Highly dynamic chromatin interactions drive neurogenesis through gene regulatory networks

Valeriya Malysheva[1,2,4], Marco Antonio Mendoza-Parra[1], Matthias Blum[1,3] and Hinrich Gronemeyer[1,4]

[1] 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.

[2] Present address:

The Babraham Institute, Babraham Research Campus, Cambridge, UK

[3] Present address:

European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[4] Corresponding authors:

Valeriya Malysheva

Email: valeriya.malysheva@babraham.ac.uk

Phone: +(44) 1223 49 6505

Hinrich Gronemeyer

Email: hg@igbmc.u-strasbg.fr

Phone: +(33) 3 88 65 34 73

Cell fate transitions are fundamental processes in the ontogeny of multicellular organisms[1] and aberrations can generate pathologies[2]. Cell fate acquisition is a highly complex phenomenon which involves a plethora of intrinsic and extrinsic instructive signals that direct the lineage progression of pluripotent cells. Previously, we defined the dynamic gene regulatory networks underlying neuronal differentiation induced by the morphogen all-trans retinoic acid (RA)[3]. Here we reveal the signal-propagating role of the chromatin interactome in the commitment and propagation of the initiating signal in early neurogenesis by reconstructing dynamic loop-enhanced Gene Regulatory Networks (eGRNs) that integrate transcriptome, chromatin accessibility and long-range chromatin interactions in a temporal dimension. We observe a highly dynamic re-wiring of chromatin interactions already at very early stages of neuronal differentiation. Long-range chromatin interactions are massively reorganized; only 30% of the initial interactome is conserved through cell differentiation, while new interactions are established already 6 hours after induction of neurogenesis. By integration of chromatin interactions together with temporal epigenome and transcriptome data, we identify a group of key regulatory elements that respond to and propagate the initial signal. Our data reveal an enormous capacity of the morphogen to reorganize long-range chromatin interactions by "reading" distant epigenetic signals and chromatin accessibility to drive cell fate acquisition. These results suggest that the differential establishment of chromatin contacts directs the acquisition of cell fate.

To understand the molecular mechanisms by which cell fate decision signals are propagated in neuronal cell fate induction by all-*trans* retinoic act (RA) and to predict its key response regulators/enhancers, we developed an integrative regulatory network approach that combines i) temporal transcriptome data for RA–induced neurogenesis mouse stem cells using the P19 system[3], ii) transcription factor-target gene (TF-TG) relationships, recovered from the CellNet[4] platform, iii) TF binding sites from TF-targeted ChIP-seq data present in the public domain, extracted from the qcGenomics[5] collection that contains >8,000 TF-targeted mouse ChIP-seq data sets for >600 TFs, iv) temporal data of accessible regulatory elements from FAIRE-seq analysis[3] and v) chromatin architecture dynamics evaluated by HiC[6] (for details see Methods).

Examination of the higher-order chromatin structure at sub-chromosomal scale showed moderate reorganization with the majority of topologically-associated domain (TAD) borders maintained (Fig. 1a), in agreement with previous studies reporting largely cell-invariant TADs[7–9]. However, while overall TAD sizes remained largely constant (Extended Data Fig. 1a), we observed the reorganization of 40% of TADs (Fig. 1a, b left panel). Moreover, numerous chromatin structure changes became apparent when we followed the fate of all reproducibly identified loops individually (Fig. 1b right panel, c). Interestingly, during neuronal differentiation of P19, the fraction of long-range interactions increases (Extended Data Fig. 1b). Thus, a subset of TADs undergoes concerted, domain–wide rearrangements together with internal changes of interaction frequencies and internal contact rearrangements as an early response to the morphogen.

Monitoring the dynamics of long-range chromatin loops emanating from the promoters of differentially expressed genes (DEGs), we observed a massive dynamic reorganization of these interactions during P19 cell differentiation. Indeed, only 32% of the initial pool of DEG promoter-associated loops pre-existed before RA treatment and were maintained, while 15,879

loops were newly formed, in part transiently, during the first 48h of treatment (Fig. 1d). Thus, at the 48h time point about the same amount of DEG loops (16,410) exist as before RA-treatment (16,911) but the large majority has been dynamically remodelled.

To comprehensively map the cell fate decision processes controlled by chromatin reorganization and epigenome changes, we matched experimentally identified TF (ChIP-seq) binding sites that display 'open' chromatin (FAIRE-seq) in P19 cells within promoter-interacting regions (PIRs) of long-range chromatin loops emanating from the promoters of DEGs. We complemented these interactions with TF-TG associations involving DEGs (Fig. 2a). Using this approach, we reconstructed a loop-enhanced GRN (eGRN) and identified long-range chromatin interactions of DEG promoters with enhancers and the cognate TF(s) involved in the regulation of DEGs through chromatin interaction.

To establish the temporal evolution of neurogenesis-relevant gene-regulatory information we used TETRAMER[3] (www.ngs-qc.org/qcgenomics). This computational approach evaluates the coherence of TF-TG relationships with the temporal evolution of transcription activation[3], purging irrelevant cell type-restricted information imported from the CellNet database and chromatin interactions that serve structural rather than signalling roles. At the same time, it evaluates the ability of the reconstructed eGRN to reconstitute the temporal transcriptional regulatory cascade resulting in the acquisition of a neuronal cell fate by evaluating the capacity of every node to induce the cell-fate specific programs.

