Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis

Marco-Antonio Mendoza-Parra*, Valeriya Malysheva, Mohamed Ashick Mohamed Saleem, Michele Lieb, Aurelie Godel and Hinrich Gronemeyer*

Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Equipe Labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique UMR 7104, Institut National de la Santé et de la Recherche Médicale U964, University of Strasbourg, Illkirch, France.

***Corresponding authors:**

Marco Antonio Mendoza-Parra; E-mail: marco@igbmc.fr;
Hinrich Gronemeyer; E-mail: hg@igbmc.u-strasbg.fr;
Phone: +(33) 3 88 65 34 73; Fax: +(33) 3 88 65 34 37

Supplementary Information

Supplemental Figure S13: Signal propagation approach for identifying key TFs involved in cell fate acquisition.

Supplemental Figure S14: Neuronal cell fate induction by CRISPR/dCas9 activation in P19 cells.

Supplementary Figure S15: CRISPR/dCas9-mediated transcription activation of neuronal factors in F9 cells.

Supplementary Figure S16: Reprogramming of F9 cells towards the neuronal lineage.

Supplemental Figure S17: Gene regulatory wiring reconstructed from the reduced Gene regulatory network governing the RA-driven cell-fate decisions in F9 and P19 cells.

Supplemental Figure S18: Comparison of negative RA-mediated gene regulation sion in mouse ES, F9 and P19 cells.

Supplemental File S1: Reconstructed Gene regulatory network in Cytoscape format. This file is accessible through the Cytoscape environment and displays (i) the temporal changes in mRNA expression levels in the context of the GRN for a total of 2981 nodes and 44931 edges and (ii) a reduced GRN composed of 115 nodes and 732 edges displaying relevant « hubs » and « bottlenecks », which correspond to nodes retained after applying a topological ranking score (for details see materials and methods).

Supplemental Methods

Rationale for using EC cells

The first demonstration of the existence of pluripotent stem cells in mammals came from transplantation experiments with early mouse embryos in ectopic sites, where they generated teratocarcinomas. These tumors contained embryo carcinoma cells (EC), from which a single cell could regenerate a tumor containing both EC and differentiated progeny. Moreover, when injected back into the embryo, karyotypically normal EC cells could contribute to many different cell types in the resulting chimeras[1]. Some EC cells, like P19, undergo initial steps of differentiation *in vitro* and not only form neurons and astrocytes but express also central nervous system (CNS) region-specific genes in presence of the morphogen RA[2-3]. Thus, RA induces in P19 cells important regulatory events of early CNS organogenesis, making them an attractive experimental model for research on cell fate decisions.

Nowadays ES or iPS cells are considered more attractive models than EC cells, which are tumorigenic. However, the overexpression of c-MYC in classical iPS protocols suggests that induced pluripotency and oncogenic tumorigenesis may be related processes [4]. Importantly, from an experimental point of view salient features of neurogenesis are very efficiently recapitulated in P19 cells, as RA induced in >90% of cells the expression of NESTIN and TUBB3, markers of multi-potent neuronal precursors and post-mitotic neurons, respectively. In contrast, RA induced NESTIN in <60% (TUBB3 in <30%) mouse ES cells and even less induction was seen in RA-primed iPS cells [5]. Thus, our results suggest that P19 cells are more committed towards the neuronal fate than ES cells.

RT-qPCR

The following primers have been used:

Neurog1: Forward: CCTTCTTTGTGACTGGCTCA

Reverse: CCCTTTTCCAAACCACACTG

NeuroD Forward: CGCAGAAGGCAAGGTGTC

Reverse: TTTGGTCATGTTTCCACTTCC

For assessing the relative gene expression all qPCR measurements were normalized relative to the constitutively expressed *36b4* mRNA levels assessed with the following primers:

36b4 Forward: AATCTCCAGAGGCACCATTG

Reverse: CCGATCTGCAGACACACACT

Transcriptomics

For transcriptomics analysis by microarrays hybridization, 250ng of extracted RNA was used for cDNA reverse transcription including a biotin-labelling procedure then hybridized on Affymetrix GeneChip® Mouse Gene 1.0 ST array (Affymetrix, Santa Clara, CA, USA) as described in the manufacture's protocols (GeneChip® whole transcript sense target labelling assay manual: P/N 701880 Rev.4). The arrays were washed and stained with streptavidine-phycoerythrin in an Affymetrix GeneChip® Fluidics station 450 using the script protocol F450-0007, then scanned with the Affymetrix Gene Chip Scanner 3000 7G. Expression values were generated with the Affymetrix software Expression Console version 1.1 using sketch quantile normalization and median polish summarization as in Robust Multi-array Average (RMA) [6].

