended to pathological cases of dedifferentiation, notably cancer. On the one hand, cancer cell chromatin is globally decondensed, with demethylated DNA and acetylated nucleosomes, except at the level of tumor supressors. On the other hand, the aggressiveness of cancer is correlated with the degree of dedifferentiation and cell state plasticity, allowing for example cells originating from the mammary epithelium to forget their initial identity, escape hormonal control and acquire migratory properties [16]. These systematic correlations prompted us to look for an underlying principle rooted in the physics of genetic networks. The dedicated tool for this purpose is naturally the Waddington epigenetic landscape, long envisioned as the ideal framework

for conceptualizing cell differentiation and development. An epigenetic landscape, in the sense initiated by Waddington [17], is a n-dimensional potential surface shaped by the mutual compatibility or incompatibility of the concentrations of the n cellular components. Indeed, the basic principle of cellular systems is that the different macromolecules can not be present in arbitrary relative concentrations in the cell, because of the internal constraints of reticulated networks. The most direct interactions are mediated by transcription factors (TFs) and the most widely studied interaction networks are gene regulatory networks (GRN). Epigenomic and transcriptomic profiles found in large datasets emerge from such underlying circuits. Under certain conditions which are fulfilled in living systems, including positive loops and nonlinear interactions, several steady states can coexist in the landscape and the system is said to be multistable. In this picture of generalized interactions, the existing cellular phenotypes correspond to the possible discrete combinations, like barcodes, of cellular components defining the bottom of the basins in the landscape. These local minima are steady states in which the different nodes of the network remain stable. All the other combinations, falling on the ridges or the sides of the mountains, are unstable and automatically pulled down by restoring forces to a basin of attraction located in the vicinity. This view illuminated our understanding of development and cellular differentiation, conceived as emerging from GRNs and biochemical circuits [18, 19, 20]. In this framework, cellular differentiation is underlain by phase space translocation of gene regulatory systems from median attractors with generalized gene expression, to border attractors with selective gene expression. The former are supposed metastable and less resistant to fluctuations whereas the latters are classically considered stable [21], but experiences show that reprogramming differentiated cells generally remains possible and that conversely, multipotent cells can persist indefinitely in culture and in the organisms [21, 22]. Consistent with these observations, a combination of experimental and theoretical approaches reveals a new general principle governing cellular differentiation, in which stemness attractors dominated by median attractors, remain stable as long as the ratio of basal vs regulated transcription is high. In turn, the lateral attractors with selective gene expression and characterizing terminally differentiated phenotypes, progressively deepen when lowering basal expression.

Results

To examine the relation between basal expression and chromatin compaction, we developed cellular/molecular instruments using mutant versions of the myocardin-related TF (MRTFA/MKL1).
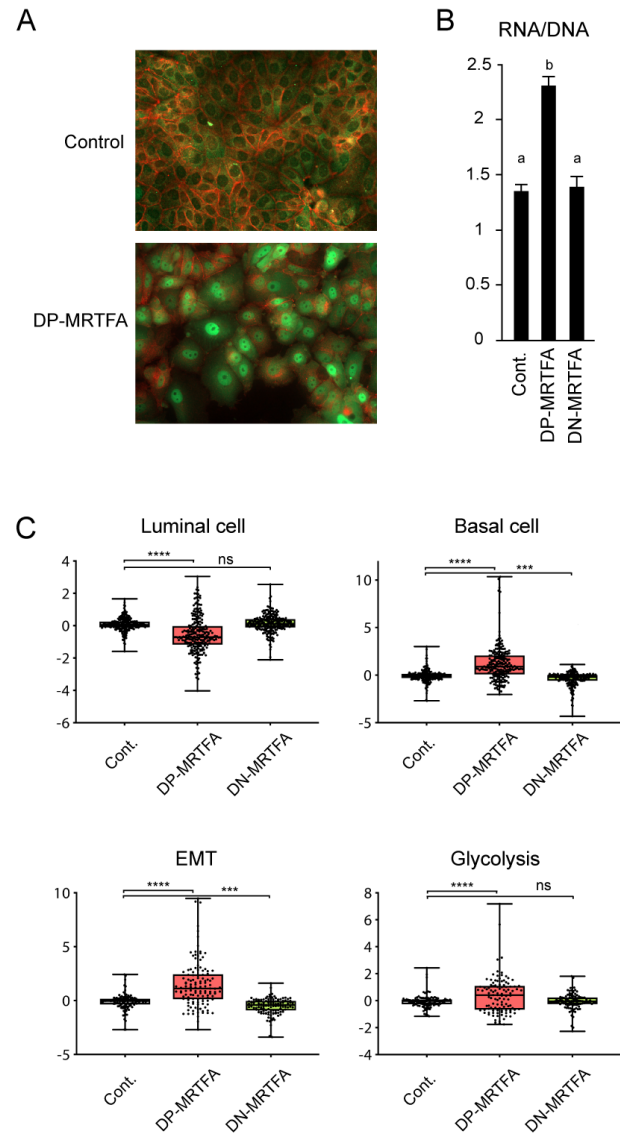
Fig. 1 (**A**) Forced dedifferentiation of MCF7 cells using DP-MRTFA. Immunofluorescence image of E-cadherin (red) and MRTFA (green) in control and DP-MRTFA expressing MCF7 cells. DP-MRTFA causes the dismantlement of pseudoepithelial intercellular contacts, with loss of pericellular E-Cadherin. (**B**) Compared overall transcription of control and MRTFA cells, determined by the ratio of cellular content in RNA over DNA. Columns with different superscripts differ significantly ($p < 0.05$). (**C**) Boxplot of the log2 mean expression values obtained from microarray experiment on MCF7 control, DP-MRTFA and DN-MRTFA cells for luminal cell, basal cell and EMT signature genes. *P-value < 0.05, **P-value < 0.01 and ***P-value < 0.001 with a t-test for comparisons. Error bars represent SD.
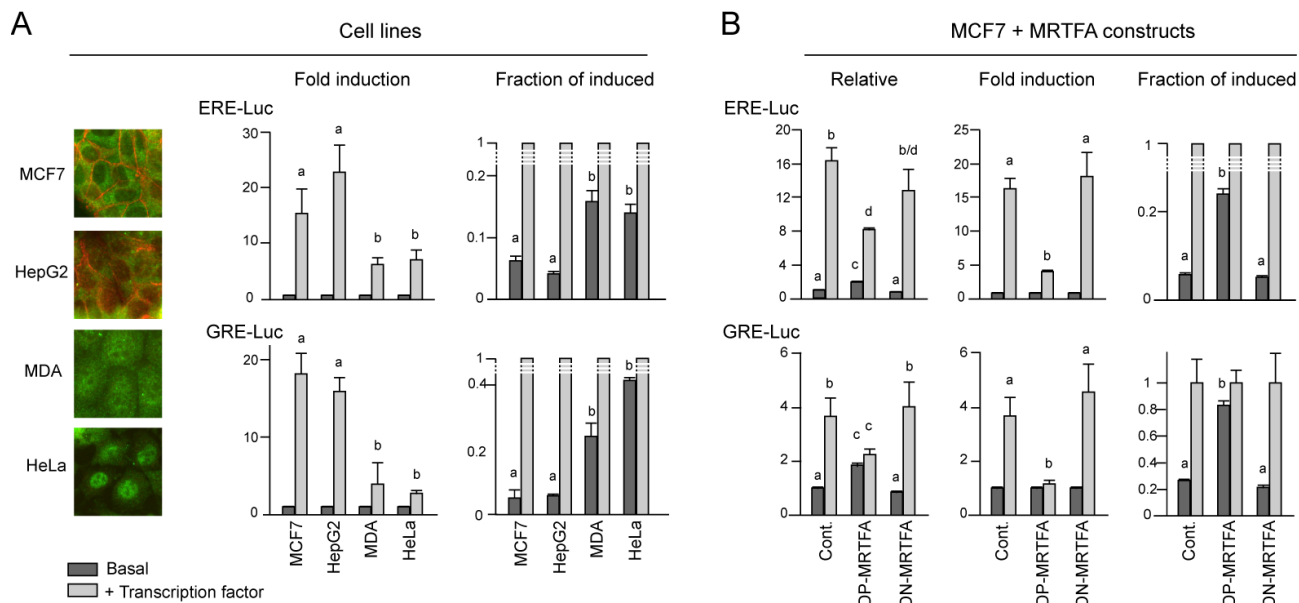
Fig. 2. Estrogen or glucocorticoid receptor-mediated transcription (**A**) in cell lines characterized by their differentiated (MCF7, HepG2) or dedifferentiated (MDA, HeLa) phenotypes or (**B**) in MCF7 cells expressing or not mutant MRTFA constructs. The insets pictures in Panel **A** show the cytoimmunostaining of E-cadherin (red) and endogenous MRTFA (green) in the four cell lines used. In MDA and HeLa cells, pericellular E-cadherin is absent and MRTFA is mainly nuclear. All cells were transfected with the ERE-LUC or GRE-LUC reporter constructs, together with CMV-βGal and 50 ng of empty, ER or GR expression vectors. MCF7 were further transfected with 100 ng MRTFA expression vectors. Cells were then treated with 10 nM of estradiol or dexamethasone. 36 hours after transfection, luciferase activities were measured and normalized with β-galactosidase. Data correspond to the average +/- SEM of at least three separate transfection experiments. Results are expressed as either fold induction, by setting the uninduced expression level to 1, or as the fraction of the induced level, obtained by setting to 1 the induced expression. Relative expression, in the left histograms of Panel **B**, is fold changes above levels measured in control MCF7 cells. Columns with different superscripts differ significantly ($p < 0.05$).

