

# Monocytic THP-1 Cells Diverge Significantly From Their Native Counterparts: A Comparative Examination of The Chromosomal Conformations And Transcriptomes

Yulong Liu

Shanghai Jiao Tong University <https://orcid.org/0000-0003-4933-782X>

Hua Li

Shanghai Jiao Tong University

Daniel M Czajkowsky (✉ [dczaj@sjtu.edu.cn](mailto:dczaj@sjtu.edu.cn))

Shanghai Jiao Tong University <https://orcid.org/0000-0002-2745-9546>

Zhifeng Shao

Shanghai Jiao Tong University

---

## Brief report

**Keywords:** Hi-C, Chromatin conformation, Monocytes, Macrophages, Transcriptome

**Posted Date:** May 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-483700/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Immortalized cell lines have long been used as model systems to systematically investigate biological processes under controlled and reproducible conditions, providing insights that have greatly advanced cellular biology and medical sciences. Recently, the widely used monocytic leukemia cell line, THP-1, was comprehensively examined to understand mechanistic relationships between the 3D chromatin structure and transcription during the trans-differentiation of monocytes to macrophages. To corroborate these observations in primary cells, we analyze *in situ* Hi-C and RNA-seq data of human primary monocytes and their differentiated macrophages in comparison to that obtained from the monocytic/macrophagic THP-1 cells. Surprisingly, we find significant differences between the primary cells and the THP-1 cells at all levels of chromatin structure, from loops to topologically associated domains to compartments. Importantly, the compartment-level differences correlate significantly with transcription: those genes that are in A-compartments in the primary cells but are in B-compartments in the THP-1 cells exhibit a higher level of expression in the primary cells than in the THP-1 cells, and vice versa. Overall, the genes in these different compartments are enriched for a wide range of pathways, and, at least in the case of the monocytic cells, their altered expression in certain pathways in the THP-1 cells argues for a less immune cell-like phenotype, suggesting that immortalization or prolonged culturing of THP-1 caused a divergence of these cells from their native counterparts. It is thus essential to reexamine phenotypic details observed in cell lines with their native counterparts so as to ensure a proper understanding of functional cell states *in vivo*.

## Background

Immortalized cell lines are one of the pillars of contemporary cell biology, providing essentially unlimited access to experimental material and highly pure cell populations [1, 2]. Their use has unquestionably advanced our understanding of molecular mechanisms underlying biological processes in many areas of biomedical research[3-7]. A notable recent example is the comprehensive examination of the changes in chromatin structure that are associated with the changes in transcription during the monocyte-to-macrophage differentiation in the model monocytic cell line THP-1, which was originally derived from an acute myeloid leukemia patient [3]. One of the critical discoveries in this work is that the transcription factor AP-1, observed to be up-regulated during this transition, plays a fundamental role in both static and dynamic chromosome structure, specifically at the level of loops [8].

In parallel, this monocyte-to-macrophage trans-differentiation process was also studied in primary human monocytes using *in situ* Hi-C and RNA-seq [9]. Intriguingly, the up-regulation of AP-1 during this transition was not confirmed in these primary cells. This observation raises the question of to what extent do the THP-1 cells and their differentiated macrophages resemble their native counterparts, a question with broad implications for other immortalized/native cell pairings as well.

To shed light on this issue, we compared the chromatin structures of the monocytic/macrophagic THP-1 cells and their primary counterparts by reanalyzing these recently available Hi-C data. Unexpectedly, we

find that their chromatin structures are dramatically different at all levels of conformation, from loops to topologically associated domains (TADs) to compartments. Further, the differences in structure at the compartment-level correlate significantly with differences in the expression of genes localized within these regions, including those that are upregulated in the primary cells that are associated with typical functions of monocytic immune cells. Thus, overall, these results reveal a divergence of the THP-1 cells during immortalization or prolonged culturing from their native counterparts, underscoring the requirement to ultimately validate observations obtained from immortalized cultured cells with native cells for a faithful characterization of functional states *in vivo*.

