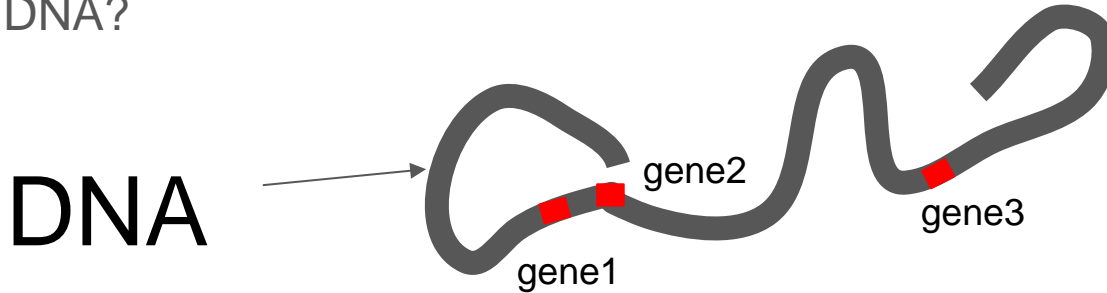# The relationship between gene expression correlation and 3D genome organization

Jason Yang, Martin Falk, Sameer Abraham
Prof. Leonid Mirny
MIT PRIMES
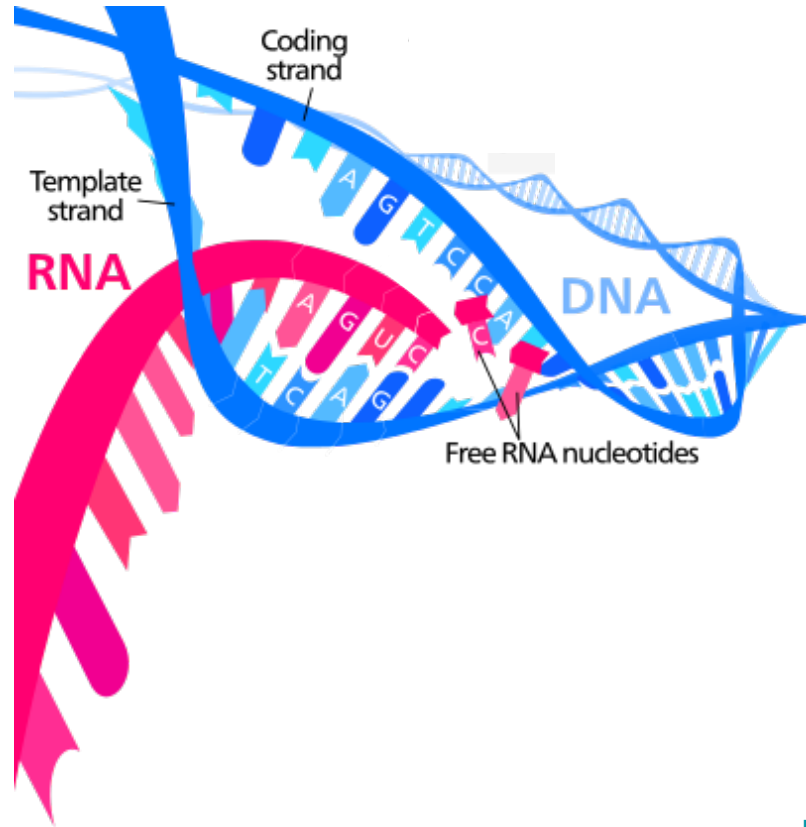
Big Question

Do gene expressions across cells correlate when genes are close to each other along DNA?

DNA

gene1

gene2

gene3

	cell1	cell2	cell3	cell4	cell5	cell6	cell7
gene1	59	62	86	70	14	37	93
gene2	60	55	91	67	11	44	94
gene3	26	95	84	78	12	46	89

