dcHiC: differential compartment analysis of Hi-C datasets

Jeffrey Wang[1,2,#], Abhijit Chakraborty[1,#]*, Ferhat Ay[1,3]*

[1]La Jolla Institute for Immunology, La Jolla, California, USA.

[2]The Bishop's School, La Jolla, California, USA.

[3]School of Medicine, University of California San Diego, La Jolla, California, USA.

*co-corresponding authors: abhijit@lji.org, ferhatay@lji.org

Abstract

Compartmental organization plays a role in important biological processes. However, its comparative analysis has been mainly limited to pairwise comparisons of contact maps with focus on finding compartment flips (e.g., A-to-B). Here, we introduce *dcHiC*, which utilizes Multiple Factor Analysis and a multivariate distance measure to systematically identify all compartmentalization differences among multiple contact maps. Evaluating dcHiC on three different collections of Hi-C data, we show its effectiveness and sensitivity in detecting biologically meaningful differences associated with cellular identity, gene expression, and lamin association. By providing a multivariate formulation, dcHiC immediately expands compartment analysis to new modalities in comparative genomics.

BACKGROUND

The three-dimensional organization of chromatin in the nucleus has been of interest to scientist for more than a century now. The observation that different chromosomes occupy a defined space in the nucleus dates back to Carl Rabl's work in animal cells in 1885 [1]. Since then, many experimental techniques have been developed to image and map chromatin allowing us to look at chromatin organization at an ever-increasing resolution. The greatest strides in this area have been made in the past decade following the advent of genome-wide conformation capture techniques. We now know that interphase chromosomes are folded into multiple layers of hierarchical structures with each layer contributing to establishment and maintenance of the epigenetic landscape that controls cellular state and function.

Among these, the megabase-scale compartmental organization of eukaryotic genomes has been shown to play a critical role in transcription, DNA replication, accumulation of mutations, and DNA methylation [2-12]. In broad terms, two types of compartments divide the genome into regions of open and active chromatin (compartment A) versus inactive and closed chromatin (compartment B) [13]. Further analysis of each compartment revealed subsets of regions with markedly different properties within each class called subcompartments [14, 15] as well as to a putative third class (intermediate or I) that is at the interface between A and B and is reorganized in tumors [9].

The main method to extract compartment information has been to analyze high-throughput chromosome conformation capture (Hi-C) contact maps using Principal Components Analysis (PCA) [13, 16, 17]. Briefly, this process involves distance normalization (Observed/Expected for each genomic distance) of the Hi-C contact map for each chromosome at a particular resolution (generally between 100kb to 1Mb) followed by transformation into a correlation matrix, where each entry *(i, j)* denotes the correlation of row *i* and row *j* (or column *i* and *j* since symmetric) of the distance-normalized Hi-C map. The eigendecomposition of the correlation matrix provides the eigenvectors, and the first eigenvector or principal component (PC1) typically represents the genomic compartments A and B. In the case that PC1 corresponds to chromosome arms or other broad patterns in the Hi-C map (e.g., copy number differences), the second principal component (PC2) is likely to represent A and B compartments. The A/B compartment labels are assigned to the positive/negative stretches of the selected PC even though depending on the implementation of eigendecomposition; it may be necessary to re-orient these assignments correctly using GC content or gene density.

Regardless if one is interested in the two major compartments or their more nuanced subsets, the magnitude and sign of eigenvalues derived from PCA have been the major determinants of compartment type. However, the standard PCA is limited in analyzing each Hi-C contact map individually and to date, there is no method to systematically compare compartmentalization across multiple Hi-C datasets. This is becoming an obstacle in analysis of the ever-increasing

chromatin conformation data, either from Hi-C or its variants [18-25], generated across many cell types and conditions [26]. Technical challenges such as selecting the correct PC representing compartments, differences in the orientation of the sign of PC values, and their scaling across different datasets become larger problems while comparing many Hi-C contact maps. Thus far, comparative compartment analysis has been mainly limited to examining compartment flips between two Hi-C maps at a time [27, 28].

Here, we introduce *dcHiC* (differential compartment analysis of Hi-C), a method that identifies statistically significant differences in compartmentalization among two or more contact maps, including changes that are not accompanied by a compartment flip. Our method employs Multiple Factor Analysis (MFA) [29], an extension of PCA, which combines measurements from more than one Hi-C contact map before performing generalized PCA on the concatenated set of maps **(Figure 1)**. *dcHiC* utilizes the normalized component scores computed by MFA to derive a multivariate distance measure [30] and employs Independent Hypothesis Weighting (IHW) [31] to estimate the statistical significance of compartment differences. In order to assess the biological significance of the identified differences, we applied *dcHiC* to several different collections of Hi-C datasets across various biological conditions, including mouse neuronal development (n=3), mouse hematopoiesis (n=10), and a set of primary human tissues and cell lines (n=11). Analyzing each Hi-C contact map at 100kb resolution, we identified a total of 981, 2569, and 2837 significant compartmentalization differences (FDR < 0.01) in the respective scenarios, covering about 4-10% of the genomes. *dcHiC* identified compartment changes for loci involving critical genes known to associate with cellular identities such as *Dppa2/4* and *Zfp42* for mouse embryonic stem cell differentiation, *Sox6*, *Meis1* and *Myc* in mouse hematopoiesis, along with human *Pcdh* and *CD84* genes within differential A compartments in the relevant tissue samples. Our results also highlight the importance of the generally ignored compartmentalization changes within the same compartment type (A or B) and the correlation of these changes with cell-type specific differences in lamin association and gene expression. Overall, *dcHiC* provides an integrative framework and an easy-to-use tool for comparative analysis of Hi-C maps that identifies biologically relevant differences in compartmentalization across multiple cell types. With immediate application to hundreds of publicly available Hi-C datasets, dcHiC will play an essential role in providing deeper insights into dynamic genome organization and its downstream effects.

RESULTS

dcHiC identifies compartments consistent with PCA-based approach

dcHiC uses Multiple Factor Analysis (MFA) to compare multiple Hi-C contact maps in an integrated way to identify all regions with statistically significant differences in their compartmentalization **(Figure 1)**. Although a group of data can be combined and analyzed by PCA, MFA is a systematic extension of PCA tailored to handle multiple data sets at once. MFA assigns an equal weighting to each defined group of variables to calculate the final product space [29, 32]. To demonstrate the correspondence between compartment scores from our

method and commonly used PCA-based approaches, we first compared dcHiC results against PC1 values from HOMER, a package representative of PCA-based methods. **Figure 2A-B** show the comparison of dcHiC and HOMER compartment scores in mouse embryonic stem cells (mESC) and neuronal progenitor cells (NPC) at 100Kb resolution. We observed a strong correlation (>0.95) between the final HOMER and dcHiC compartment scores for both cell types, reflecting the equal adeptness of MFA in assigning compartment scores and labels.

Eigendecomposition of Hi-C data assigns A/B compartments that correspond to euchromatin and heterochromatin states of the genome. Association with the nuclear lamina (or radial position in general) is another strong indicator of a broad-level chromatin state with heterochromatin localizing at the periphery and euchromatin at the nucleus center. Such organization is a conserved feature of eukaryotic genomes across most cell types except special cases [33, 34]. Here we used lamin B1 association profiles of mESC and NPC [35] as an independent measure of compartmentalization and compared with dcHiC scores. As expected, our compartment scores showed a strong negative correlation with lamin association, confirming the previous findings [27, 36] in both mESC and NPC samples **(Figure 2C-D)**. We further plotted the compartment score distribution of dcHiC, PC1 values from HOMER, and lamin B1 association scores highlighting the tight correlation among these three tracks at a large-scale **(Figure 2E-F, Supplemental Table 2)**. These results establish that both the MFA-based (dcHiC) and the PCA-based (HOMER) approaches accurately capture compartment patterns. However, dcHiC produces readily scaled scores and aligned orientation of compartment signs, which becomes important for the downstream comparative analysis, especially when multiple different Hi-C maps are compared.

Pairwise differential compartment analysis of mouse neuronal differentiation

Previous studies have reported substantial compartment switches during mouse embryonic cells (mESC) to neuronal progenitor cell (NPC) transition, a well-studied *in vitro* differentiation system [37, 38]. These differences have been studied further using replication timing profiling, lamin association mapping, and fluorescence *in situ* hybridization (DNA FISH) [8, 27, 36]. Therefore, we choose these two cell types to demonstrate dcHiC's utility in a pairwise comparison setting to replicate known compartment flips and identify significant changes that do not involve flips from one compartment type to another.

Overall, dcHiC identified 1369 pairwise significant differential compartments at 100kb resolution, covering 5.3% of the genome. For mESC and NPC, these differences constituted 52% and 63% of A (active) compartments, respectively. The differential compartments are further subdivided into two categories: flipping (A↔B) or matching (A→A or B→B) compartments. We observed a total of 657 (2.56% of the genome) significant differences belonging to the flipping class, and 712 (2.77%) are within matched compartments **(Figure 3A)**. We further classified changes within the matched compartments based on their compartment scores in mESC or NPC. **Figure 3B** shows the compartment score, lamin B1 association, and gene expression profiles of all six types of compartmental changes. The results demonstrate that the reported

regions show significant shifts in their compartment scores and that these are accompanied by changes in the expected direction for lamin association and gene expression measurements. For instance, regions that show a significant change from higher B compartment scores in mESC to a lower B compartment score in NPC are lamin associated regions with very low gene expression in mESC. For NPC, the same regions have significantly weakened lamin association accompanied by detectable gene expression levels. These results demonstrate that significant compartmentalization differences captured by dcHiC, in both switching or matching compartments, exhibit relevant changes in their radial position and transcriptional status.

To better annotate the compartmentalization changes, we also provide a module that performs GO term enrichment analysis [39] for the genes overlapping dcHiC reported regions. To demonstrate this feature, we perform GO term enrichment analysis on differential regions among mESC A compartments with lower compartment scores in NPC (either A or B in NPC; **Figure 3C**) and vice versa for the NPC A compartments **(Figure 3D).** The list of significant biological processes includes specific GO terms such as neuron differentiation, development, and neurogenesis for genes overlapping with A compartment regions in NPC that dcHiC deemed to have a significantly smaller PC score in mESC. For similar regions in mESC, we observe relevant but broad ontology terms such as cell cycle and chromosome organization. These results suggest that having an integrated GO term enrichment module may help interpret compartment changes at a high level (**Supplemental Table 3**).