The reconstructed eGRN revealed temporally evolving regulatory landscapes underlying RA-induced neuronal cell differentiation initiated by RA binding to RAR-RXR retinoid receptor heterodimers[10]. Signal propagation from the activated RAR-RXRα towards the final phenotype gave rise to an eGRN comprising 2,548 nodes and PIRs and 10,858 edges, which acted as

direct mediators linking TFs and cognate DEGs; many of these involved critical long-range chromatin interactions (Fig. 2b; Supplementary Information File 1).

Importantly, this analysis predicted many previously discovered drivers of neurogenesis as key TFs, validating the accuracy of the integrative method[3,11–15]. In addition to key TFs, this approach enables the identification of PIRs/enhancers that are crucial for propagating the transcription regulatory cascades towards neurogenesis (Fig. 3). This is due to their association with master TFs, supporting a model in which master TFs act through long-range chromatin interactions. In addition, TETRAMER-based analysis of the reconstructed eGRN divided signal propagating enhancers in 2 groups – those with high yield of final phenotype reconstitution and high signal propagation (>91%) and those with very low (<2%) propagation efficacy. Notably, there are many more of such low efficacy "supportive" than "key" enhancers. This suggests a concept of "driver" enhancers governing the expression of crucial TFs through looping whose action can be supported by "supportive enhancers" (Fig. 2b). Potentially, the action of low efficacy supportive enhancers could, at least partially, compensate for the absence of a key regulator, which mediates the action of a critical TF through chromatin looping, in case of failure (e.g., due to a mutation). Alternatively, the very high numbers of loops between DEG promoters and "supportive enhancers" and their enormous dynamics and combinatorial potential invite speculations about a possible role in the adaptive evolution of this system in the absence of any mutational burden.

Several examples of propagation of the initial RA signal through key TFs and enhancers looping towards DEGs, as predicted from the reconstructed eGRN, are illustrated in Figure 3 and described below. In particular, RA induces TAL2 and GBX2 expression (Fig. 3a, Supplementary Information File 2), which is essential for midbrain neurogenesis[11,12]; both factors have been shown in our previous study to act as master regulators of neuronal cell fate

acquisition[3]. Together, through CellNet-predicted association, they activate their target gene Meis1, which controls the neurogenic cascade[13]. MEIS1 in turn binds to regulatory regions on chromosome 2 according to publicly available ChIP-seq experiments performed in P19 cells (Fig. 3b, binding sites marked by arrows). Notably, these regulatory elements become accessible (as seen from FAIRE-seq) only at 48h after RA treatment and MEIS1 binds these regions at the same time point. In turn, these MEIS1-bound regulatory elements interact through looping with the Pax6 promoter region, which encodes one of key TFs involved in the development of neural tissues[14], and is followed by upregulation of Pax6 expression. Only at this time point the corresponding chromatin interactions take place (Fig. 3b; loops marked by arrows; see Extended Data Fig. 2a for validations by the '3C' method). This mechanism is likely to explain the recently reported regulatory effect of MEIS1 on Pax6 in the development of cerebellar granule cells[16]. The signal propagation proceeds further towards interneuron-related Nr2f2 and Nr2f1[17] through predicted TF-TG interactions of PAX6, as well as through the long-range chromatin interaction, as PAX6 binds to a regulatory element that loops towards the promoter of Nr2f2 gene. Altogether this example shows how the dynamic regulation of transcriptome, chromatin accessibility and chromatin organization are elegantly orchestrated for efficient propagation of signal(s) driving neuronal cell fate acquisition.

Similarly, through a complex cascade of regulatory events, RA activates expression of ASCL1 (Fig. 3c), a TF well-known to be essential for neurogenesis[15]. ASCL1 binds to a range of regulatory elements in neuronal progenitor cells (Fig. 3d) on different chromosomes[18], propagating the signal towards its target genes Dbx1, Sox11 and Foxp2, all involved in neuronal fate specification[19–23], through chromatin looping as predicted by our eGRN (Figs. 3c, 3d; Extended Data Fig. 2b; Supplementary Information File 3).

We have previously predicted and validated DMRT1 as a key factor involved in retinoid-induced neurogenesis[3]. Here, using the eGRN-based signal propagation we predict that DMRT1 propagates the RA signal via loop formation to upregulate Mafb expression (Fig. 3e, 3f; Extended Data Fig. 2c; Supplementary Information File 4). MAFB is an important regulatory node in the neurogenesis-specific eGRN as it acts on several regulatory elements across the genome and induces the expression of several TF genes, including those of Foxp2 and Zfp516 that had already pre-established chromatin interactions between their gene promoters and cognate MAFB-bound enhancers. Thus, the signal can be propagated through pre-established loops, as well as through the newly formed loops in response to external cell fate-instructive signals.

In order to verify whether these chromatin structure dynamics is specific to neurogenesis cell fate acquisition, we conducted the same integrative studies for the differently committed ES-like F9 cells that undergo RA-induced endodermal differentiation under conditions where P19 cells differentiate along the neuronal lineage. Remarkably, all key loops of the Pax6 interaction landscape (Fig. 3a, Supplementary Information File 1) are absent in F9 during the entire period of RA treatment (Supplementary Information File 5). At the same time, Sox11 (Fig. 3c) and Zfp516 (Fig. 3e) interactomes are shared between the two cell fates at least at the early stages of differentiation (Supplementary Information File 5). These partially shared interactomes may reflect their similar stem cell origin during the non-differentiated state. However, the observed differences in chromatin interactions (RA-Gbx2-Pax6 subnetwork) suggest that the difference in acquired cell fate is critically encoded in the chromatin structure, which would explain the divergent RA-induced cell fates of F9 and P19 cells in view of their nearly identical epigenomes, chromatin accessibility landscapes and transcriptomes in the untreated state (data not shown). Examination and comparison of TADs between F9 and P19 in non-differentiated state highlights major differences in their chromatin structure, supporting the hypothesis that

cell fate predisposition may be written down in the 3D structure of their genomes (Extended Data Figures 3a, 3b). Overall, these observations suggest that (i) the chromatin structure encodes cell fate-specific information and (ii) its reorganization is crucial for proper signal propagation and cell lineage progression.