RNA-seq

Measure counts of RNA transcripts of a certain gene in a cell to determine number of times that gene is expressed in that cell
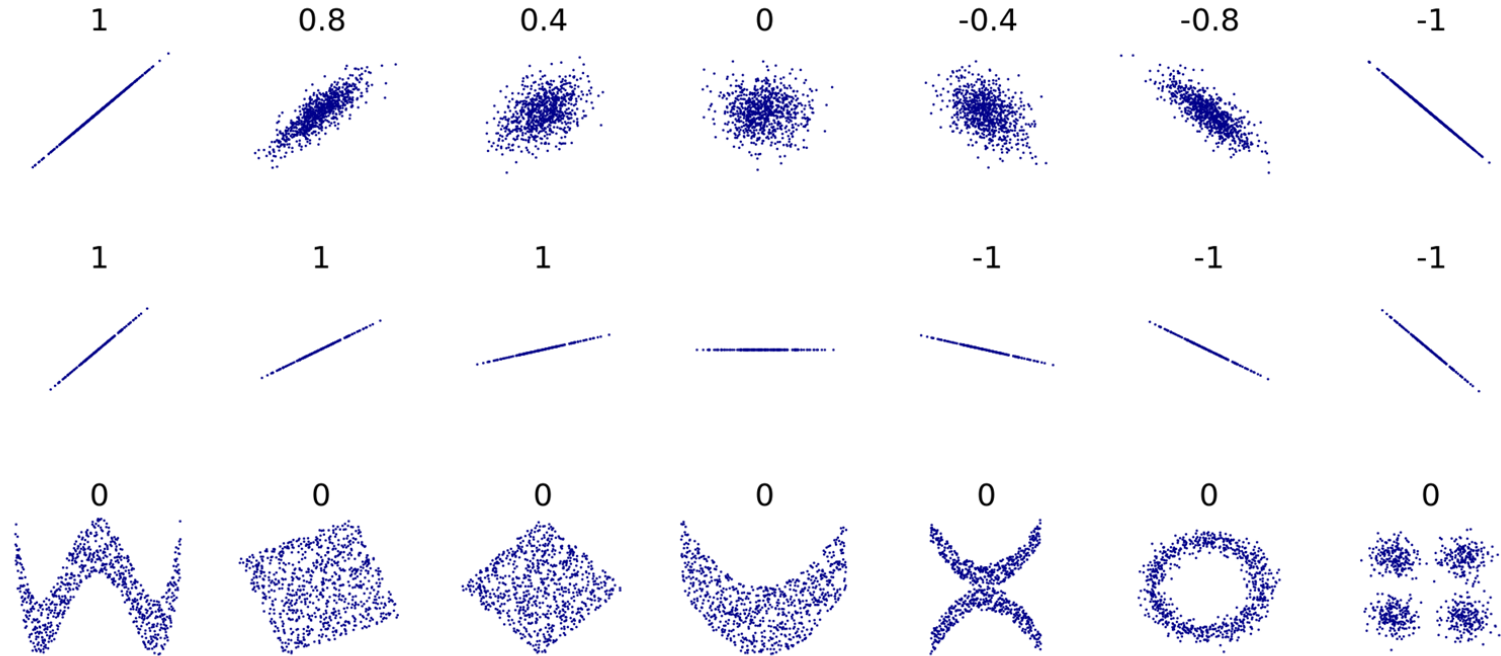
Example of RNA-seq table

	cell1	cell2
gene1	92	3
gene2	76	66
gene3	78	49
gene4	27	41

\# of times gene1 has been expressed
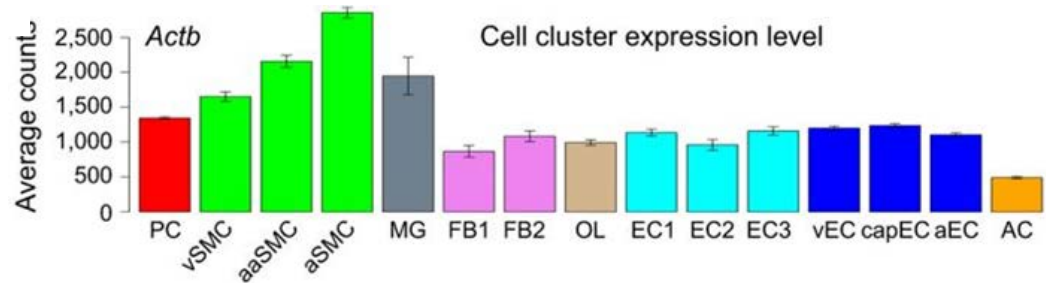in cell1

Pearson Correlation

measuring the strength of the linear relationship between two variables

The data

"Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types" (GSE98816; Sci Data. 2018 Aug 21;5:180160 He et. al)

scRNA-seq on 19937 genes of 3186 mouse brain cells, which was used to determine 15 cell types

PC - Pericytes
SMC - Smooth muscle cells
MG - Microglia
OL - Oligodendrocytes
FB - Vascular Fibroblast-like cell
EC - Endothelial cells
AC- Astrocytes

v- venous
cap - capillary
a - aterial
aa - arteriolar
1,2,3 - subtypes
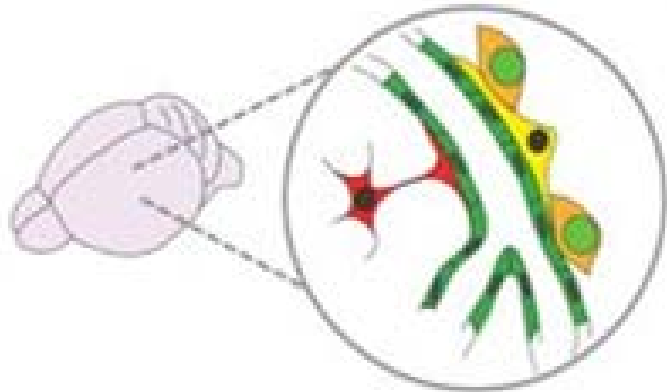
https://www.nature.com/articles/sdata2018160
https://figshare.com/articles/Description_of_each_single-cell_samples/6170075/1