Next, we investigate compartmental changes in genomic regions encompassing important developmentally regulated genes **(Figure 3E-J).** The murine *Dppa2/4* domain on chromosome 16 has three active genes in the mESC stage, and two of them, *Dppa2* and *Dppa4*, have been identified as pluripotency markers only expressed in mESCs by previous studies [36, 40]. When co-expressed with other transcription factors, *Dppa2* and *Dppa4* are shown to accelerate reprogramming kinetics, activate mESC enhancers, and convert somatic cells into induced pluripotent stem cells (iPSCs) [40]. Consistent with previous findings, dcHiC reports a significant shift from A (active) to B (repressed) compartment during mESC to NPC differentiation **(Figure 3E)**. The transition is also detectable from lamin B1 association profiles. This can be attributed to a large shift of the *Dppa2/4* domain from the mESC nucleus interior to the NPC periphery, an observation also confirmed by FISH experiments [36]. Notably, dcHiC reported the Dppa2/4 compartmentalization change as the most significant difference in chromosome 16. Likewise, dcHiC also detected a differential compartment covering another key developmentally regulated gene, *Zfp42* (also known as *Rex1*) **(Figure 3F)**. *Zfp42* is a zinc finger protein primarily expressed in undifferentiated stem cells and used as a pluripotency marker [41]. We observed that the spatial organization of *Zfp42* followed a similar pattern as that of the *Dppa2/4* domain, where the compartmental transition accompanying differentiation correlated with the suppression of gene expression and a validated (FISH) change in the radial positioning of the locus [36]. We also detected significant compartmental differences in the opposite direction (i.e., B in mESC to A in NPC) encompassing known marker genes of neuronal differentiation **(Figure 3G-H)** such as Pleiotrophin (*Ptn*) and EPH Receptor B1 (*Ephb1*) [42, 43]. Previous studies

on *Ptn* and *Ephb1* have demonstrated a significant subnuclear repositioning of the genes from the periphery to the nucleus center during mESC-NPC differentiation [36]. In contrast, a housekeeping gene highly expressed in both cell types (Actin Beta or *Actb*) has invariant compartmentalization with no significant difference reported by dcHiC **(Supplementary Figure S1A).**

Even though the above examples highlight cases where gene expression is tightly correlated with compartment changes, this is not necessarily the case for all genes. For instance, the transcription factor *Pou5f1/Oct4* is a pluripotency marker gene and is critical for the self-renewal of undifferentiated embryonic stem cells [44]. Although *Pou5f1/Oct4* is expressed at much lower levels in NPC, it has been shown to have invariable radial positioning during differentiation [36]. We reconfirm this result using our dcHiC calls, where we observe the *Pou5f1* gene in a conserved A compartment in both cell types with no significant difference in compartmentalization **(Supplementary Figure S1B).**

Multi-cell type differential compartment analysis of mouse neuronal differentiation

The same *in vitro* system used for differentiating mESC to NPC also allows further differentiation of NPCs to cortical neurons (CNs) [38]. This developmental lineage provides us with a system to demonstrate how dcHiC uses MFA and a multivariate distance measure to simultaneously compare the compartmentalization of more than two cell types. For such multiway comparisons, dcHiC provides a quick and straightforward approach to detect outliers in compartment scores among all samples instead of the traditional paradigm of multiple pairwise comparisons—an approach that becomes difficult to interpret with increasing sample size.

Applying dcHiC to 100kb resolution Hi-C data from mESC, NPC, and CN samples, we identified a total of 981 significant differential compartments covering about 3.8% of the genome. Compartment A constituted 42.3%, 47.1%, 49% of the respective genomes, while 4.9%, 5%, and 3.6% of these regions in the A compartment exhibited statistically significant changes for mESC, NPC, and CN lineages **(Figure 4A). Figure 4B** illustrates the changes in the compartment score patterns of all differential regions, while **Figure 4C** shows the breakdown of the fraction of differential compartments during the differentiation course. Notably, we observed that a large fraction of significant differences in compartmentalization were within the same compartment for both mESC to NPC (59%) and NPC to CN (71%).

Previous work has shown that compartmental dynamics are strongly correlated with gene expression and histone modifications [2, 5, 45]. We observed similar trends within the significant differential compartments identified by dcHiC **(Figure 4D)**. Among the bins with a significant compartment change, those that change from A to B have lower gene expression than bins that change within the A compartment. Similarly, bins that change from B to A have marginally higher expression than B compartment changes. A similar trend holds for the

average number of peaks (MACS *p-value* < 1e-5; Methods) within 100kb bins for active histone modifications H3K27ac and H3K4me1, while the heterochromatin mark, H3K9me3, shows an opposite trend **(Figure 4D)**. The modifications associated with active enhancers H3K27ac (also at active promoters) and H3K4me1 showed a sharp decrease in the average number of peaks in A to B transitions and a sharp increase for B to A compared to the changes in a promoter-associated histone mark H3K4me3.

With the dcHiC differential results, we also performed a functional enrichment analysis of the genes overlapping with the A compartments with higher compartment scores in CN than both NPC and mESC **(Figure 4E)**. As expected, we observed enrichment of genes associated with neuron generation, differentiation, and development in differential compartments with the highest A compartment score in CN. Using multiway comparisons, we also observed significant compartment score shifts within the same compartment that encompass important genes and would not have been captured by pairwise comparisons focusing on compartment flips **(Figure 4F, Supplemental Figure S2, Supplemental Table 4)**. **Figure 4F** demonstrates one such example of this: *Top1, a* gene that lies within a 100kb bin with a significant (FDR<0.01; 3-way comparison) A to A compartment switch identified by dcHiC. *Top1* is ubiquitously expressed across cells, as it encodes for the topoisomerase enzyme, a critical piece of cellular machinery [**46**] However, recent studies have shown that *Top1* also regulates gene expression in neurons and that inhibiting the gene irreversibly impairs synaptic function [47, 48]. As such, *Top1* is in the A compartment in all three cell types but has the strongest compartmentalization in CNs. We also observed a significant change in compartment associated with the *Rmst* gene, a long non-coding RNA that binds with Sox2 transcription factor and regulates neural fate (**Figure 4G**). Previous human studies have confirmed its brain-specific expression and its role in neurogenesis [49]. The cortical neuron subtype marker gene *Dcx* (or Doublecortin) and *Vldlr* (very low-density lipoprotein receptor) are two other interesting examples (**Supplemental Figure S2A, S2B**). *Dcx* is a known marker for differentiating neurons and appears to direct neuronal migration by regulating the microtubular organization and stability [50] while *Vldlr* is considered to be essential for the correct formation of layers in the cerebral cortex [51]. All of these were in the B compartment for mESC and the A compartment in the two other cell types. While the traditional method of looking for A/B sign changes would capture *Rmst*, *Dcx,* and *Vldlr* using pairwise comparisons, the significant changes within A compartment for *Top1* would have been missed completely.

Differential compartments in mouse hematopoietic system

The hematopoietic system is a developmentally regulated and well-characterized cell differentiation model system [52, 53]. The lineage provides an opportunity to understand the dynamic changes associated with the evolving chromatin structures during differentiation in detail. Hematopoiesis is a complex biological process that involves multiple cell-types and requires a systematic approach to study the dynamic genome organization. A recent study by Zhang et al. 2020 [28] provided a detailed overview of the spatiotemporal dynamics of

chromatin architecture during hematopoiesis. The study involved the classic hematopoietic model with ten primary stem, progenitor, and differentiated cell populations from mouse bone marrow **(Figure 5A-B)**.

Using the Hi-C data from this system, we carried out a multivariate differential analysis through dcHiC at 100Kb resolution. We observed very similar proportions for A and B compartments except for two cell types (granulocyte-macrophage progenitor (GMP) and megakaryocyte-erythrocyte progenitor (MEP)), where the primary cells showed a lower frequency of A compartments **(Figure 5A)**. We detected a total of 2,569 differential compartments across the ten cell types encompassing many of the genomic regions previously shown to undergo dynamic changes [28]. Another observation was an overall higher proportion of A compartments in the differential regions other than MEPs. The hierarchy of the hematopoietic system enabled us to map the A to B transitions between successive cell types. **Figure 5B** shows the dynamic changes in compartments during lineage differentiation within differential regions and throughout the genome. Although most of the changes are transitions within the same compartment, we observed an increased tendency of B to B transitions (~11%) in the late differentiating hematopoietic cells (MEP, MKP, MK, GMP, GR). For compartment flips, we observed an increasing pattern of A to B compartment transitions in the differential regions starting from the top of the hierarchy up to the point where CMP branches into GMP and MEP. Interestingly, we found the largest proportion of A to B transitions during CMP transition into either MEP or GMP, likely representing the locking of lineage choice by shutting down specific genes. The B to A compartment transition in the differential regions does not show any gradual pattern; however, we did notice an overall higher proportion of B to A shifts in the terminally differentiating myeloid and granulocyte lineages. These differential results suggest gradual transitions into the B compartment earlier on, followed by transitions into A during later stages of lineage commitment **(Figure 5B)**.

Next, we investigated the genes overlapping with the differential compartments involved in lineage differentiation and chromatin dynamics in hematopoiesis [54]. **Figure 5C** and **Supplementary Figure S3** show a set of critical genes that overlap with differential compartments from our multivariate analysis. A few notable examples include *Meis1*, a transcription factor required for maintenance of hematopoiesis under stress and over the long term **(Supplementary Figure S3A)** [55], as well as *Runx2* and *Sox6* essential for progenitor cell differentiation (**Supplementary Figure S3B, Figure 5C)** [56, 57]. We also identified *Myc* **(Supplementary Figure S3C)**, known for its role in balancing hematopoietic stem cell self-renewal and differentiation [58] adjacent to a differential compartment encompassing the *Pvt1* gene **(Supplementary Figure S3C)**. The long non-coding RNA *Pvt1* is known to harbor intronic enhancers that interact with *Myc* and promote *Myc* expression during tumorigenesis [59]. Previous analysis by Lara-Astiaso et al. [54] reported a set of critical genes for hematopoietic lineage differentiation; we identified 12 of the 13 at an *FDR* value < 0.01 (**Supplemental Figure S4**). We also found supporting evidence for some examples of lineage-specific genes highlighted by Zhang et al., 2020 [28]. The authors showed that increased gene-body

associating domains (GAD) scores are linked to active transcription and indicate cell-type specific features. We identified 9 out of 13 such differential GAD genes between ST-HSC and GR as part of multivariate differential compartments across the lineage (**Supplemental Figure S5**). One such example is the *Abca13* gene, where we observe it exclusively in the A compartment for GR and in B compartment for other cell types **(Figure 5D)**.

Finally, gene enrichment analysis of the differential compartments that are solely positive (A compartment) and have highest compartment scores in the granulocyte arm (GMP & GR) compared to the other cell types highlights the effect of compartments on the lineage-specific gene sets **(Figure 5E)**. Our results show significant enrichment of GO biological processes involved in leukocyte activation and neutrophil activation, as expected for the GMP arm (**Supplemental Table 5**). We also observed that genes overlapping with LT and ST-Hematopoietic stem cells (HSC) specific differential compartments were solely enriched in regulating hormone metabolic processes. HSCs produce all mature blood cells throughout their life, and a dynamic balance between quiescence and differentiation is a feature of HSCs [60]. We found the luteinizing hormone/choriogonadotropin receptor gene (*Lhcgr*) as part of the hormone regulatory process and that exclusively overlaps with differential A compartments in LT and ST-HSCs (**Supplementary Figure 3D**). A previous study [61] identified the *Lhcgr* gene as highly critical in regulating the HSC homeostasis and showed that its loss leads to abnormally elevated hematopoiesis and leukocytosis.

