**DATA GENERATION COMPONENT**

**Title: Spatiotemporal Regulation of 3D Genome Organization and Function in Human Cells**

**PROJECT SUMMARY**

*En route* to the establishment of the refined genome mapping–3D modeling NPS platform based on the ChIA-PET and RICh-PET methods, we will produce substantial, large volumes of high-quality 3D genome mapping data. In parallel with the progress of our technology development efforts, we will take a phased approach to producing high-quality 3D genome mapping data from established cell culture lines representing major development/differentiation lineages and from primary human cells with a focus on the human immune system. We will also generate corresponding transcriptomic and epigenomic datasets from the same cell types. Such data, along with the integration of existing ENCODE and GWAS datasets, will provide thorough functional annotation support to the 3D models predicted from the ChIA-PET data. To start, we will use established cell lines representing major human progenitor and lineage cell types that can be tested in large numbers (108 cells), both to optimize 3D mapping data production parameters and establish data quality standards including reproducibility, sensitivity, and specificity **(Aim 1)**. These data will enable the construction of pilot 3D maps representing major human development and cellular differentiation lineages (embryonic stem, endoderm, mesoderm, and ectoderm), as well as comparative analyses to elucidate the structural changes that occur during major cellular differentiation events **(Aim 1)**. With the improvements to our mapping technology in place, i.e., requiring fewer numbers of cells (107-6 cells per experiment), we will expand our data generation efforts to study primary human blood–derived immune cells, focusing on key immune cell lineages **(Aim 2).** The 3D genome datasets generated in this Aim will illuminate features underpinning the nucleome structural changes that occur during immune cell differentiation. Finally, to further demonstrate the utility of our technology platform for addressing important biological questions, we will investigate changes in immune cell 3D genome structure linked with three major well-characterized immune events of immune responses **(Aim 3).** Completion of these three aims will generate a broad, comprehensive 3D genome mapping data covering cellular differentiation across large, medium, and fine time-scales. These datasets will be valuable for studying topological changes of 3D genome organization in space and in time with well-defined cellular phenotypes and functions, laying the foundation for future of 4D Nucleome studies. Furthermore, the availability of these datasets will empower the research community with new resources with which to investigate the complexity of higher-order genome organization, to ultimately accelerate our understanding of genome organization and function in health and disease.

**SPECIFIC AIMS**

This **Data Generation** component is organized aroundtwo major objectives in line with the overarching goals of the proposal: 1) to establish a reproducible, robust technology platform that capitalizes on cell sources that can be obtained in large cell numbers and 2) to generate pilot 3D genome maps from human cells, including never-before-tested primary cell types, and over various time-scales of development and activation. The current ChIA-PET and RICh-PET protocols are ready to produce high-quality chromatin interaction data from human cell lines in which large quantities of cells (108 per experiment) can be obtained, including those that represent major cellular development/differentiation lineages (e.g., embryonic, mesoderm, endoderm, ectoderm progeny cell types). Thus, we will begin generating pilot 3D maps in major human cell lines, both to optimize data production parameters and establish data quality standards. With the planned improvements to our mapping technologies in place, which will allow us to generate ChIA-PET and RICh-PET data from fewer numbers of cells (107-6 cells per experiment), we will extend our data generation efforts to human primary cells from *in vivo* systems. Cells of the blood hematopoietic lineage are ideal for studying the relationship between genome topological structure and transcriptional regulation during cell differentiation, and will also enable us to test if our 3D technologies are capable of detecting topological changes during induction from the naïve to the activated state. Completion of our three Specific Aims will yield a broad range of 3D genome mapping data covering cellular differentiation states over large, medium, and fine time-scales. These datasets will be valuable for studying topological changes of 3D genome organization in space and time, in cell types with well-defined cellular phenotypes and functions, thus laying the foundation for future explorations of the 4D Nucleome. Furthermore, the availability of these datasets will have immediate impact on the large research community and will advance our understanding of genome organization and its functions in healthy and diseased systems.

**Aim 1: Generate data for constructing pilot 3D maps of major human cell lineages.** First, we will use established cell lines that can supply large numbers of cells (108) and thus allow us to further optimize 3D mapping data-production parameters and to establish data quality standards. Such optimized protocols will be tested by using a cohort of cell lines (H1-hESC, GM12878, EndoC-ßH1, IMR90 and Hi-neurons) representative of human embryonic stem cells and major human cell lineages (mesoderm, endoderm, and ectoderm) during cellular development. We chose these cell lines as the representatives of cell lineages as there are already large ENCODE data available for these lines and of our biological interests. These data will enable us to build pilot 3D maps in major human cell lineages. Finally, we will perform comparative analyses of these five pilot 3D maps to elucidate the structural changes occurring during major cellular differentiation events.

**Aim 2:** **Generate data for constructing 3D maps of major human blood cell lineages.** With our improved mapping technology platform in place, that will be capable of generating high-quality data from fewer numbers of cells (107-6 cells per experiment), we will expand our data generation efforts to study primary human cells. We chose blood-derived immune cells because of the ease with which they can be collected compared to other types of human primary cells and because of their importance in human health. We will specifically focus on the following immune cell lineages: CD34+ hematopoietic stem and progenitor cells (HSPCs); lymphoid lineage cell types, including B, T and natural killer (NK) cells; and myeloid lineage–derived neutrophils and monocytes. Some subtypes of blood cells, including naïve and memory B (CD19+) cells, naïve and memory CD4+ T cells, and naïve and memory CD8+ T cells, can also be purified in the range of millions (1-7x106), which will be sufficient for ChIA-PET and RICh-PET analysis using the improved protocols. The 3D genome datasets generated in this aim will provide a detailed view of nucleome structural changes throughout the immune cell differentiation paradigm.

**Aim 3:** **Perform 3D genome mapping of immune cell responses to activation signals.** To further demonstrate the utility of our technology platform for addressing important biological questions, we will investigate changes in immune cell 3D genome structure linked with three major well-characterized immune events: **1)** The cytokine-driven switch of monocyte phenotype and function to antigen-presenting dendritic cells (DCs) or to antigen-degrading macrophages; **2)** The maturation of DCs that control immunity (mature DCs) and tolerance (immature DCs); and **3)** The activation and differentiation of resting naïve CD4+ T cells into functionally diverse progeny cells with distinct transcription factor expression and cytokine production patterns including: T helper 1 (Th1) cells, Th2 cells, and T regulatory (Treg) cells. *In vitro* induction of these cell differentiations are well defined, highly controlled, and robust. Therefore, these experiments provide an ideal system for generating high-quality 3D genome mapping data for 4D nucleome analysis at fine time-scale.