What we did

Plot each pair of genes as a point on the xy-plane

x=DNA base distance between pair of genes (distance between midpoints of both genes)
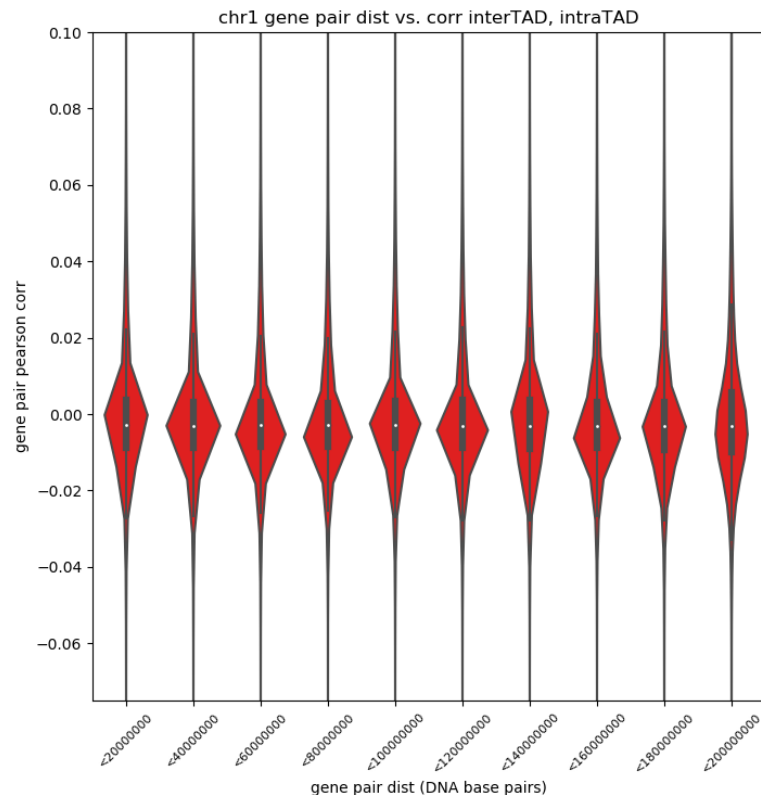
y=Pearson correlation of expression between the pair of genes across all brain cells

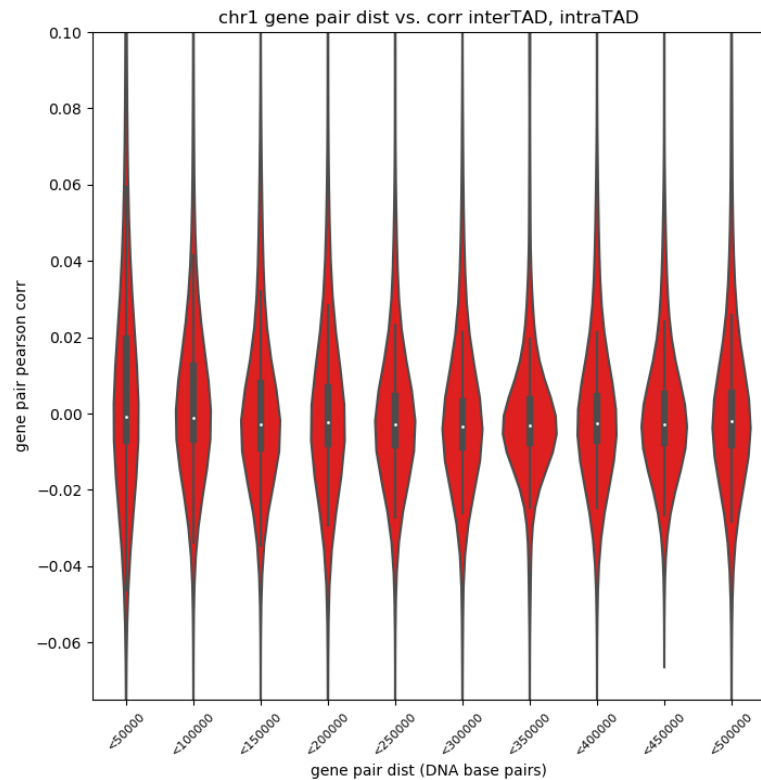Only consider gene pairs within a single chromosome