Differential Hi-C compartments in a compendium of human cell Hi-C data

We also applied the dcHiC approach to various human Hi-C datasets, including several primary tissues (Brain Astrocytes, Hippocampus, Liver, Bladder, Pancreas, Ovary, Endometrial, and Naïve Germinal B cells) and karyotypically normal cell lines (GM12878, IMR90, and HMEC). Comparing all these cell types simultaneously, our results reiterated the trend that differential compartments in the majority of the cases encompass functionally critical genes specific a specific cell type. Overall, we observed a lower proportion of A compartments, constituting about 45-50% of the genome. However, the proportion of differential calls in each compartment varied, with some cell types showing a higher fraction of differential A compartments compared to B **(Figure 6A)**. **Figure 6B** and **Supplemenal Table 6** show the GO term enrichment in differential regions with the highest A compartment score in the hippocampus, highlighting relevant biological processes. An example of such a region encompasses clustered protocadherin (Pcdhs), a neuronal recognition molecules family. Pcdh genes help in forming neuronal circuits through diversification and combinatorial effects of cell surface proteins. Previous studies have shown that *Pcdha1* helps organize serotonergic axons [62, 63] and *Pcdhb1* to modulate neuronal ensemble activity in the hippocampus [64]. As expected, we observed *Pcdha1* and *Pcdhb1* as part of a significantly different region with highest A compartment score in the hippocampus tissue (**Figure 6C**).

Figure 6D shows the GO term enrichment in differential regions with the highest A compartment scores in naïve B cells and GM12878 (a cell line derived from B cells) compared to other cell types. As expected, adaptive immune response and lymphocyte related functions are significantly enriched for genes in these regions (**Supplemenal Table 7**). One such example is *CD84*, a cell surface receptor that plays a critical role in modulating adaptive immune regulation [65, 66]. B cells activated by T helper cells undergo various states, including the formation of germinal centers (GCs), and previous studies have shown that *CD84* helps maintain B cell tolerance in GCs and prevent autoimmunity [67]. Figure 6E shows the CD84 gene overlapping with significantly different Hi-C bins and part of the highest A compartment scores in both naïve germinal B cells and GM12878 cell line.

GO term enrichment analysis for all other samples using their respective genes in differential A compartment bins demonstrated, in almost all cases, the top enrichment terms reflected the broad functions of the individual samples (**Supplemental Table 8-17**). The results depicted here reconfirm that open compartments of the human genome have a contributory rather than a deterministic effect on the biological functions based on the cell and tissue types. The results also reflect the flexibility and aptness of dcHiC in integrating multiple Hi-C data originating from a diverse set of cell types and still providing us with a meaningful outcome for downstream analysis.

CONCLUSIONS

Within the past decade, Hi-C sequencing has rapidly been adopted to study the 3D chromatin structure. Compartment analysis has been a fundamental tenet of this rapid rise; yet, thus far, it is still limited to the basic comparison of signs between different principal components, typically between two cell types at a time. This paper presented a new application, dcHiC, to compare Hi-C datasets. As an extension of the existing PCA-based method, dcHiC utilizes MFA to retrieve weighted and scaled component scores, facilitating comparative analysis across multiple datasets in an integrated manner. Using a combination of an outlier and replicate distance measures, dcHiC identifies biologically relevant differential compartments irrespective of their signs and assigns rankings with statistical confidence scores. Along with conventional pairwise differential calls, dcHiC allows a single multivariate differential comparison of Hi-C datasets, facilitating the users with a straightforward approach to analyze the data and avoid combinatorial problems involving finding a grouping.

We applied dcHiC to various biological scenarios, ranging from neuronal and hematopoietic stem cell differentiation in mice to primary tissues and normal cell lines in humans. Our results confirmed that dcHiC detects known compartmental changes among cell types, including those previously validated to play a role in neuronal and hematopoietic stem cell differentiation. Across and within compartmental changes, dcHiC accurately identified differential compartements with results that compellingly align with LaminB1, gene expression, and

methylation data. Furthermore, using both mice and human datasets, we also demonstrated that dcHiC systematically identifies critical biological marker genes and can recover cell-specific functions from differential compartment analysis alone. In mice datasets, dcHiC captured differential compartments overlapping with the genes vital for pluripotent and neuronal stem cell differentiation. These included pluripotency marker genes like *Dppa2/4* and *Zfp42* in mESC, and neuronal differentiation marker genes like *Ptn, Epbh1* in NPC, and *Dcx* in CN, all of which overlapped with differential compartments. Supported by multiple independent lines of biological evidence, these differential results suggested a lineage-specific compartmental gene regulatory mechanism associated with them. The hierarchical mouse hematopoietic stem cell differentiation model, consisting of ten different cell-types, also provided a unique opportunity to demonstrate the utility of dcHiC. A straightforward multivariate differential comparison of the hematopoietic system revealed previously known lineage-specific critical genes encompassing the differential compartments. Notably, we identified vital transcription factors like *Myc, Sox6, Runx2, Meis1*, and the highly crucial *Lhcgr* gene by solely analyzing the differential calls. Our functional GO enrichment analysis of gene sets overlapping with the differential compartment reconfirmed that compartments play a contributory role in determining the accessibility of genes and modulating their expression to specific cell-types. We observed a significant and relevant enrichment of biological processes of differential compartments like neuron differentiation in NPC and CN cell-types, adaptive immune response in naïve germinal B cells, and embryonic development in endometrial tissue. Taken together, dcHiC represents a significant leap in Hi-C compartment calling.

Chromatin compartment analysis is critical to genome organization. The framework we developed here provides a systematic way to identify differential compartments and visualize these differences in different scenarios, including time-series data and many-to-many comparisons. With over 522 publicly available Hi-C datasets in the 4D Nucleome Data Portal and hundreds of others in the Gene Expression Omnibus, dcHiC will play an essential role in enhancing our knowledge about high-level genome organization changes and their downstream effects on cellular function.

Methods

Compartment identification using Multiple Factor Analysis

To weigh each dataset equally in a group comparison, dcHiC uses Multiple Factor Analysis [29], an extension of PCA, for an unbiased assessment of multiple datasets. MFA can be thought of as a "weighted PCA" in which every data set is normalized for its biases. In the case of Hi-C data, the input to MFA is K distance normalized correlation tables $(X_1, X_2, X_3 \,..\, X_K)$. The first step of MFA is to perform a standard PCA on each data table and expressed via its SVD as:

$$X_K = U_K \cdot \Gamma_K \cdot {V_K}^T \quad with\ {U_K}^T \cdot U_K = {V_K}^T \cdot V_K = I \tag{1}$$

The matrices U_K and V_K store the left and right singular vectors of the table X_K. The singular values of the table X_K are stored in the diagonal matrix Γ_K. In MFA, the weight of a table is obtained from the first singular value of its PCA. The weight α_K is equal to the inverse of the first squared singular value γ_K.

$$\alpha_K = \frac{1}{\gamma_K^2} \tag{2}$$

In the next step, the K tables are normalized and merged to create a concatenated table $\overline{X}$, followed by its SVD:

$$\overline{X} = [(\sqrt{\alpha_1} \cdot X_1) \cdots (\sqrt{\alpha_K} \cdot X_K)] \tag{3}$$

$$\overline{X} = \overline{P} \cdot \Delta \cdot \overline{Q}^T \quad with\ \overline{P}^T \cdot \overline{P} = \overline{Q}^T \cdot \overline{Q} = I \tag{4}$$

The loading of the ith table is obtained by:

$$Q_i = \frac{1}{\sqrt{\alpha_i}} \cdot \overline{Q_i} \tag{5}$$

And the partial factor scores of a table is obtained by:

$$F_i = K \cdot \alpha_i \cdot X_i \cdot Q_i \tag{6}$$

The partial factor scores can be used to plot the observations as done in standard PCA, for which each column represents a dimension.

Outlier detection using Mahalanobis distance

The distance measure for features is critical for any classification method. Mahalanobis' distance (M) is a multivariate statistical measure of the extent to which the cases are marked as outliers, based on a Chi-square distribution [68]. The Mahalanobis distance of a point *x* and a distribution *D* is defined as:

$$D_M^2 = (x - \mu_i)^T \cdot \Sigma_i^{-1} \cdot (x - \mu_i) \quad (7)$$

Where Σ_i^{-1} is represents the inverse of the covariance matrix of variables, and μ is the center of the distribution. The drawback of Mahalanobis distance is the equal weighting of the variance normalized squared distances of features. To boost up the outlier detection, we use a feature weighted Mahalanobis distance [69] as:

$$D_M^{Weighted} = \{w \cdot D_M^2 \mid w[x] \geq 0, \sum_{x=1}^{N} w[x] = N\} \quad (8)$$

where

$$w[x] = \left\{ \frac{N \cdot d[x]}{\sum_{p=1}^{N} d[p]} \;\middle|\; x, p \in N \right\} \quad (9)$$

and

$$d[x] = \sum_{p=1}^{N} |D_M^2[x] - D_M^2[p]| \quad (10)$$

The approach provides more weight to the features that are distant from others than to features that are closer together. The significance of the corresponding weighted distance is calculated from the critical Chi-square distribution table as $\chi^2(D_M^{Weighted}, df)$.

Independent Hypothesis Weighting using pairwise replicate variation

Further, dcHiC applies Independent Hypothesis Weighting (IHW) approach to adjust the p-values using pairwise replicate variation measure as a covariate [31]. The pairwise replicate variation in partial factor scores or component scores is learned by dcHiC during the MFA step. dcHiC first uses linear regression to fit an expected line between component scores of replicates, followed by measuring the average shortest distance and its deviation of each point from the fitted line. This provides the background model parameters to quantify the sample-wise component score variation as Z-scores. Later, the Z-score is used as a covariate to adjust the p-values using the IHW approach.

Initial data processing, result generation, and visualization

dcHiC accepts multiple Hi-C data formats as input, including sparse matrix, hic, and cool formatted interaction files with a prefixed resolution defined by the user. Once set, dcHiC can perform chromosome wise MFA and outlier detection in serieal or parallel mode. The differential Hi-C bins are scanned against the gene coordinates of the respected genome defined by the user using 'bedtools map' function [70]. The unique overlapping set of genes are then extracted and are used for GO term enrichment analysis using the ToppGene suite API function [39]. dcHiC also generates a Javascript based dynamic IGV-HTML page to visualize the compartments and differential results, with an option to add additional one-dimensional

epigenetic tracks like RNA-seq or ChIP-seq data. For calling ChIP-seq peaks from samples (**Supplemental Table 1**), we first mapped the respective fastq files on the mm10 genome using the bowtie2 [71] mapper and generated the corresponding bam files (MAPQ > 20). The aligned files were then used as input to the MACS2 program [72] to call peaks (*p-value* < 1e-5) against their respective controls. The continuous ChIP-seq peaks were then merged and the unique set was mapped to the 100Kb differential compartments to calculate the average number of peaks displayed in the Figure 4D.

AVAILABILITY OF DATA AND MATERIALS

A Python/R implementation of dcHiC is freely available at github.com/ay-lab/dcHiC. This application is compatible with Hi-C data in HiC-Pro, *.hic*, and *.cool* formats. The data used in this study are available at the following GEO accession numbers: GSE96107, GSE152918, GSE142216, GSE152918, GSE142216, GSE87112, GSE63525, GSE43070, GSE84022, GSE105194, GSE105710, GSE105988. These are also available in **Supplemental Table 1**. All reported compartments for all cell lines, multivariate differential scores, RNA-Seq, and ChIP-Seq data used in this manuscript can be viewed interactively at: ay-lab.github.io/dcHiC. These standalone HTML files employ dcHiC's visualization utility.

