

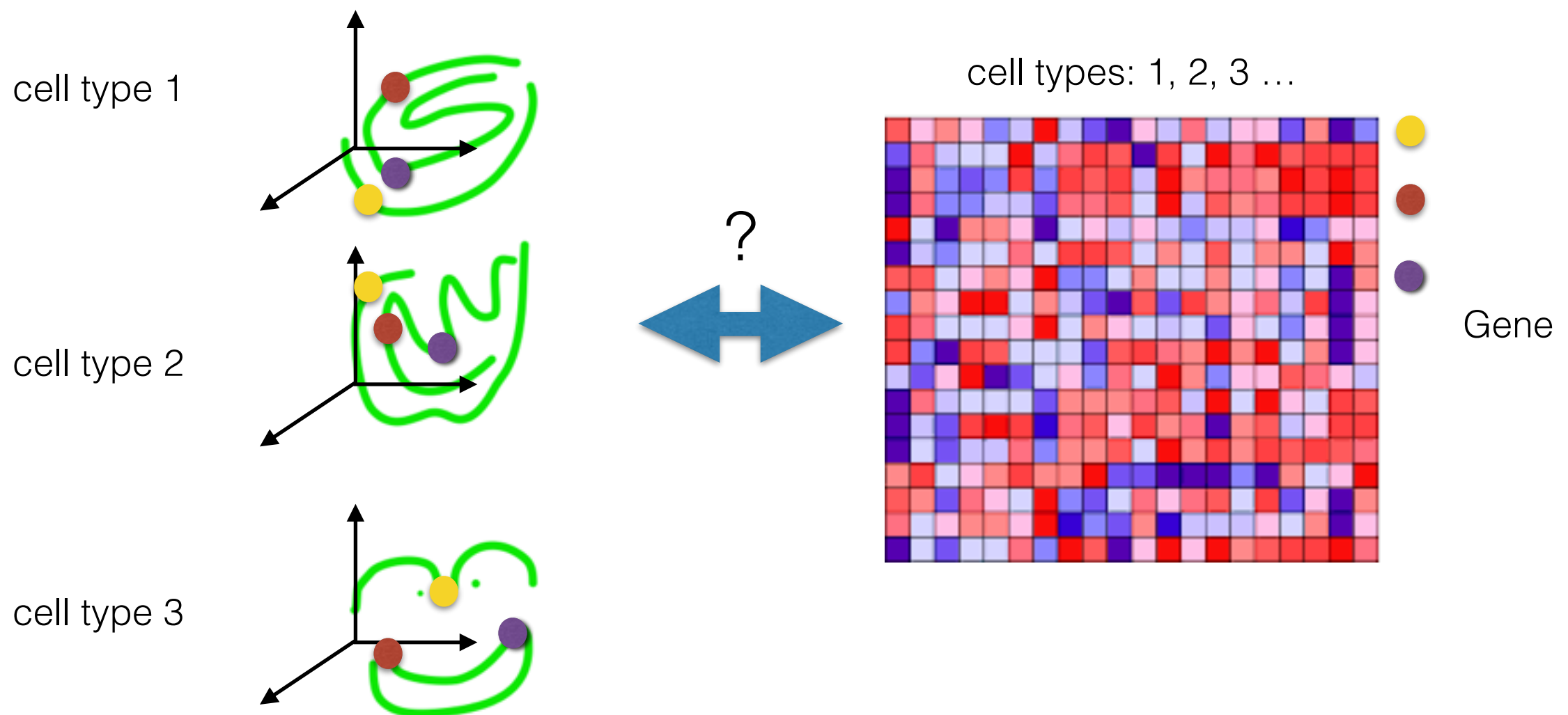
How the spatial organization of genes shapes their expression patterns, or vice versa?