**RESEARCH STRATEGY**

**SIGNIFICANCE**

Advances in genomic and epigenomic techniques have enabled identification of large numbers of regulatory elements including enhancers and non-coding RNAs (ncRNAs) that may function over long genomic distances. A growing body of evidence suggests that spatial genome conformation, i.e., the chromatin folds and loops that bring distal linear regulatory elements into close 3D spatial proximity, is important for remote transcriptional regulation. Meanwhile, recent GWAS studies have identified a large number of SNPs conferring genetic risk to common diseases ([1](#_ENREF_1)), the vast majority of which are located within non-coding regions and overlap transcriptional enhancers in physiologically relevant cell types ([2-6](#_ENREF_2)). These studies underscore the importance of regulatory element (dys)function in disease pathophysiology. Identifying the target genes controlled by such regulatory elements remains an important challenge and an area of active investigation.

ChIA-PET is one of a few methods that can identify long-range chromatin interactions between specific regulatory elements (e.g., enhancers and promoters) with high genome-wide precision and resolution ([7](#_ENREF_7), [8](#_ENREF_8)). For this reason, ChIA-PET was implemented by ENCODE ([6](#_ENREF_6)) to generate chromatin connectivity maps from human cells. The ChIA-PET datasets available *via* ENCODE ([11](#_ENREF_11)) have been extensively explored by us ([9](#_ENREF_9), [10](#_ENREF_10)) and many others to advance their studies ([12](#_ENREF_12), [13](#_ENREF_13)) (see letters from Drs. Gallagher, Leonard, and O’Shea). We are thus confident that the large volumes of high quality ChIA-PET data generated herein will continue to spearhead studies of genome dysregulation and disease by researchers across a broad range of fields.

Recent advances in RNA biology strongly point to a role for non-coding (nc) RNAs, particularly long ncRNAs (lncRNAs), in higher-order genome regulation ([14](#_ENREF_14)). RICh-PET (see **Overall Fig. 6**) is an all-to-all method designed to detect all lncRNAs and their chromatin interaction target sites. Although process optimizations are ongoing, both ChIA-PET and RICh-PET are ready to produce high-quality data from large numbers of cells for pilot 3D genome maps that reflect chromatin interactions mediated by protein factors and lncRNAs. Many well-established cell lines and primary blood cell types, such as B-cell and T-cell populations, provide sufficient material for high-quality ChIA-PET and RICh-PET data production. Although the main focus of the 4DN project is on technology development, we expect that the availability of such datasets and pilot 3D maps will prompt the research community to develop tools for data analysis/mining, and applications to biological questions of interest, thus collectively, accelerating our understanding of fundamental biology.

We have selected primary human blood–derived immune cells as the cell class to test the refinements to our technology platform and generate novel data. The human immune system critically protects us from pathogens and, when dysfunctional, leads to infectious or autoimmune diseases. The hematopoietic system employs sophisticated regulatory mechanisms to finely balance tolerance and immunity. Recent advances in genomics and epigenomics allow us to map putative regulatory elements mediating transcriptional responses ([15](#_ENREF_15), [16](#_ENREF_16)), and recent studies suggest that long-range interactions between such regulatory elements modulate transcriptional activities during immune responses ([9](#_ENREF_9)). However, many aspects of this regulation are entirely unclear. Therefore, detailed architectural maps of human immune cells will provide a foundation to better understand cell differentiation during the immune response to pathogens or self.

**INNOVATION**

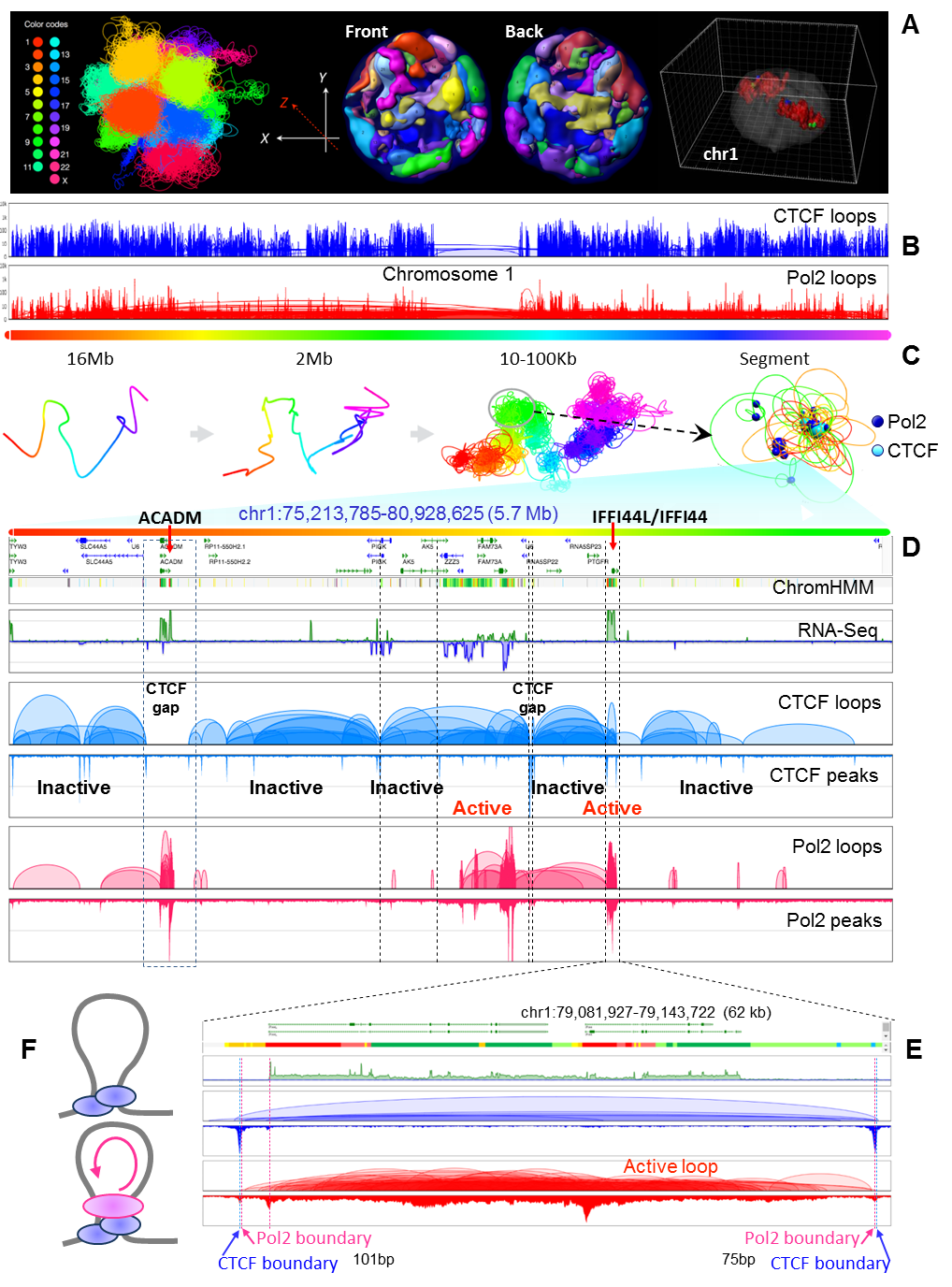
The uniqueness of this data generation strategy lies in the **innovative tools and approaches** we propose to generate **novel and highly valuable 3D genome datasets** from human cell lines and primary cells. ChIA-PET stands apart from all other mapping technologies owing to the number of distinct datasets a single experiment can generate, namely: i) protein binding sites, ii) enriched specific chromatin interactions, iii) non-enriched topological chromatin contacts and iv) haplotype-specific interactions. This multiplicity of data makes ChIA-PET the most inclusive technology for producing high-content 3D genome mapping data, which will be invaluable for the construction of structural 3D models and functional annotation of genome organization. RICh-PET is a novel method for studying the regulatory roles of lncRNAs through the unbiased, genome-wide detection of lncRNAs and chromatin interactions. The RICh-PET data generated from cell lines and primary blood cells will facilitate a richer understanding of lncRNA function in the 3D nucleome space, thereby adding a unique dimension of insight into 3D genome regulation. Finally, our studies will yield the first 3D maps of primary human immune cells and, importantly, novel insight into how 3D genome structures impact transcription regulation during the differentiation of hematopoietic stem cells (HSPCs) to fully differentiated lymphocytes. These longitudinal studies lay the foundation for future investigations of the 4D nucleome.