FIGURE LEGENDS

Figure 1: **Outline of the method.** An overall schematic of the dcHiC approach, with Multiple Factor Analysis followed by multivariate statistical measures.

Figure 2: Comparison of compartment scores and Lamin B1 profile. (A, B) The comparison of dcHiC and HOMER compartments scores in mESC and NPC cell lines. **(C, D)** Lamin B1 correlation profiles with dcHiC compartment scores in mESC and NPC cell lines. **(E, F)** A visual comparison of dcHiC, HOMER compartment scores and Lamin B1 profiles in mESC (chromosome 16) and NPC (chromosome 9) for representative chromosomes.

Figure 3: The pairwise differential compartment analysis of mESC and NPC cell lines. (A) The total number of differential compartmental changes observed across different compartment classes. **(B)** The compartment score distribution, gene expression, and Lamin B1 profile changes between different compartment types. **(C, D)** Gene Ontology enrichment of biological processes in mESC and NPC enriched A compartments. **(E-H)** Examples of genes that lie in regions with significant compartment changes.

Figure 4: Multivariate differential compartment analysis of mESC, NPC, and CN cell lines. (A) Summary of overall compartment decomposition and significant compartment changes (in parentheses) observed across the three cell lines. **(B)** A parallel coordinate plot showing the change in the differential compartment scores during mESC to CN lineage differentiation**. (C)** The frequency of differential compartmental changes across different conditions. **(D)** Changes

in gene expression and histone modifications across significant differential compartments. **(E)** GO enrichment results for CN-enriched differential compartments. **(F, G)** Significant compartment changes in regions encompassing *Top1* and *Rmst* genes.

Figure 5: Multivariate differential compartment analysis of mouse hematopoiesis. (A) Summary of overall compartment decomposition and significant compartment changes (in parentheses) observed across 10 cell types. **(B)** Hierarchical classification of the hematopoietic stem cell differentiation model. Pie charts adjacent to the arrows represent the different classes of compartmental changes observed during differentiation. Gray boxes are showing the genes encompassing differential compartments previously known to play a critical role during differentiation. **(C, D)** Significant compartment changes in regions encompassing *Sox6* and *Abca13* genes. **(E)** GO enrichment of differential compartments with highest positive compartment scores in GMP and GR cells.

Figure 6: Multivariate differential compartment analysis of primary human tissues and cell lines. (A) Summary of overall compartment decomposition and significant compartment changes (in parentheses) observed across 11 tissue samples and cell lines. **(B)** GO enrichment of differential compartments with highest positive compartment scores in hippocampus. **(C)** A snapshot of the region encompassing a cluster of protocadherin family (PCDH) genes. **(D)** GO enrichment of differential compartments with highest positive compartment scores in Naïve Germinal B cells and GM12878 cell line. **(E)** A snapshot of the region encompassing CD84, an important immune cell gene that is a member of the signaling lymphocyte activation molecule (SLAM) family.

REFERENCES

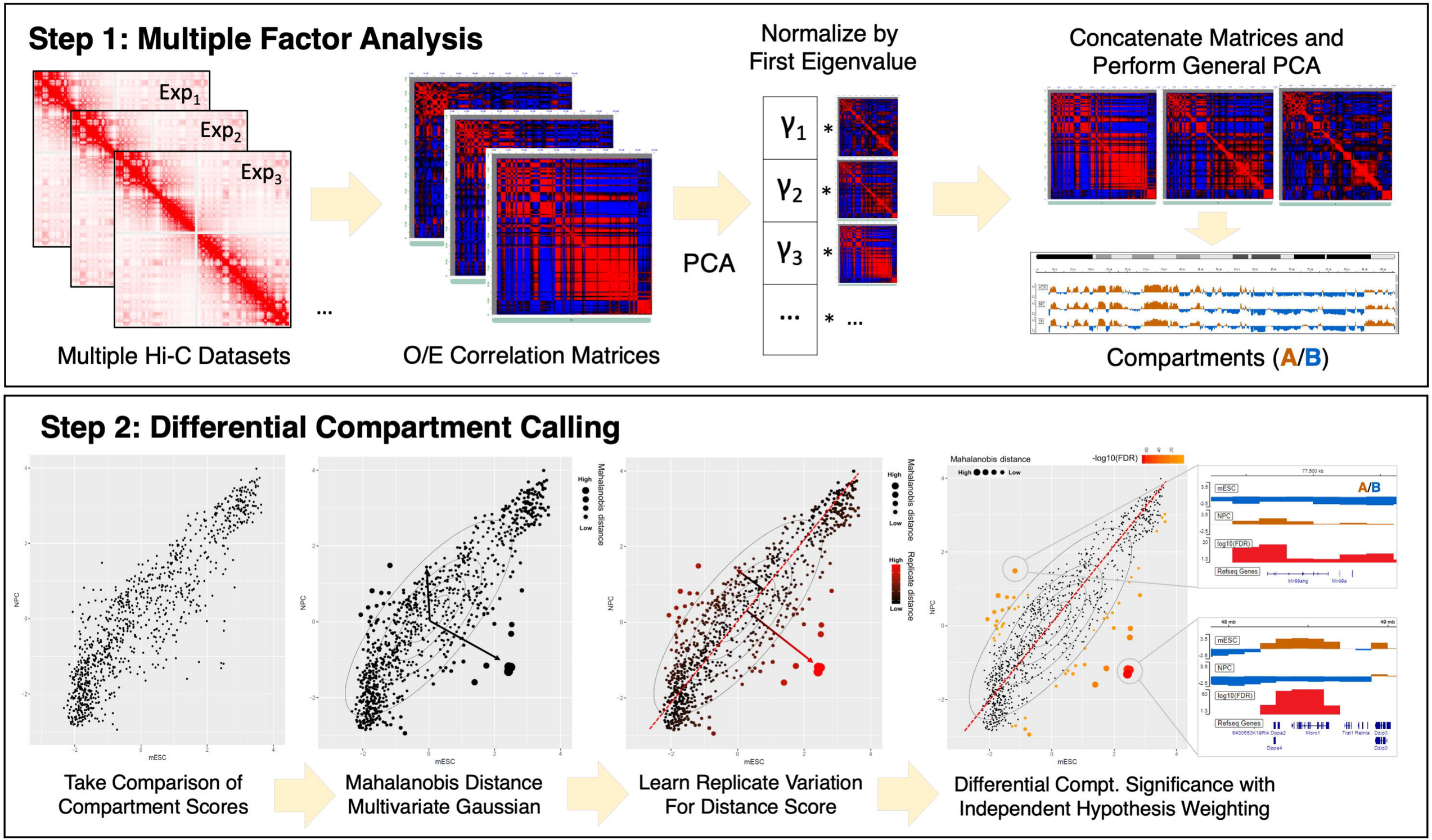
1. C R: **Über Zelltheilung.** *Morphologisches Jahrbuch Gegenbaur C (ed)* 1885, **10:**214–330.
2. Oudelaar AM, Higgs DR: **The relationship between genome structure and function.** *Nat Rev Genet* 2020.
3. Brackey CA, Marenduzzo D, Gilbert N: **Mechanistic modeling of chromatin folding to understand function.** *Nat Methods* 2020, **17:**767-775.
4. Finn EH, Misteli T: **Molecular basis and biological function of variability in spatial genome organization.** *Science* 2019, **365**.
5. Zheng H, Xie W: **The role of 3D genome organization in development and cell differentiation.** *Nat Rev Mol Cell Biol* 2019, **20:**535-550.
6. Schoenfelder S, Fraser P: **Long-range enhancer-promoter contacts in gene expression control.** *Nat Rev Genet* 2019, **20:**437-455.
7. Marchal C, Sima J, Gilbert DM: **Control of DNA replication timing in the 3D genome.** *Nat Rev Mol Cell Biol* 2019, **20:**721-737.
8. Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, et al: **Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication.** *Cell* 2019, **176:**816-830 e818.

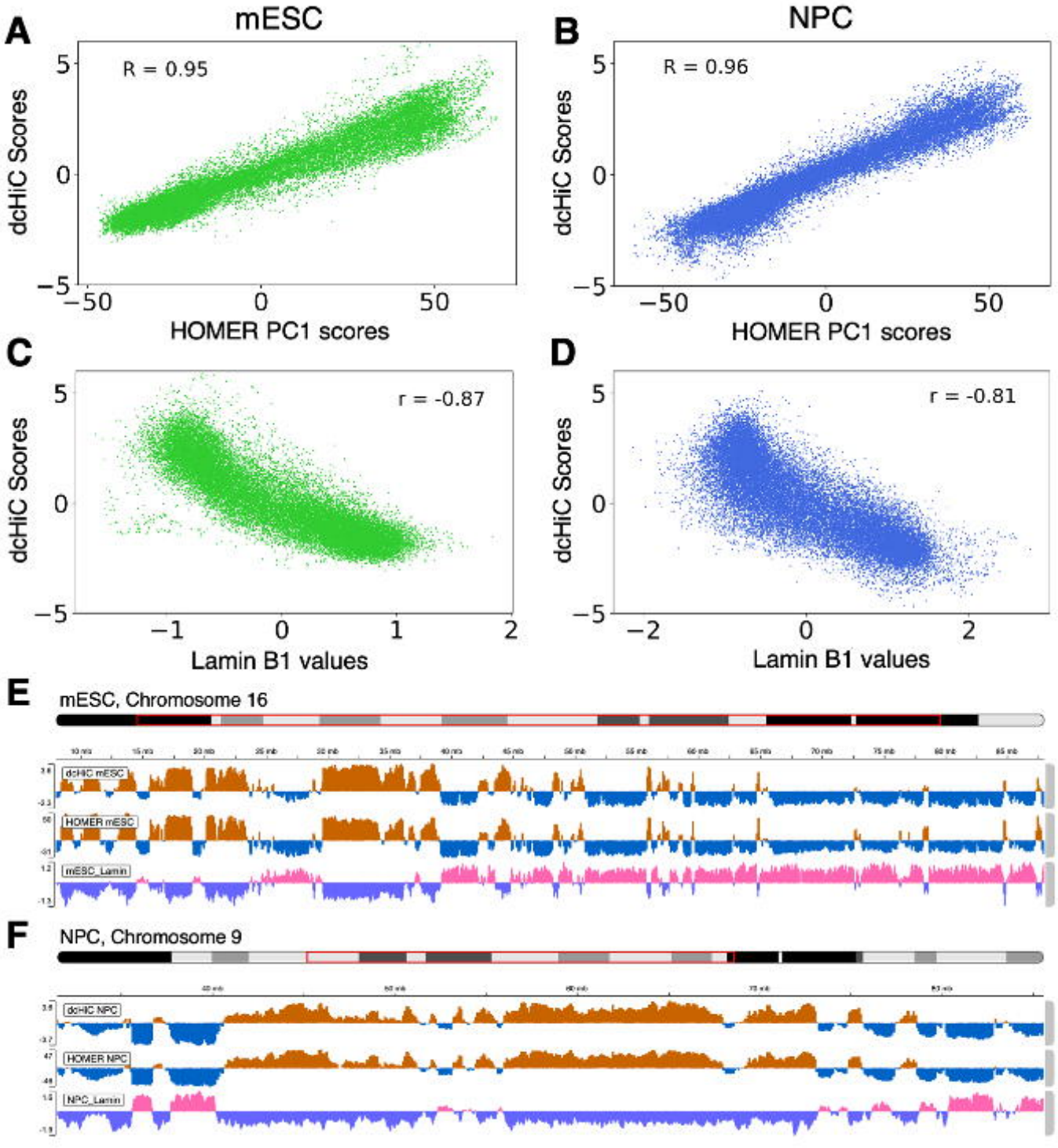
9. Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, Chen JH, Zou LS, Drier Y, Hecht V, et al: **Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer.** *Cell* 2020, **182:**1474-1489 e1423.
10. De S, Michor F: **DNA secondary structures and epigenetic determinants of cancer genome evolution.** *Nat Struct Mol Biol* 2011, **18:**950-955.
11. McLaughlin K, Flyamer IM, Thomson JP, Mjoseng HK, Shukla R, Williamson I, Grimes GR, Illingworth RS, Adams IR, Pennings S, et al: **DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency.** *Cell Rep* 2019, **29:**1974-1985 e1976.
12. Madakashira BP, Sadler KC: **DNA Methylation, Nuclear Organization, and Cancer.** *Front Genet* 2017, **8:**76.
13. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326:**289-293.
14. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell* 2014, **159:**1665-1680.
15. Xiong K, Ma J: **Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions.** *Nat Commun* 2019, **10:**5069.
16. Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, et al: **Transcription Elongation Can Affect Genome 3D Structure.** *Cell* 2018, **174:**1522-1536 e1522.
17. Zheng X, Zheng Y: **CscoreTool: fast Hi-C compartment analysis at high resolution.** *Bioinformatics* 2018, **34:**1568-1570.
18. Krietenstein N, Abraham S, Venev SV, Abdennur N, Gibcus J, Hsieh TS, Parsi KM, Yang L, Maehr R, Mirny LA, et al: **Ultrastructural Details of Mammalian Chromosome Architecture.** *Mol Cell* 2020, **78:**554-565 e557.
19. Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, Hesson J, Cavanaugh C, Ware CB, Krumm A, et al: **Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution.** *Methods* 2018, **142:**59-73.
20. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.** *Nat Biotechnol* 2011, **30:**90-98.
21. Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, Ren B: **Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq.** *Cell Res* 2016, **26:**1345-1348.
22. Li X, Luo OJ, Wang P, Zheng M, Wang D, Piecuch E, Zhu JJ, Tian SZ, Tang Z, Li G, Ruan Y: **Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions.** *Nat Protoc* 2017, **12:**899-915.
23. Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, et al: **Complex multi-enhancer contacts captured by genome architecture mapping.** *Nature* 2017, **543:**519-524.
24. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, et al: **Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus.** *Cell* 2018, **174:**744-757 e724.
25. Zheng M, Tian SZ, Capurso D, Kim M, Maurya R, Lee B, Piecuch E, Gong L, Zhu JJ, Li Z, et al: **Multiplex chromatin interactions with single-molecule precision.** *Nature* 2019, **566:**558-562.
26. Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, Mirny LA, O'Shea CC, Park PJ, Ren B, et al: **The 4D nucleome project.** *Nature* 2017, **549:**219-226.