Chromatin immunoprecipitation and FAIRE assays

Cells were were fixed with 1% paraformaldehyde (Electron Microscopy Sciences) for 30 minutes at room temperature. ChIP assays were performed following previously described conditions: chromatin sonication and immunoprecipitation in lysis buffer (50mM Tris–Cl pH8, 140mM NaCl, 1mM EDTA, 1% Triton, 0.1% Na-deoxycholate) complemented with protease inhibitor cocktail (Roche 11873580001); 2x washes with lysis buffer; 2x washes with lysis buffer containing 360mM NaCl; 2x washes with washing buffer (10mM Tris–Cl pH8, 250mM LiCl, 0.5% NP-40, 1mM EDTA, 0.5% Na-deoxycholate); 2x washes with 1x TE; elution at 65°C; 15 min in elution buffer (50mM Tris–HCl pH8, 10mM EDTA, 1% SDS). RXRα has been immunoprecipitated with purified polyclonal antibodies generated by immunization of rabbits with the following peptide:

mRXRA: PB105 (MDTKHFLPLDFSTQVNSSSLNSPTGRGC).

RNA polymerase II (sc-9001 H-224), H3K27me3 (Upstate 07-449) and H34me3 (Abcam 8580) antibodies were purchased from their corresponding commercial suppliers. RXRA ChIP assays were performed with $6x10^6$ cells per time point; while histone modification marks were evaluated with $2x10^6$ cells. FAIRE assays were performed as described previously [7-8]. Briefly, F9 or P19 cells were fixed with 1% para-formaldehyde (Electron Microscopy Sciences) for 10 minutes at room temperature. $1x10^6$ cells diluted in a volume of 300 µl lysis buffer were extracted with 1 volume of a phenol / chloroform / isoamylalcohol preparation (50/48/2%, respectively; termed 'PCI'). The interphase/organic fraction has been washed a second time with 150 µl lysis buffer, then recovered together with the first extract. The recovered aqueous phases have been extracted with the same PCI preparation other 3 times, then treated with Proteinase K (Roche ref: 03115852001); 37°C during 2 hours, followed by an ethanol precipitation in presence of Glycogen (MP ref: Glyco001).

All ChIP and FAIRE assays were validated using positive and negative controls. ChIP validation assays were performed by quantitative real-time PCR (qPCR, Roche instrument LC480 light cycler) using Quantitect kit (Qiagen). Specifically, RXRα ChIP assays were validated based on previously described enrichment information; H3K27m3, H3K4me3 and RNA polymerase II enrichment performance was monitored at promoter regions of genes responding to the RA treatment. Finally, FAIRE assays were validated by taken RXRα enrichment information as well as promoter regions of genes responding to the RA treatment. ChIP validation assays were performed by quantitative real-time PCR (qPCR, Roche instrument LC480 light cycler) using Quantitect kit (Qiagen).

Massive parallel sequencing

10 ng of the ChIPed DNA wasused to prepare multiplexed sequencing libraries (NEXTflexTM ChIP-seq Bioo Scientific; ref: 514120). Sequencing was done on the Illumina instrument HiSeq2000 (4 ChIP samples per lane). Regular Illumina pipelines were used for image processing and base calling. Sequence files were then aligned to the mouse genome assembly following by default parameters (mm9; Bowtie).

Enrichment pattern detection and intensity profile normalisation

Enriched sites in all ChIP-seq and FAIRE-seq datasets were identified with MeDiChISeq [9], a regression-based approach, which after a learning process defines a representative binding pattern from the dataset analysed [9]. Specifically, we applied a P-value confidence cut off of $10^{-2.5}$ and a peak intensity threshold defined by a random enrichment background model (Poisson distribution; P-value 0.995).