No significant pattern found

Violin plot of chromosome 1

chr1 gene pair dist vs. corr interTAD, intraTAD
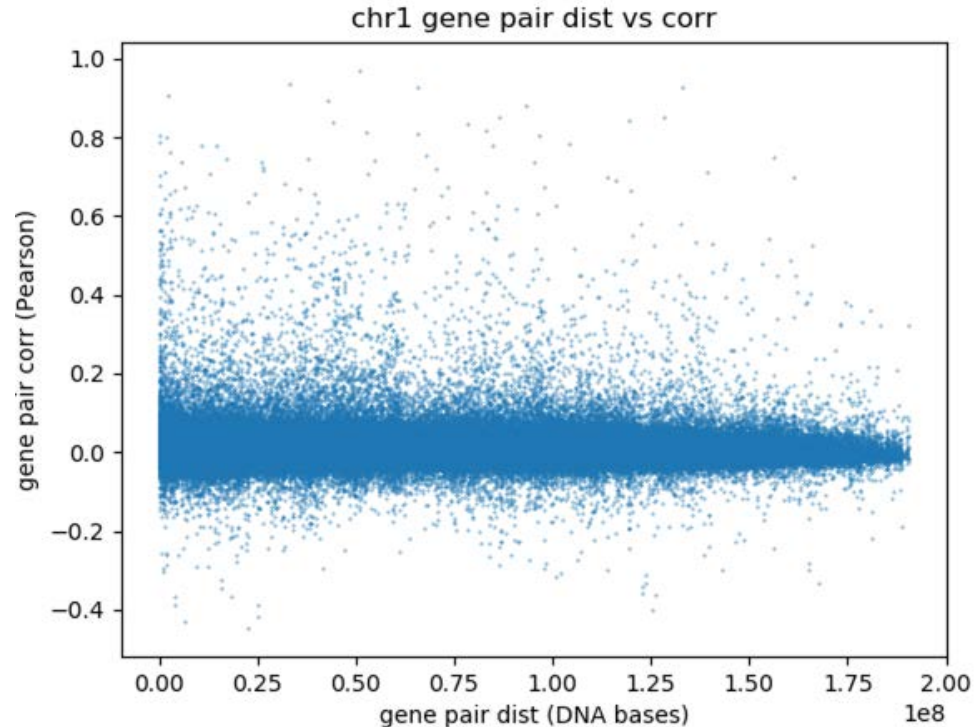
Violin plot of chromosome 1 (closer range)

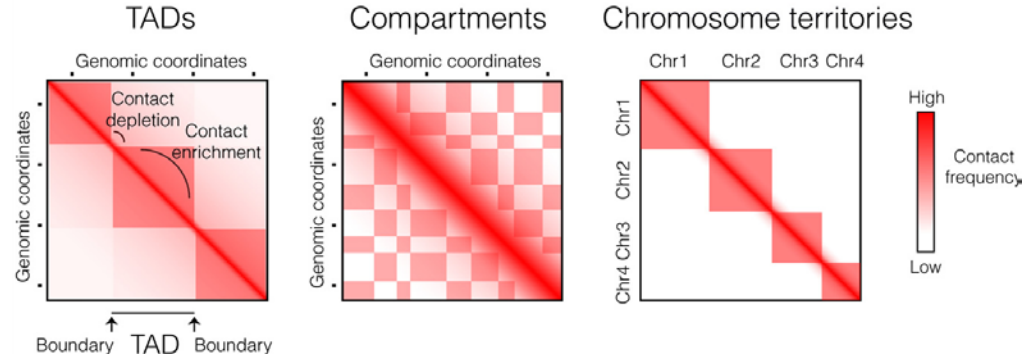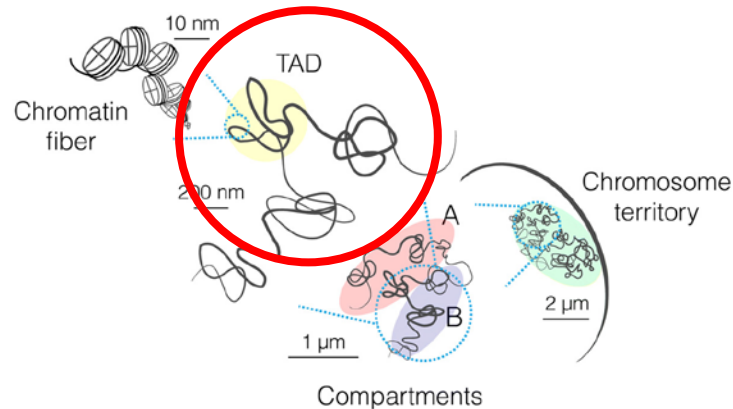chr1 gene pair dist vs. corr interTAD, intraTAD

What about TADs (Topologically Associated Domains)?

Regions of DNA where loci inside them contact each other frequently compared to loci outside of TADs