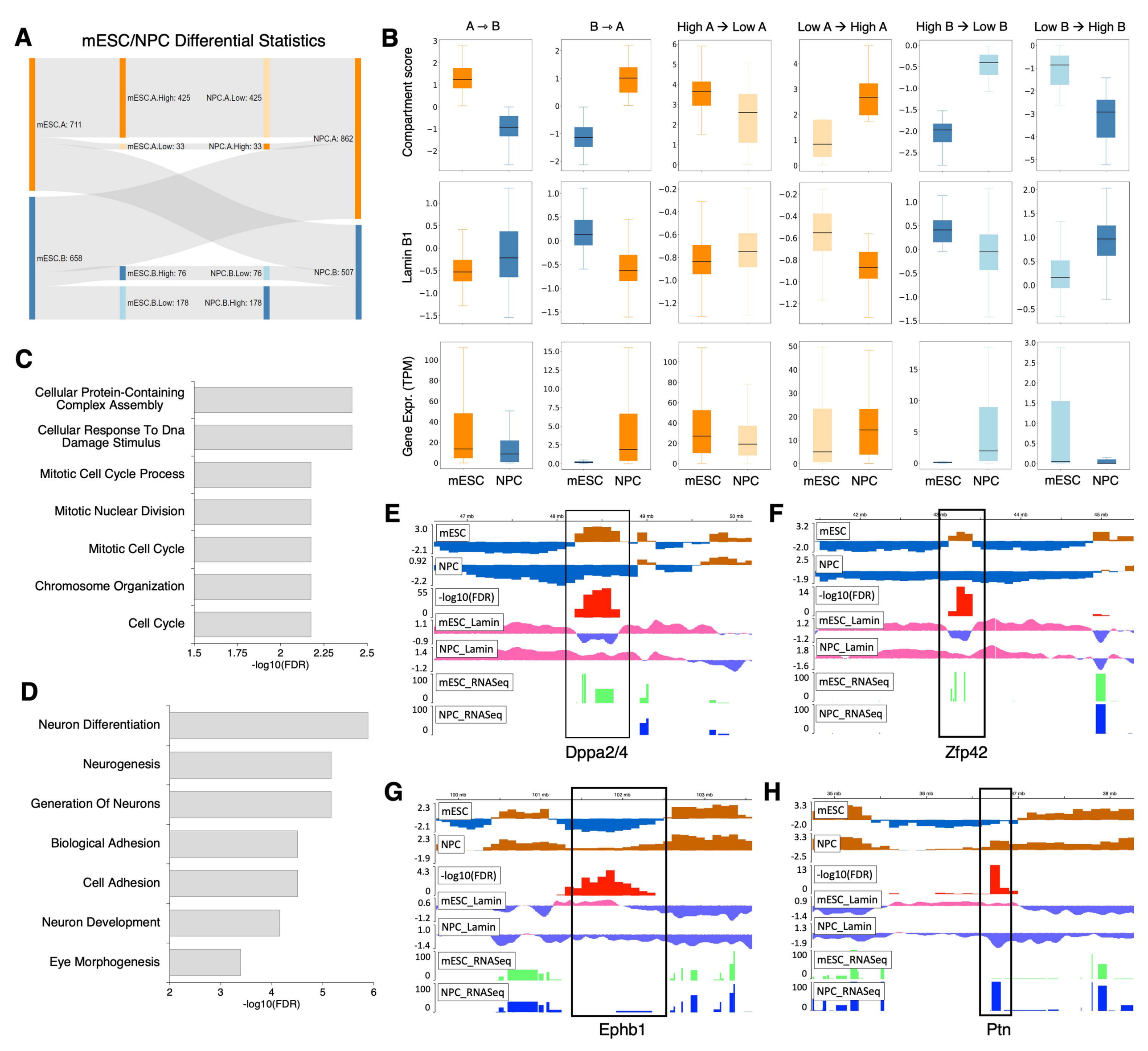
27. Miura H, Takahashi S, Poonperm R, Tanigawa A, Takebayashi SI, Hiratani I: **Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization.** *Nat Genet* 2019, **51:**1356-1368.
28. Zhang C, Xu Z, Yang S, Sun G, Jia L, Zheng Z, Gu Q, Tao W, Cheng T, Li C, Cheng H: **tagHi-C Reveals 3D Chromatin Architecture Dynamics during Mouse Hematopoiesis.** *Cell Rep* 2020, **32:**108206.
29. Abdi HW, L.J.; Valentin, D: **Multiple factor analysis: principal component analysis for multitable and multiblock data sets.** *Wiley Interdisciplinary Reviews: Computational Statistics* 2013, **5:**149-179.
30. Maesschalck RDJ-R, D.; Massart, D.L.: **The Mahalanobis distance.** *Chemometrics and Intelligent Laboratory Systems* 2000, **50:**1-18.
31. Ignatiadis N, Klaus B, Zaugg JB, Huber W: **Data-driven hypothesis weighting increases detection power in genome-scale multiple testing.** *Nat Methods* 2016, **13:**577-580.
32. Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC: **Dimension reduction techniques for the integrative analysis of multi-omics data.** *Brief Bioinform* 2016, **17:**628-641.
33. Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, Joffe B, Dekker J, Fudenberg G, Solovei I, Mirny LA: **Heterochromatin drives compartmentalization of inverted and conventional nuclei.** *Nature* 2019, **570:**395-399.
34. Kim Y, Zheng X, Zheng Y: **Role of lamins in 3D genome organization and global gene expression.** *Nucleus* 2019, **10:**33-41.
35. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M, et al: **Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.** *Mol Cell* 2010, **38:**603-613.
36. Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, Fussner E, Bazett-Jones DP, Plath K, Dalton S, et al: **Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis.** *Genome Res* 2010, **20:**155-169.
37. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al: **Chromatin architecture reorganization during stem cell differentiation.** *Nature* 2015, **518:**331-336.
38. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G: **Multiscale 3D Genome Rewiring during Mouse Neural Development.** *Cell* 2017, **171:**557-572 e524.
39. Chen J, Bardes EE, Aronow BJ, Jegga AG: **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.** *Nucleic Acids Res* 2009, **37:**W305-311.
40. Hernandez C, Wang Z, Ramazanov B, Tang Y, Mehta S, Dambrot C, Lee YW, Tessema K, Kumar I, Astudillo M, et al: **Dppa2/4 Facilitate Epigenetic Remodeling during Reprogramming to Pluripotency.** *Cell Stem Cell* 2018, **23:**396-411 e398.
41. Scotland KB, Chen S, Sylvester R, Gudas LJ: **Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation.** *Dev Dyn* 2009, **238:**1863-1877.
42. Qin EY, Cooper DD, Abbott KL, Lennon J, Nagaraja S, Mackay A, Jones C, Vogel H, Jackson PK, Monje M: **Neural Precursor-Derived Pleiotrophin Mediates Subventricular Zone Invasion by Glioma.** *Cell* 2017, **170:**845-859 e819.
43. Arvanitis DN, Behar A, Tryoen-Toth P, Bush JO, Jungas T, Vitale N, Davy A: **Ephrin B1 maintains apical adhesion of neural progenitors.** *Development* 2013, **140:**2082-2092.
44. Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, Ng HH: **Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells.** *Mol Cell Biol* 2005, **25:**6031-6046.
45. Bonev B, Cavalli G: **Organization and function of the 3D genome.** *Nat Rev Genet* 2016, **17:**661-678.