MRTFA has been shown to participate to a transcriptional cocktail of stemness in breast cancer cells [23] and to erase the initial differentiation status of cells [24]. But as for most biological molecules, the function of MRTFA is finely regulable, for instance by its level of expression, subcellular location and interaction partners, making it difficult to manipulate. We showed that it is possible to impose to MRTFA clear-cut functions by deletion of specific interaction domains. Overexpression of a dominant positive mutant version (DP-MRTFA) devoid of cytoplasm-anchoring domain, constitutively nuclear and transcriptionally active, leads to global chromatin decondensation and induces stem cell marks such as bivalent chromatin [16]. By contrast, an other mutant, dominant-negative (DN-MRTFA) devoid of transactivation domain, tightens chromatin and strengthens the differentiated phenotype [16]. As shown in Fig.1A, marked phenotypic changes are induced by DP-MRTFA, with disruption of pericellular E-cadherin and of the pseudo-epithelial structure of cell monolayers.

Dedifferentiation, chromatin decondensation and basal expression are coupled phenomena.

Global transcriptomic studies allowed to identify the modifications of gene expression occuring in these cells in term of signatures (Fig.1C). DP-MRTFA caused a loss of differentiation characteristics accompanied by a clear emergence of basal cell and epithelial to mesenchymal transition (EMT) signatures, which characterize mammary stem cells and epithelial de-differentiation respectively. A phenomenon regularly associated with EMT: a switch of energy metabolism to glycolysis, is also obtained with DP-MRTFA. For comparison, DN-MRTFA induces no such modifications of gene expression. The changes induced by DN-MRTFA are globally inverse to those of DP-MRTFA, but moderate. These less visible phenotypic and genetic changes could be due to the fairly differentiated nature of the starting MCF7 cells. DP-MRTFA expressing cells specifically contain large amounts of RNA per cell (mRNA + rRNA + small RNAs), as mea-

3

sured using the perchloric acid precipitation procedure, indicative of increased overall transcription (Fig.1B). This hypertranscription, which is another feature common to stem cells [25], could be due either to a strong increase in the specific expression of certain genes, or to a global increase of non-specific gene expression. It is technically challenging to decide between these two possibilities and to compare gene expression between cells, because basal expression is generally unnoticed in current experimental approaches. It is ignored in transcriptome-type techniques where the results are expressed per unit mass of RNA, and/or are calibrated using gene expression supposed invariant between the situations. Since transfected DNA is packaged into nucleosomal structures similar to native chromatin [26], transient expression assays are expected to incorporate basal expression, but they have also some pitfalls: (i) It is first necessary to ensure the equivalence of transfection efficiency between the different culture conditions or the cell types to be compared. (ii) It is then necessary to find the most appropriate point of reference to quantify transcriptional changes. (i) The first requirement can in principle be satisfied by co-transfection of a neutral expression vector, provided its expression is not influenced by the more or less permissive nuclear context of each cell. A suitable internal control for this purpose is a strong promoter capable of abstracting itself from repressive contexts [27], such as the cytomegalovirus promoter (CMV) selected here. (ii) The second point is more subtle. Transcriptional results are generally presented as "fold induction" by arbitrarily setting the uninduced condition to 1, but this presentation could introduce a bias in the interpretation of the results, since the basal (uninduced) expression level depends on the cellular context. To highlight this phenomenon, in Fig.2, the same results are presented in different manners. In these experiments, the transcriptional induction of reporter vectors is directed by either estrogen receptor (ERE) or glucocorticoid receptor response elements (GRE). ERE- and GRE-driven reporter plasmids were transfected with or without expression vectors of their respective inducers: ERα with estradiol for ERE-Luc, and GR with dexamethasone for GRE-Luc. The cells tested where either human cell lines with different degrees of differentiation (Fig.2A) and MCF7 cells expressing MRTFA constructs (Fig.2B). The left histograms in Fig.2A show the results traditionally presented in fold induction, after fixing the basal expression level to 1 for each cell type. Presented in this way, the results suggest that ERα and GR appear less potent in dedifferentiated cells; but setting the induced level to 1, rather suggests that basal expression strongly increases in these cells. It is all the more difficult to decide which conclusion is the right one, that the compared cellular contexts largely differ. To bypass this problem of comparison, the cotransfection procedure was then applied to MCF7 cells only, which allowed to verify that the dedifferentiating

construct DP-MRTFA actually increases basal expression (Fig.2B). The reinterpretation of results in term of variation in basal expression instead of modified induction, is unusual in the literature, which could explain why the role of basal expression is generally overlooked. Comparison of Fig.1 and Fig.2B shows that the increase vs decrease of basal expression is correlated with the tendency of dedifferentiation vs differentiation.
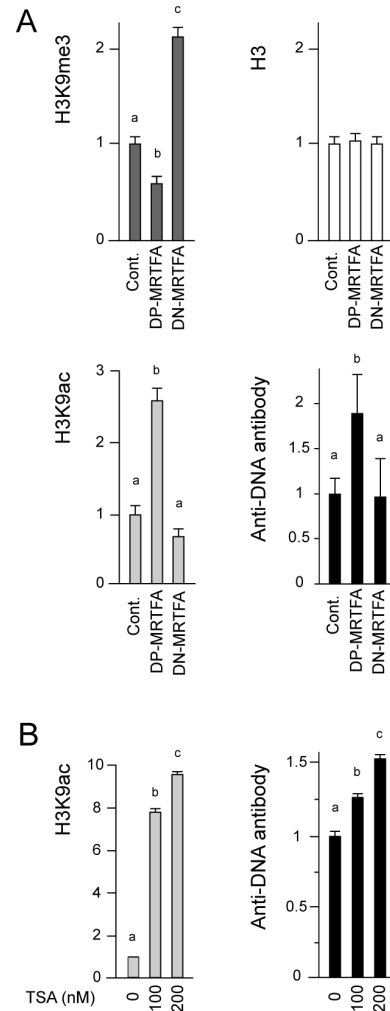

Fig. 3. (**A**) MRTFA constructs alter the global levels of H3K9 trimetylation, H3K9 acetylation and DNA accessibility, evaluated by fluorescent staining using anti-dsDNA antibody, in MCF7 sub-clones 48 hours after tetracycline treatment to induce MRTFA transgenes expression. (**B**) TSA treatment induces H3K9 acetylation and DNA accessibility. Results are expressed as the fold change above levels measured in untreated MCF7 cells. Columns with different superscripts differ significantly ($p < 0.05$).

When looking for a mechanism possibly underlying both chromatin decondensation and increase of basal expression,

the most obvious candidate is the degree of chromatin acety-lation. Quantification of the antagonistic chromatin marks H3K9ac and H3K9me3 shows a marked increase in acetylation in DP-MRTFA expressing cells and inverse variations in DN-MRTFA expressing cells (Fig.3A). As a control, the inhibitor of histone deacetylases HDAC trichostatin A (TSA), induces potent H3K9 acetylation, as expected (Fig.3B). Histone acetylation has long been shown to alleviate electrostatic interactions between nucleosomes and DNA, mechanically causing chromatin loosening, which is in turn expected to promote the accessibility of DNA to large proteins such as TFs. To test this hypothesis, we quantified the general accessibility of DNA with a large DNA binding molecule, an antibody directed against double stranded DNA. Remarkably, a significant increase of accessibility was obtained with DP-MRTFA (Fig.3). This observation is consistent with the established importance of histone acetylases in stem cells [28] and with the availability of their cosubstrate, acetyl-CoA [2]. Acetyl-CoA actually drops when switching the energetic metabolic from glycolysis to oxydative phosphorylation (oxphos) during differentiation [2]. An inverse switch towards glycolysis is precisely observed in DP-MRTFA-expressing cells (Fig.1C).

Chromatin hyperacetylation is sufficient to induce some dedifferentiation characteristics

To determine whether the triple relationship between (i) chromatin hyperacetylation, (ii) dedifferentiation and (iii) basal expression, is fortuitous or causal, we tested if chromatin acetylation caused by artificial drug treatment can induce some characteristics of DP-MRTFA-expressing cells. As shown in Fig.4, mechanical opening of chromatin using TSA, turns to be capable in itself to reproduce certain properties of DP-MRTFA cells, including a rise in basal expression (Fig.4A). In this respect, comparison of transcriptional induction by ERα in the presence or absence of TSA clearly confirms the misleading character of representations in fold induction in this context. When setting the control to 1, TSA seems to inhibit the activity of ERα (Fig.4A, middle histograms). But when setting the induced level to 1 (right histograms), its becomes clear that this drop in fold induction may instead be due to a strong increase in basal expression. TSA treatment also induces clear phenotypic (Fig.4B) and genetic (Fig.4C) changes. 24-hour treatment with 200 nM TSA disrupts cell-cell contacts and downregulates E-cadherin. TSA does not cause nuclear accumulation of endogenous MRTFA as potent as for the mutant construct DP-MRTFA, but it is however significant, particularly for larger cells, and a strong perinuclear accumulation of MRTFA is observed. A signature analysis was then conducted to determine the transcriptomic changes induced by TSA in MCF7 cells.
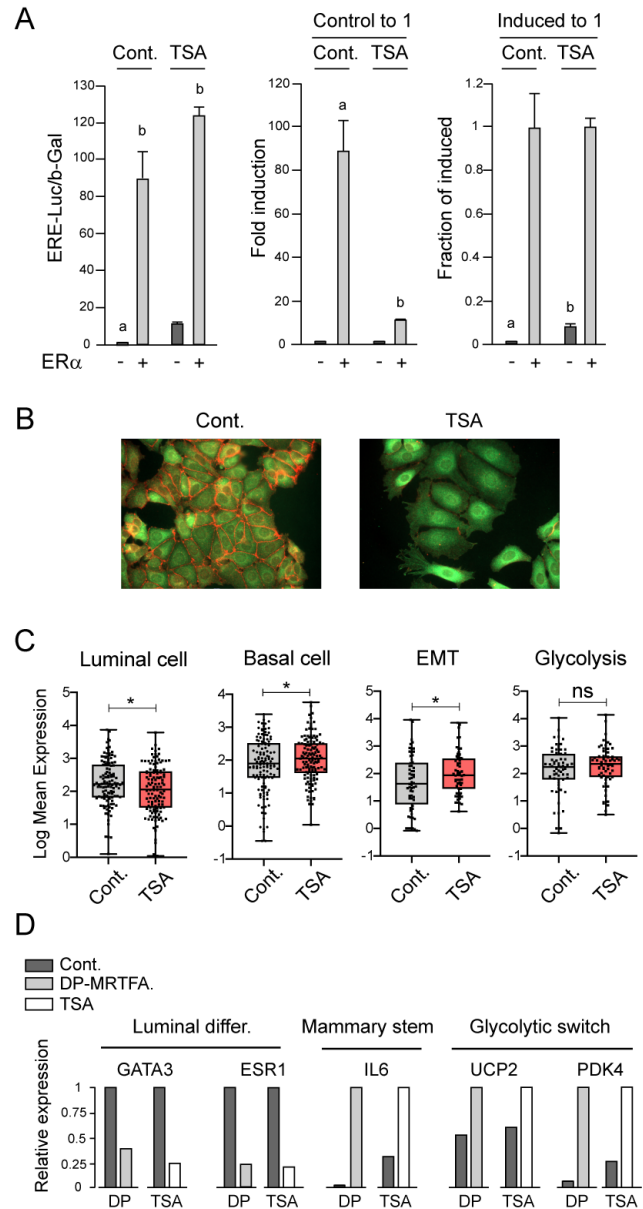
Fig. 4. Effects of TSA on basal gene expression and transcriptomic reprogramming in MCF7 cells. (**A**) Stimulatory effect of ERα on ERE-driven transcription, in presence or absence of 500nM TSA. The transfection experiment was performed as described in Fig.2. Left histograms: relative values, setting to 1 the control in absence of both TSA and ERα. Middle histograms: normalization of the results by setting the control to 1 ("fold induction"). Right histograms: normalization by setting the induced level to 1. Columns with different superscripts differ significantly ($n=3$, $p < 0.05$) (**B**) Phenotypic changes induced by TSA including disruption of intercellular contacts with loss of E-cadherin (red), and a tendency of perinuclear and intranuclear accumulation of endogenous MRTFA (green). (**C**) Gene expression signatures of TSA-treated MCF7 cells