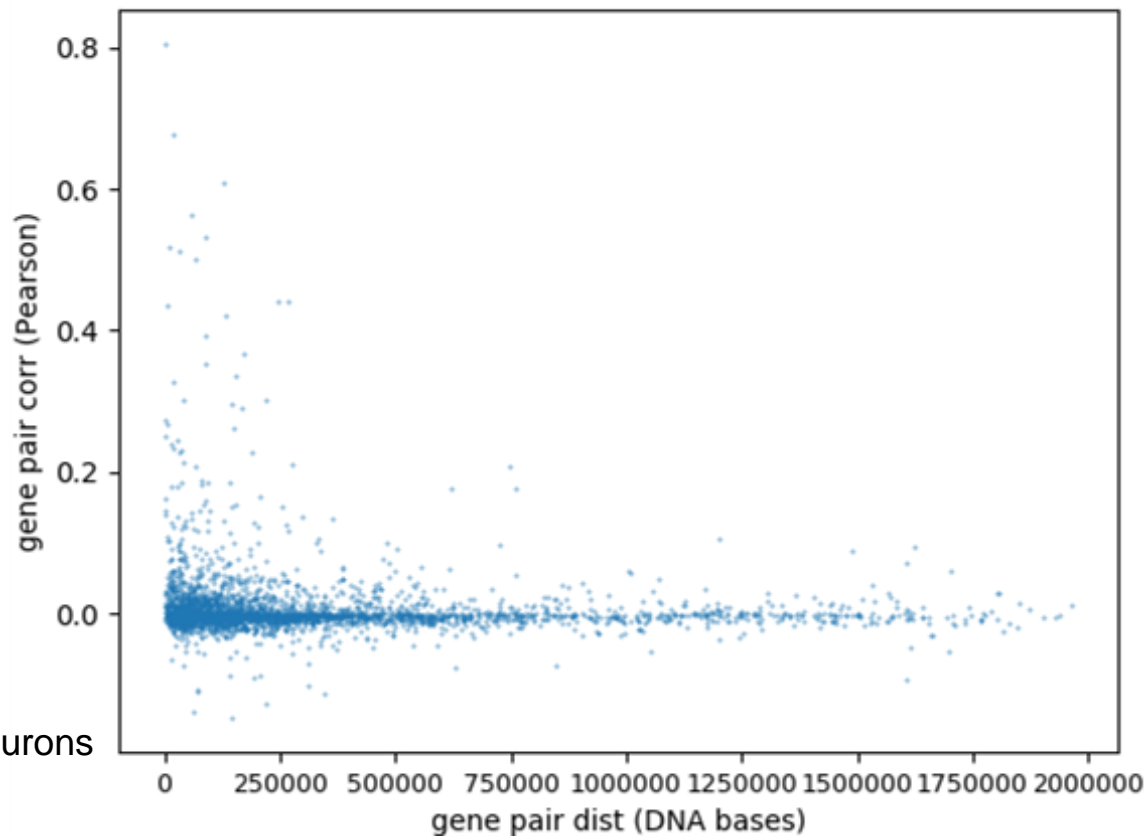Shown in Hi-C as small, bright squares along the diagonal

Smaller than compartments, which are groups of regions of DNA whereby loci in the same compartment contact each other than loci in different compartments