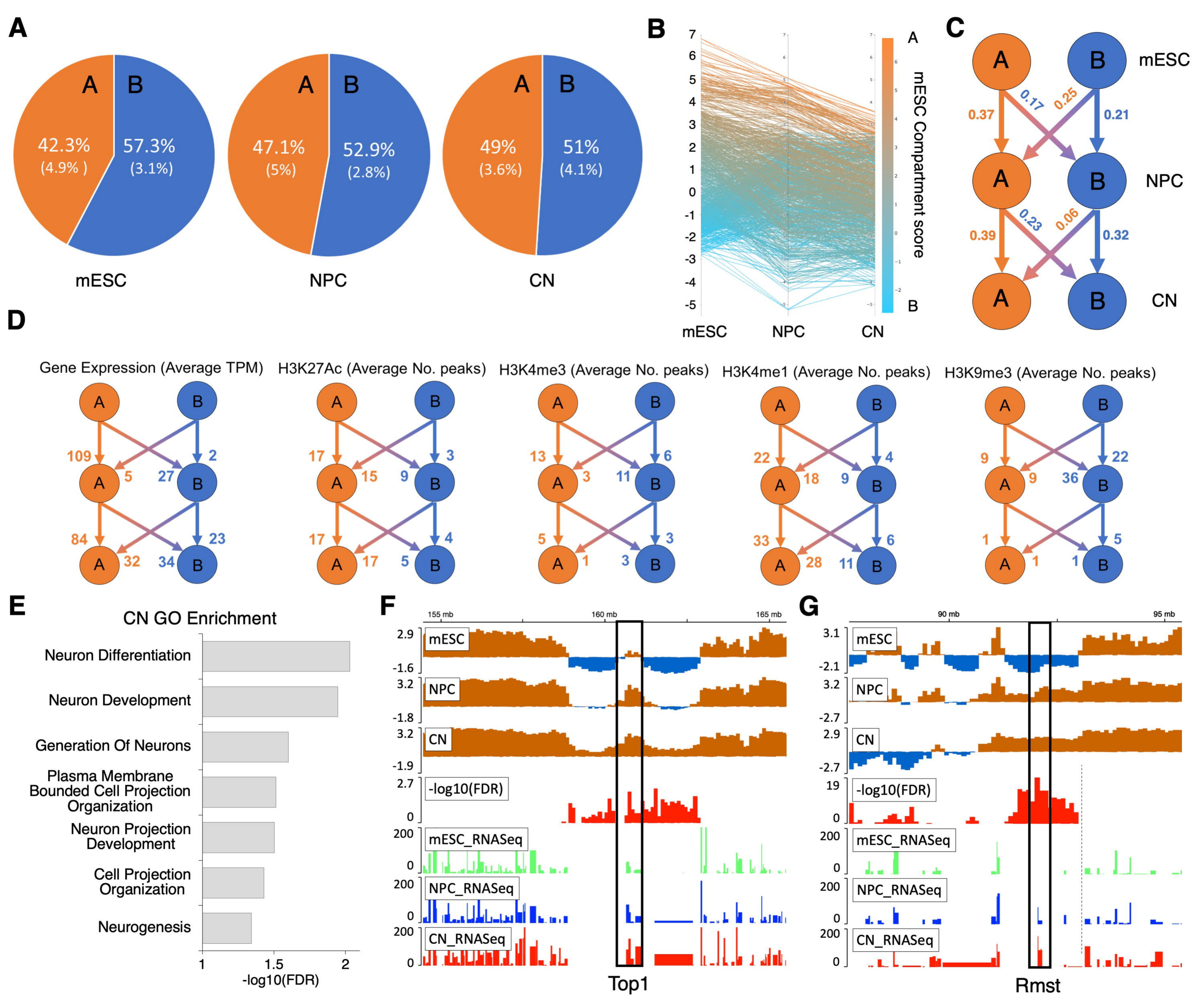
46. Morham SG, Kluckman KD, Voulomanos N, Smithies O: **Targeted disruption of the mouse topoisomerase I gene by camptothecin selection.** *Mol Cell Biol* 1996, **16:**6804-6809.
47. Mabb AM, Simon JM, King IF, Lee HM, An LK, Philpot BD, Zylka MJ: **Topoisomerase 1 Regulates Gene Expression in Neurons through Cleavage Complex-Dependent and -Independent Mechanisms.** *PLoS One* 2016, **11:**e0156439.
48. Plaschkes I, Silverman FW, Priel E: **DNA topoisomerase I in the mouse central nervous system: Age and sex dependence.** *J Comp Neurol* 2005, **493:**357-369.
49. Ng SY, Bogu GK, Soh BS, Stanton LW: **The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis.** *Mol Cell* 2013, **51:**349-359.
50. Gleeson JG, Lin PT, Flanagan LA, Walsh CA: **Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons.** *Neuron* 1999, **23:**257-271.
51. Hack I, Hellwig S, Junghans D, Brunne B, Bock HH, Zhao S, Frotscher M: **Divergent roles of ApoER2 and Vldlr in the migration of cortical neurons.** *Development* 2007, **134:**3883-3891.
52. Yoder MC: **Embryonic hematopoiesis in mice and humans.** *Acta Paediatr Suppl* 2002, **91:**5-8.
53. Doulatov S, Notta F, Laurenti E, Dick JE: **Hematopoiesis: a human perspective.** *Cell Stem Cell* 2012, **10:**120-136.
54. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, et al: **Immunogenetics. Chromatin state dynamics during blood formation.** *Science* 2014, **345:**943-949.
55. Unnisa Z, Clark JP, Roychoudhury J, Thomas E, Tessarollo L, Copeland NG, Jenkins NA, Grimes HL, Kumar AR: **Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress.** *Blood* 2012, **120:**4973-4981.
56. de Bruijn M, Dzierzak E: **Runx transcription factors in the development and function of the definitive hematopoietic system.** *Blood* 2017, **129:**2061-2069.
57. Dumitriu B, Patrick MR, Petschek JP, Cherukuri S, Klingmuller U, Fox PL, Lefebvre V: **Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development.** *Blood* 2006, **108:**1198-1207.
58. Wilson A, Murphy MJ, Oskarsson T, Kaloulis K, Bettess MD, Oser GM, Pasche AC, Knabenhans C, Macdonald HR, Trumpp A: **c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation.** *Genes Dev* 2004, **18:**2747-2763.
59. Jin K, Wang S, Zhang Y, Xia M, Mo Y, Li X, Li G, Zeng Z, Xiong W, He Y: **Long non-coding RNA PVT1 interacts with MYC and its downstream molecules to synergistically promote tumorigenesis.** *Cell Mol Life Sci* 2019, **76:**4275-4289.
60. Heo HR, Chen L, An B, Kim KS, Ji J, Hong SH: **Hormonal regulation of hematopoietic stem cells and their niche: a focus on estrogen.** *Int J Stem Cells* 2015, **8:**18-23.
61. Peng YJ, Yu H, Hao X, Dong W, Yin X, Lin M, Zheng J, Zhou BO: **Luteinizing hormone signaling restricts hematopoietic stem cell expansion during puberty.** *EMBO J* 2018, **37**.
62. Katori S, Hamada S, Noguchi Y, Fukuda E, Yamamoto T, Yamamoto H, Hasegawa S, Yagi T: **Protocadherin-alpha family is required for serotonergic projections to appropriately innervate target brain areas.** *J Neurosci* 2009, **29:**9137-9147.
63. Chen WV, Nwakeze CL, Denny CA, O'Keeffe S, Rieger MA, Mountoufaris G, Kirner A, Dougherty JD, Hen R, Wu Q, Maniatis T: **Pcdhalphac2 is required for axonal tiling and assembly of serotonergic circuitries in mice.** *Science* 2017, **356:**406-411.
64. Asai H, Ohkawa N, Saitoh Y, Ghandour K, Murayama E, Nishizono H, Matsuo M, Hirayama T, Kaneko R, Muramatsu SI, et al: **Pcdhbeta deficiency affects hippocampal CA1 ensemble activity and contextual fear discrimination.** *Mol Brain* 2020, **13:**7.

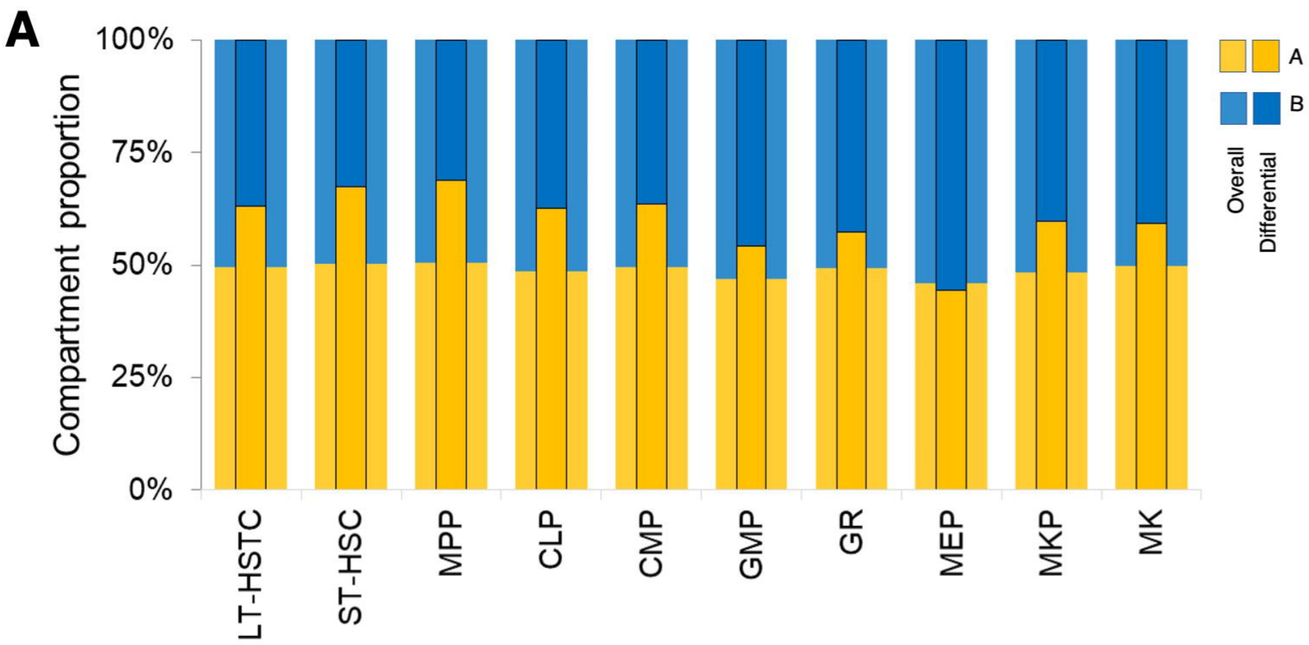
65. Tangye SG, van de Weerdt BC, Avery DT, Hodgkin PD: **CD84 is up-regulated on a major population of human memory B cells and recruits the SH2 domain containing proteins SAP and EAT-2.** *Eur J Immunol* 2002, **32:**1640-1649.
66. Martin M, Romero X, de la Fuente MA, Tovar V, Zapater N, Esplugues E, Pizcueta P, Bosch J, Engel P: **CD84 functions as a homophilic adhesion molecule and enhances IFN-gamma secretion: adhesion is mediated by Ig-like domain 1.** *J Immunol* 2001, **167:**3668-3676.
67. Wong EB, Soni C, Chan AY, Domeier PP, Shwetank, Abraham T, Limaye N, Khan TN, Elias MJ, Chodisetti SB, et al: **B cell-intrinsic CD84 and Ly108 maintain germinal center B cell tolerance.** *J Immunol* 2015, **194:**4130-4143.
68. Brereton RG: **The Mahalanobis distance and its relationship to principal component scores.** *Journal of Chemometrics* 2015.
69. Wölfel ME, H.K.: **Feature weighted mahalanobis distance: Improved robustness for Gaussian classifiers.** *IEEE* 2005, **13th European Signal Processing Conference**.
70. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.
71. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-359.
72. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9:**R137.