In summary, we have reconstructed an eGRN for neuronal cell fate acquisition that reveals the connectivity between TFs, regulatory elements and target genes, integrating the information transfer through critical large distance chromatin interactions. Moreover, we provide a map of key drivers of early neurogenesis, whose connectivity and contribution to the signal propagation can be easily explored via Cytoscape. We also provide a method for a comprehensive integrative analysis of multi-scale data for the discovery of key driver genes and enhancers and describe the large class of "supportive" enhancers, which may contribute to response robustness and/or adaptive evolution. Our approach is a powerful integrative analysis that can be applied to any cell model or process for the discovery of key regulatory nodes crucial for dynamic physiological or pathological processes.

METHODS

Cell culture. F9 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 4.5 g/l glucose; P19 cells were grown in DMEM supplemented with 1 g/l glucose, 5% FCS and 5% delipidated FCS. Both media contained 40 μg/ml Gentamicin. F9 or P19 EC cells were cultured in monolayer on gelatine-coated culture plates (0.1%). For cell differentiation assays, RA was added to plates to a final concentration of 1μM for different exposure times.

Transcriptome and Epigenome assays. We have previously described the data on

transcriptome dynamics and chromatin immunoprecipitation assays[3]; they are available from the Gene Expression Omnibus (GEO) database (GSE68291). ChIP-seq profiles for MEIS1, ASCL1 and MAFB are accessible from GEO (GSM2188919, GSM2188920, GSM2188924, GSM1187228 and GSM1964739).

HiC experiments. Chromatin organization has been assessed by HiC[6]. The original HiC protocol has been improved, increasing the ligation yields and modifying the steps that favour chromatin de-crosslinking, while keeping the conventional HiC workflow. Briefly, cells were crosslinked with 1% PFA (Electron Microscopy Sciences) for 30 min at room temperature, after which cells were washed twice in ice-cold PBS. Cells were collected by centrifugation (1000 r.p.m. for 10 min at 4°C), washed once with PBS followed by a second centrifugation step. The supernatant was removed and cell pellets were snap frozen in liquid nitrogen and stored at −80°C. Per HiC sample 20-25M cells were used. Cell pellets were incubated in 1 ml ice-cold lysis buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% NP-40, protease inhibitor cocktail (Roche) for 15 min on ice. The cells were Dounce homogenized (30 strokes, 1 min pause on ice, 30 strokes). After centrifugation (2500 r.p.m. for 5 min at 4°C), nuclei were washed once with NEBuffer 2 and resuspended in 250 µl of NEBuffer 2. The sample was split into aliquots of 50 µl. Each sample was further complemented with 307 µl of NEBuffer2, 38 µl of 1% SDS (0.1% final concentration), incubated for 10 min at 65°C and put on ice directly afterwards. To quench SDS, 44 µl of 10% Triton X-100 was added. The restriction digest was set up directly afterwards and performed overnight at 37°C with agitation (750 r.p.m.) with *HindIII* (NEB; 20 µl of 20 U/µl). Using biotin-14-dCTP (Life Technologies), dATP, dGTP and dTTP (all at a final concentration of 28 µM), the *HindIII* restriction sites were then filled in with Klenow (Thermo Fisher Scientific) for 75 min at 37°C mixing at 750 r.p.m. for 5 s each 5 min. 3C sample was generated from 5M cell aliquots in parallel with HiC by replacing the fill-in mix with corresponding amount of nuclease-free water and treated in same way as HiC

sample. The fill-in reaction was directly followed by overnight ligation at 16°C in a total volume of 8.2 ml ligation buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, 100 µg/ml BSA, 1% Triton X-100, 10,000 units T4 DNA ligase (NEB 400U/µl)) per 5 million cells starting material. After ligation, reverse crosslinking (65°C overnight in the presence of Proteinase K (Roche)) was followed by two sequential phenol/chloroform/isoamylalcohol extractions. Purified solution was then concentrated on 30K AMICON 15 ml centrifugal filter (span at 3000 r.p.m.) to 300 µl and DNA was precipitated in ethanol for 1 hour at −80°C. The DNA was span down (13,000 r.p.m. for 30 min at 4°C). The pellets were washed twice in 70% ethanol, dried and resuspended in 100 µl TE buffer. HiC samples were combined and washed twice on Amicon 0.5 ml centrifugal filter unit with TE and concentrated to 100 µl in case of HiC sample and to 25-30 µl in case of 3C sample. DNA concentration was determined using a Nanodrop. The efficiency of biotin incorporation was assayed by amplifying a ligation product, followed by digestion with *HindIII* or *NheI*. If the ligation product was fully or almost fully digested (Extended Data Fig. 4) by NheI, HiC sample was used further.

To remove biotin from non-ligated fragment ends, 40 µg of Hi-C library DNA were incubated with T4 DNA polymerase (NEB) for 4 hours at 20°C, followed by phenol/chloroform purification and DNA precipitation overnight at −20°C or for 1 hour at -80°C. After DNA pellet resuspension, the sonication was carried out to generate DNA fragments with a size peak around 200 bp (Covaris Sonolab 7 settings: duty factor: 10%; peak incident power: 175W; cycles per burst: 200; time: 60 sec).