KKY, Aug 2015

A mapping between 2 spaces

real physical space

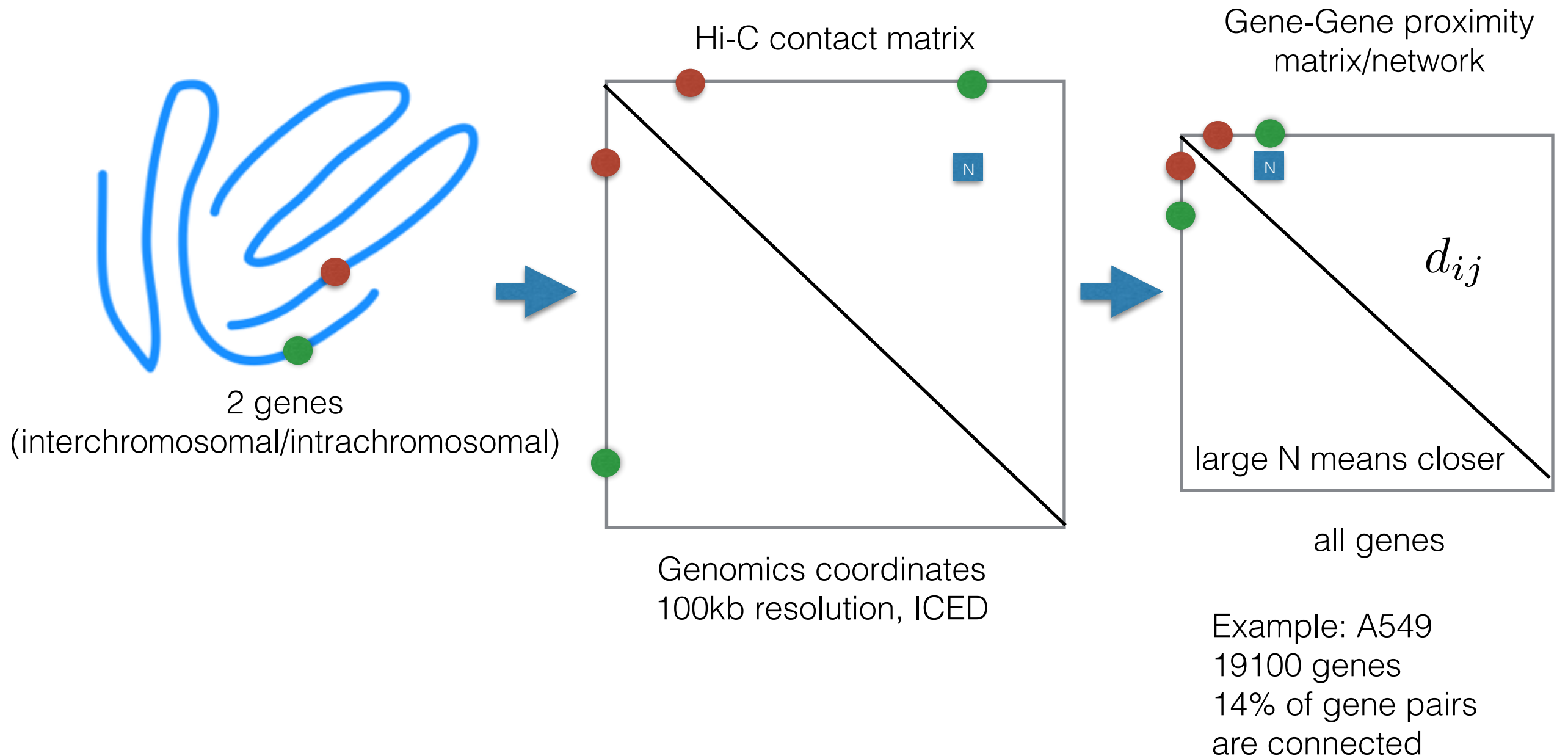
abstract expression space



ENCODE3 Hi-C data

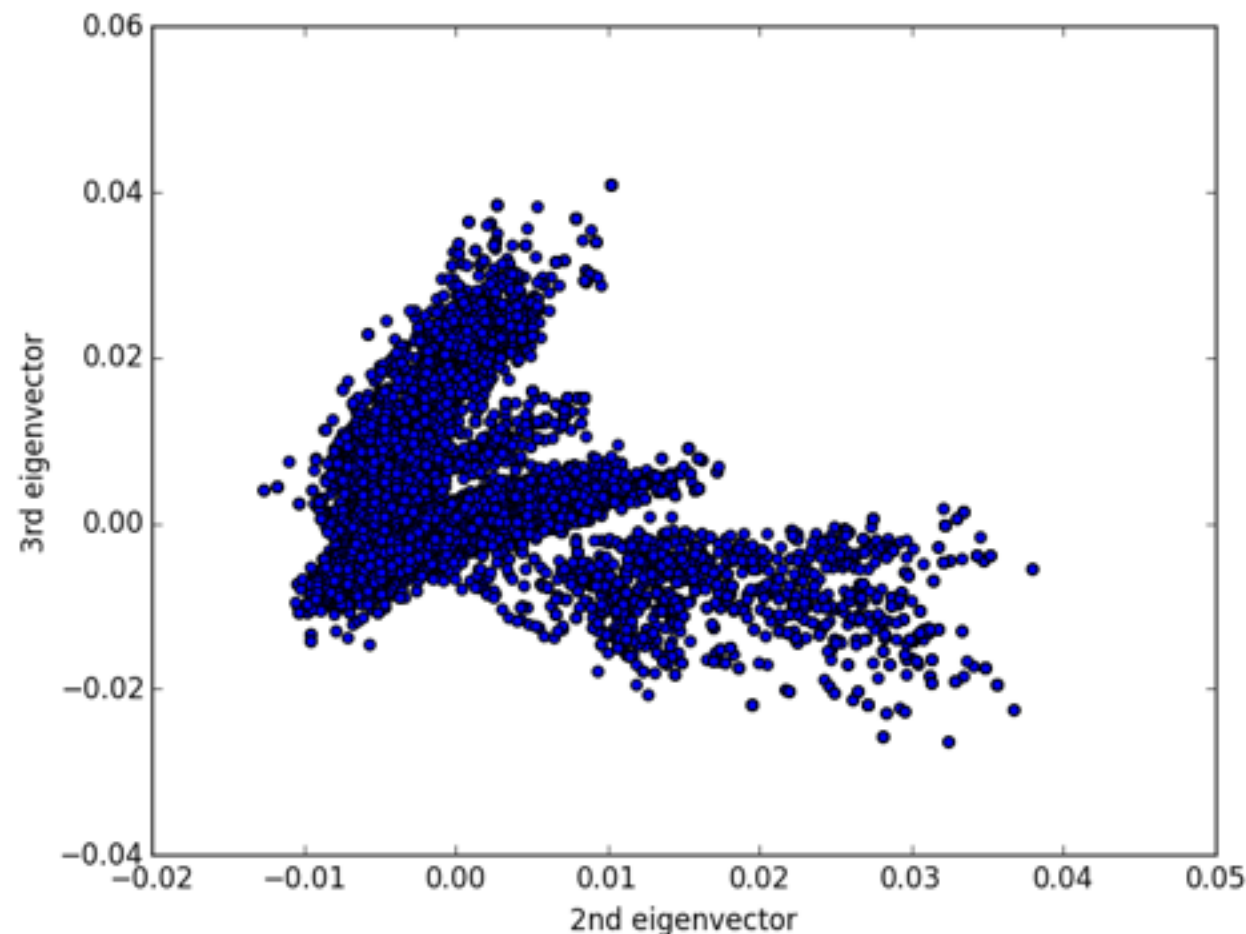
- Dekker Lab
- 12 completed cell lines: A549, Caki2, G401, LnCAP, NCI-H460, Panc1, PRMI-7951, SJCRH30, SK-MEL-5, SK-N-DZ, SK-NM-C, T470. 2 replicates per cell lines
- Multiple bin sizes: 10mb, 2.5mb, 1mb, 500kb, 250kb, 100kb, 40kb
- Raw counts and “ICED”
- In progress: HAc (AdrenocorPcal carcinoma) HA-s (Astrocytes spinal cord) HBVP (Brain vascular pericytes) DLD1 (Colon epithelial), ACHN (Kidney epithelial), HHSEC (HepaPc sinusoidal endothelial), HBMEC (BrainMicrovascularendothelial), HCMEC (Immortalized HBMEC)

Gene-Gene Proximity Network



What can we learn from the GGP network?