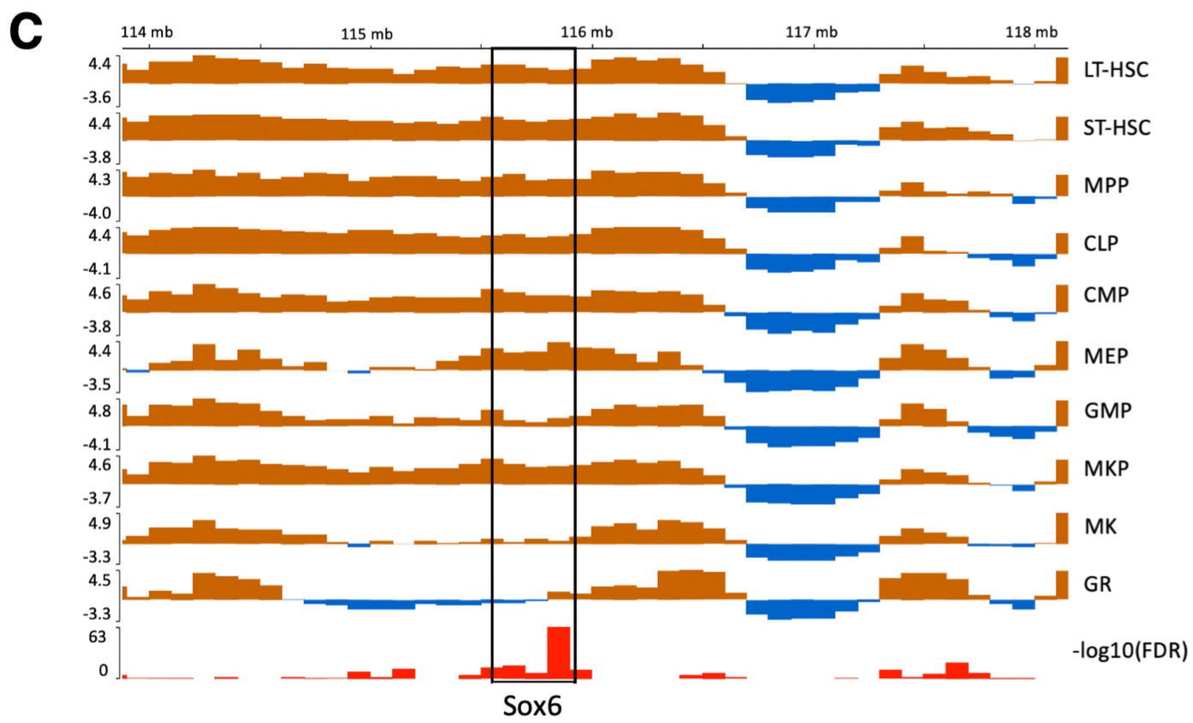

Step 1: Multiple Factor Analysis
Exp1
Exp2
Exp3
...
Multiple Hi-C Datasets
O/E Correlation Matrices
PCA
Normalize by First Eigenvalue
γ1 *
γ2 *
γ3 *
... * ...
Concatenate Matrices and Perform General PCA
Compartments (A/B)
Step 2: Differential Compartment Calling
NPC
mESC
Mahalanobis distance
High
Low
Replicate distance
-log10(FDR)
A/B
mESC
NPC
log10(FDR)
Refseq Genes
Take Comparison of Compartment Scores
Mahalanobis Distance Multivariate Gaussian
Learn Replicate Variation For Distance Score
Differential Compt. Significance with Independent Hypothesis Weighting

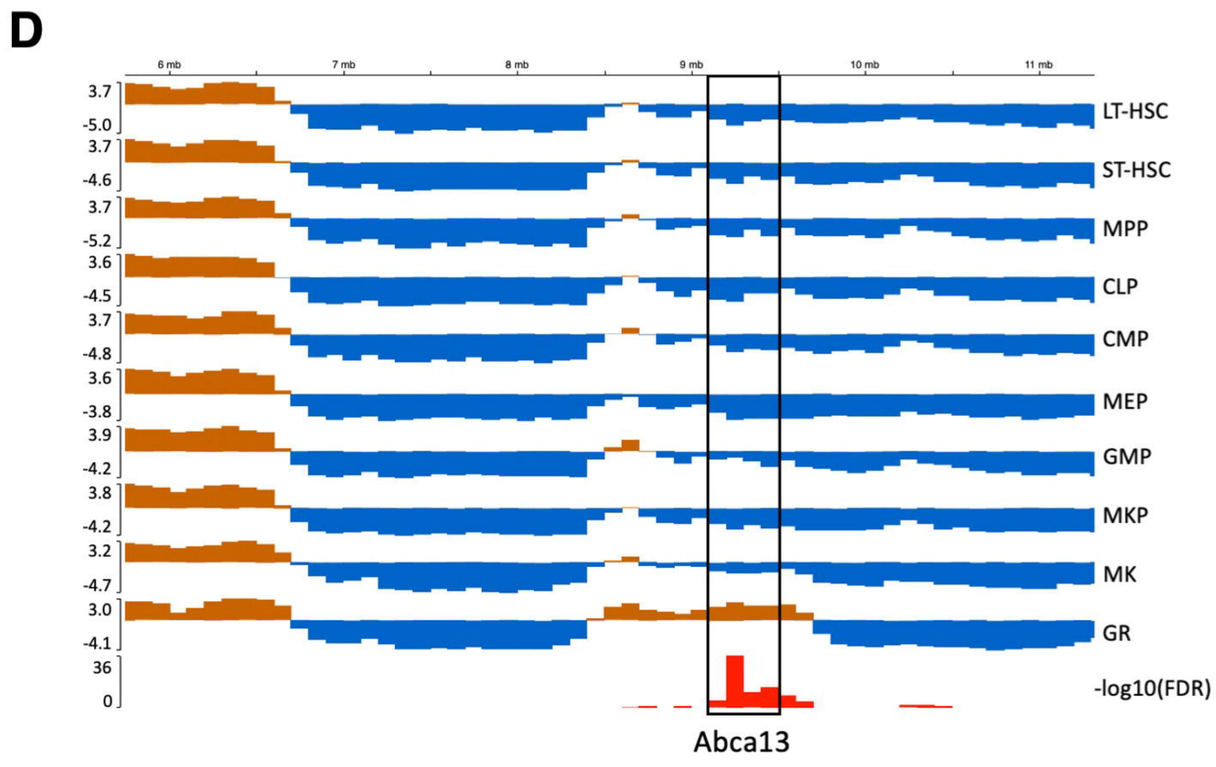

A
mESC
R = 0.95
dcHiC Scores
HOMER PC1 scores
B
NPC
R = 0.96
dcHiC Scores
HOMER PC1 scores
C
r = -0.87
dcHiC Scores
Lamin B1 values
D
r = -0.81
dcHiC Scores
Lamin B1 values
E
mESC, Chromosome 16
dcHiC mESC
HOMER mESC
mESC_Lamin
F
NPC, Chromosome 9
dcHiC NPC
HOMER NPC
NPC_Lamin

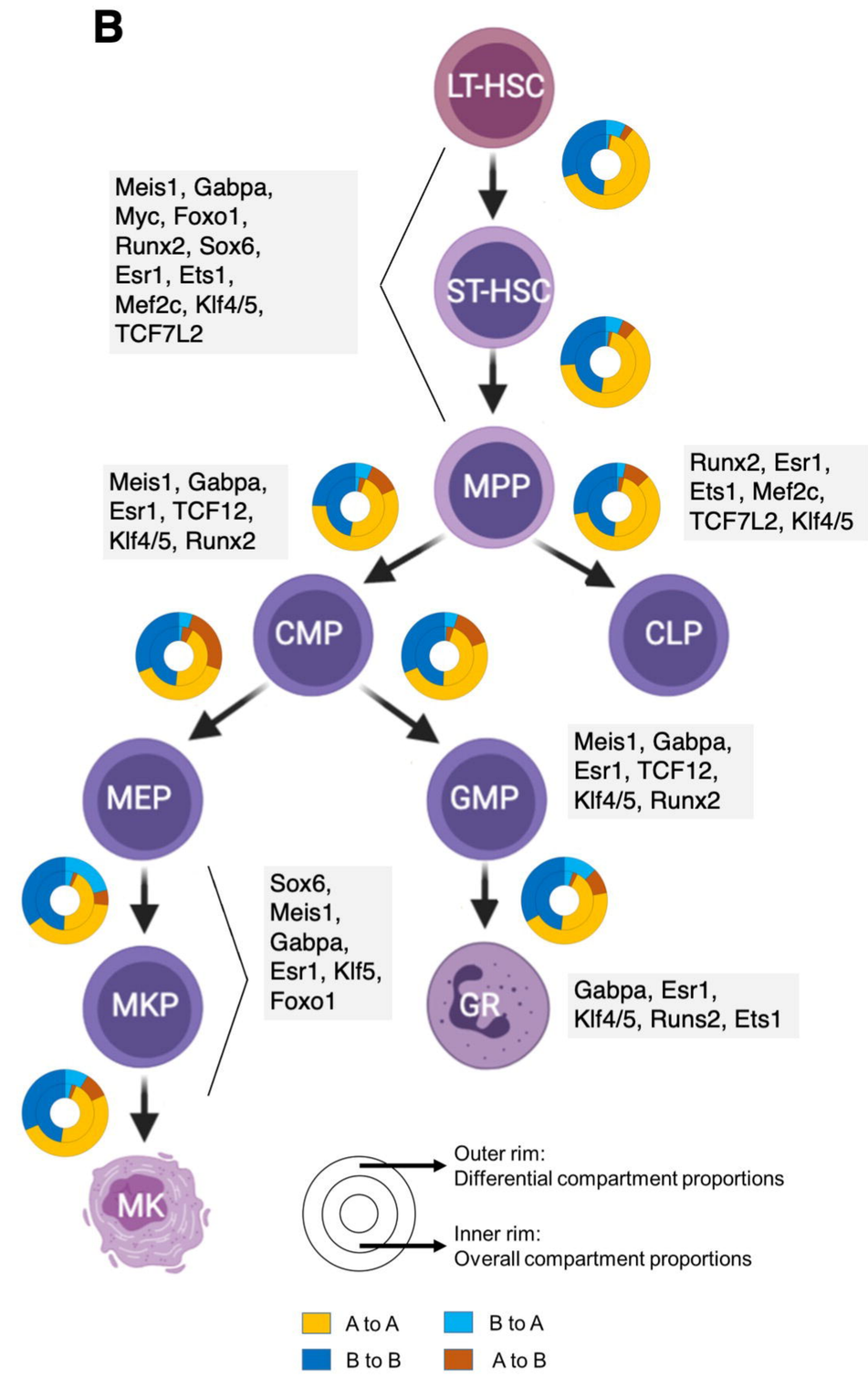

A
mESC/NPC Differential Statistics
mESC.A: 711
mESC.A.High: 425
NPC.A.Low: 425
mESC.A.Low: 33
NPC.A.High: 33
NPC.A: 862
mESC.B: 658
mESC.B.High: 76
NPC.B.Low: 76
NPC.B: 507
mESC.B.Low: 178
NPC.B.High: 178
B
A → B
B → A
High A → Low A
Low A → High A
High B → Low B
Low B → High B
Compartment score
Lamin B1
Gene Expr. (TPM)
mESC
NPC
C
Cellular Protein-Containing Complex Assembly
Cellular Response To Dna Damage Stimulus
Mitotic Cell Cycle Process
Mitotic Nuclear Division
Mitotic Cell Cycle
Chromosome Organization
Cell Cycle
-log10(FDR)
D
Neuron Differentiation
Neurogenesis
Generation Of Neurons
Biological Adhesion
Cell Adhesion
Neuron Development
Eye Morphogenesis
-log10(FDR)
E
mESC
NPC
-log10(FDR)
mESC_Lamin
NPC_Lamin
mESC_RNASeq
NPC_RNASeq
Dppa2/4
F
Zfp42
G
Ephb1
H
Ptn