A double size selection using AMPure XP beads (Beckman Coulter) was performed to select 100-400 bp DNA fraction. The ratio of AMPure XP beads solution volume to DNA sample volume was adjusted to 0.7:1. After incubation for 10 min at room temperature, the sample was transferred to a magnetic separator (DynaMag-2 magnet; Life Technologies), and the supernatant was transferred to a new 1.5 ml tube, while the beads were discarded. The ratio of

AMPure XP beads solution volume to DNA sample volume was then adjusted to 1.1:1 final. After incubation for 15 min at room temperature, the sample was transferred to a magnet (DynaMag-2 magnet; Life Technologies) and washed twice with 80% ethanol. The DNA was eluted in 100 µl of TLE (10 mM Tris-HCl pH 8.0; 0.1 mM EDTA).

Biotinylated ligation products were isolated using MyOne Streptavidin C1 Dynabeads (Life Technologies) on a DynaMag-2 magnet (Life Technologies) in binding buffer (5 mM Tris pH8, 0.5 mM EDTA, 1 M NaCl) for 45 min at room temperature. After one wash in binding buffer and two washes in wash buffer (5 mM Tris, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) the DNA-bound beads were resuspended in a final volume of 55.5 µl of nuclease-free water and used for end repair, adaptor ligation and PCR amplification with NebNext Ultra library preparation kit for Illumina (NEB #E7370S) following the manufacture protocol. Note that after adaptor ligation, beads were washed twice in washing buffer, twice in 10 mM Tris-HCl pH 8.0, resuspended in 25 µl of 10 mM Tris-HCl pH 8.0 followed by PCR. For the PCR amplification 5 µl DNA-bound beads were used with Index primer and universal PCR primer diluted 1/10. Only five PCR cycles were made to obtain the HiC library that was purified with AMPure XP beads at 1:1 ratio, followed by paired-end high-throughput sequencing on HiSeq2500.

HiC data processing. Raw paired-end sequencing reads were mapped against the mouse genome (mm9) with Bowtie2. Low quality reads (MAPQ<=10), PCR duplicates and interactions falling within the same restriction fragment were filtered out. Hi-C contact maps were constructed at 5 kb resolution, then normalized by matrix balancing using the ICE algorithm. Importantly, to avoid the contamination of the downstream analysis by over-normalized interactions, the cut-off for the extremely low and high occupancy bins has been adjusted individually for each sample, instead of using fixed thresholds proposed in the ICE method. To define the upper and lower outliers cut-off, we evaluated the distribution of total

counts per bin and applied the modified z-score (outliers > 3.5) to fix the upper cut-off and the first local minima to define the lower cut-off (Extended Data Fig. 5a, b). Statistically significant interactions were identified by Fit-Hi-C[24]. To define reproducible interactions between biological replicates we used sdef[25] R package. Only interactions that appear in both biological replicates, pass the cut-off defined by sdef and have a normalized count greater than or equal to 3 in both replicates were considered as significant and have been used for further integrative analysis and GRN reconstruction.

Chromatin structure, epigenome and transcriptome integration. We have associated open-chromatin regions - defined by FAIRE-seq assay - to the promoter-associated distal GAPs. FAIRE localization sites were then compared with a comprehensive collection of TF ChIP-seq assays retrieved from the public domain[5]. Note that the TF collection in use in this study includes a large number of datasets in addition to those provided by the ENCODE consortium; thus, our analysis is a comprehensive comparative study not only because of the large number of datasets used but also with respect to the diversity of the cellular systems considered. Transcriptome, RXR binding sites from ChIP-seq, TF annotations from public datasets and HiC long-range chromatin interactions were integrated and visualized using the Cytoscape platform (version 2.8.3). The signal propagation was performed multiple times using TETRAMER tool[3] and a randomized network approach as control.

REFERENCES

1. Bedzhov, I., Graham, S. J. L., Leung, C. Y., Zernicka-goetz, M. & Zernicka-goetz, M. Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. **369**, (2014).
2. Reynaud, D. *et al.* Article IL-6 Controls Leukemic Multipotent Progenitor Cell Fate

and Contributes to Chronic Myelogenous Leukemia Development. *Cancer Cell* 661–673 (2011). doi:10.1016/j.ccr.2011.10.012

3. Mendoza-Parra, M.-A. *et al.* Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis. *Genome Res.* gr.208926.116 (2016). doi:10.1101/GR.208926.116

4. Cahan, P. *et al.* CellNet: Network Biology Applied to Stem Cell Engineering. *Cell* **158,** 903–915 (2014).

5. Mendoza-Parra, M. A., Van Gool, W., Saleem, M. A. M., Ceschin, D. G. & Gronemeyer, H. A quality control system for profiles obtained by ChIP sequencing. *Nucleic Acids Res.* **41,** (2013).

6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-.).* **326,** 289–93 (2009).

7. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–80 (2012).

8. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

9. Dixon, J. R. *et al.* Chromatin Architecture Reorganization during Stem Cell Differentiation. *Nature* **518,** 331–336 (2015).

10. Gronemeyer, H., Gustafsson, J.-åke & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* **3,** 950–964 (2004).

11. Nakayama, Y., Kikuta, H., Kanai, M. & Yoshikawa, K. Gbx2 functions as a transcriptional repressor to regulate the specification and morphogenesis of the mid – hindbrain junction in a dosage- and stage-dependent manner. *Mech. Dev.* **130,** 532–552 (2013).