analysed from the data of [45]. (**D**) Parallel effects of DP-MRTFA expression (DP) and TSA treatment on relative changes obtained in transcriptomic arrays for selected gene markers (maximal signals set to 1).

As shown in Fig.4C, TSA has significant effects on gene expression, although less marked than with DP-MRTFA, on the decrease in luminal signature and the increase of signatures of basal cells, EMT and glysolysis. This transcriptomic reshaping is illustrated in Fig.4D by the changes in expression of a selection of well-identified marker genes. A decrease in GATA3 and ERα, involved in mammary luminal differentiation, a strong increase in the mammary stem cell marker IL6 [29, 23], and upregulation of genes involved in the metabolic switch to glycolysis (UCP2 and PDK4) (Fig.4D). Considering the systematic correlation observed between basal expression and cellular differentiation, it is now of interest to determine if this association is merely phenomenological or reflects a fundamental property of biological systems. To this end, we tested the influence of basal expression on particular GRNs clearly identified as regulators of cellular differentiation.

Role of basal expression on GRN multistability

The impact of basal expression on differentiation is tested in the framework of Waddington landscapes using two simple model systems, unidimensional and bidimensional.

Principle of differential GRN modeling.

A fundamental property of living systems is the permanent renewal of all their constituents, through continuous cellular refuelling with matter and energy. In this highly dynamic picture, the concentration of each constituent x results from the relative synthesis (S) and removal (R)

$$\frac{dx(t)}{dt} = S(t) - R(t) \tag{1a}$$

Synthesis can itself be split into basal synthesis (S_b), independent of the specific regulators of the considered gene, and activated synthesis (S_a) triggered by combinations of TFs, ncRNAs and virtually all the other components of the network in an indirect manner.

$$S(t) = S_b(t) + S_a(t) \tag{1b}$$

The removal of molecules also results from a basal mechanism (R_b) generally approximated as an exponential decay,

but in addition there is also the possibility of an active degradation (R_a) by specific actors such as ubiquitin ligases for proteins.

$$R(t) = R_b(t) + R_a(t) \tag{1c}$$

Both basal and specific synthesis will be considered, but for removal, we will only retain, as in most studies, an expotential decay $R(x,t) = r\ x(t)$. Basal synthesis will be reduced to a basal frequency of transcription initiation $S_b(t) = b$, whereas activated synthesis $S_a(t)$ is the product of the maximal transcription initiation frequency a and a fractional promoter occupation function f, ranging from 0 to 1, saturable and generally nonlinear, of potentially all system's components converging to TFs. The global evolution equation of the component x_j thus reduces to

$$\frac{dx_j}{dt} = b + a\ f\ [x_1, \ldots, x_n] - r\ x_j \tag{1d}$$

The functions f mediating the interdependence of the different constituents of the system, impose a collective organization where only certain combinations of concentrations can remain stable. The steady states at which all constituent concentrations are mutually compatible, define the possible cell types generated by the system. Starting from given ingredients, a system is multistable when multiple such steady states can coexist. In each steady state, a set of compatible relative concentrations defines a cellular state. Multistability is a preeminent feature of living systems and is synonymous to the capacity of differentiation. It can be obtained when (i) the system is open, subject to permanent constituent renewal, (ii) at least one positive circuit is included [30] and (iii) velocities of either synthesis or removal are nonlinearly dependent on constituent concentrations. To test the effect of basal expression on the structure of the epigenetic landscape, we used abundantly documented paradigms of multistable circuits, consisting of one or two genes. Such minimalist circuits may appear ridiculously small compared to complete cellular systems, but they actually underly real cases of bipotent progenitor differentiation. In addition, they have the practical advantage to be representable in the form of 2D and 3D landscapes.

Single gene circuit: The self-regulated gene encoding a dimerizable TF.

A single autoregulated gene (Fig.5A) which is certainly one of the simplest possible circuits, is nethertheless sufficient to give rise to bistability, provided the consitions listed above are fulfilled. Despite its simplicity, this minimalist circuit is actually encountered in nature, as for example: (i) For the auroregulated gene ComK in the context of the sporulation

6

of the baterium B. subtilis [31], which is a sort of bacterial differentiation. (ii) In vertebrates, it is involved in the vitellogenesis memory effect, evidenced in all egg-laying vertebrates tested, from fishes to birds [32]. As simple as it is, this system exhibits the principle proposed here. Increasing basal expression shifts from a single attractor to two different states, of low and high gene expression, which can be regarded as a minimalist type of cellular differentiation. This one-dimensional circuit is important to consider because it can give by integration a genuine Waddington landscape. The potential function for this unidimensional landscape can be straightly calculated by integration of the product evolution function (synthesis minus removal) [33, 34]. Assuming a time scale separation between the DNA/TF interactions and gene expression dynamics, the rigorous modeling of this minimalist gene circuit, which distinguishes the monomer, dimer and total concentrations of the TF [35], reads

$$\frac{d[TF]_{\text{tot}}}{dt} = b + a\frac{[TF_2]}{K + [TF_2]} - r[TF]_{\text{tot}} \qquad (2a)$$

where b is the basal expression rate, a is the maximal rate of activated expression, K is the constant of dissociation from DNA, and $[TF_2]$ is the concentration of the TF dimer, related to its total concentration $[TF]_{\text{tot}}$ through

$$[TF_2] = \left(1 + 4D[TF]_{\text{tot}} - \sqrt{1 + 8D[TF]_{\text{tot}}}\right)/8D \qquad (2b)$$

where D is the homodimerisation constant. When replacing $[TF]_{\text{tot}}$ by x, the differential equation becomes

$$\frac{dx}{dt} = b + a\frac{1 + 4Dx - \sqrt{1 + 8Dx}}{1 + 8KD + 4Dx - \sqrt{1 + 8Dx}} - rx \qquad (2c)$$

The balance between synthesis and removal terms (in short the right hand side in the last equation Eq.(2c)), is shown in Fig.5B for the following set of parameter values: $D = 1.5$; $K = 0.2$; $a = 19$ and $r = 15$. The effect of basal expression can be straightly understood by considering the evolution function dx/dt represented in Fig.5B. Depending on the relative values of the production and removal functions, dx/dt can cross several times the null line 0, thereby yielding several possible steady states. To obtain such curves, degradation is generally exponential, with a flux proportional to the product concentration, while synthesis follows a saturable and sigmoidal function of the TFs. This sigmoidicity can be due to a variety of reasons, including TF dimerisation [36, 33, 35, 37], or sequestration by a "poison partner", as in the next example. The resulting landscapes obtained by integration of the evolution function are shown in Fig.5C using the same parameters as above and for various values of b. The blue curves at the bottom of valleys stand for attractive steady state and the green one for the repulsive one.
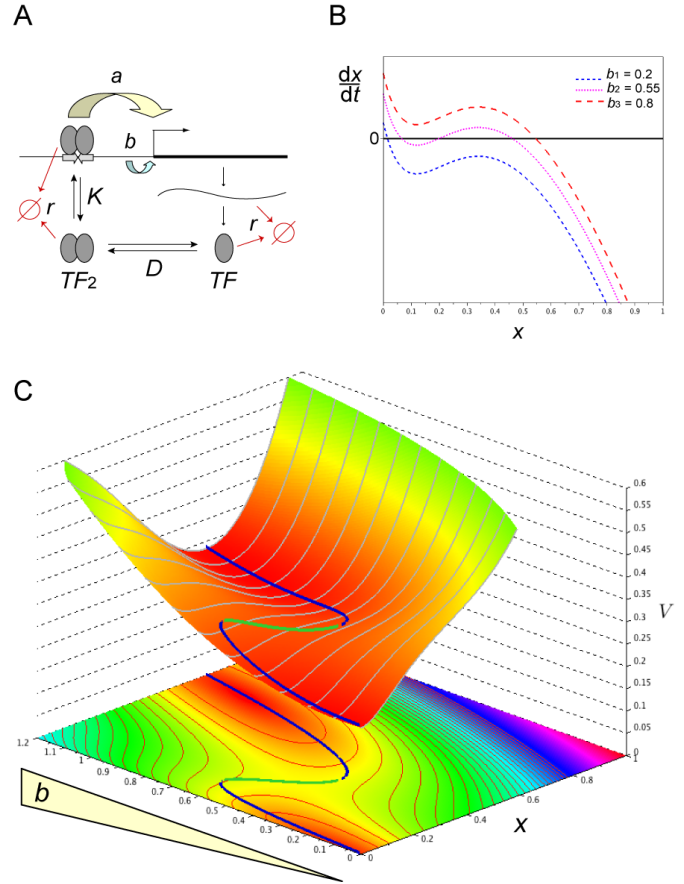
Fig. 5. Unidimensional gene circuit. (A) Minimalist bistable system with a single gene activated by its own product. Gene expression is the resultant of two initiation frequencies: b, basal (independent of the TF) and a, activated by the TF. (B) Simple mechanism to eliminate or create bistability by modifying basal expression. The number of times the function dx/dt crosses the line 0, corresponding to the equivalence between synthesis and removal, gives the number of steady states. A unique crossing is obtained for low and high values of b whereas two stable steady states are obtained for intermediate values of b. (C) Evolution of Waddington landscapes when increasing basal expression. The Waddington landscapes correspond to slices of this 3D plot (transversal curves in the x direction).