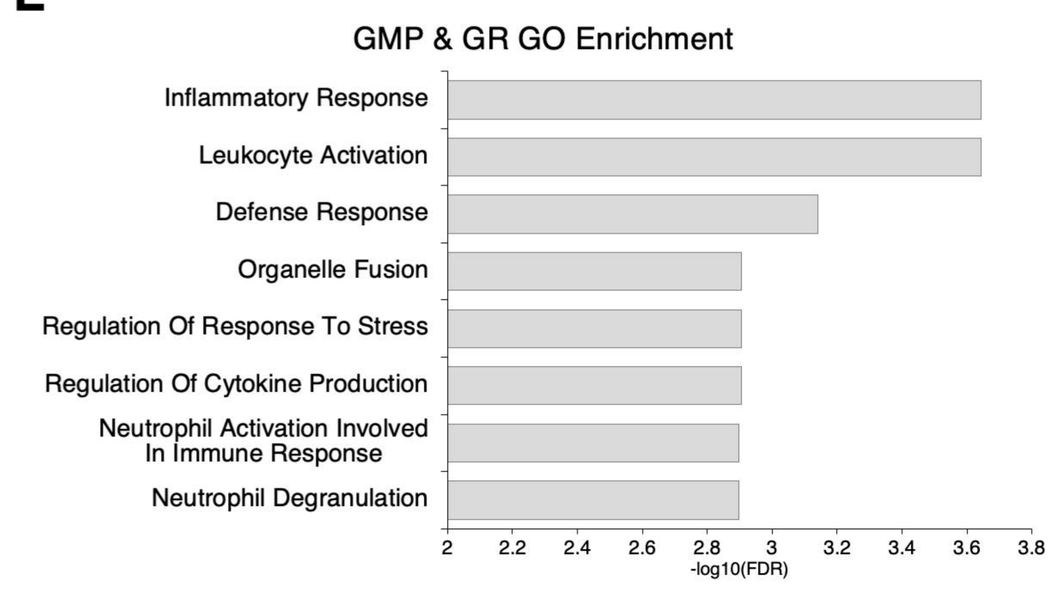

A
A
B
42.3%
(4.9%)
57.3%
(3.1%)
mESC
A
B
47.1%
(5%)
52.9%
(2.8%)
NPC
A
B
49%
(3.6%)
51%
(4.1%)
CN
B
mESC Compartment score
A
B
mESC
NPC
CN
C
A
B
mESC
0.37
0.17
0.25
0.21
A
B
NPC
0.39
0.23
0.06
0.32
A
B
CN
D
Gene Expression (Average TPM)
H3K27Ac (Average No. peaks)
H3K4me3 (Average No. peaks)
H3K4me1 (Average No. peaks)
H3K9me3 (Average No. peaks)
109
5
27
2
84
32
34
23
17
15
9
3
17
17
5
4
13
3
11
6
5
1
3
3
22
18
9
4
33
28
11
6
9
9
36
22
1
1
1
5
E
CN GO Enrichment
Neuron Differentiation
Neuron Development
Generation Of Neurons
Plasma Membrane Bounded Cell Projection Organization
Neuron Projection Development
Cell Projection Organization
Neurogenesis
1
1.5
2
-log10(FDR)
F
155 mb
160 mb
165 mb
mESC
NPC
CN
-log10(FDR)
mESC_RNASeq
NPC_RNASeq
CN_RNASeq
Top1
G
90 mb
95 mb
mESC
NPC
CN
-log10(FDR)
mESC_RNASeq
NPC_RNASeq
CN_RNASeq
Rmst

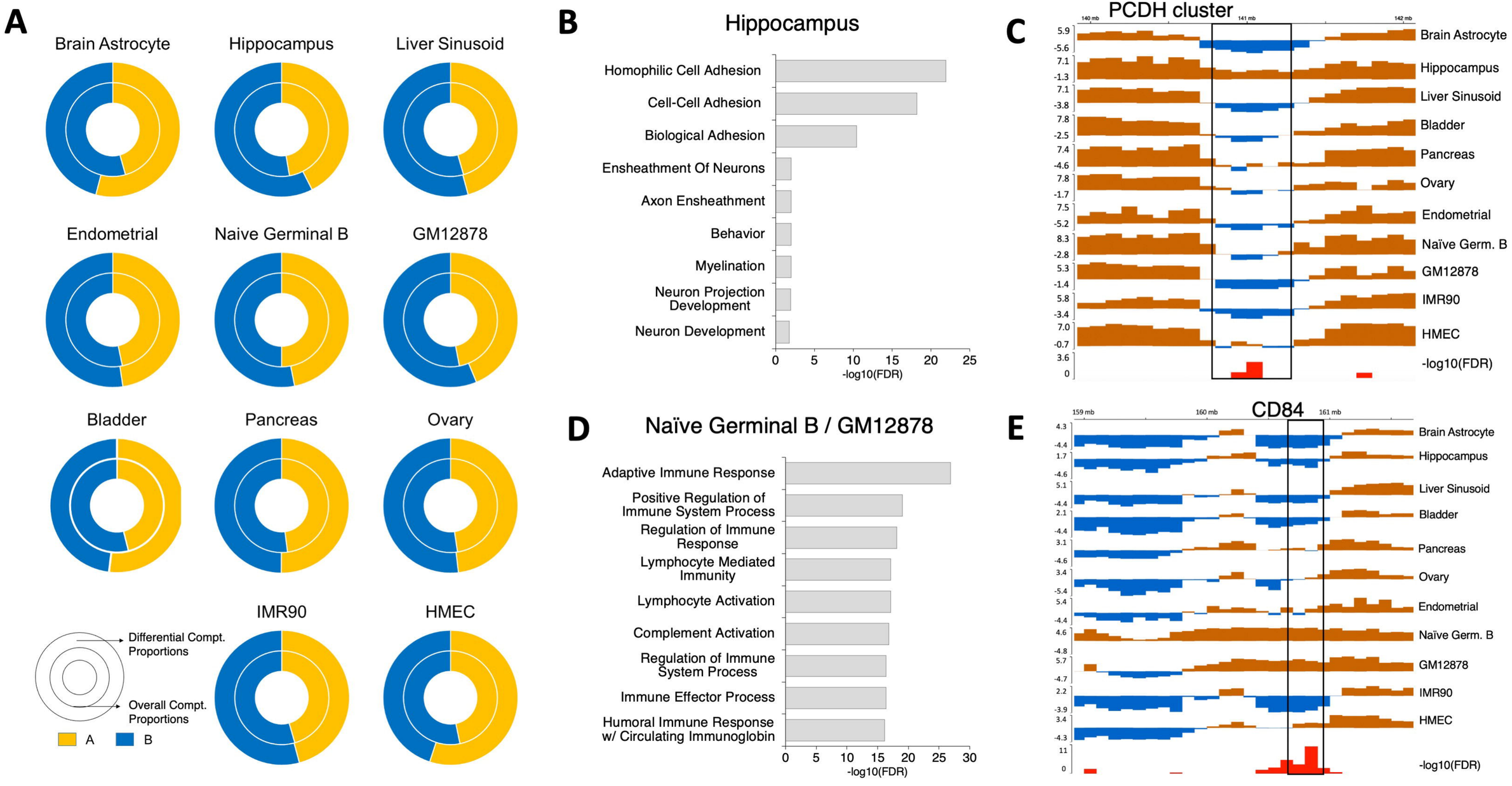

A
100%
75%
50%
25%
0%
Compartment proportion
LT-HSTC
ST-HSC
MPP
CLP
CMP
GMP
GR
MEP
MKP
MK
A
B
Overall
Differential


C
114 mb
115 mb
116 mb
117 mb
118 mb
LT-HSC
ST-HSC
MPP
CLP
CMP
MEP
GMP
MKP
MK
GR
-log10(FDR)
Sox6


D
6 mb
7 mb
8 mb
9 mb
10 mb
11 mb
LT-HSC
ST-HSC
MPP
CLP
CMP
MEP
GMP
MKP
MK
GR
-log10(FDR)
Abca13


B
LT-HSC
ST-HSC
MPP
CMP
CLP
MEP
GMP
MKP
GR
MK
Meis1, Gabpa, Myc, Foxo1, Runx2, Sox6, Esr1, Ets1, Mef2c, Klf4/5, TCF7L2
Meis1, Gabpa, Esr1, TCF12, Klf4/5, Runx2
Runx2, Esr1, Ets1, Mef2c, TCF7L2, Klf4/5
Meis1, Gabpa, Esr1, TCF12, Klf4/5, Runx2
Sox6, Meis1, Gabpa, Esr1, Klf5, Foxo1
Gabpa, Esr1, Klf4/5, Runs2, Ets1
Outer rim:
Differential compartment proportions
Inner rim:
Overall compartment proportions
A to A
B to A
B to B
A to B


E
GMP & GR GO Enrichment
Inflammatory Response
Leukocyte Activation
Defense Response
Organelle Fusion
Regulation Of Response To Stress
Regulation Of Cytokine Production
Neutrophil Activation Involved In Immune Response
Neutrophil Degranulation
2
2.2
2.4
2.6
2.8
3
3.2
3.4
3.6
3.8
-log10(FDR)


A
Brain Astrocyte
Hippocampus
Liver Sinusoid
Endometrial
Naive Germinal B
GM12878
Bladder
Pancreas
Ovary
IMR90
HMEC
Differential Compt. Proportions
Overall Compt. Proportions
A
B
B
Hippocampus
Homophilic Cell Adhesion
Cell-Cell Adhesion
Biological Adhesion
Ensheathment Of Neurons
Axon Ensheathment
Behavior
Myelination
Neuron Projection Development
Neuron Development
0 5 10 15 20 25
-log10(FDR)
C
PCDH cluster
140 mb
141 mb
142 mb
Brain Astrocyte
Hippocampus
Liver Sinusoid
Bladder
Pancreas
Ovary
Endometrial
Naïve Germ. B
GM12878
IMR90
HMEC
-log10(FDR)
D
Naïve Germinal B / GM12878
Adaptive Immune Response
Positive Regulation of Immune System Process
Regulation of Immune Response
Lymphocyte Mediated Immunity
Lymphocyte Activation
Complement Activation
Regulation of Immune System Process
Immune Effector Process
Humoral Immune Response w/ Circulating Immunoglobin
0 5 10 15 20 25 30
-log10(FDR)
E
CD84
159 mb
160 mb
161 mb
Brain Astrocyte
Hippocampus
Liver Sinusoid
Bladder
Pancreas
Ovary
Endometrial
Naïve Germ. B
GM12878
IMR90
HMEC
-log10(FDR)