## Results And Discussion

The recent publication of *in situ* Hi-C data of the monocytic THP-1 cells and their differentiated macrophage induced by the addition of phorbol myristate acetate (PMA), together with that of the *in situ* Hi-C study of human primary monocytes and the granulocyte-macrophage colony stimulating factor (GM-CSF)-induced macrophages, provides a unique opportunity to validate the observations obtained from this commonly used immortalized cell line with its native counterpart.

Following a well-established method [10], we first compared larger-length scale, compartment-level interactions, after first verifying that the A-compartments are generally enriched for active genes and the B-compartments are enriched for inactive genes for each of the four cell types (Supplementary Fig. S1) [11]. We found that there is a significant difference (20%) in compartment-level interactions between the primary monocytes and monocytic THP-1 cells (Supplementary Fig. S2a), which is consistent with previous analysis [9]. A comparable difference (19%) in compartment-level interactions was also found between the primary, GM-CSF-induced macrophages and the macrophagic THP-1 cells (Supplementary Fig. S2b). Such a difference in compartment-level features is generally considered as evidence of substantially reorganized chromatin [12, 13], similar to what is observed following, for example, differentiation of embryonic stem cells [13]. Thus, these results indicate that these immortalized cells have, in fact, evolved to a fundamentally different chromatin structure from their native counterparts.

To determine whether these longer-range structural differences are consequential for gene expression, we compared the level of gene expression within those regions that are in different compartments in the primary and THP-1 cells. For the monocytic-type cells, those genes that are found in the A-compartments in the primary cells but in the B-compartments in the THP-1 cells exhibit a higher level of expression in the primary monocytes than in the THP-1 cells (Fig. 1a). These genes are enriched for many different gene ontological terms, suggesting that there are many biological pathways affected by these structural differences (Fig. 1b, Supplementary Table 1). However, it is worth noting that many terms are associated with basic immune functions, such as the innate immune response, positive regulation of chemokine production, and defense response to viruses, all important processes in monocytes (Fig. 1b). For example, several genes that encode for human alpha defensins (DEFA1, DEFA1B, and DEFA3), which are involved in host defense [14, 15], all exhibit essentially no expression in the monocytic THP-1 cells where they are in the B-compartment, but have significant expression in primary monocytes where they are in the A-

compartment (Fig. 1c). This suggests that the immortalized cells have lost at least some of their original monocytic functions as a consequence of, or in association with, these alterations in chromatin structure. In addition, the genes in the B-compartments in the primary monocytes that have relocated to the A-compartments in the THP-1 cells are found to be upregulated in the THP-1 cells (Fig. 1d). These THP-1 up-regulated genes are also enriched for many biological pathways, including response to starvation, cell adhesion, and response to nutrients (Fig. 1e, Supplementary Table 1). It is possible that these genes exhibit greater expression in the THP-1 cells owing to their highly proliferative nature in culture, while for the primary cells, since proliferation is largely absent, these genes are not needed and are thus inactive.

Similarly, when the primary, GM-CSF-induced macrophage cells and the macrophagic THP-1 cells were compared, we also found significant relocation of genes between different compartments (Supplementary Fig. S3a, b). Likewise, the genes in these different compartments are enriched for many biological pathways, suggesting that these structural changes are also associated with differences in the functional states between these cells (Supplementary Fig. S3c, d; Supplementary Table 1). Thus, like the monocytic-type cells, these results suggest that the macrophagic THP-1 cells are also fundamentally different from their native counterpart.