The values of b for which such a repulsive steady state is present, precisely correspond to the ones for which two attractors coexist. In its dependence in the parameter b, the system exhibits two connected saddle-node bifurcations. We emphasize that the parameter b is not driven by any dynamic or modeling considerations. This is precisely the point to understand how its (slow) evolution may affect the whole stability properties of the (fast) system. The potential scalar function $V(b, x)$ in Fig.5C is obtained directly by integrating the opposite of the right-hand side in the evolution

equation Eq.(2c):

$$V = -b\,x + \frac{r\,x^2}{2} - a \int_0^x \frac{1 + 4D - \sqrt{1+8Dy}}{1 + 8KD + 4Dy - \sqrt{1+8Dy}}\,dy + V_o(b). \quad (3)$$

This potential provides a full characterization of attractive of repulsive steady states in the following sense: the equilibria appear then respectively as (local) minimizers and maximizers of $V(b, \cdot)$. Actually, we introduce an artificial additive potential $V_o(b)$, depending only on b, and therefore not affecting the local minimizers and maximizers in the x-direction, but only their level. This allows to obtain a readable graphical representation. The idea behind this choice is to normalize integration constants so that any of the minimizers levels are more or less independant of b. For simplicity, we only choose a hand-designed polynomial corrective term

$$V_o(b) = 0.33\ (b - 1.2) + 0.3\ b\ (b - 1.2) - 0.05\ b^2\ (b - 1.2)^2.$$

The influence of basal expression on the multistability of this toy system is well illustrated in Fig.5C, which shows a single attractor of zero expression for $b = 0.2$, two attractors for $b = 0.55$ and a single profound attractor for $b = 0.8$.

Two gene circuit: Re-modeling the celebrated GATA-PU system.

Mutual repression has long been envisioned as a stereotyped multistability switch motif [38]. The most popular tristable two-gene landscape is generated by a circuit with mutually repressing and self-activating genes [20, 21, 34, 39, 40, 41]. Remarkably, this model corresponds to real cases of bipotent progenitor differentiation, including: (i) the balance between red and white blood cells resulting from a choice between GATA1/2 and PU.1 [46], (ii) the muscular vs vascular differentiation of somite cells determined by the Pax3:Foxc2 balance [47], (iii) the ectodermal vs mesendodermal differentiation depending on the Sox2:Oct4 circuit [48]. In all these cases, upon differentiation the system evolves from a central attractor where the antagonistic genes are coexpressed, to one of two lateral attractors of exclusive gene expression. Luminal mammary differentiation certainly obeys the resolution of a circuit of this type, but since it is still unclear whether ESR1 and GATA3 form a positive [49] or negative [50] loop, we will rather use the well established GATA1/2 and PU.1 involved in blood cell differentiation. In the initial models of the GATA1/2:PU.1 circuit, self-activation and mutual repression were disconnected as depicted in the scheme 1. The variables x and y are understood as total GATA1/2 and PU.1 protein concentrations by assuming that translation is not rate-limiting. This modeling is widespread in the literature [34, 39, 40] even if the basal activity is not explicitly written (set to 1) or interpreted as such.
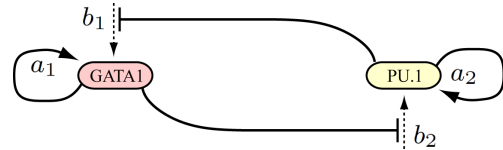

Scheme 1. Traditional interpretation of transcriptionally independent self-stimulation and reciprocal inhibition.

The classical formulation of this scheme is

$$\frac{dx}{dt} = a_1\ \frac{x^{n_1}}{K_1^{n_1} + x^{n_1}} + b_1\ \frac{K_1^{n_1'}}{K_1^{n_1'} + y^{n_1'}} - r_1\ x \quad (4a)$$

$$\frac{dy}{dt} = a_2\ \frac{y^{n_2}}{K_2^{n_2} + y^{n_2}} + b_2\ \frac{K_2^{n_2'}}{K_2^{n_2'} + x^{n_2'}} - r_2\ y \quad (4b)$$

But in this scheme, basal expression would not be an independent parameter, as defined in Eq.(1d), but would be regulated by system constituents. In fact, based on the physical and functional interactions between GATA1/2 and PU.1 described in the seminal article of [46], which are recalled in Fig.6A, the mutual inhibition between the two genes does not proceed through reduction of some basal level (Scheme 1), but through preventing the self-stimulations of GATA1/2 and PU.1. This scheme thus includes a genuine basal expression frequency (Scheme 2). A key parameter for modeling the revised mechanism of inhibited activation, is the molecular association between GATA1/2 and PU.1 (Fig.6A) [46]. This association corresponds to a mutual sequestration preventing PU.1 from (i) stimulating its own gene and (ii) inhibiting the GATA1/2 gene, and vice versa (Scheme 2).
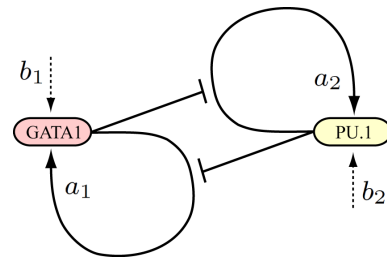
Scheme 2. Present interpretation based on reciprocal inhibition of self-stimulation, in which basal transcription frequencies are not regulated by specific factors.

The set of equations corresponding to scheme 2 reads

$$\frac{dx}{dt} = b_1 + a_1\ \frac{x_f}{K_1 + x_f} - r_1\ x \quad (5a)$$

$$\frac{dy}{dt} = b_2 + a_2\ \frac{y_f}{K_2 + y_f} - r_2\ y \quad (5b)$$

where x_f and y_f are the concentrations of molecules not mutually interacting in $x \bullet y$ complexes. Given the time

8

scale separation between molecular interactions (very fast) and gene expression dynamics (much slower) the free concentrations are simply given by non-differential, algebraic equations.

$$x_f = x - x \bullet y \qquad (5c)$$

$$y_f = y - x \bullet y \qquad (5d)$$

where the complex is given by

$$x \bullet y = \frac{1}{2}\left[(D + x + y - \sqrt{(D + x + y)^2 - 4xy}\right] \qquad (5e)$$

where D is the equilibrium dimerisation constant between x and y. The exponents n in the classical modeling of Eq.(4) are Hill coefficients describing molecular cooperativity, whose values are generally chosen for convenience. Arbitrarily increasing Hill's coefficients is indeed an easy way to accentuate the relief of epigenetic landscapes, but this twist is poorly justifiable in practice in absence of precise quantitative data. By contrast, the simple mechanism of mutual sequestration is both biologically relevant and sufficient to provide the nonlinearity necessary for multistability, whether or not the TFs work as monomers or preformed dimers. Concretely, the production functions in the differential equations are unchanged but the TF concentrations should just be replaced by their free concentrations (x_f and y_f). The roles of GATA1/2 and PU.1 are supposed symmetrical with identical parameters for both genes ($a_1 = a_2$, $b_1 = b_2$, $K_1 = K_2$ and $r_1 = r_2$). Unlike the unidimensional evolution system Eq.(3), not any differential system in higher dimension may be described from a simple scalar-valued potential function. When a Lyapunov function exist however, it provides directly a scalar characterization of attractive behavior of steady states and thus enables the possibility to draw a landscape. This is the case for example for any gradient-like systems. More generally, one may try to take into account the hamiltonian part of the dynamic, however this is not clear how to use the Hodge-Helmholtz-like decompositions to draw a Waddington landscape [51, 20, 52]. An alternative approach is based on the probabilistic point of view and concerns the large deviations theory for invariant measures of stochastic reaction-diffusion processes [53]. In the phase space (x, y), we consider many trajectories of the deterministic system Eq.(5) perturbated with a small brownian motion. Roughly speaking, many of these trajectories accumulates in large time close to attractive steady states or and then evolves finally only through a fine balance between the diffusion process and the inherent deterministic dynamic. The density of trajectories at any point of the phase space becomes independant of time, this is called the invariant measure of the stochastic process. By simulating

some probabilistic processes or by computing the solutions of a related reaction-diffusion partial differential equations, we can determine this mean equilibrium density of trajectories at any point. This quantity is a new real-valued function revealing the attractive points and/or the attractive limit cycles of the dynamical system. The probability in Fig.6B is obtained using a finite difference scheme to solve the partial differential equation

$$\frac{\partial p}{\partial t} + X(x,y)\frac{\partial p}{\partial x} + Y(x,y)\frac{\partial p}{\partial y} = \epsilon\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right), \qquad (6)$$

over the domain $[0, 10] \times [0, 10]$ in (x, y) and over $[0, 10]$ in the time variable. The initial data is set to a uniform density $p(t = 0, x, y) = 1$. The dynamical field $X(x, y), Y(x, y)$ correspond to the respective right-hand sides in Eq.(5), and the diffusion parameter is set to $\epsilon = 0.025$. For the biological parameters, we use the following set of values: $a_1 = a_2 = 10$; $K_1 = K_2 = 6$, $r_1 = r_2 = 1$ and $D = 1$. In Fig.6B, the basal expression rate is set either to a low value ($b = 1$) for which bistability is present, or to a higher value ($b = 4$) with then monostability. The background color field represents the intensity of the density p and the black curves figure the deterministic trajectories, solution to the dynamical system Eq.(5). Almost every of these trajectories goes in large time to one (of the two) attractive points. The computational phase space domain $[0, 10] \times [0, 10]$ is chosen sufficiently large so that we can prove it consists in an invariant domain for the dynamical system: no trajectory escape the domain. This property is useful to design convenient boundary conditions when solving Eq.(5). The same strategy is used for the Fig.6C. In summary, rigorous landscape treatment of the celebrated GATA1/2:PU.1 differentiation circuit clearly shows that a simple change in basal expression level can modify the fate of the system. For $b = 1$, two cell types coexist (Fig.6B, left panel), whereas for $b = 4$, a single "indecise" cell type exists, with equivalent coexpression of GATA1/2 and PU.1. The projection plot of Fig.6C shows the reshaping of Waddington landscape triggered by b. With the set of parameters used, the transition from bistability to monostability occurs at $b = 2.54$.