- A lot of topological features could be examined. Densely connected regions refer to closely localized genes - transcriptional factories?



Standard spectral clustering:
Project the network onto a few
eigenvectors of the diffusion matrix.

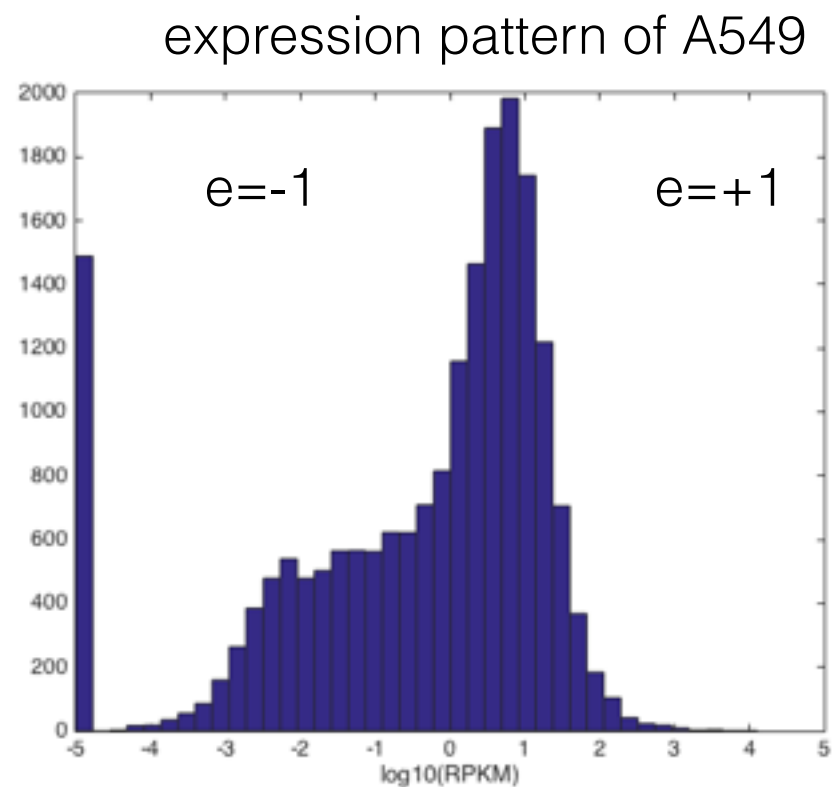
Comparison of GGP networks between 12 cell types

- We have been working on the comparison of networks:
 - Network rewiring
 - OrthoClust, multi-layers network clustering
 - Compare regulatory networks of worm, fly, human
 - BrainSpan, co-expression networks in different parts of the brain
 - Tissue specific PPI networks
- Gene-Gene proximity, conserved? specific?
 - what's the proper distance metric?
 - a favor of structural alignment

Gene-Gene proximity versus gene-gene expression

- To what extent spatially localized genes tend to be expressed or not expressed together?
- RNA-Seq data for matched cell lines: [A549](#), [Caki2](#), [G401](#), [LNCaP](#), [NCI-H460](#), [Panc1](#), [PRMI-7951](#), [SJCRH30](#), [SK-MEL-5](#), [SK-N-DZ](#), [SK-NM-C](#), [T470](#)