**APPROACH**

***Strategy.*** We will use a phased approach to produce reliable 3D mapping data from cell lines and primary blood immune cells during their differentiation into hematopoietic lineage cell types. Known differences in the phenotype and function of these cells could also help validate the technology developed in this 4DN program.

Our strategy is to use CTCF and RNAPol2 ChIA-PET to generate comprehensive 3D genome mapping data, including datasets for protein factor binding sites, specific chromatin interactions, topological contacts and haplotype-specific interactions (see **Overall Fig. 3)**. CTCF was selected owing to its known role in nuclear architecture ([17](#_ENREF_17)). A recent study using super-deep sequencing of HiC libraries revealed that most interaction anchors are associated with CTCF binding sites ([18](#_ENREF_18)). RNAPol2 is involved in most interactions involving transcription. Thus, targeting these two factors will enable us to capture most of the structural and functional interactions associated with transcriptional regulation, and to address the relationship between topological structure and transcriptional function. We will also use ChIA-PET to explore other protein factors relevant to specific cell types and biological questions. Candidates include the cohesion protein Med21 and Nanog in embryonic stem (ES) cells; NFKB and STAT in immune cells; and some repressive factors such as EZH2. We will also generate RICh-PET datasets, to understand what regulatory roles lncRNAs play in the 3D nucleome, as well as supportive genomic datasets, i.e., RNA-seq, ChIP-seq of histone marks and ATAC-seq for open chromatin sites, to help annotate the 3D genome. To aid in haplotype analysis, we will conduct whole genome sequencing (WGS) or SNP arrays when no genome sequencing data is available. We will use such data to identify heterozygous SNPs in each sample, and use ChIA-PET interaction data to experimentally phase thegenomes and identify allele-specific chromatin interactions.

**Figure 1. 3D genome models derived from ChIA-PET data. (A)** 3D model of a representative GM12878 cell nucleome (left), front and back views of 3D-FISH images of chromosome territories (middle), and two copies of chr1 territory (right). **(B)** CTCF (blue) and Pol2 (red) ChIA-PET mapping tracks of chr1. **(C)** 3D interaction models of chr1 at various resolutions (bin size 16Mb, 2Mb, and 10-100Kb), and the 3D model of an enlarged segment (5.7Mb) with CTCF and Pol2 binding (far right). **(D)** Zoomed-in view of the 5.7Mb segment showing detailed profiling tracks of chromatin states (ChromHMM, color bars indicate regions of various transcription properties: yellow-enhancer, red-promoter, green-gene body, blue-insulator, grey-heterochromatin), gene expression (strand specific RNA-Seq), and ChIA-PET data for CTCF (blue)- and Pol2 (red)-mediated loops and binding peaks. Based on ChromHMM, RNA-seq and Pol2 ChIA-PET, transcriptionally active and inactive domains are demarcated (between dotted lines). Two examples of actively expressed genes are indicated (red arrows). ACAMD is in a CTCF-gap region, and IFFI44L/IFFI44 in a CTCF-loop region. **(E)** Zoomed-in view of the IFFI44L/IFFI44 locus (62Kb), showing the outmost Pol2 loop boundaries are in very close proximity (101 and 75bp) to the CTCF boundaries. **(F)** Suggested topological model for CTCF that defines loop space for Pol2 to operate.