12. Achim, K. *et al.* The role of Tal2 and Tal1 in the differentiation of midbrain

GABAergic neuron precursors. *Biol. Open* 990–997 (2013). doi:10.1242/bio.20135041

13. Marcos, S. *et al.* Meis1 coordinates a network of genes implicated in eye development and microphthalmia. *Development* 3009–3020 (2015). doi:10.1242/dev.122176

14. Kallur, T., Gisler, R., Lindvall, O. & Kokaia, Z. Molecular and Cellular Neuroscience Pax6 promotes neurogenesis in human neural stem cells. *Mollecular Cell. Neurosci.* **38,** 616–628 (2008).

15. Huang, H. S., Turner, D. L., Thompson, R. C. & Uhler, M. D. Ascl1-induced neuronal differentiation of P19 cells requires expression of a specific inhibitor protein of cAMP-dependent protein kinase Holly. *J. Neurochem.* **120,** 667–683 (2012).

16. Owa, T. *et al.* Meis1 Coordinates Cerebellar Granule Cell Development by Regulating Pax6 Transcription , BMP Signaling and Atoh1 Degradation. *J. Neurosci.* **38,** 1277–1294 (2018).

17. Hu, J. S. *et al.* Coup-TF1&2 (Nr2f1 and Nr2f2) control subtype and laminar identity of MGE- derived neocortical interneurons. *Development* **144,** 2837–2851 (2017).

18. Wapinski, O. L. *et al.* Hierarchical mechanisms for transcription factor-mediated reprogramming of fibroblasts to neurons. *Cell* **155,** (2013).

19. Inamata, Y. & Shirasaki, R. Dbx1 triggers crucial molecular programs required for midline crossing by midbrain commissural axons. *Development* **141,** 1260–1271 (2014).

20. Karaz, S. *et al.* Neuronal fate specification by the Dbx1 transcription factor is linked to the evolutionary acquisition of a novel functional domain. *Evodevo* **7,** 1–13 (2016).

21. Wang, Y., Lin, L., Lai, H., Parada, L. F. & Lei, L. Transcription Factor Sox11 is Essential for both Embryonic and Adult Neurogenesis. *Dev. Dyn.* **242,** 638–653 (2013).

22. Tsui, D., Vessey, J. P., Tomita, H., Kaplan, D. R. & Miller, F. D. FoxP2 Regulates

Neurogenesis during Embryonic Cortical Development. *J. Neurosci.* **33,** 244–258 (2013).

23. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171,** 557–572.e24 (2017).

24. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24,** 999–1011 (2014).

25. Blangiardo, M., Cassese, A. & Richardson, S. sdef: An R package to synthesize lists of significant features in related experiments. *BMC Bioinformatics* **11,** (2010).

Data availability. Sequencing data reported in the current study were deposited at the Gene Expression Omnibus repository (GEO) of the National Center for Biotechnology Information under the GEO accession number GSE111875.

Acknowledgements. We thank Dr. Mikhail Spivakov for advice concerning the HiC data analysis. We thank as well the IGBMC Microarray and Sequencing platform (France Génomique consortium—ANR-10-INBS-0009). This study was supported by AVIESAN-ITMO Cancer, the Ligue National Contre le Cancer, the Institut National du Cancer (INCa) and the Agence Nationale de la Recherche (ANRT-07-PCVI-0031-01, ANR-10-LABX- 0030-INRT and ANR-10-IDEX-0002-02).

Author contributions. H.G. designed the study; V.M. performed the experiments; V.M. and M.B. carried out computational analysis; V.M. reconstructed the GRN with the help of M.A.M.P.; V.M. and H.G. wrote the manuscript.

FIGURE LEGENDS

Figure 1. Dynamics of chromatin organization during the early stages of retinoic acid (ATRA) induced neuronal differentiation of P19 cells. **a.** TAD dynamics. The majority of TADs conserve the TAD borders, however, several hundred were reorganized resulting in an increased number of unique TADs 48h after morphogen treatment. **b.** HiC heatmaps with corresponding directionality indexes (DIs) displaying a TAD split at chromosome 6 (yellow arrows) after 48h of treatment with ATRA (left panel) and changes in the inner structure of TADs and overall loop dynamics (black and yellow arrows) after 6h and 48h of ATRA-induced neurogenesis (right panel) **c.** Genome-wide dynamics of long-range chromatin interactions. **d.** Dynamics of long-range chromatin loops connecting promoters of differentially expressed genes (DEGs) with their putative regulatory elements

Figure 2. Reconstructed loop-enhanced Gene Regulatory Network (eGRN) of early ATRA-induced neurogenesis. **a.** Scheme of data integration: transcriptome, CellNet TF-TG correlations, chromatin accessibility, TF-targeted ChIP-seqs and HiC long-range interactions. PIR, promoter-interacting region. **b.** Reconstructed eGRN. Nodes represent genes (red and blue nodes corresponding to up- and downregulated genes, respectively, at 48h of ATRA treatment) and regulatory elements (key elements, dark green; supportive, light green). Node color reveals gene expression ratio at 48h after RA treatment relative to non-differentiated sample as indicated. The size of nodes corresponds to the yield of signal propagation as predicted by TETRAMER (see Methods) **c.** Schematic representation of elements in the reconstructed GRN (b).

Figure 3. eGRN subnetworks (**a**, **c** and **e**) propagating the initial driver signal – ATRA, activating RAR-RXRa - to the target nodes of the final phenotype (48h) through the key regulatory TFs and key regulatory elements. The color of genes shows gene expression levels

at 48h after RA treatment; the color of enhancers reflects their "driver" or supportive role in signal propagation. Unidirectional arrows reveal the flux of the initial signal from the source node to the target node. Figures **b, d** and **f** give integrative views of the ChIP-seq profiles of the indicated TFs, the dynamics of chromatin accessibility (FAIRE-seq), RNA polymerase II recruitment (Pol II), and illustrates the state/occupancy of regulatory elements (black arrows) with key TFs regulating the signal propagation predicted by TETRAMER.