Gene-Gene proximity versus Gene-Gene expression

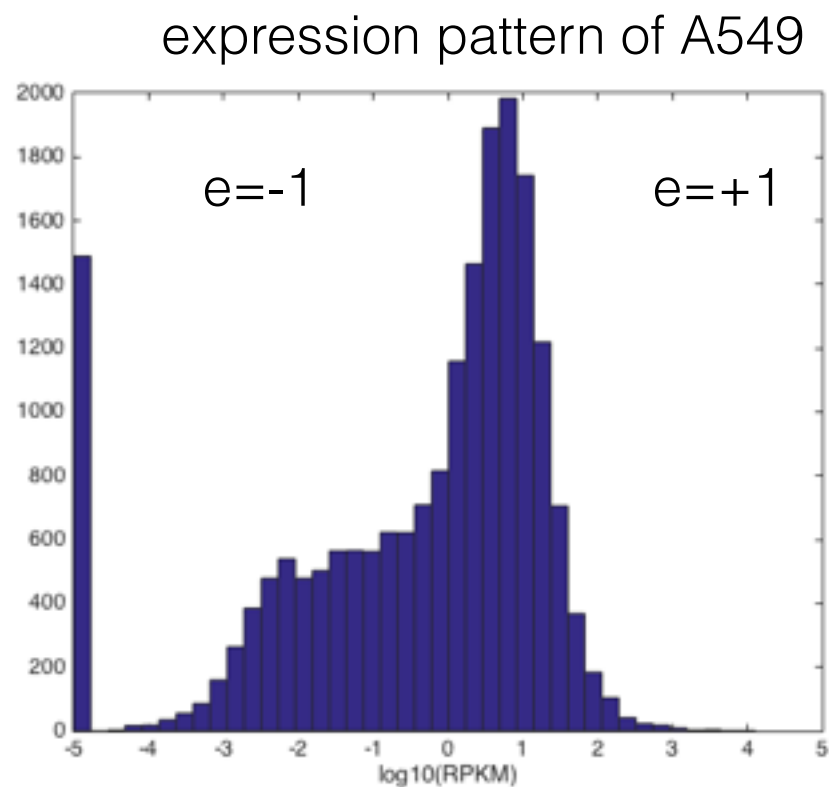


$$H = - \sum_{ij} d_{ij} e_i e_j$$

a low energy state means co-expressed genes are co-localized

Graph partition problem