We next examined for differences in the more local-level interactions of loops and TADs, the latter of which are very highly conserved (> 92%) between many cell types [16, 17]. Since statistically valid comparisons of such high-resolution features requires data at a comparable sequencing depth, we first obtained multiple (5) down-sampled datasets of the original THP-1 data, each at a similar level (90 million reads) as the primary cell datasets (see Methods). We used TopDom to identify the TADs within each dataset at the highest map resolution possible with this data (40 kb, see Methods) and calculated the extent to which the TAD locations were identical between different datasets. We found that there was good agreement among the down-sampled monocytic THP-1 datasets ( $74.1 \pm 0.4\%$ ), with the less than perfect agreement between these datasets consistent with the lowered sampling depth. In contrast, only  $26.1 \pm 0.3\%$  of the TADs in the primary monocytes were the same as those in the down-sampled THP-1 datasets, which is significantly different ( $p < 0.001$ , normality test) from that within the down-sampled datasets (Supplementary Fig. S4a). Similarly, comparing just the down-sampled macrophagic THP-1 datasets, there was good agreement in the identified TADs ( $71.2 \pm 0.6\%$ ). However, only  $51.5 \pm 0.3\%$  of the TADs in the primary, GM-CSF-induced macrophage cells were the same as those in the down-sampled macrophagic THP-1 cells (Supplementary Fig. S4b), which is also significantly different ( $p < 0.001$ , normality test) from the measurements of just the down-sampled datasets. Similar differences, for both monocytic and macrophagic-type cells, were also observed at lower map resolutions as well (Supplementary Fig. S4c, d). This degree of difference is truly exceptional inasmuch as, for example, the differentiation of embryonic stem cells was associated with changes in chromatin structure at the compartment-level but with essentially no changes at the TAD-level [13]. This underscores the magnitude of the structural changes in the chromatin that the THP-1 cells have undergone from their presumed original state.

We last compared the differences at the loop-level, using the program Peakachu, which has been shown to robustly identify loops at a 10 kb resolution even for data that is at a lower sequencing depth (down to 30 million reads) than our primary cell data [18]. This method computes the probability that there is a loop in each pixel in the Hi-C map using a machine-learning framework, with bona-fide loops identified as those pixels with probability values above 0.97, a threshold value recommended by the original authors (see Methods) [18]. With the same consideration as that for the TAD analysis, we first compared the number of loops identified among the THP-1 down-sampled datasets, and found excellent agreement among the monocytic THP-1 ( $885 \pm 8$ ) and the macrophagic THP-1 ( $826 \pm 17$ ) down-sampled datasets (Supplementary Fig. S5). By contrast, we found that there are 451 and 997 loops in the primary monocytes and primary, GM-CSF-induced macrophages, respectively. Thus, there is a statistically significant difference in the number of loops between the primary and cell lines with both types of cells ( $p < 0.001$ , normality test), with a quite pronounced increase of nearly 2-fold in the number of loops in the monocytic THP-1 cells.

We finally examined the similarity in the loop locations between the different datasets (Fig. 2, Supplementary Fig. S6). For this, we compared the extent to which the loops identified in each dataset (down-sampled THP-1 or primary cells) are at the same location as in the full THP-1 datasets. Using the threshold cutoff value of 0.97, the loops identified in the down-sampled THP-1 datasets were indeed highly similar to those in the full datasets, with  $88.3 \pm 0.6\%$  and  $89.9 \pm 1.4\%$  of loops identical in the monocytic and macrophagic cells, respectively. However, at the same threshold, there was significantly less ( $p < 0.001$ , normality test) identity of loop locations between the primary monocytes (80.1%) and the primary, GM-CSF-induced macrophage cells (80.1%) with the corresponding full THP-1 datasets (Fig. 2c). A similar level of difference between the down-sampled datasets and the primary cell datasets was observed over a broad range of threshold values (Fig. 2c), showing the robustness in this difference. Thus, in both the number and location of the loops, there are significant differences between the primary cells and the THP-1 cell line.

## Conclusion

Although immortalized cell lines are unquestionably a staple of present biological research, it is also essential to understand the extent to which they faithfully reflect the properties of native cells. Here, we show that there is a surprisingly large degree of differences in the chromatin structures of the THP-1 cell line and their primary cell counterparts at all levels of chromatin structure. The degree of differences in the compartments ( $\sim 20\%$ ) indeed reflects a significant change in chromosome structure, especially among ostensibly highly related cells (that is, they are both monocytic cells), since at this level, most pairwise comparisons of different cell types reveal fewer differences [13]. Yet, in previous studies that identified this high degree of compartment-level differences, the structures at the TAD-level were nonetheless found to be nearly identical [16, 17]. However, we found that there was a drastic difference at the TAD level (19% and 48% different from the down-sampled datasets themselves) between these primary cells and the THP-1 cells, demonstrating that these cells are exceptionally different in chromatin structure at this level. Finally, even at the level of loops, we found that there are significant differences in

both their number and location between these cells. With such wide-ranging differences in structure, it is perhaps not surprising that AP-1, which was found to play a significant role in the formation of loops during the trans-differentiation of THP-1 cells [8], is not up-regulated in the primary cells [9]. Thus, the critical components and molecular mechanisms underlying the structural changes in the primary cells during trans-differentiation remain to be discovered.