EXTENDED DATA

Extended Data Figure 1. TAD and chromatin loops length in early RA-induced neurogenesis. TAD size is largely conserved during early neurogenesis (**Fig. 1a**), while chromatin interactions tend to span longer distances (**Fig. 1b**).

Extended Data Figure 2. 3C validation of PIR – promoter interactions. 3C-PCR validation of several long-range chromatin interactions predicted as key for the signal propagation, represented in **Fig. 3** of the main text. Corresponding primers are provided in Extended Data File 7.

Extended Data Figure 3. Comparison of chromatin structure between F9 and P19 at non-differentiated state. TAD border comparison shows large differences in higher-order chromatin structure (**Fig. 3a**). **Fig. 3b** shows an example of TAD and inner TAD structure differences in non-differentiated cell lineages.

Extended Data Figure 4. HiC and 3C digestion tests. Digestion of a PCR amplicon generated from non-neighboring restriction fragments in both 3C and Hi-C sample. The amplicon was digested with *HindIII*, *NheI*, or not digested (0). High digestion rates with the *NheI* in the HiC sample indicates high filling-in and ligation efficiency. Representative cases for P19 (**Fig. 4a**) and F9 (**Fig. 4b**) samples are shown.

Extended Data Figure 5. HiC normalization adjustment. ICE bias vs raw count distribution when applying **a.** the standard ICE thresholds and **b.** sample-specific thresholds, showing that low count bins are not transformed into bins with artificially high count, enabling further downstream analysis.

SUPPLEMENTARY INFORMATION

Supplementary Information File 1. Full reconstructed eGRN of early RA-induced neurogenesis in an interactive Cytoscape file format.

Supplementary Information File 2. Subnetwork of full eGRN showing the transfer of signal from RA through Meis1, Pax6 and their regulatory elements (presented on **Fig. 3a** of the main text) in an interactive Cytoscape format.

Supplementary Information File 3. Subnetwork of full eGRN showing the transfer of signal from RA through Gbx2, Ascl1 and their regulatory elements (presented on **Fig. 3c** of the main text) in an interactive Cytoscape format.

Supplementary Information File 4. Subnetwork of full eGRN showing the transfer of signal from RA through Dmrt1, Mafb and their regulatory elements (presented on **Fig. 3e** of the main text) in an interactive Cytoscape format.

Supplementary Information File 5. F9 chromatin interactions of Pax6, Sox11 and Zfp516 promoters.

Supplementary Information File 6. HiC sample processing statistics reflecting the quality and complexity of the data.

Supplementary Information File 7. PCR primers used for 3C validations of key regulatory interactions described in the current study.