Gene-Gene proximity versus Gene-Gene expression

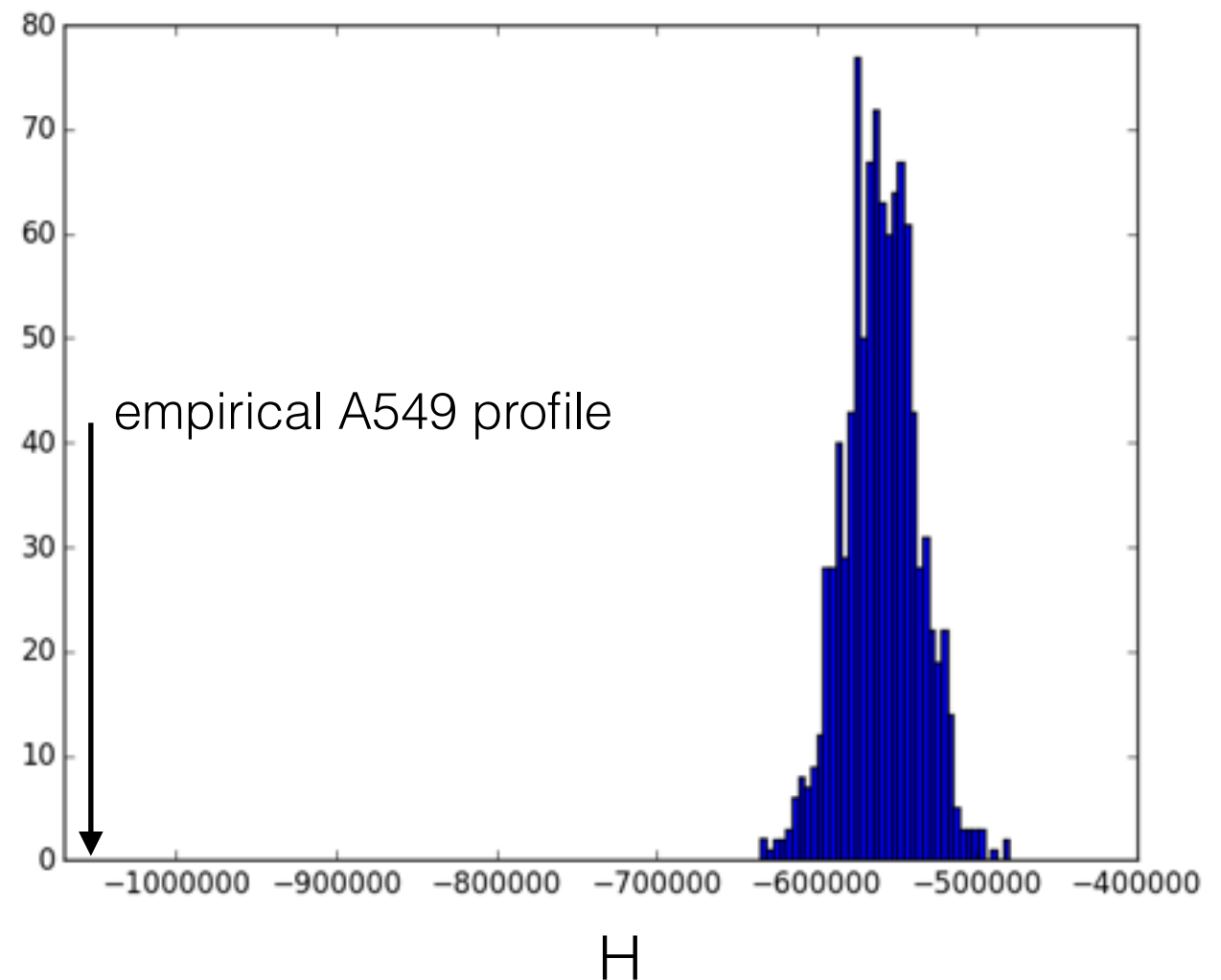


$$H = - \sum_{ij} d_{ij} e_i e_j$$

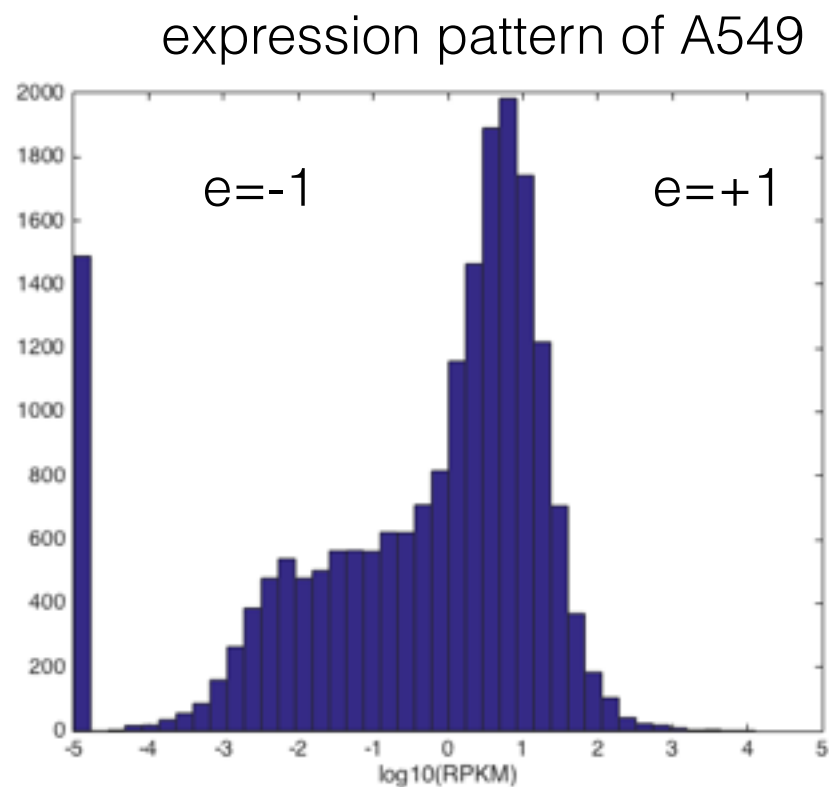
a low energy state means co-expressed genes are co-localized