General principle of bifurcation by resolution of conflicting circuits.

The cellular differentiation tree in development proceeds through a cascade of successive bifurcations [54], each one coinciding with the resolution of a gene conflict, of which a perfect example is the battle between the GATA1/2 and PU.1 genes. The present results show that a high level of basal expression alleviates the impact of their mutual repression and allows the coexpression of both antagonistic genes in progenitor cells. Lowering basal expression increases the intensity of the fight and accelerates its resolution, achieved

9

when one of the genes loses the fight.
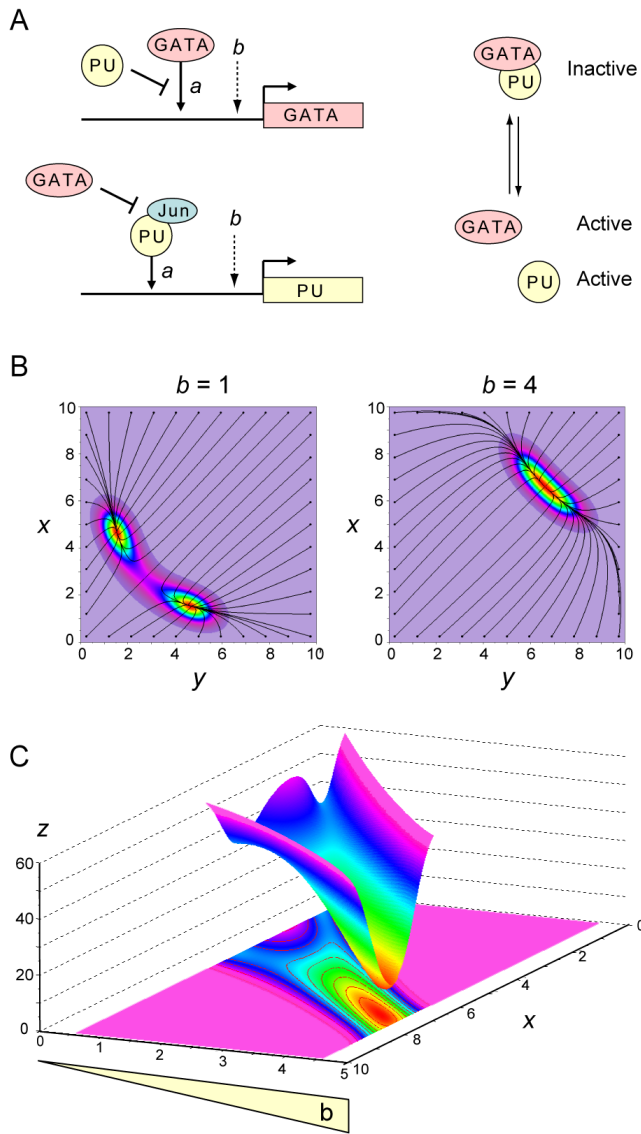
Fig. 6. The GATA-PU system. (**A**) Summary of the molecular mechanisms involved in the self-stimulation and mutual repression of the GATA1/2 and PU.1 factors, described in the original article of [46]. The physical interaction between GATA1/2 and PU.1 is the keystone of this model since only the proteins which are not mutually interacting, are active. (**B**) Waddington landscapes shaped by the GATA1/2 (x): PU.1 (y) circuit, according to the scheme 2, with low (left) or high (right) basal expression b. (**C**) Projection of steady states on the axis of one of the variables as a function of b, showing the switch from bi- to monostability when increasing b, for fixed parameters a and K.

Model insertion in the current picture of cellular differentiation

The present model of differentiation centered on the role of basal expression, makes it possible to weave an integrated picture combining several known properties of cellular differentiation.

Orientation of differentiation

The mechanism of differentiation proposed here is attractive in that it is generic and valid for all cell lineages. The orientation of progenitors committed to differentiate towards a particular destination attractor is supposed to result from either (i) stochastic fluctuations, favored by the low number of certain molecules like mRNAs and allowing the cellular system to jump between adjacent attractors with a waiting time exponentially dependent on the height of the saddle point between them; or (ii) instructive exogenous inputs, like EPO vs G-CSF for the red vs white blood cells, or Spemann's organizers during organogenesis, transiently altering the initial steady state. The strong heterogeneity detected in single cell transcriptomic analyses supports an important role for the first mechanism. In fact, entrusting developmental bifurcations at random is not actually a risk since any imbalance in the number of cells falling in the final attractors can be corrected a posteriori by selective proliferation and/or apoptosis, to restore the appropriate partitioning of cellular masses.

Paradoxical decrease of regulated transcription when opening chromatin

b and a are used here as independent parameters and only b is modified in the simulations shown above, but a mechanical link can exist between them, through which when one decreases, the other increases. In addition to the enhancers present in regulated genes, the genome contains a multitude of non-specific TF binding sites, generally unoccupied in heterochromatin. Hence, these cryptic sites are logically exposed during dedifferentiation and can trap certain TFs, thereby reducing their free concentration and their recruitment at enhancers (thus reducing a). Such a titration mechanism, which is not necessarily valid for all types of TF, has already been invoked for example to explain how the general TF TATA-binding protein (TBP) whose concentration is limiting [55], puts in competition all its target genes [56]. Since the chromatin of dedifferentiated cells is more accessible for proteins, as examplified by anti-DNA antibodies (Fig.3), or DNAseI [15], it seems logical that the access of TFs is also favored. The b/a ratio is therefore expected to increase upon chromatin loosening in two ways: (i) by allowing the generalized access of transcriptional ma-

10

chineries to a wide variety of genes and (ii) by sequestering TFs in bulk DNA sites. This defocusing of TFs from enhancers is therefore capable at the same time to cause an increase of b and a decrease of a, which act in concert to strongly increase the ratio b/a and desimprint previous gene regulatory circuits. Simultaneous enhancer weakening and genome opening, is the ideal scenario for reprogramming systems, by erasing pre-printed circuits in the face of the emergence of new actors.

Hypothetical origin of bivalent chromatin

A particularity of histone acetylation, contrary to other histone modifications is that its role is unambiguous. For comparison, the effect of histone methylation depends on the lysine residue. Methylation of H3K4 has a permissive role and that of H3K9 methylation has a repressive effect on trancription. By contrast, histone acetylation is always permissive for TF binding regardless of the target lysine, including H3K4 in the promoter of active genes [57]. Note that the methylation of H3K4 in the transcribed regions of active genes is precisely stimulated by acetylated substrates [58]. A subtlety however remains to be explained: in the so-called bivalent chromatin of stem cells, acetylated H3K9 which is permissive, coexists with deacetylated and methylated enhancer H3K27 (E-H3K27) which is repressive. A simple hypothesis to explain this apparent paradox, is based on the observation that acetylation of E-H3K27 results from the docking of HAT by TFs at the level of enhancers, thereby generating a positive loop in which TF binding stimulates E-H3K27 acetylation, which in turn favors enhancer accessibility to TFs. In hyperacetylated chromatin, TF-binding motifs previously cryptic in heterochromatin become exposed and can trap TFs, reducing their free concentration and consequently their availability for binding to enhancers (Model formalized in the S.I.). Then, E-H3K27 methylases like polycomb could complete the system by methylating poorly occupied E-H3K27, as verified in [59], thus precluding their reacetylation.

Why some genes are repressed in the context of globally open chromatin

A long standing enigma about cancer cell chromatin is while it is largely released, certain genes, encoding for instance tumor repressors, are closed. Similar situations are found here. In particular, TSA treatment alone is capable of both decondensing chromatin and repressing genes involved in mammary epithelial differentiation such as GATA3 and ERα. Candidate mechanisms to explain this, include transcriptional repressors as the zinc-figers SNAI1 (Snail) and/or SNAI2 (Slug), known to selectively repress differentiation

genes, for instance muscular [60]. They are strongly upregulated in the TSA treatment by 22- and 19-fold for SNAI1 and SNAI2 respectively, and in cancer, such as during the mammary hormonal escape, where the SNAI repress ERα and E-cadherin [61]. Another excellent molecular candidate for repressing differentiation genes after chromatin opening is the polycomb system mentioned above, which could simply validate the lower occupancy of enhancers by TFs [59] (model in the S.I.) and proceed to their closure.

Interplay between acetylation, metabolism and dedifferentiation

The link between chomatin loosening and basal expression is likely to be mediated by histone acetylation, itself depending on the cellular amount of acetyl-CoA, which ultimately results from the type of energetic metabolism of the cell. This relation singularly concretizes the intimate relationship between metabolism and differentiation long anticipated by Warburg [62]. Warburg noticed that glycolysis is predominant in "less structured" (understand less differentiated) cells. The activity of acetylation enzymes is critically dependent on acetyl-CoA as a source of acetyl groups. Precisely, the concentration of acetyl-CoA has been shown much higher in undifferentiated cells with high glycolytic activity [2] in full agreement with the present results, including induction of glycolysis (Fig.1C) and H3K9 acetylation (Fig.3A) upon dedifferentiation.