Figure 1

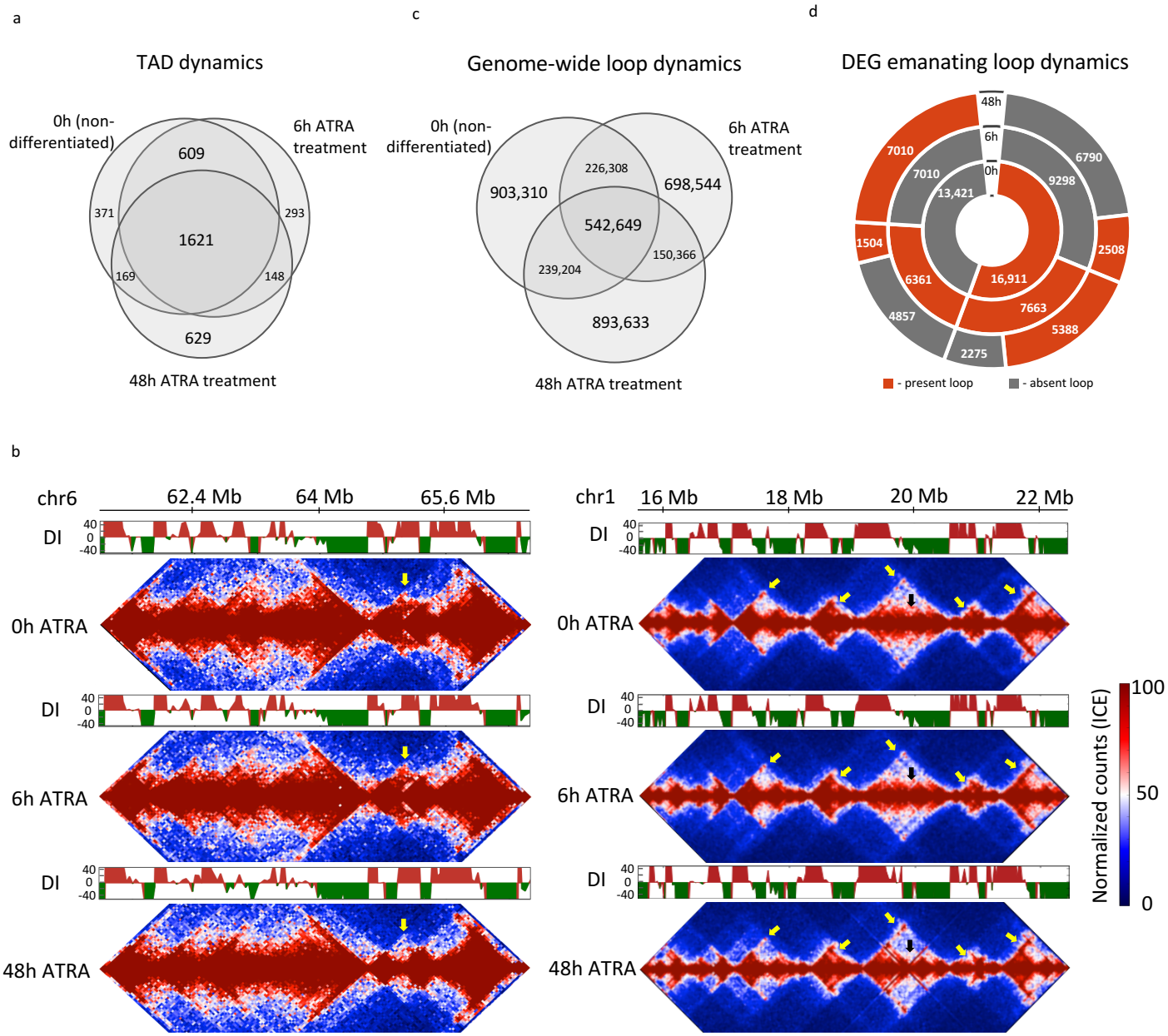

a
TAD dynamics
0h (non-differentiated)
6h ATRA treatment
48h ATRA treatment
609
371
293
1621
169
148
629
c
Genome-wide loop dynamics
0h (non-differentiated)
6h ATRA treatment
48h ATRA treatment
903,310
226,308
698,544
542,649
150,366
239,204
893,633
d
DEG emanating loop dynamics
48h
6h
0h
7010
7010
13,421
6790
9298
2508
1504
6361
16,911
7663
5388
4857
2275
- present loop
- absent loop
b
chr6
62.4 Mb
64 Mb
65.6 Mb
chr1
16 Mb
18 Mb
20 Mb
22 Mb
DI
0h ATRA
6h ATRA
48h ATRA
Normalized counts (ICE)
100
50
0

Figure 2

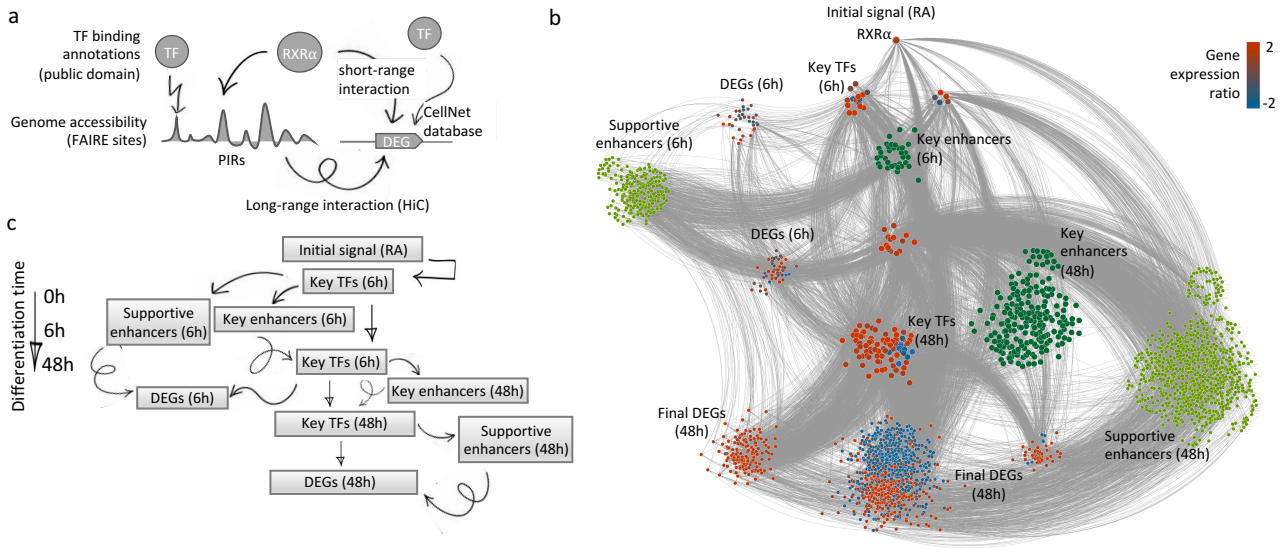
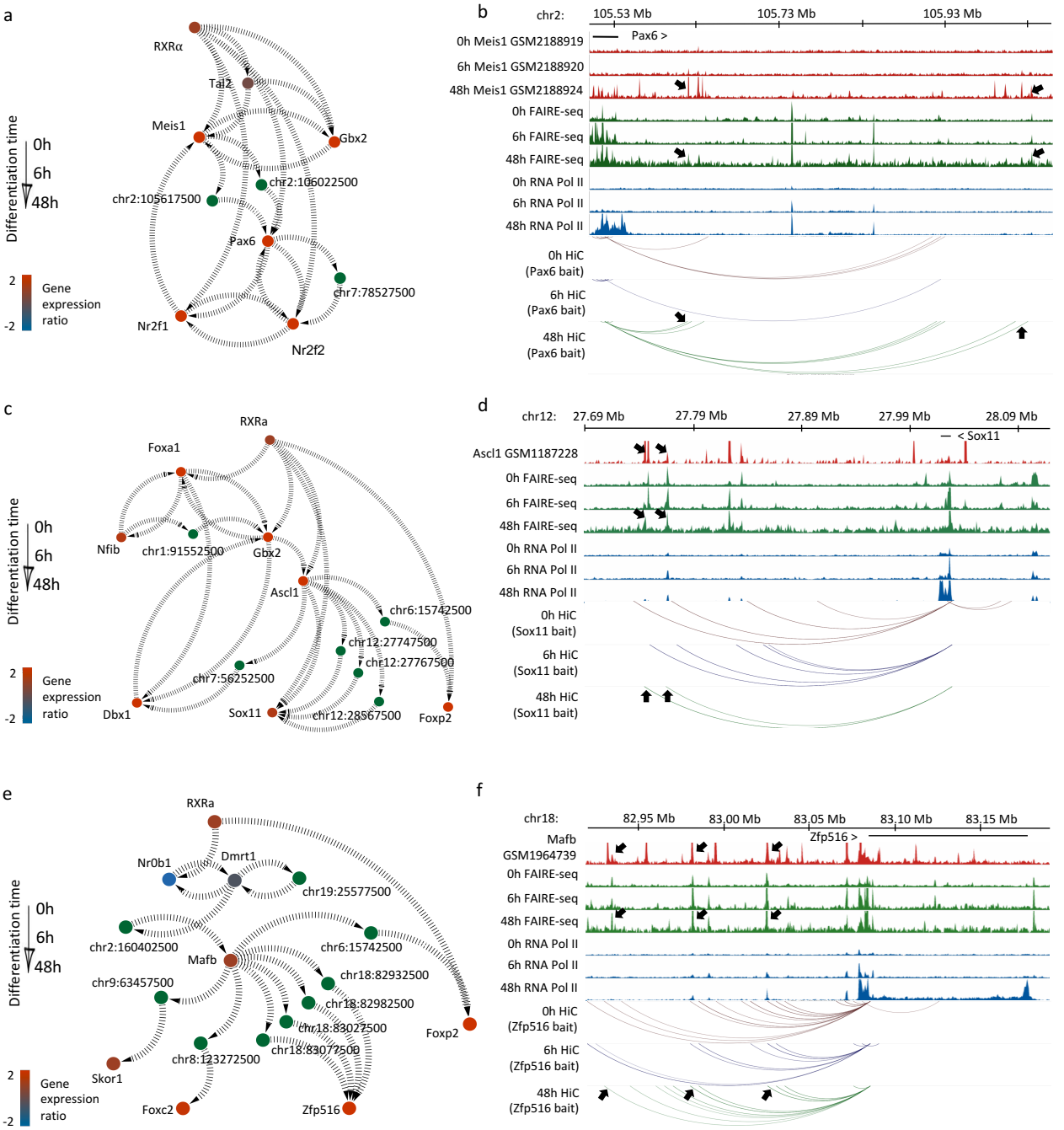