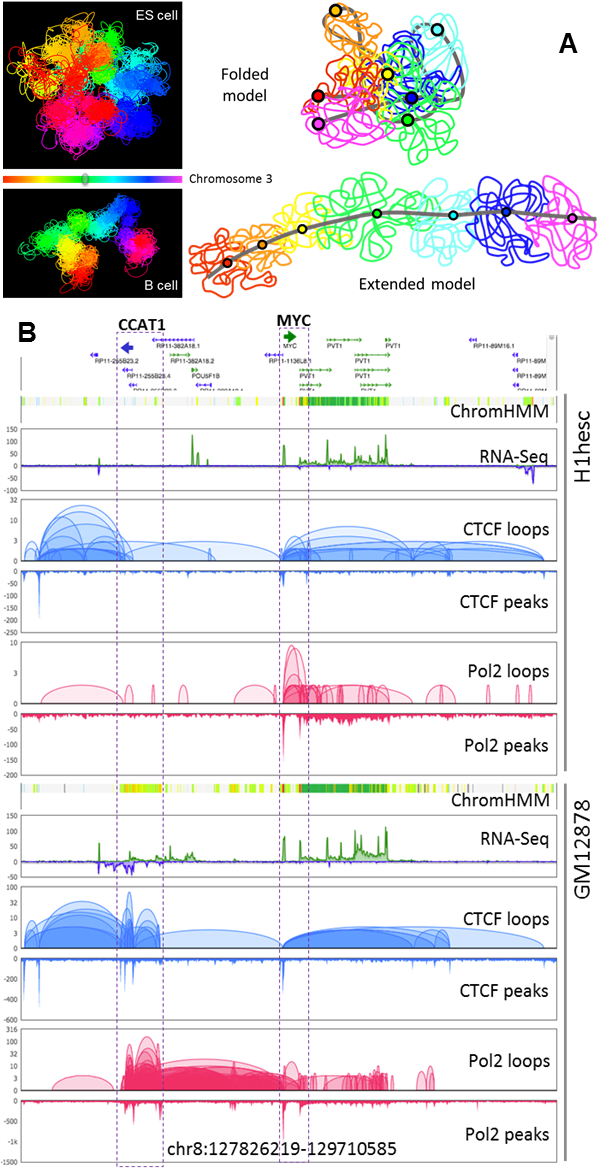


***Preliminary*** ***results.*** We have generated high-quality CTCF and Pol2 ChIA-PET data for a number of euploid cell lines such as ES, GM12878, and HUVEC, using our improved ChIA-PET method (see **Overall Figs. 1-4** for some preliminary results). In general, CTCF strongly interacts with binding sites involved in looping structures across the entire genome, but with punctuated gaps.

This CTCF-defined “loop-gap-loop” profile correlates well with the topologically associating domains (TADs) (see **Overall Fig. 2)**. Using the combined CTCF and Pol2 ChIA-PET data from GM12878 cells and our 3D genome modeling platform (3D-NOME, see **Data Analysis and Modeling)** we were able to construct some prototype 3D models **(Fig. 1A, left).** To evaluate this predicted 3D genome model, we generated a number of 3D-FISH images from GM12878 nuclei **(Fig. 1A, middle).** We are in the process of comparing the predicted models using data derived from millions of cells with the 3D nuclear images obtained from individual cells. Using ChIA-PET data we were also able to generate a 3D model for individual chromosomes **(Fig. 1B-C),** which can be compared with the nuclear image obtained by 3D-FISH **(Fig. 1A, right).** We are developing computational programs to capture the 3D coordinates and use them to further refine the prediction model and the methods (see **Data Analysis** **and Modeling** for details). Using different bin sizes of interaction data, 3D models can be generated at different resolutions **(Fig. 1C).** For instance, detailed topological structures and associated protein binding profiles can be revealed at particular segments. Such analysis demonstrated that CTCF binds extensively along the genome and forms the CTCF-defined “loop-gap-loop” landscape, which is punctuated by Pol2-associated transcriptional chromatin interaction loops. When correlated with epigenomic chromatin states (ChromHMM) and transcriptional output (RNA-Seq), CTCF-defined genome structure can be categorized into active and inactive transcriptional domains **(Fig. 1D).** Most Pol2 loops are circumscribed within the CTCF-defined loops, and some are tightly bound within the CTCF-defined space boundaries within 100bp margins (e.g., the *IFFI44L* locusin **Fig. 1E).** These data suggest a working topological model whereby CTCF binding defines the chromatin structure and RNAPol2 enters the space for function **(Fig. 1F).** The data also reveal exceptions, such as at the *ACADM* locus **(Fig. 1D).**

In our preliminary comparative analysis of topological structures in different cell types, we uncovered high-level structural variations in chromosomal conformation and, at a more detailed level, of specific promoter-enhancer interactions. For instance, in human ES cells, chr3 adopts a round shape whereas this chromosome in B cells has an extended shape **(Fig. 2A).** We are now validating this observation by 3D-FISH in ES and B cells.

Figure 2. Differential 3D chromatin interactions. (A) 3D models of chr3 in ES and B cells (left) and draws of folded and extended chromosomal conformations. (B) Chromatin properties at the *MYC* locus (1.9Mb) in ES and B cells. Track information order is the same as in Fig 1. Differential chromatin loops are highlighted in dotted boxes.



At the local interaction level, CTCF loops are mostly conserved between different cell types whereas RNAPol2-associated interactions tend to be cell-type-specific. At the *CCAT1*-*MYC* loci, for instance, CTCF binding and looping profiles are similar in H1hESC and GM12878 cells. However, the *CCAT1* locus displays B-cell-specific CTCF interactions absent in ES cells, indicating that cell-type-specific CTCF loops can be identified. In contrast, RNAPol2-associated chromatin interactions differ substantially between the two cell lines. Although *MYC* is expressed at the same level in ES and B cells, the *MYC* promoter appears to interact mostly with its downstream domains in ES cells, but with upstream enhancers, and forming a complex with the *CCAT1* promoter, in B cells. This regulatory cluster might facilitate co-transcriptional regulation of these two genes. In support of this idea, *MYC* and *CCAT1* loci are actively transcribed in B cells, whereas only *MYC* is expressed in ES cells. Correspondingly, the mouse *MYC* promoter is also linked to lineage-specific enhancers in B and ES cells ([9](#_ENREF_9)).

**Aim 1:** **Generate data for constructing pilot 3D maps of major human cell lineages.**

Our objectiveis two-pronged: 1) to establish a reproducible and robust technology platform, capitalizing on cellular sources that can be obtained in large cell numbers; 2) to generate pilot 3D genome maps from human embryonic cells as well as progeny cell types of mesoderm, endoderm, ectoderm, and endoderm lineage. The methodologies established here will be applied to work discussed in the following aims.

***Cell culture lines***: We will focus on five well-characterized euploid human cell lines representing the cellular origin of human tissues and some major lineages. These are: 1) H1 ES, 2) GM12878 (B lymphocyte, mesoderm lineage), 3) EndoC-ßH1 (pancreatic beta cells, endoderm lineage), 4) IMR90 (fetal lung fibroblasts, endoderm lineage), and 5) H1-neurons (ectoderm lineage) cells. Sustainable supplies will facilitate development of our technology platform for 3DN and 4DN analysis. We will also consider fresh human primary cells (e.g., hepatocytes) and major types of blood cells (e.g., B- and T-cell populations).