Conclusions

Functional correlations between chromatin loosening, dedifferentiation and basal gene expression, reflect a universal mechanism in which a decrease of basal expression systematically leads to differentiation and conversely, increasing basal expression associated to H3K9 acetylation, opens the way to reprogramming. The release of chromatin repression and the increase in non-specific gene expression naturally participate to the high entropy of the less organized undifferentiated cells. This study points out the importance of the basal expression level in GRNs, which is currently neglected in both experimental and theoretical approaches. It is often omitted in theoretical studies, as shown here for the previous modeling of the GATA1/2:PU.1 circuit, and eliminated during standardization steps in transcriptomic and epigenomic studies. The largest datasets generated by high throughput, multiplexed or single cell approaches, are unable to provide relevant information on the basal expression level. Reintroduction of this overlooked parameter allows to propose a unifying explanation to multiple observations including (i) the wide open chromatin of stem cells [1, 2], their generalized low level of gene expression [5, 6], and (ii) the influence of chromatin on their differentiation

[7, 9, 10, 63]. These different results are not only reconciled but in addition, make it possible to develop a model of cellular differentiation/dedifferentiation in the spirit of Waddington, that is to say based more on a physical principle than on specific genes. Waddington epigenetic landscapes are shown here to be structured by the level of basal gene expression. Two concurrent views of Waddington landscapes coexist: (i) a single rigid landscape specific to the genome of each organism, whose different basins correspond to the different possible cell types in the organism, or (ii) a deformable landscape whose attractors and their depth can vary with parameter adjustments. The new mechanism proposed here clearly belongs to the latter category as it predicts that the landscape is shaped by the degree of basal expression as shown in Fig.5C and Fig.6C, in such a way that the undecided progenitor attractors remain profound as long as the cells retain their basal expression and open chromatin. This mechanism is strongly consistent with gene expression specificities of stem cells. As basal expression is progressively reduced, differentiation attractors emerge sequentially. In this gradual process, the initial commitment of totipotent cells could be triggered by a modest reduction of basal expression, while terminal differentiation of bipotent progenitors requires a strong reduction of basal expression and chromatin closure. Terminally differentiated cells with robustly imprinted circuits have a tightly packed chromatin enriched in H3K9me3, with some islets of accessibility to TFs at the level of H3K27 acetylated enhancers. The sharply partitioned chromatin of differentiated cells may ensure the persistence of well-focused specific circuits. Conversely, chromatin hyperacetylation could be responsible for a rise in basal expression and expose newly accessible binding sites for TFs, defocusing them from enhancers. A remarkable property of the present model is that differentiation and stemness attractors do not coexist at a given moment, so that stem cells cannot accidentally fall into a differentiation attractor. Conversely, the strength of established heterochromatin in differentiated cells is a powerful barrier against the risk of de-differentiation, since stem cell attractors no longer exist in that state. However, pathological or age-related loss of H3K9me3 can unlock the system and re-open the road to dedifferentiation. Hence, mechanisms ensuring the maintenance of H3K9me3 [64, 8] are essential for longevity and cancer prevention. For example, the better heterochromatinized mammary cells of formerly gestating women are less prone to cancerization, even after menopause [65]. Although the word epigenetics was first introduced in the context of the gene networks conceived by Waddington, this term was then hijacked by researchers working on chromatin, who restricted the term epigenetics to chromatin "marks" [41]. Strikingly, the present theory merges these two views by confering to chromatin epigenetics a driver role in Waddington epigenetics.

Materials and methods

Plasmids, MCF7 subclones, antibodies

The following constructs used in this study: pCR ERα, pSG-GR, pCR-DP-MRTFA (ΔN200), pCR DN-MRTFA (ΔC301), ERE-Luc (C3-Luc) and GRE-tk-LUC, are described in [16]. pCMV-galactosidase is from Promega. The stably transfected MCF7 control, DP-MRTFA and DN-MRTFA subclones are described in [16]. The following primary antibodies were used: anti-E-cadherin (ab15148; Abcam), anti-MKL1 (sc21558; Santa Cruz Biotechnology), anti-histone H3 (E173-58; Epitomics), anti-H3K9ac (histone H3 acetylated at Lys9; ab10812; Abcam), anti-H3K9me3 H3K9me3 (histone H3 trimethylated at Lys9; ab8898, Abcam), anti-double stranded DNA (ab27156, Abcam). Secondary antibodies conjugated to Alexa Fluor 488 or 594 were from Invitrogen.

Cell culture, transfection and reporter assays

HepG2, HeLa, MCF7, MDA-MB231 (MDA), MCF7 control and constitutively expressing MRTFA constructs (T-Rex system, Invitrogen) were grown in DMEM (Dulbeccos modified Eagles medium; Invitrogen) supplemented with 10% fetal bovine serum (FBS, Biowest) and antibiotics (Invitrogen) at 37C and 5% CO_2 humidified atmosphere. Prior all transfections and treatments, the medium was replaced with phenol red-free DMEM (Invitrogen) containing 2.5% charcoal-stripped FCS (Biowest). Expression of the MRTFA proteins of interest was induced by a 48h treatment of MCF7 sub-clones with tretracyclin. Cells were treated for 24 h when required with ligands (10 nM estradiol or dexamethasone) or ethanol (vehicle control). The treatment with trichostatin A (TSA 647925, Merck) was performed for 24 hours at 100, 200 or 500 nM. Transfection experiments were carried out exactly as previously described [16]. RNA vs DNA content ratios were determined using the $HClO_4$ hydrolysis method [42].

Immonohistochemistry

Cells were grown on 10-mm-diameter coverslips in 24-well plates in DMEM containing 2 % charcoal-stripped FBS. A suprimer (and treated with trichostatin A (TSA, 647925, Merck) for 24 hours at 100, 200 or 500 nM). Cells were fixed with 4% PFA (paraformaldehyde) for 10 min and permeabilized in PBS- 0.3%Triton X-100 for 10 min. Incubation with the primary antibody (1:1000 dilution) was performed overnight at 4C. Secondary antibodies conjugated to Alexa Fluor were incubated for 1 h at room temperature. After washing in PBS, the cover slides were mounted in Vectashield$^{\circledR}$ medium with DAPI (Vector) and images were

obtained with an Imager.Z1 ApoTome AxioCam (Zeiss) epifluorescentmicroscope and processedwith AxioVision Software. For each coverslips, 10 to 20 pictures were randomly taken. Pictures were visually screened in blind condition and deleted if artefactual fluorescent aggregates were present or in case of focus problems. For each picture, fluorescence values of each nucleus were obtained in an automatic manner using a homemade plugin working on Fiji [43]. Briefly, each nucleus was identified using the DAPI labelling and, after background subtraction, total fluorescence of each nucleus was extracted from the picture obtained with the fluorescent antibody. For each condition, the mean of fluorescence intensities of more than a thousand cells were calculated.

Transcriptomic data

The microarray data on MRTFA cell lines have been submitted to the NCBI Gene Expression Omnibus website under accession No. GSE107924. Gene signatures were obtained from the publicly available database MsigDb from GSEA. The luminal and basal signatures were extracted from [44] (curated gene sets). A) Comparison of the transcriptional signature of the two DP-MRTFA1 and Dn-MRTFA clones in comparison to MCF7 control cells. B) Comparison of the four transcriptional signatures between TSA treated versus vehicle treated (Cont.) MCF7 cells. Data were obtained from by TempO-Seq targeted whole transcriptome profiling GEO accession: GSE91395 [45].

Acknowledgments

References

[1] Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M, Open chromatin in pluripotency and reprogramming, Nat. Rev. Mol. Cell Biol. 12 (2011) 36-47

[2] Moussaieff A, Rouleau M, Kitsberg D, Cohen M, Levy G, Barasch D, Nemirovski A, Shen-Orr S, Laevsky I, Amit M, Bomze D, Elena-Herrmann B, Scherf T, Nissim-Rafinia M, Kempa S, Itskovitz-Eldor J, Meshorer E, Aberdam D, Nahmias Y, Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells, Cell Metab. 21 (2015) 392-402

[3] Meshorer E, Misteli T. Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. 7 (2006) 540-546

[4] Ugarte F, Sousae R, Cinquin B, Martin EW, Krietsch J, Sanchez G, Inman M, Tsang H, Warr M, Passegué E,

Larabell CA, Forsberg EC, Progressive chromatin condensation and H3K9 methylation regulate the differentiation of embryonic and hematopoietic stem cells. Stem Cell Rep. 5 (2015) 728-740

[5] Hipp JA, Hipp JD, Atala A, Soker S, Functional genomics: new insights into the 'function' of low levels of gene expression in stem cells. Curr. Genomics 11 (2010) 354-358

[6] Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH, Gingeras TR, Misteli T, Meshorer E, Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2 (2008) 437-447

[7] Matoba S, Liu Y, Lu F, Iwabuchi KA, Shen L, Inoue A, Zhang Y, Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell 159 (2014) 884-895

[8] Becker JS, Nicetto D, Zaret KS, H3K9me3-dependent heterochromatin: Barrier to cell fate changes. Trends Genet. 32 (2016) 29-41

[9] Antony J, Oback F, Chamley LW, Oback B, Laible G, Transient JMJD2B-mediated reduction of H3K9me3 levels improves reprogramming of embryonic stem cells into cloned embryos. Mol. Cell Biol. 33 (2013) 974-983

[10] Wei J, Antony J, Meng F, MacLean P, Rhind R, Laible G, Oback B KDM4B-mediated reduction of H3K9me3 and H3K36me3 levels improves somatic cell reprogramming into pluripotency. Sci. Rep. 7 (2017) 7514

[11] Hezroni H, Tzchori I, Davidi A, Mattout A, Biran A, Nissim-Rafinia M, Westphal H, Meshorer E, H3K9 histone acetylation predicts pluripotency and reprogramming capacity of ES cells. Nucleus 2 (2011) 300-309

[12] Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA, Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat. Biotechnol. 26 (2008) 795-797

[13] Pace L, Goudot C, Zueva E, Gueguen P, Burgdorf N, Waterfall JJ, Quivy JP, Almouzni G, Amigorena S, The epigenetic control of stemness in CD8+ T cell fate commitment. Science 359 (2018) 177-186