a
TF binding annotations (public domain)
TF
RXRα
TF
short-range interaction
CellNet database
Genome accessibility (FAIRE sites)
DEG
PIRs
Long-range interaction (HiC)
b
Initial signal (RA)
RXRα
Key TFs (6h)
DEGs (6h)
Supportive enhancers (6h)
Key enhancers (6h)
Gene expression ratio
2
-2
DEGs (6h)
Key enhancers (48h)
Key TFs (48h)
Final DEGs (48h)
Supportive enhancers (48h)
Final DEGs (48h)
c
Differentiation time
0h
6h
48h
Initial signal (RA)
Key TFs (6h)
Supportive enhancers (6h)
Key enhancers (6h)
Key TFs (6h)
DEGs (6h)
Key enhancers (48h)
Key TFs (48h)
Supportive enhancers (48h)
DEGs (48h)

Figure 3


a
RXRα
Tal2
Meis1
Gbx2
chr2:106022500
chr2:105617500
Pax6
chr7:78527500
Nr2f1
Nr2f2
Differentiation time
0h
6h
48h
2
-2
Gene expression ratio
b
chr2:
105.53 Mb
105.73 Mb
105.93 Mb
Pax6 >
0h Meis1 GSM2188919
6h Meis1 GSM2188920
48h Meis1 GSM2188924
0h FAIRE-seq
6h FAIRE-seq
48h FAIRE-seq
0h RNA Pol II
6h RNA Pol II
48h RNA Pol II
0h HiC (Pax6 bait)
6h HiC (Pax6 bait)
48h HiC (Pax6 bait)
c
RXRa
Foxa1
Nfib
chr1:91552500
Gbx2
Ascl1
chr6:15742500
chr12:27747500
chr12:27767500
chr7:56252500
Dbx1
Sox11
chr12:28567500
Foxp2
Differentiation time
0h
6h
48h
2
-2
Gene expression ratio
d
chr12:
27.69 Mb
27.79 Mb
27.89 Mb
27.99 Mb
28.09 Mb
< Sox11
Ascl1 GSM1187228
0h FAIRE-seq
6h FAIRE-seq
48h FAIRE-seq
0h RNA Pol II
6h RNA Pol II
48h RNA Pol II
0h HiC (Sox11 bait)
6h HiC (Sox11 bait)
48h HiC (Sox11 bait)
e
RXRa
Nr0b1
Dmrt1
chr19:25577500
chr2:160402500
Mafb
chr6:15742500
chr18:82932500
chr9:63457500
chr18:82982500
chr18:83027500
chr18:83072500
chr8:123272500
Foxp2
Skor1
Foxc2
Zfp516
Differentiation time
0h
6h
48h
2
-2
Gene expression ratio
f
chr18:
82.95 Mb
83.00 Mb
83.05 Mb
83.10 Mb
83.15 Mb
Zfp516 >
Mafb GSM1964739
0h FAIRE-seq
6h FAIRE-seq
48h FAIRE-seq
0h RNA Pol II
6h RNA Pol II
48h RNA Pol II
0h HiC (Zfp516 bait)
6h HiC (Zfp516 bait)
48h HiC (Zfp516 bait)