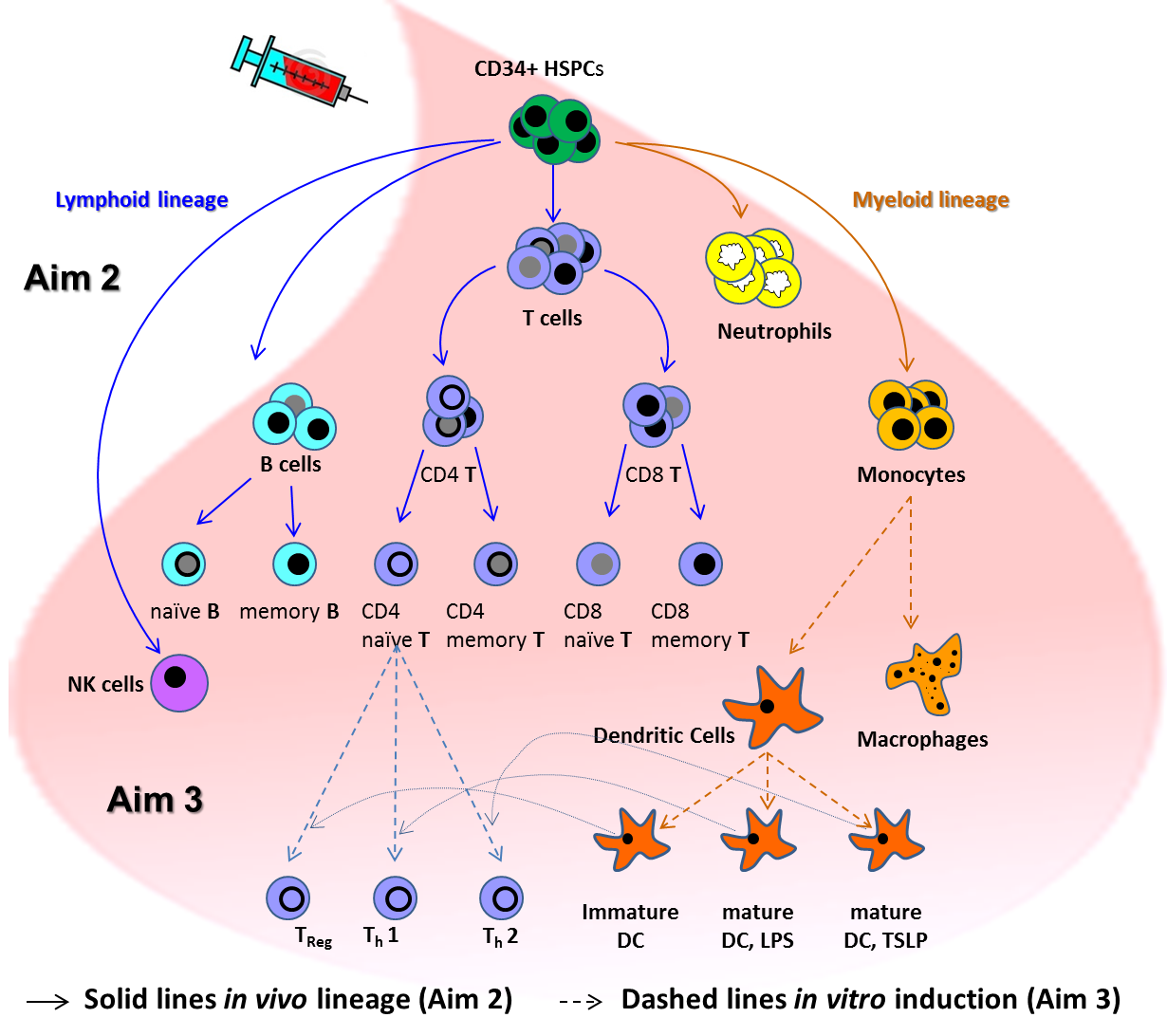
***Experimental design***: Our initial focus will be on GM12878 and H1 ES cells to establish major technical specificities and data analysis parameters that we will then extend to the other cell lines. Using established protocols, we will use 108 cells for two biological replicates of CTCF and Pol2 ChIA-PET library construction and sequencing. Once a library is constructed, we routinely use MiSeq to generate a test run of sequencing data (about 10-20 million reads). If this test run data confirms high library quality (high rate of enrichment of ChIP, high percentage of interligation, high library complexity) we will generate one lane of HiSeq sequencing run, usually to generate 108 reads. We will then combine the MiSeq and HiSeq sequencing data into one single library. Reads that pass quality control (QC) will be processed for downstream data analysis to call protein factor binding sites and to identify specific and non-specific interactions. Another 108 cells will be used to generate two RICh-PET data replicates to identify ncRNA-chromatin interactions using a library sequencing method similar to that used to generate the ChIA-PET library. In addition, ~50 million cells will be used to generate replicate RNA-seq, ATAC-seq and histone ChIP-seq data sets for analysis. Furthermore, 1-5ug genomic DNA will be used for WGS-seq analysis or SNP arrays to identify heterozygous SNPs for haplotype-specific chromatin interaction analysis. These comprehensive datasets for each cell type will provide sufficient data for generating the pilot 3D maps of human genomes. As in our preliminary results **(Figs. 1,2),** we will use 3D-NOME (see **Data Analysis and Modeling**) to process the data and build the pilot 3D maps of these representative cell types. We will also start comparative 4D nucleome analysis among the different pilot 3D maps to identify lineage-specific features of 3D genome structures and transcriptional functions.

***Expected results and potential pitfalls***: We have obtained triplicate RNAPol2 and CTCF ChIA-PET datasets from GM12878 cells, duplicate datasets for H1 ES cells and related transcriptomic and epigenomic mapping data. Our preliminary analyses show the ChIA-PET data are of high quality (**Overall Fig. 4-5)** for 3D genome mapping and modeling **(Fig. 1-2).** Although the optimized protocols for these two cell lines may not be immediately applicable to others due to cellular variations, we do not anticipate any substantial technical issues when fine-tuning the experimental conditions for the other three cell lines. Once all data are generated, we will conduct thorough analyses for each dataset and comparative analyses between different datasets, which will help us to assess technical properties such as reproducibility, detection sensibility and specificity. Moreover, we will attempt to build pilot 3D maps for these five cell types and compare the 3D models to investigate potential structural changes over major cell differentiation lineages **(Fig. 2).** These efforts will also help us to assess potential challenges and complexities we may encounter in large-scale 4D nucleome studies.

**Aim 2:** **Generate data for constructing 3D maps of major human blood cell lineages**

With the planned improvements to our mapping technology that will enable ChIA-PET sequencing with fewer numbers of cells (1-10 million cells per experiment), we will extend our data generation plan to study human primary blood-derived cells. The objective is to establish 3D maps as a function of differentiation of the major immune cell lineages from HSPCs (Fig 3). Major cell subtypes that can be isolated in large numbers (108) are lymphoid lineage B and T cells, as well as myeloid lineage neutrophils and monocytes. Some subtypes, including natural killer (NK) cells, naïve and memory B (CD19+) cells, naïve and memory CD4+ T cells, and naïve and memory CD8+ T cells, can also be purified in the tens of millions (1-7x107) ([19](#_ENREF_19)), which will be sufficient for ChIA-PET and RICh-PET analysis using the improved protocols.