With the aim of comparing in multiple profiles not only the enrichment patterns at defined genomic regions but also the associated read count intensity levels, we first normalized the profile amplitudes. As we demonstrated in a previous comparative study of RNA polymerase II ChIP-seq profiles the superiority of non-linear normalisation [10], we have developed and used a new generalized method for normalizing histone modification marks (H3K4me3, H3K27me3), RNA Polymerase II profiles and FAIRE-seq enrichment patterns, which was used in a recent study [11]. This new quantile-based approach normalises first all datasets associated to a given target (five time points) and subsequently the normalised target datasets are brought to the same scale by z-score normalization. A detailed description of this quantile normalization procedure for a variety of ChIP-seq and enrichment-related NGS datasets will be available under the name of 'Epimetheus', a user-friendly dedicated normalization tool (Epimetheus; manuscript in preparation).

Dynamic regulatory maps and RA-driven gene regulatory network reconstruction

GRN reconstruction was done such that (1) the TFs listed in the GRN collection of CellNet were associated to genes differentially expressed in either F9 or P19 cells and (2) the recovered TF-TG interactions were enriched by at least 10% in the DREM-predicted co-expression paths. Using these criteria DREM predicted several co-expression paths from P19 or F9 transcriptomes by computing co-expression bifurcation points (BPs), which correspond to the time points at which a subset of the co-expressed genes changes in their transcriptional behaviour relative to the other members of the co-expression path.

As part of the visualisation options in Cytoscape, the changes in expression levels per nodes were coloured in a heatmap format, such that dynamic changes become apparent (**Supplemental file S1).**

Furthermore, a two-step GRN reduction process was applied by using topological metrics of the cytoscape plugin Cytohubba [12]. Specifically, clustering coefficient metrics were first used to extract the top50 most clustered nodes (plus first neighbour) for each gene regulatory program (common or F9/P19-specific). These three populations were combined into a single network to which a second layer of topological metrics reduction was applied. This reduction is based on scoring for bottleneck nodes, as previous reports demonstrated that in addition to highly connected nodes ("hubs"), "bottleneck" nodes (defined as those interconnecting highly connected nodes or hubs in the system) can represent highly relevant components of a system [13]. Particularly, in signal transduction systems, bottlenecks can correspond to essential entities required in the system for the continuous flow of the signal transduction. We have thus reconstructed a reduced GRN composed of 80 nodes and 626 edges, to which a ranking colour code (heatpmap) displaying the hub importance metrics has been attributed. This network reveals major nodes with potential functional roles in RA-dependent cell fate acquisition and contains four sub-networks; two regulate cell differentiation (pluripotency and HOX factors), two others specify neuronal or endodermal (regulatory) factors. All sub-networks are interconnected revealing the connectivity between common and cell-fate specific programs. Moreover, this network reveals how key factors, whose implication in cell fate specification was known or is described here, are temporally connected to a plethora of other genes involved in maintenance of pluripotency or cell cycle regulation, in addition to neurogenesis. The reduced GRN displays also relevant F9-specific factors that are at the basis of endodermal fate acquisition. Finally, the organisation of the reduced GRN and its visualisation has been performed with the cytoscape package Cerebral [14] **(Supplemental Fig. S17; Supplemental File S1)**.

Targeted gene knockouts with CRISPR/Cas9 system

P19 cells were transfected with a pair of double nickase plasmids encoding a D10A mutation in Cas9 nuclease and a 20nt guide RNA targeting the following genes: *Gbx2* (Santa Cruz Biotech: sc-420503-NIC), *Tal2* (Santa Cruz Biotech: sc-423262-NIC). P19 cells were transfected with Lipofectamine 2000 (Invitrogen; Cat. No. 11668-027); 18h later the medium was replaced and supplemented with puromycin (200ng/ml). After one week of selection, cells were transferred to 96 well plates for single colony screening. Single cell-derived cultures were treated with ATRA during 24 hours and evaluated for loss of

gene expression by qPCR. Positive candidates were re-evaluated in the context of a complete kinetics (0, 2, 6, 24, 48 and 72h of ATRA treatment) and further characterised for the transcriptional expression of other neuronal-specific genes.