Graph partition problem

Distribution of H by shuffling the expression profile of A549



Gene-Gene proximity versus Gene-Gene expression

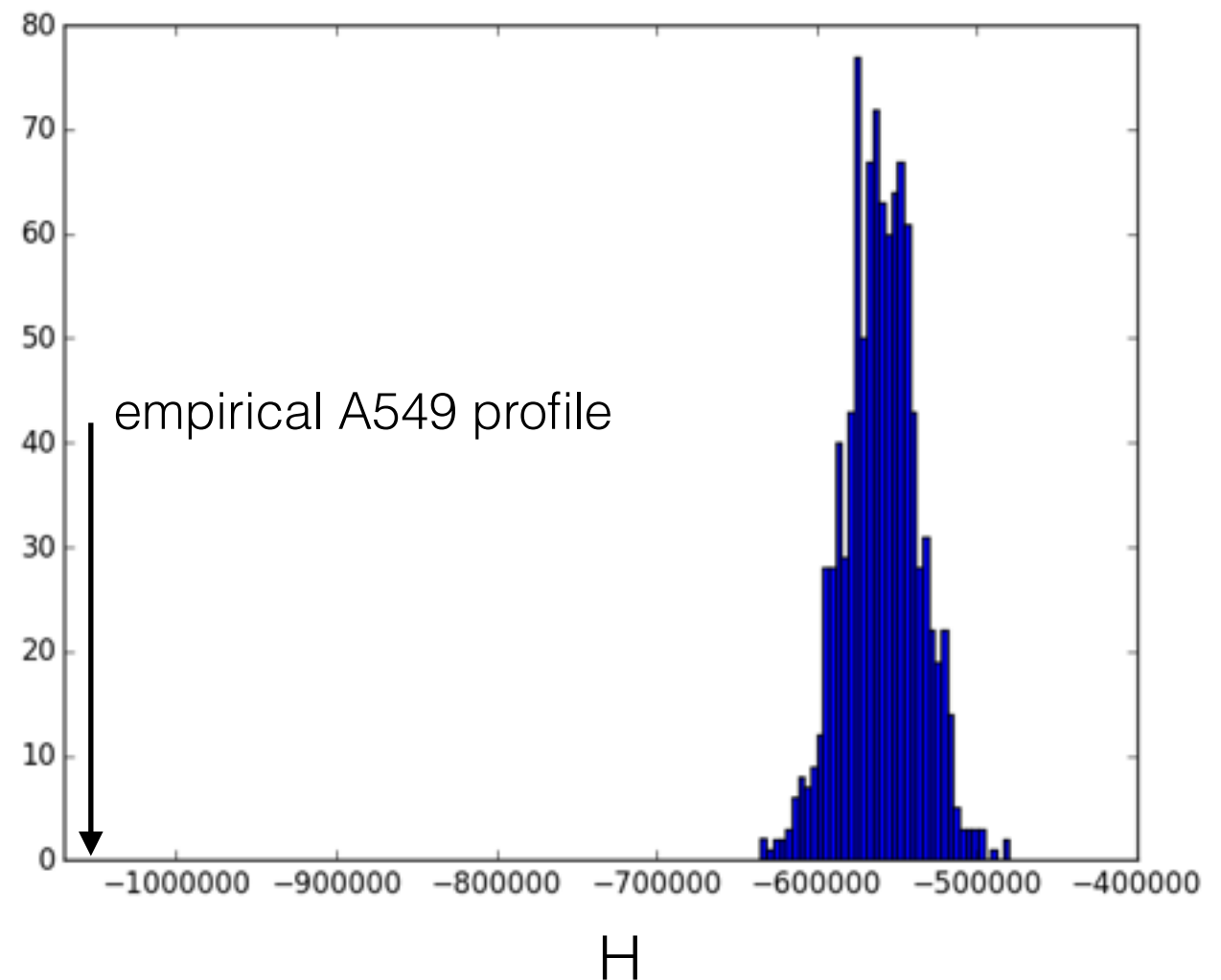


$$H = - \sum_{ij} d_{ij} e_i e_j$$

a low energy state means co-expressed genes are co-localized

Graph partition problem

Distribution of H by shuffling the expression profile of A549



Too naive null model.
Better alternative by incorporating
other genomics features?

Gene-Gene proximity versus Gene-Gene expression

- A lot more could be explored with the expression profiles of other cell types.
- Given a pair of genes, $\delta r \propto \delta e$?
- Given the configuration for cell type A, what's the energy if one plugins the expression profile of cell type B?
- For a particular expression profile, how optimal is the current spatial configuration?