Gene pairs intra- and inter- TADs

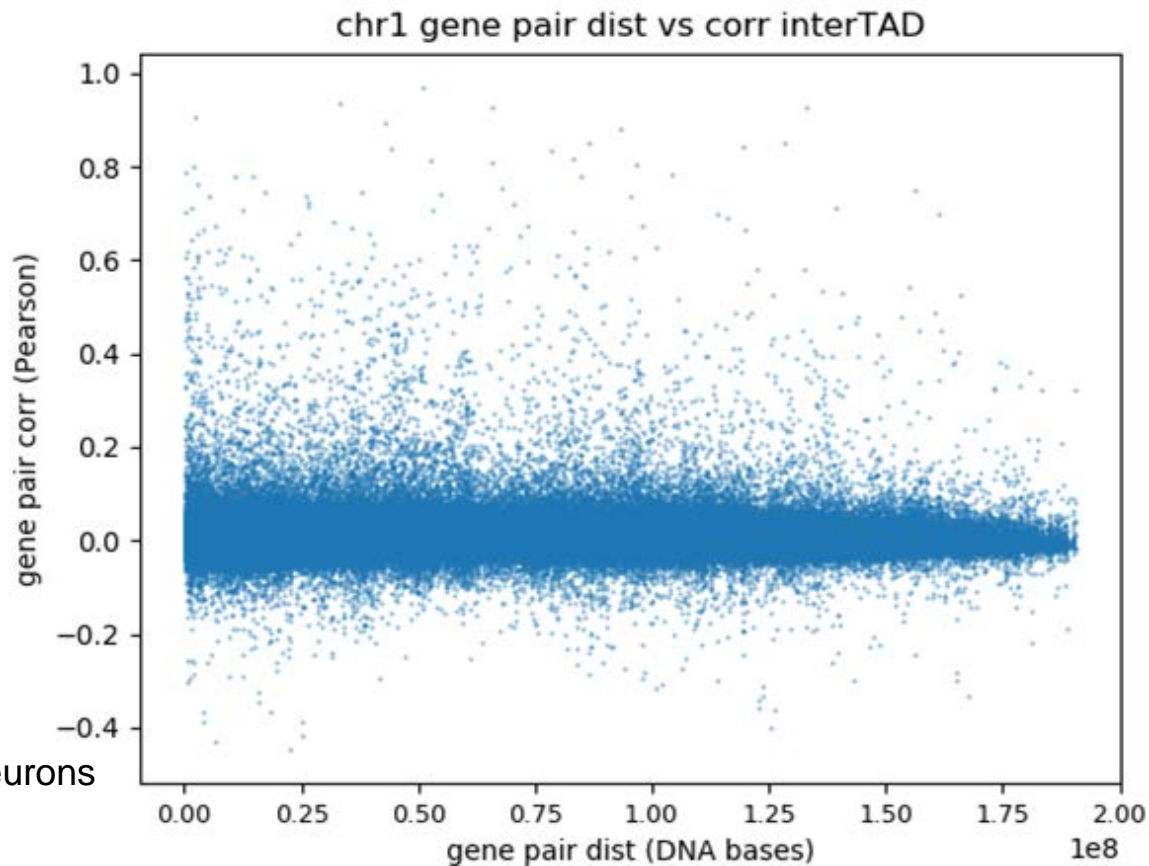intraTAD: gene pairs
within a single TAD

chr1 gene pair dist vs corr intraTAD

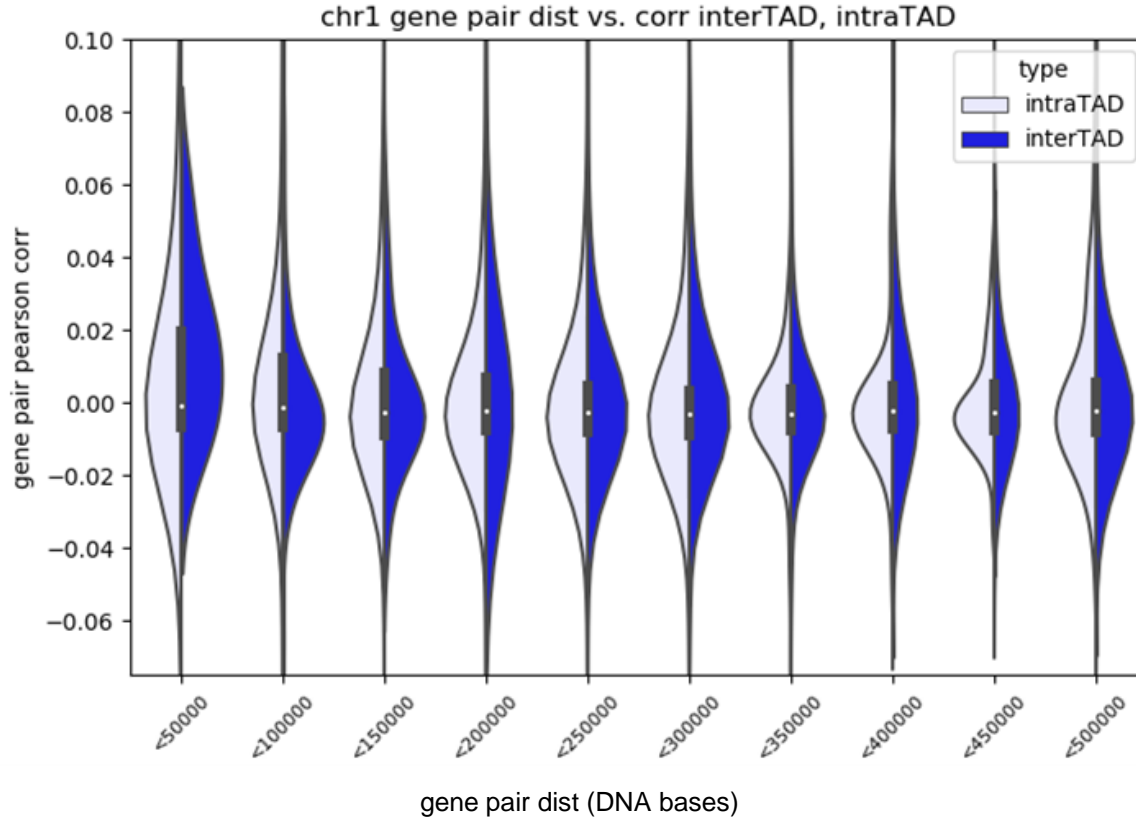TADs are of retinal neurons
from Falk et. al.

Gene pairs intra- and inter- TADs

interTAD: gene pairs
between two TADs

TADs are of retinal neurons
from Falk et. al.