The functional significance of these structural differences between the THP-1 and the primary cells is evidenced by the clear correlation between the compartment and gene expression differences, with a significant enrichment of many different biological pathways in all different compartments, including those associated with immune-cell processes. We note that the lower sequencing depth of the primary cells prevents a comprehensive analysis of relationships between differences at the loop or TAD level and gene expression in the primary and THP-1 cells. In addition, the differences in the induction mechanisms (PMA versus GM-CSF) might have contributed to some of the differences observed between the macrophagic-type cells. Nonetheless, the differences observed between the primary monocytes and the THP-1 cells can only be owing to the evolution of the THP-1 cells during immortalization or prolonged culturing from their original native monocytic cell. Thus, our results indicate that the THP-1 cell line has lost some of its native monocytic features in chromatin structure, likely associated with changes in the activation of specific pathways, that lead to a differing phenotype. Hence, while immortalized cells are, without question, useful for analysis of biological processes within cells, there should be caution when attributing physiological significance to observations obtained with these cells that have not been directly validated with native cells [17].

## Materials And Methods

### *Hi-C data processing*

All Hi-C reads were mapped to the human hg19 reference genome iteratively using Bowtie2 as described in [19]. Reads mapped to multiple genomic locations and with low mapping quality (MAPQ < 30) were removed. Read pairs that mapped to the same restriction fragment and PCR duplicates were further removed. To eliminate the bias introduced by extreme read coverage, Hi-C reads mapped to the blacklist features defined in the ENCODE project were also removed [20]. Finally we generated Hi-C contact matrices at 10 kb, 40 kb, and 100 kb resolution and normalized using cooler [21].

### *mRNA-Seq quality control and alignment of sequencing data*

Raw reads were processed with Cutadapt (v1.18) [22] and Trimmomatic (v0.38) [23] to remove sequencing adapters, short reads (length < 35 bp) and low quality reads. FastQC was then used to ensure high quality reads [24]. The clean reads were mapped to the human genome (hg19) with the HISAT2 software [25] with same parameter in [26]. Gene expression levels were calculated based on FPKM (fragments per kilobase of exon per million fragments mapped) with StringTie [25]. The GO functional enrichment analysis was performed with DAVID [27].

### *Identification of compartments*

Compartments were annotated using Juicer [10] at 200 kb resolution. The signs of the eigenvectors that indicate different types of compartments were determined according to the Pearson correlation coefficient (PCC) between the eigenvector and gene density as described in [28]. Eigenvectors are multiplied by  $-1$  if the PCC is negative. Positive eigenvectors indicate A-compartments and negative eigenvectors indicate B-compartments.

#### *Identification of TADs and loops*

The identification of TADs is based on the 40 kb resolution normalized matrix, which is the highest resolution the our dataset could achieve [29]. TADs were annotated using the TopDom R package [30] with a window size of 20, which is within the recommended range (5–20). All loops were annotated in the 10 kb resolution normalized Hi-C matrix. The loops were annotated using Peakachu, which has been shown to be a reliable loop caller and is robust to sequencing depth [18]. For all loop predictions in the work, the cutoff threshold was 0.97 as recommended by the authors of this software.

#### *The comparison of similarity of loop locations*

The loop anchors were defined in the 10 kb resolution map at either the start or end points of the loops. Due to the sparsity introduced by down-sampling from the full dataset, the loop center in the down-sampled dataset may have shifted 1 to 2 bins (each bin is 10 kb) compared to the loop centers in full dataset. Thus, identical loops in the different datasets were defined as those loops with both anchors within 20 kb of each other. The “identical percentage” is the number of identical loops divided by the number of all loops in the down-sampled or primary cell datasets.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Competing interests**

No Competing Interest.

### **Funding**

This work was supported by the following grants: National Key R&D Program of China (Grant No. 2018YFC1003500), Medical Engineering Cross Fund of Shanghai Jiao Tong University (ZH2018ZDB02) and National Natural Science Foundation of China (Grant Nos. 31971151 and 81627801).