Fig 3. Human major immune cell differentiation lineages



***Experimental design***: We will first analyze blood cells that can be isolated in large numbers (107-108 cells) from a single blood donor. We will obtain HSPCs (108 cells per donor) isolated from mobilized blood through our collaboration with Dr. Patrick Gallagher (Yale Medical School, see letter) and a commercial source (Allcells, Alameda, CA). For other blood cell types, we can easily obtain large numbers (107-108 cells) of neutrophils, monocytes, B cells and CD4+ and CD8+ T-cell subsets. The cells will be purified to 95% purity from blood buffy coats (obtained from NYC Blood Bank or from blood apheresis), as we described ([19](#_ENREF_19)). Neutrophils will be enriched by Percoll centrifugation and processed immediately. Fractionated mononuclear cells will be further subdivided by bead-based purification and serial positive selection and negative depletion as described ([20](#_ENREF_20)). To account for individual variations for each type of blood cells, we will collect samples from three healthy donors for analysis.

Each blood sample will be used for ChIA-PET and RICh-PET analysis. As in Aim 1, we will generate additional datasets (transcriptome and epigenome datasets) from the same primary cell samples, enabling us to correlate 3D genome structures with functional output. For haplotype and allele specificity analysis, we will also generate WGS sequencing or SNP array data to define heterozygous SNPs from each blood donor. Thus, with the SNP and ChIA-PET data, we will be able to identify individual specific haplotype information and obtain 3D mapping information for all 46 chromosomes. The 3D mapping data and matched transcriptome and epigenome data generated will undergo 3D modeling by 3D-NOME and functional annotation analysis, as well as 4DN studies as described in Aim 1.

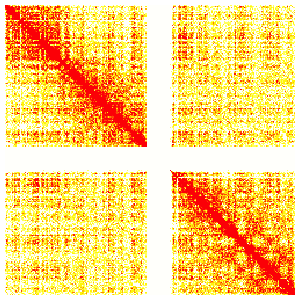
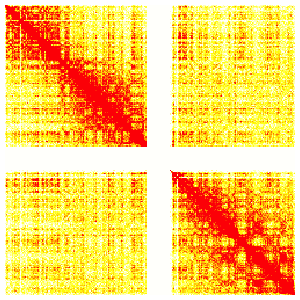
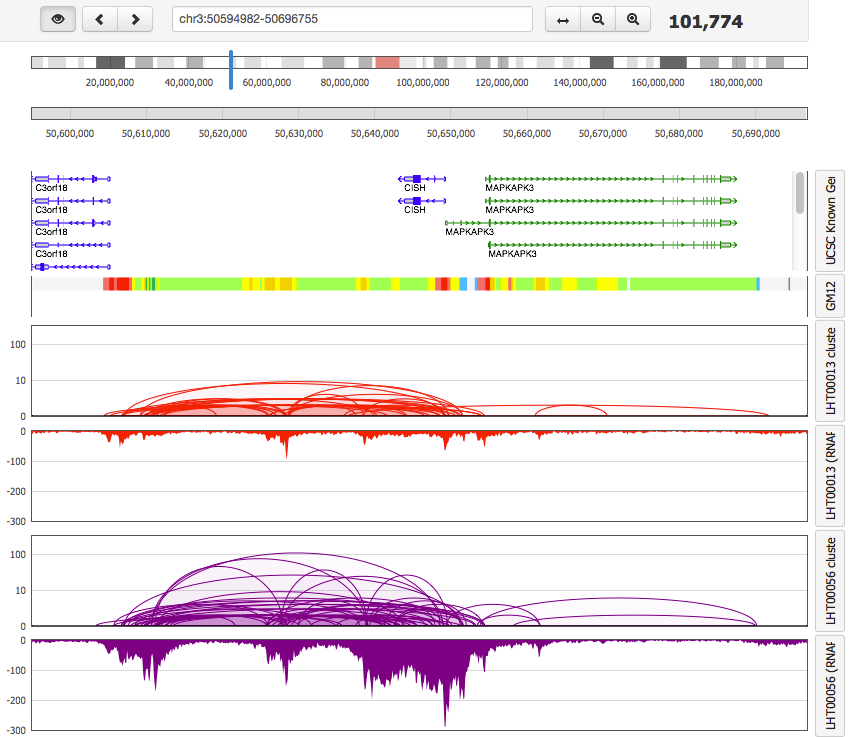


Fig 4. Pol2 ChIA-PET data from primary T cells

(A) Contact frequency heatmaps of Chr1 from T cell before and after IL2 induction. (B) At the *CISH* locus, IL2 induction showed significant increase of Pol2 binding and interaction between distal enhancer and promoter.

**A**

**B**

Chr1 R=0.72

T cell Rp1

T cell Rp2

T cell control

T cell IL2 induction

***CISH***

***Expected results/Pitfalls***: Because of the phenotypic and functional differences, we expect to observe 3D genome organizations that are distinct between major populations of blood cells and between major subsets of cells. This will further validate our technology platform. Indeed, using our current protocol, we can already reliably generate quality and biologically insightful ChIA-PET data from 108 B cells and T cells. For example, RNAPol2 ChIA-PET data from 108 primary T cells from a blood donor showed reliable reproducibility **(Fig. 4A).** Furthermore, detailed examinations at specific loci showed significant increases (>10-fold) of RNAPol2 binding and RNAPol2-mediated long-range chromatin interactions between distal enhancers and target gene promoters, e.g., at the *CISH* (cytokine inducible SH2-containing protein) locus **(Fig. 4B).** Major blood cell–lineage separation was also substantiated by studies from ImmGen consortium ([15](#_ENREF_15)), which systematically analyzed 260 subsets of mouse leukocytes by microarray ([21](#_ENREF_21)). We have many years of experience in isolating, culturing and characterizing blood cell subsets from different sources, and thus expect no technical difficulties in isolating blood cells for these analyses ([19](#_ENREF_19), [23](#_ENREF_23)). To ensure the expected range in recovered cell numbers, we will prepare cells from five donors to ensure high-quality samples from three donors for analysis.