Gene pairs intra- and inter- TADs with violin plot

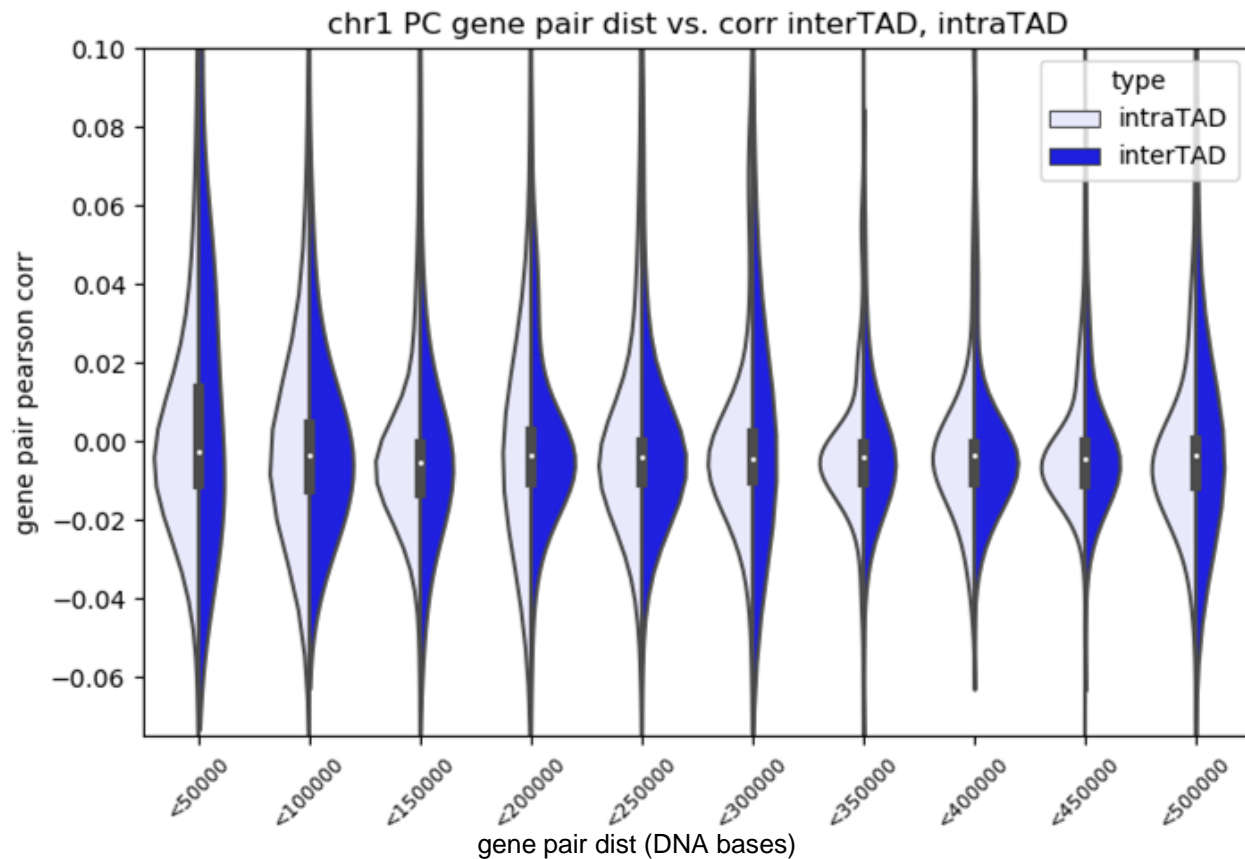chr1 gene pair dist vs. corr interTAD, intraTAD

intra-TAD on left side of each violin (light blue)

inter-TAD on right side of each violin (dark blue)

gene pair dist (DNA bases)

Gene expression correlation within a single cell type

chr1 PC gene pair dist vs. corr interTAD, intraTAD

PC cell type from He et. al.; cell type containing the highest number of cells

Gene expression correlation within a single cell type

chr1 capilEC gene pair dist vs. corr interTAD, intraTAD
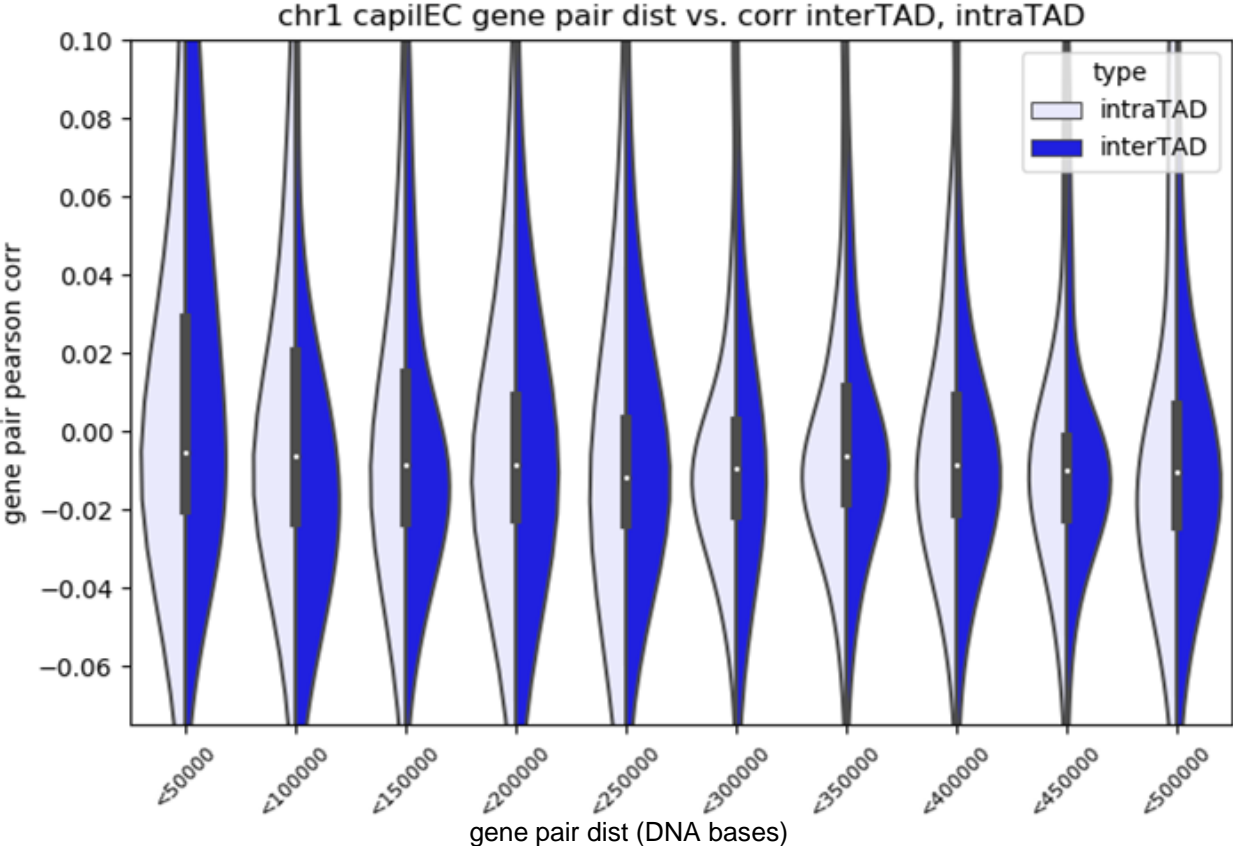
capilEC cell type from He et. al.; cell type containing the 2nd highest number of cells