### **Availability of data and materials**

The data of the THP-1 cells were obtained from the Gene Expression Omnibus (GEO) with series number GSE96800 (RNA-seq) and Sequencing Read Archive (SRA) with accession number PRJNA385337 (Hi-C). The data of the primary cells were obtained from ArrayExpress with following codes: E-MTAB-8261 (RNA-seq) and E-MTAB-8262 (Hi-C).

## Abbreviations

Not applicable.

## Authors' contributions

Z.S. and D.M.C conceived and designed the project. Y.L performed the data analysis. Y.L, D.M.C and Z.S wrote the manuscript. H.L, D.M.C and Z.S are senior authors of this manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Dr. Qiu Sun and Chuansheng Hu for useful discussions.

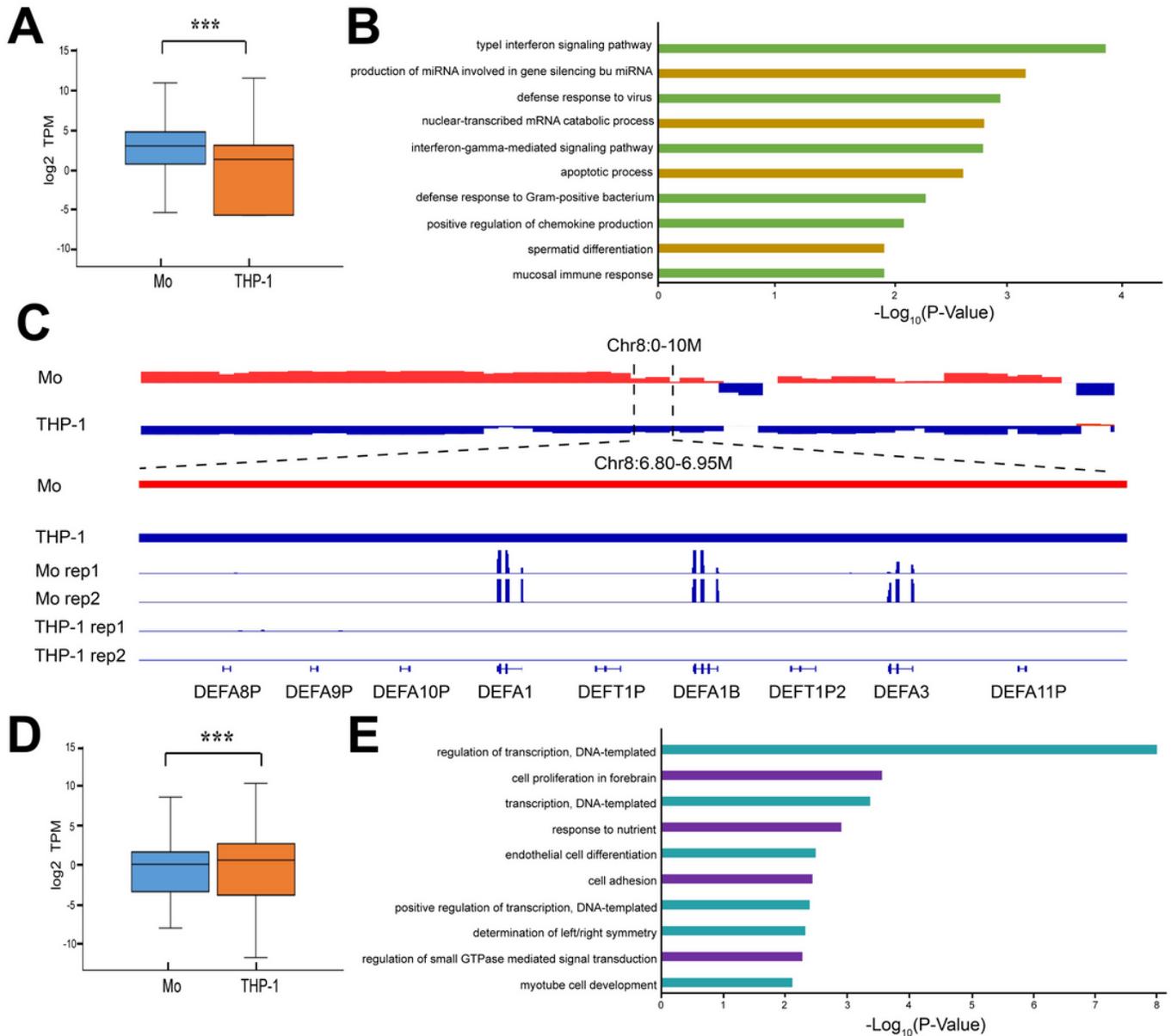
## References

1. Kaur, G. and J.M. Dufour, Cell lines: Valuable tools or useless artifacts. 2012, Taylor & Francis.
2. Masters, J.R., Human cancer cell lines: fact and fantasy. *Nature reviews Molecular cell biology*, 2000. **1**(3). 233-236.
3. Auwerx, J., The human leukemia cell line, THP-1: a multifaceted model for the study of monocyte-macrophage differentiation. *Experientia*, 1991. **47**(1). 22-31.
4. Chanput, W., J.J. Mes, and H.J. Wichers, THP-1 cell line: an in vitro cell model for immune modulation approach. *International immunopharmacology*, 2014. **23**(1). 37-45.
5. Marion, M.-J., O. Hantz, and D. Durantel, The HepaRG cell line: biological properties and relevance as a tool for cell biology, drug metabolism, and virology studies, in *Hepatocytes*. 2010, Springer. 261-272.
6. Masters, J.R., HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer*, 2002. **2**(4). 315-319.
7. Thomas, P. and T.G. Smart, HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal of pharmacological and toxicological methods*, 2005. **51**(3). 187-200.
8. Phanstiel, D.H., et al., Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Molecular cell*, 2017. **67**(6). 1037-1048. e6.
9. Zhang, Z., et al., Massive reorganization of the genome during primary monocyte differentiation into macrophage. *Acta biochimica et biophysica Sinica*, 2020. **52**(5). 546-553.
10. Durand, N.C., et al., Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems*, 2016. **3**(1). 95-98.

11. Lieberman-Aiden, E., et al., Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 2009. **326**(5950). 289-293.