[14] Lyssiotis CA, Walker J, Wu C, Kondo T, Schultz PG, Wu X, Inhibition of histone deacetylase activity induces developmental plasticity in oligodendrocyte precursor cells. Proc. Natl. Acad. Sci. USA 104 (2007) 14982-14987

[15] Szabó Jr G, Damjanovich S, Sümegi J, Klein G, Overall changes in chromatin sensitivity to DNase I during differentiation. Exp. Cell Res. 169 (1987) 158-168

[16] Flouriot G, Huet G, Demay F1, Pakdel F1, Boujrad N1, Michel D, The actin/MKL1 signalling pathway influences cell growth and gene expression through large-scale chromatin reorganization and histone post-translational modifications. Biochem. J. 461 (2014) 257-268

[17] Waddington CH, The strategy of the genes. Routledge Taylor and Francis Group. London and New York 1957

[18] Kauffman SA, The origins of order. Oxford University Press, New York, NY. 1993

[19] Huang S Back to the biology in systems biology: What can we learn from biomolecular networks? Gen. Proteom. 2 (2004) 279-297

[20] Huang S, Guo YP, May G,and T. Enver T, Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev. Biol. 305 (2007) 695-713

[21] Moris N, Pina C, Arias AM, Transition states and cell fate decisions in epigenetic landscapes. Nat. Rev. Genet. 17 (2016) 693-703

[22] Nichols J, Smith A, Pluripotency in the embryo and in culture. Cold Spring Harb. Perspect. Biol. 4 (2012) a008128

[23] Kim T, Lim DS, The SRF-YAP-IL6 axis promotes breast cancer stemness. Cell Cycle 15 (2016) 1311-1312

[24] Ikeda T, Hikichi T, Miura H, Shibata H, Mitsunaga K, Yamada Y, Woltjen K, Miyamoto K, Hiratani I, Yamada Y, Hotta A, Yamamoto T, Okita K, Masui S Srf destabilizes cellular identity by suppressing cell-type-specific gene expression programs. Nat. Commun. 9 (2018) 1387

[25] Turner BM, Open chromatin and hypertranscription in embryonic stem cells. Cell Stem Cell 2 (2008) 408-410

[26] Cereghini S, Yaniv M, Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. EMBO J. 3 (1984) 1243-1253

[27] Koutroubas G, Merika M, Thanos D, Bypassing the requirements for epigenetic modifications in gene transcription by increasing enhancer strength. Mol. Cell. Biol. 28 (2008) 926-938

[28] Ryall JG, Cliff T, Dalton S, Sartorelli V (2015) Metabolic reprogramming of stem cell epigenetics. Cell Stem Cell 17: 651-662

[29] Iliopoulos D, Hirsch HA, Wang G, Struhl K, Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proc. Natl. Acad. Sci. USA 108 (2011) 1397-1402

[30] Kaufman M, Soulé C, Thomas R, A new necessary condition on interaction graphs for multistationarity J. Theor. Biol. 248 (2007) 675-685

[31] Dubnau D, Losick R, Bistability in bacteria. Mol. Microbiol. 61 (2006) 564-572

[32] Nicol-Benoit F, Amon A, Vaillant C, le Goff P, le Dréan Y, Pakdel F, Flouriot G, Valotaire Y, Michel D, A dynamic model of transcriptional imprinting derived from the vitellogenesis memory effect. Biophys. J. 101 (2011) 1557-1568

[33] Ferrell JE Jr, Bistability, bifurcations, and Waddington's epigenetic landscape. Curr. Biol. 22 (2012) R458-R466

[34] Zhou JX, Aliyu MD, Aurell E, Huang S,) Quasi-potential landscape in complex multi-stable systems. J. R. Soc. Interface 9 (2012) 3539-3553

[35] Nicol-Benoit F, le Goff P, Michel D, Drawing a Waddington landscape to capture dynamic epigenetics Biol. Cell 105 (2013) 576-584

[36] Keller A, Model genetic circuits encoding autoregulatory transcription factors. J. Theor. Biol. 172 (1995) 169185

[37] Zhang K, Sasai K, Wang J, Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. Proc. Natl. Acad. Sci. USA 110 (2013) 1493014935

[38] Monod J, Jacob F, Teleonomic mechanisms in cellular metabolism, growth and differentiation. Cold Spring Harbor Symposia on Quantitative Biology 26 (1961) 389-401

[39] Bhattacharya S, Zhang Q, Andersen ME, A deterministic map of Waddington's epigenetic landscape for cell fate specification. BMC Syst. Biol. 5 (2011) 85

[40] Wang J, Zhang K, Xu L, Wang E, Quantifying the Waddington landscape and biological paths for development and differentiation. Proc. Natl. Acad. Sci. USA 108 (2011) 8257-8762

[41] Huang S, The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? Bioessays 34 (2012) 149-157

[42] Munro A, Fleck HN, The determination of nucleic acids. In Methods of Biochemical Analysis. Glick D. Ed. 1966

[43] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A, Fiji: an open-source platform for biological-image analysis. Nat. Meth. 9 (2012) 676-682

[44] Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adélaïde J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D, Bertucci F, Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene 25 (2006) 2273-2284

[45] Yeakley JM, Shepard PJ, Goyena DE, VanSteenhouse HC, McComb JD, Seligmann BE A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One 12 (2017) e0178302

[46] Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS, Auron PE, Tenen DG, Sun Z, Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc. Natl. Acad. Sci. USA 96 (1999) 8705-8710

[47] Lagha M, Brunelli S, Messina G, Cumano A, Kume T, Relaix F, Buckingham ME, Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell 17 (2009) 892-899

[48] Thomson M, Liu SJ, Zou LN, Smith Z, Meissner A, Ramanathan S, Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145 (2011) 875-889

[49] Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M, Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. Cancer Res. 67 (2007) 6477-6483

[50] Lee J, Tiwari A, Shum V, Mills GB, Mancini MA, Igoshin OA, Balázsi G (2014) Unraveling the regulatory connections between two controllers of breast cancer cell fate. Nucleic Acids Res. 42: 6839-6849

[51] Wang J, Landscape and flux theory of non-equilibrium dynamical systems with application to biology. Adv. Phys. 64 (2015) 1-137

[52] Zhou P, Li T, Construction of the landscape for multistable systems: Potential landscape, quasi-potential, A-type integral and beyond. J. Chem. Phys. 144 (2016) 094109

[53] Freidlin MI, Wentzell AD, Random perturbations. In: Random Perturbations of Dynamical Systems, Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics) 260 Springer, New York, NY. 1998

[54] Foster DV, Foster JG, Huang S, Kauffman SA, A model of sequential branching in hierarchical cell fate determination. J. Theor. Biol. 260 (2009) 589-597

[55] Colgan J, Manley JL, TFIID can be rate limiting in vivo for TATA-containing, but not TATA-lacking, RNA polymerase II promoters. Genes Dev. 6 (1992) 304-315

[56] Muldrow TA, Campbell AM, Weil PA, Auble DT, MOT1 can activate basal transcription in vitro by regulating the distribution of TATA binding protein between promoter and nonpromoter sites. Mol. Cell. Biol. 19 (1999) 2835-2845

[57] Guillemette B, Drogaris P, Lin HH, Armstrong H, Hiragami-Hamada K, Imhof A, Bonneil E, Thibault P, Verreault A, Festenstein RJ, H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. PLoS Genet. 7 (2011) e1001354

[58] Nightingale KP, Gendreizig S, White DA, Bradbury C, Hollfelder F, Turner BM, Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. J. Biol. Chem. 282 (2007) 4408-4416

[59] Berrozpe G, Bryant GO, Warpinski K, Spagna D, Narayan S, Shah S, Ptashne M,) Polycomb responds to low levels of transcription. Cell. Rep. 20 (2017) 785-793

[60] Soleimani VD, Yin H, Jahani-Asl A, Ming H, Kockx CE, van Ijcken WF, Grosveld F, Rudnicki MA, Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. Mol. Cell 47 (2012) 457-468

[61] Dhasarathy A, Kajita M, Wade PA, The transcription factor snail mediates epithelial to mesenchymal transitions by repression of estrogen receptor-alpha. Mol. Endocrinol. 21 (2007) 2907-2918

[62] Warburg O, On the origin of cancer cells Science 123 (1956) 309-314

[63] Jullien J, Vodnala M, Pasque V, Oikawa M, Miyamoto K, Allen G, David SA, Brochard V, Wang S, Bradshaw C, Koseki H, Sartorelli V, Beaujean N, Gurdon J, Gene resistance to transcriptional reprogramming following nuclear transfer is directly mediated by multiple chromatin-repressive pathways. Mol. Cell 65 (2017) 873-884

[64] Jehanno C, Flouriot G, le Goff P, Michel D, A model of dynamic stability of H3K9me3 heterochromatin to explain the resistance to reprogramming of differentiated cells. Biochim. Biophys. Acta 1860 (2017) 184-195

[65] Russo J, Santucci-Pereira J, de Cicco RL, Sheriff F, Russo PA, Peri S, Slifker M, Ross E, Mello ML, Vidal BC, Belitskaya-Lévy I, Arslan A, Zeleniuch-Jacquotte A, Bordas P, Lenner P, Ahman J, Afanasyeva Y, Hallmans G, Toniolo P, Russo IH, Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. Int. J. Cancer 131 (2012) 1059-1070

[66] Pandey R, Dou Y, H2A.Z sets the stage in ESCs. Cell Stem Cell 12 (2013) 143-144

Appendix

A Parsimonious model of bivalent chromatin

An intriguing specificity of the so-called bivalent chromatin of stem cells is the presence of repressive marks (low acetylation and high methylation of enhancer H3K27, E-H3K27) and permissive marks (H3K9 acetylated and H3K4 methylated). A speculative hypothesis to explain bivalent chromatin may be based on the wide range decompaction of chromatin by acetylation of H3K9, which could be indirectly responsible for the relative closure of enhancers. Enhancers are precisely characterized, when active, by acetylated H3K27. The bivalent marks would therefore reflect a relative increase in the b/a ratio of basal to regulated transcription frequencies.

Model ingredients.