CRISPR/dCas9 transcriptional activation and immunohistochemistry

P19 cells were transfected with CRISPR/dCas9 activation plasmids targeting the following factors: *Tal2* (Santa Cruz Biotech: sc-423262-ACT), *Gbx2* (Santa Cruz Biotech: sc-420503-ACT), *Tshz1* (Santa Cruz Biotech: sc-431049-ACT), *Lhx2* (Santa Cruz Biotech: sc-421425-ACT), *Dmrt1* (Santa Cruz Biotech: sc-420576-ACT). P19 cells, grown in P19-DMEM on coverslips, were transfected with Lipofectamine 2000 (Invitrogen; Cat. No.11668-027). 18h later the medium was removed and replaced with P19-DMEM complemented with antibiotics (puromycin 300ng/ml, hygromycin 400ug/ml, blasticidine 5ug/ml), and cells were treated with either BMS961 or ethanol. After another 3 days, the medium with antibiotics was replaced by P19-DMEM. At day 6 the cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences), followed by 3 x 5 min washes in PBS. Cells were permeabilised (Triton 0.1% in PBS; 15 min at room temperature) and blocked (10% heat inactivated FCS in PBS) during 1 h at room temperature. Cells were washed 3 x 5 min in permeabilization buffer, then incubated with the primary antibodies anti-beta III Tubulin/anti-TUBB3 (Abcam: ab14545) or anti-MAP2 (ab32454). After one hour incubation, cells were washed 3 x 10 min with permeabilisation buffer followed by incubation with a secondary antibody (Donkey anti-mouse IgG (H+L) Antibody Alexa 555: Invitrogen A-31570; Donkey anti-rabbit IgG (H+L) antibody Alexa 488; Invitrogen A-21206) and/or DAPI (Invitrogen: D3571). After 1 hour at room temperature, cells were washed for 3 x 10 min in permeabilisation buffer, twice with milli Q water and finally mounted on microscope slides.

References

1 Rossant, J. in *Essentials of stem cell biology* eds R. Lanza & A Atala) 35-38 (Academic Press, 2014).

2 Soprano, D. R., Teets, B. W. & Soprano, K. J. Role of retinoic acid in the differentiation of embryonal carcinoma and embryonic stem cells. *Vitam Horm* **75**, 69-95, doi:S0083-6729(06)75003-8 [pii]

10.1016/S0083-6729(06)75003-8 (2007).

3 Varga, B. V. *et al.* Generation of diverse neuronal subtypes in cloned populations of stem-like cells. *BMC Dev Biol* **8**, 89, doi:10.1186/1471-213X-8-89 (2008).

4 Riggs, J. W. *et al.* Induced pluripotency and oncogenic transformation are related processes. *Stem Cells Dev* **22**, 37-50, doi:10.1089/scd.2012.0375 (2013).

5 Sartore, R. C. *et al.* Retinoic acid-treated pluripotent stem cells undergoing neurogenesis present increased aneuploidy and micronuclei formation. *PLoS One* **6**, e20667, doi:10.1371/journal.pone.0020667

PONE-D-10-02885 [pii] (2011).

6 Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264, doi:10.1093/biostatistics/4.2.249 (2003).

7 Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**, 877-885, doi:gr.5533506 [pii]

10.1101/gr.5533506 (2007).

8 Simon, J. M., Giresi, P. G., Davis, I. J. & Lieb, J. D. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc* **7**, 256-267, doi:nprot.2011.444 [pii]

10.1038/nprot.2011.444 (2012).

9 Mendoza-Parra, M. A., Nowicka, M., Van Gool, W. & Gronemeyer, H. Characterising ChIP-seq binding patterns by model-based peak shape deconvolution. *BMC Genomics* **14**, 834, doi:10.1186/1471-2164-14-834 (2013).

10 Mendoza-Parra, M. A., Sankar, M., Walia, M. & Gronemeyer, H. POLYPHEMUS: R package for comparative analysis of RNA polymerase II ChIP-seq profiles by non-linear normalization. *Nucleic Acids Res* **40**, e30, doi:10.1093/nar/gkr1205 (2012).

11 Chaligne, R. *et al.* The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Res* **25**, 488-503, doi:10.1101/gr.185926.114 (2015).

12 Chin, C. H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* **8 Suppl 4**, S11, doi:1752-0509-8-S4-S11 [pii]

10.1186/1752-0509-8-S4-S11 (2014).

13 Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. & Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* **3**, e59, doi:06-PLCB-RA-0302R2 [pii]

10.1371/journal.pcbi.0030059 (2007).

14 Barsky, A., Gardy, J. L., Hancock, R. E. & Munzner, T. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics* **23**, 1040-1042, doi:btm057 [pii]

10.1093/bioinformatics/btm057 (2007).