12. Barutcu, A.R., et al., Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome biology*, 2015. **16**(1). 1-14.
13. Dixon, J.R., et al., Chromatin architecture reorganization during stem cell differentiation. *Nature*, 2015. **518**(7539). 331-336.
14. Duits, L.A., et al., Expression of  $\beta$ -defensin 1 and 2 mRNA by human monocytes, macrophages and dendritic cells. *Immunology*, 2002. **106**(4). 517-525.
15. Saahil, F., D.W. El-Naccache, and T.L. Chang, Multifaceted immune functions of human defensins and underlying mechanisms. *Seminars in Cell and Developmental Biology*, 2018. **88**. S1084952117304159-.
16. Anthony D., et al., A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*, 2016.
17. Liu, X., et al., Epithelial Cells in 2D and 3D Cultures Exhibit Large Differences in Higher-order Genomic Interactions. *Genomics, Proteomics & Bioinformatics*, 2021.
18. Salameh, T.J., et al., A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nature communications*, 2020. **11**(1). 1-12.
19. Wang, Q., et al., Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nature communications*, 2018. **9**(1). 1-8.
20. Amemiya, H.M., A. Kundaje, and A.P. Boyle, The ENCODE blacklist: identification of problematic regions of the genome. *Scientific reports*, 2019. **9**(1). 1-5.
21. Abdennur, N. and L.A. Mirny, Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*, 2020. **36**(1). 311-316.
22. Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 2011. **17**(1). 10-12.
23. Bolger, A.M., M. Lohse, and B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014. **30**(15). 2114-2120.
24. Andrews, S., FastQC: a quality control tool for high throughput sequence data. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
25. Pertea, M., et al., Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, 2016. **11**(9). 1650.
26. Wang, Y., et al., Q-Nuc: a bioinformatics pipeline for the quantitative analysis of nucleosomal profiles. *Interdisciplinary Sciences: Computational Life Sciences*, 2020. **12**(1). 69-81.
27. Sherman, B.T. and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 2009. **4**(1). 44.
28. Miura, H., et al., Practical Analysis of Hi-C Data: Generating A/B Compartment Profiles: Methods and Protocols. 2018: X-Chromosome Inactivation.