In this integrated model intended to reconcile several observations with a minimum of hypotheses, stationary solutions will be obtained directly by skipping time-dependent differential equations. We will consider the existence of a single acetylation enzyme (HAT) such as CBP, capable of acetylating both H3K9 and H3K27, and a single enzyme (HDAC) capable of deacetylating them. The only difference is that in bulk chromatin, H3 lysines (H3K9, but also possibly H3K27), can be acetylated autonomously by the HAT, whereas acetylation of E-H3K27 is assisted by TFs recruiting the HAT at their target enhancers. The second postulate is that TFs have a large number of cryptic binding sites in genomic DNA that are normally not accessible in the closed chromatin of differentiated cells, but become accessible in case of generalized decompaction. Assuming that the cellular content in TF is approximately constant, this fixation will mechanically reduce its presence on enhancers, and as a consequence decrease the maintenance of their acetylated state. For simplicity, acetylation and deacetylation of H3K9 and E-H3K27 are assumed to follow traditional Michaelis-Menten velocities with the same Michaelis constant (K_{HAT} and K_{HDAC}). The enzymes are supposed to bind to both lysines, but to be significantly sequestrated by H3K9 only, considering that the E-H3K27 sites are restricted to enhancers, so that for a diffusing enzyme, the accessible concentration of H3K9 is much higher than that of E-H3K27.

Fraction of acetylated H3K9.

H3K9 acetylation and methylation are mutually exclusive marks, but we will consider here that the dynamics of H3K9

acetylation/deacetylation is fast enough compared to that of methylation, to allow considering only the unmethylated fraction of H3K9. Let us define dimensionless Michaelis constants weighted by the substrate concentrations, with

$$[\text{H3K9}]\text{tot} = N$$

$$K_A = K_{HAT}/N$$

and

$$K_D = K_{HDAC}/N$$

The maximal velocity of acetylation is

$$V_A = c_A[\text{HAT}]\text{tot}\,[\text{acetyl-CoA}]$$

where c_A is the catalytic rate, and the velocity of deacetylation is

$$V_D = c_D[\text{HDAC}]\text{tot}$$

.

Writing V the sum of maximal velocities $V = V_A + V_D$, we define fractional maximal velocities

$$\theta = \frac{V_A}{V}$$

and

$$1 - \theta = \frac{V_D}{V}$$

Using this nomenclature, the fraction of acetylated H3K9 (written ρ) and that of deacetylated H3K9 ($1 - \rho$) are given by the traditional zero-order mechanism formulated in Table 1.

TF sequestration by H3K9-acetylated chromatin.

Simple statistics and systematic sequencing have shown that consensual and near-consensual DNA binding sites for most TFs are widespread in the genome, but that only a few of them correspond to genuine regulatory elements or enhancers. ChIP-seq experiments confirmed that these cryptic putative binding sites are generally not occupied in native chromatin, suggesting that their accessibility is prevented by chromatin closure. Hence, we will postulate here that chromatin loosening by acetylation could render the cryptic sites accessible. In random sequences, cryptic sites are distributed on average every n nucleosomes. Their concentration is $R = [\text{H3K9ac}]/n = \rho N/n$, n being about 20 for a consensus sequence of 6 base pairs in a random sequence. If the TF binds to these sites with an average dissociation constant $K = k_d/k_a$, a fraction of the TF (of constant total concentration F) will be sequestrated, yielding only a residual free concentration f, such that

$$R = \frac{(F - f)(K + f)}{f} \tag{A.1a}$$

16

Table 1: Acetylated and deacetylated fractions of unmethylated H3K9.

$\dfrac{[\text{H3K9}]\text{ac}}{[\text{H3K9}]\text{tot}}$	$\rho = \dfrac{(1-\theta)(K_A+1) + \theta(K_D-1) - \sqrt{((1-\theta)(K_A+1)+\theta(K_D-1))^2 - 4\theta(1-2\theta)K_D}}{2(1-2\theta)}$
$\dfrac{[\text{H3K9}]}{[\text{H3K9}]\text{tot}}$	$1-\rho$

giving

$$f = \frac{1}{2}\left(F - K - R + \sqrt{(F+K+R)^2 - 4RF}\right) \quad \text{(A.1b)}$$

E-H3K27 acetylation.

The knowledge of the free concentrations of TF and enzymes finally allows to predict the acetylation status of E-H3K27. Contrary to that of H3K9, E-H3K27 acetylation is supposed to necessitate previous TF binding. In turn TF binding is poorly efficient in absence of E-H3K27ac and is then greatly facilitated by E-H3K27ac. The rates used for the different reactions are listed in Table 2.

In absence of specific data, we will arbitrarily assume that the catalytic rate and Michaelis constant of the HAT are the same for H3K9 in absence of F and for E-H3K27 in presence of TF. In turn, enzymatic sequestration is assumed to be caused by H3K9 only considering the minor contribution of the enhancers in the genomes. Using the first-order or pseudo-first order rates of Table 2, the stationary probabilities of E-H3K27 acetylation is

$$P(H3K927ac) =$$
$$\cfrac{1}{1 + \cfrac{k_d\left[N(k_d+hf)(K_A+1-\rho) + V\theta\right](1-\theta)}{hf\left[N(k_d+k_af)(K_D+\rho) + V(1-\theta)\right]\theta}} \quad \text{(A.2)}$$

Replacing in this equation f by its value given in Eq.(7b) and ρ by its value given in Table 1, allows to express $P(H3K927ac)$ as a function of the variable θ only. As represented in Fig.S1, E-H3K27 can be largely deacetylated in spite of an overall hyperacetylation in the cell. In other words, a global increase of acetylases activity can simultaneously open bulk chromatin and alter enhancers, leading to an increase of b and a decrease of a for $\theta > 0.5$ (Fig.S1).
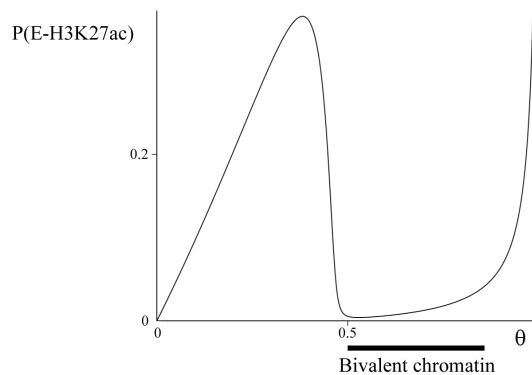

Fig. S1. Curve drawn to Eq.(A.2) for the set of parameters $k_a = k_d = 1$, $h = 0.001$, $K_a = K_d = 0.1$, $N = 3000$, $n = 10$ and $F = 100$

This model is minimalist in that it is based only on the competition between bulk histone acetylation and E-H3K27 acetylation, and recourses to very few ingredients in a field, chromatin epigenetics, which involves a lot of molecular actors. Other correlations and amplification phenomena, not incoporated here are naturally expected to complete the picture. For instance, methylation maintains non-acetylated lysines in non-acetylable form by competition, thereby locking the system. Conversely the histone variant H2A.Z, whose profile is parallel to that of H3K27ac at the level of enhancers, is also an important player in enhancer functions by causing nucleosomal depletion [66]. Certain marks are clearly correlated: H3K9 acetylation is associated to H3K4 methylation, and is closely related to DNA methylation in multiple ways including: (i) the presence of methyl-cytosine binding protein MBD1 in H3K9 methyl transferase complexes like SETDB1 and CAF1, (ii) the recruitment of HDACs by methylated DNA-bound MeCP2 and conversely (iii) the recruitment of a DNMT by H3K9me3-bound HP1. H3K9 methylation can stabilize chromatin in a non-acetylable form in post-mitotic differentiated cells. It should be noted in this respect that H3K9me3 is particularly persistent and constitutes the main lock against the risk of de-differentiation and reprogramming [7]. From the Waddington-type view, the developmental selection of the genes to close in the course of differentiation, proceeds more

17

Table 2: E-H3K27 acetylation/deacetylation cycle assuming that the HAT is recruited by the TF. f is the concentration of free TF (not sequestrated by cryptic elements) $k_a f$ and k_d are the pseudo-first order and first order rate constants of TF binding to and dissociation from acetylated chromatin respectively. The binding of TF to the deacetylated enhancers is supposed possible, but with a constant h lower than k_a.

Transition	Rate
E-H3K27 $\rightarrow$ E-H3K27-F	hf
E-H3K27-F $\rightarrow$ E-H3K27	k_d
E-H3K27-F $\rightarrow$ E-H3K27ac-F	$\dfrac{c_A[\text{free HAT}]}{K_{\text{HAT}}} = \dfrac{V}{N}\dfrac{\theta}{K_A + 1 - \rho}$
E-H3K27ac-F $\rightarrow$ E-H3K27ac	k_d
E-H3K27ac $\rightarrow$ E-H3K27ac-F	$k_a f$
E-H3K27ac $\rightarrow$ E-H3K27	$\dfrac{c_D[\text{free HDAC}]}{K_{\text{HDAC}}} = \dfrac{V}{N}\dfrac{1 - \theta}{K_D + \rho}$

by absence of expression than by active repression. Repressive machineries like polycomb complexes, which suppress the expression of many genes in embryonic stem cells [25], could only ratify preexisting low transcription states, as suggested in [59], thereby locking selectively the genes which have already lost their Waddington-type fight in antagonistic genetic circuits. The relative contributions of basal vs regulated expression (b/a) are the main regulator of the balance of cellular dedifferentiation/differentiation in the present model. A speculative and a general scenario depicted in Fig.S2, can thus be proposed around this central core, which connects several results of cellular biology, from metabolism to multistability, which is the fundamental hallmark of differentiation.
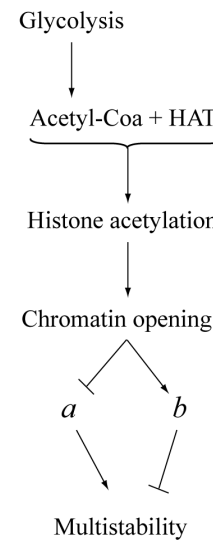
Fig. S2. Hypothetical scheme connecting metabolism to inhibition of multistability, equivalent to dedifferentiation.