What we found

No significant relationship between distance and correlation of gene pairs

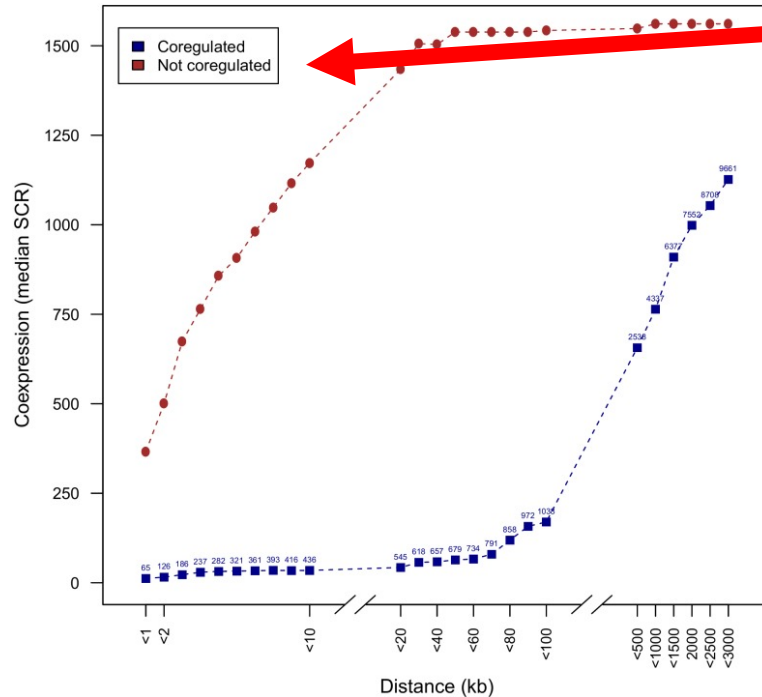Correlation is not enhanced with TADs or cell type filtering

Other people's work

- For E. Coli there is a pattern, but it is highly dependent on gene coregulation

Future Work

In the future we would like to investigate:

- Coregulation (sharing molecules that activate them)
- Enhancers (short regions of DNA that make a gene more likely to be transcribed)/insulators
- Single Cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq)
- compartments/more Hi-C data
- different ways of calculating distance between genes (3D)

Bibliography

Luecken, Malte D, and Fabian J Theis. "Current Best Practices in Single-Cell RNA-Seq Analysis: a Tutorial." Molecular Systems Biology, John Wiley & Sons, Ltd, 19 June 2019, https://www.embopress.org/doi/10.15252/msb.20188746.

Kruglyak, Semyon, and Haixu Tang. "Regulation of adjacent yeast genes." Trends in Genetics. Cell, www.cell.com/trends/genetics/comments/ S0168-9525(99)01941-1. Accessed 6 Sept. 2019.

Szabo, Quentin, et al. "Principles of Genome Folding into Topologically Associating Domains." Science Advances, American Association for the Advancement of Science, 1 Apr. 2019, https://advances.sciencemag.org/content/5/4/eaaw1668.

van Berkum, Nynke L, et al. "Hi-C: a Method to Study the Three-Dimensional Architecture of Genomes." Journal of Visualized Experiments : JoVE, MyJove Corporation, 6 May 2010, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3149993/.

Works Cited

Falk, Martin, et al. "Heterochromatin Drives Compartmentalization of Inverted and Conventional Nuclei." Nature, U.S. National Library of Medicine, 5 June 2019, https://www.ncbi.nlm.nih.gov/pubmed/31168090.

He, Liqun, et al. "Single-Cell RNA Sequencing of Mouse Brain and Lung Vascular and Vessel-Associated Cell Types." Nature News, Nature Publishing Group, 21 Aug. 2018, https://www.nature.com/articles/sdata2018160.

Pannier, Lucia, et al. "Effect of Genomic Distance on Coexpression of Coregulated Genes in E. Coli." PloS One, Public Library of Science, 18 Apr. 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5395161/.

Acknowledgments