29. Rao, S.S., et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 2014. **159**(7). 1665-1680.
30. Shin, H., et al., TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic acids research*, 2016. **44**(7). e70-e70.

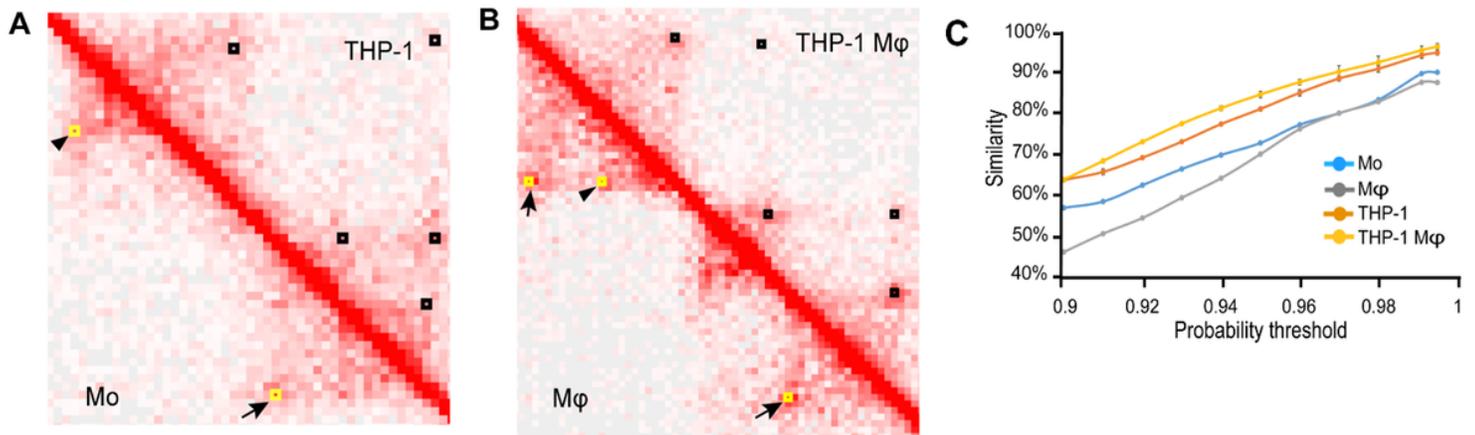
## Figures



**Figure 1**

Relationship between differences in compartments and gene expression between primary monocytes and THP-1 cells. (A) Comparison of gene expression levels at regions that switched from A- compartment in the primary monocytes to B-compartment in the monocytic THP-1 cells (“\*\*\*” represents  $p < 0.001$ , Wilcoxon rank sum test) TPM refers to transcript per million. (B) Gene ontology analysis of the genes that

reside in the compartments described in (A). Colored in green are those terms related to typical immune cell functions. (C) Example of immune cell-related genes (DEFA1, DEFA1B, DEFA3) whose changed expression is associated with the different compartments depicted in (A). A-compartments are colored in red and B-compartments are colored in blue. (D) Comparison of gene expression levels at regions that switched from B-compartment in primary monocytes to A-compartment in the monocytic THP-1 cells (“\*\*\*” represents  $p < 0.001$ , Wilcoxon rank sum test). (E) Gene ontology analysis of the genes reside in the compartments described in (D). Colored in purple are those terms related to cell proliferation.



**Figure 2**

Differences in the structure at the loop level between primary and cell lines. (A) Hi-C heat map at Chr2:99.0-99.5M (10 kb resolution) of the monocytic primary and THP-1 cells. The squared regions are the loops identified by Peakachu. The arrows point to loops the are present in both cell types. The arrowheads point to loops only in the primary cells. (B) Hi-C heat map at Chr1:174.7-175.3M (10 kb resolution) of the macrophagic primary and THP-1 cells. Annotations are the same as in (A). (C) Relationship between the identical loop percentage and the probability threshold cutoff used to filter the loops, comparing all data to the full dataset of the corresponding THP-1 cells.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplefiglegendfinal.doc](#)
- [SupplementaryTableS1.xlsx](#)