**Aim 3: Perform 3D genome mapping of immune cell responses to activation signals**

Further refinement of our mapping technology will enable ChIA-PET and RICh-PET analysis using 106 or fewer cells per experiment. With these improvements, we will be in a position to ask if our 3D technologies can detect topological changes over fine time-scales during cell-state transitions/induction. We will address these questions biologically in blood-derived immune cells stimulated to undergo differentiation from the naïve to activated state. There are three *in vitro* experimental systems representing three major immune events that are of particular interest to us. They are: 1) the cytokine-driven differentiation of monocytes to antigen-presenting dendritic cells (DCs) or antigen-degrading macrophages (MFs) ([24](#_ENREF_24), [25](#_ENREF_25)); 2) transition of DCs from immature (tolerance-inducing) to mature (immunity-inducing) cells ([26](#_ENREF_26)); and 3) the activation and differentiation of resting naïve CD4+ T cells induced by different DCs into functionally diverse progeny. This is defined by expression of unique transcription factors linked with the production of specific cytokines, including (respectively): 1) T helper 1 (Th1) cells (T-bet and Interferon [IFN]-) ([27](#_ENREF_27)), Th2 (GATA3 and Interleukin [IL-4], IL-13) ([28](#_ENREF_28)) and T regulatory cells (Treg) (FOX-P3 and IL-10) ([29](#_ENREF_29)) **(Fig. 3).**

***Experimental design*:** We will collect primary cells from three individual donors and set up *in vitro* induction conditions for the following three sets of experiments:

*1. Monocytes and monocyte-derived cells*: *In vitro*, monocytes yield DCs when stimulated with granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin-4 (IL-4) ([24](#_ENREF_24), [25](#_ENREF_25)). The addition of fibroblasts switches monocyte differentiation to MF *via* IL-6-M-CSF signaling ([24](#_ENREF_24), [25](#_ENREF_25)). These critically important immune processes impact the fate of the antigen and launching of the immune response. Thus, DCs capture and process antigens to launch T-cell responses, whereas MFs predominantly degrade captured antigens ([24](#_ENREF_24), [25](#_ENREF_25)). We will use *in vitro* culture systems to generate these two cell types from blood monocytes enriched by negative depletion and cultured in the presence of serum and cytokines as described above, and be able to collect 107 DCs and 107 MFs for ChIA-PET and RICh-PET analysis. We will also collect cells from the resting precursor stage (time 0), transient veiled stage (day 2) and fully differentiated DC stage (day 7).

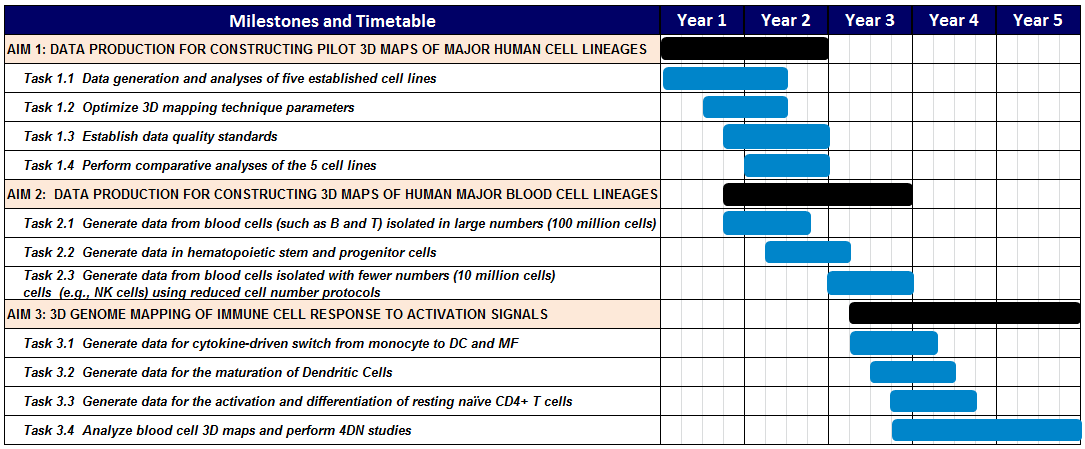
*2. DC activation:* DCs obtained at day 7 of culture are immature and thought to induce tolerance ([27](#_ENREF_27), [28](#_ENREF_28)). These immature DCs have special morphological and molecular characteristics distinct from mature DCs that can launch differentiation of antigen-specific T cells into effector cells with unique functions and cytokine profiles ([30](#_ENREF_30), [31](#_ENREF_31)). The methods are well established ([32](#_ENREF_32)). To this end, immature DCs (day 7) will be exposed for six hours to two powerful but distinct activation signals: 1) microbial LPS and the T-cell-expressed CD40 ligand, or 2) thymic stromal lymphopoietin (TSLP). Immature and mature DCs activated by LPS/CD40 ligand or TSLP will be collected in the range of 106 or more cells for 3DN and 4DN studies.

*3. Subsets of CD4+ T cells*: Upon interaction with immature or mature DCs, naïve T cells will differentiate into antigen-specific memory T cells with distinct functions ([31](#_ENREF_31)). It is well accepted that at each stage of this differentiation, transcription factors interact not only with the existing landscape of histone modifications and nucleosome packaging, but also with other bound factors, while modifying the landscape for later-to-arrive factors in ways that fundamentally affect control of gene expression ([33](#_ENREF_33)). Thus, this system possibly features distinctive genome regulation dynamics for 3DN and 4DN studies. We will generate subsets of T cells from naïve CD4+ T cells *in vitro* through co-culture with different DCs. We will then collect 106 Th1 cells through co-culture with mature, LPS and CD40 ligand–activated DCs secreting IL-12 ([34](#_ENREF_34)); 106 Th2 cells through co-culture with TSLP-activated mature DCs ([35](#_ENREF_35)); and 106 Treg cells through co-culture with immature DCs ([36](#_ENREF_36)). These T cells will be sorted to purity using specific cell-surface markers.

All cells collected from these three experimental systems will be analyzed by ChIA-PET, RICh-PET and related analyses. Collectively, these datasets will enable us to study the longitudinal changes of 3D genome structure in 4D nucleome analysis in fine time-scale.

***Expected results/Future studies***: The studies proposed herein will help validate and demonstrate the usefulness of our technology platform in using small cell numbers to address important biological questions. The successful completion of these studies will encourage the wider research community to apply the tools and concepts generated herein to their valuable biological questions. An added value of this study is that it will also help increase our understanding of how immune cell plasticity is regulated at whole genome level in space (3D) and time (4D). We hope to identify specific structural features that are important for cell plasticity that are linked with acquisition and maintenance of specific phenotypes and function. Future studies will be developed to establish links between nucleome organization and antigen specificity in health and disease.

**PROJECT TIMELINE**

****

**BIBLIOGRAPHY AND REFERENCES CITED**

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(23):9362-7. doi: 10.1073/pnas.0903103106. PubMed PMID: 19474294; PubMed Central PMCID: PMC2687147.

2. Ernst J, Kellis M. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. Genome Res. 2013;23(7):1142-54. doi: 10.1101/gr.144840.112. PubMed PMID: 23595227; PubMed Central PMCID: PMC3698507.

3. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(44):17921-6. doi: 10.1073/pnas.1317023110. PubMed PMID: 24127591; PubMed Central PMCID: PMC3816444.

4. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47. doi: 10.1016/j.cell.2013.09.053. PubMed PMID: 24119843; PubMed Central PMCID: PMC3841062.

5. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2014. doi: 10.1038/nature13835. PubMed PMID: 25363779.

6. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMC3439153.

7. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009;462(7269):58-64. doi: 10.1038/nature08497. PubMed PMID: 19890323; PubMed Central PMCID: PMC2774924.

8. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012;148(1-2):84-98. doi: 10.1016/j.cell.2011.12.014. PubMed PMID: 22265404; PubMed Central PMCID: PMC3339270.

9. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell. 2013;155(7):1507-20. doi: 10.1016/j.cell.2013.11.039. PubMed PMID: 24360274; PubMed Central PMCID: PMC3905448.

10. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013;504(7479):306-10. doi: 10.1038/nature12716. PubMed PMID: 24213634; PubMed Central PMCID: PMC3954713.

11. ENCODE-ChIA-PET. <http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=399865457_UMqQeGejbUVyoGDimzsbaDUA2KH1&c=chr1&g=wgEncodeGisChiaPet>.

12. Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. Nature. 2013;500(7461):207-11. doi: 10.1038/nature12064. PubMed PMID: 23925245; PubMed Central PMCID: PMC3740412.

13. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371-5. doi: 10.1038/nature13138. PubMed PMID: 24646999; PubMed Central PMCID: PMC4113484.

14. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual review of biochemistry. 2012;81:145-66. doi: 10.1146/annurev-biochem-051410-092902. PubMed PMID: 22663078; PubMed Central PMCID: PMC3858397.

15. Heng TS, Painter MW, Immunological Genome Project C. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol. 2008;9(10):1091-4. doi: 10.1038/ni1008-1091. PubMed PMID: 18800157.

16. Chang JT, Wherry EJ, Goldrath AW. Molecular regulation of effector and memory T cell differentiation. Nat Immunol. 2014;15(12):1104-15. doi: 10.1038/ni.3031. PubMed PMID: 25396352.

17. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet. 2011;43(7):630-8. doi: 10.1038/ng.857. PubMed PMID: 21685913; PubMed Central PMCID: PMC3436933.

18. Rao Suhas SP, Huntley Miriam H, Durand Neva C, Stamenova Elena K, Bochkov Ivan D, Robinson James T, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;In press. doi: 10.1016/j.cell.2014.11.021.

19. Yu CI, Becker C, Wang Y, Marches F, Helft J, Leboeuf M, et al. Human CD1c+ dendritic cells drive the differentiation of CD103+ CD8+ mucosal effector T cells via the cytokine TGF-beta. Immunity. 2013;38(4):818-30. doi: 10.1016/j.immuni.2013.03.004. PubMed PMID: 23562160; PubMed Central PMCID: PMC3639491.

20. Aspord C, Pedroza-Gonzalez A, Gallegos M, Tindle S, Burton EC, Su D, et al. Breast cancer instructs dendritic cells to prime interleukin 13-secreting CD4+ T cells that facilitate tumor development. J Exp Med. 2007;204(5):1037-47. PubMed PMID: 17438063.

21. Shay T, Kang J. Immunological Genome Project and systems immunology. Trends Immunol. 2013;34(12):602-9. doi: 10.1016/j.it.2013.03.004. PubMed PMID: 23631936.

22. Kim CC, Lanier LL. Beyond the transcriptome: completion of act one of the Immunological Genome Project. Curr Opin Immunol. 2013;25(5):593-7. doi: 10.1016/j.coi.2013.09.013. PubMed PMID: 24168965; PubMed Central PMCID: PMC3855358.

23. Pulendran B, Banchereau J, Burkeholder S, Kraus E, Guinet E, Chalouni C, et al. Flt3-ligand and granulocyte colony-stimulating factor mobilize distinct human dendritic cell subsets in vivo. J Immunol. 2000;165(1):566-72.

24. Chomarat P, Banchereau J, Davoust J, Palucka AK. IL-6 switches the differentiation of monocytes from dendritic cells to macrophages. Nat Immunol. 2000;1(6):510-4. PubMed PMID: 11101873.

25. Chomarat P, Dantin C, Bennett L, Banchereau J, Palucka AK. TNF skews monocyte differentiation from macrophages to dendritic cells. J Immunol. 2003;171(5):2262-9. PubMed PMID: 12928370.

26. Steinman RM, Hawiger D, Nussenzweig MC. Tolerogenic dendritic cells. Annu Rev Immunol. 2003;21:685-711. PubMed PMID: 12615891.

27. Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell. 2000;100:655-69.

28. Zheng W, Flavell RA. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell. 1997;89(4):587-96. PubMed PMID: 9160750.

29. Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol. 2003;4(4):330-6. PubMed PMID: 12612578.

30. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. Nature. 1998;392(6673):245-52.

31. Steinman RM, Banchereau J. Taking dendritic cells into medicine. Nature. 2007;449(7161):419-26. PubMed PMID: 17898760.

32. Banchereau R, Baldwin N, Cepika AM, Athale S, Xue Y, Yu CI, et al. Transcriptional specialization of human dendritic cell subsets in response to microbial vaccines. Nat Commun. 2014;5:5283. doi: 10.1038/ncomms6283. PubMed PMID: 25335753; PubMed Central PMCID: PMC4206838.

33. Rothenberg EV. The chromatin landscape and transcription factors in T cell programming. Trends Immunol. 2014;35(5):195-204. doi: 10.1016/j.it.2014.03.001. PubMed PMID: 24703587; PubMed Central PMCID: PMC4039984.

34. Trinchieri G. Interleukin-12 and its role in the generation of TH1 cells. Immunol Today. 1993;14(7):335-8. Epub 1993/07/01. PubMed PMID: 8103338.

35. Liu YJ, Soumelis V, Watanabe N, Ito T, Wang YH, De Waal Malefyt R, et al. TSLP: An epithelial cell cytokine that regulates T cell differentiation by conditioning dendritic cell maturation. Annu Rev Immunol. 2007;25:193-219.

36. Roncarolo MG, Levings MK, Traversari C. Differentiation of T regulatory cells by immature dendritic cells. J Exp Med. 2001;193(2):F5-9. PubMed PMID: 11208869.