

# Normalization of HiC data

Chromosomal Conformation Course 2017

CRG, Barcelona

[enrique.vidal@crg.eu](mailto:enrique.vidal@crg.eu)

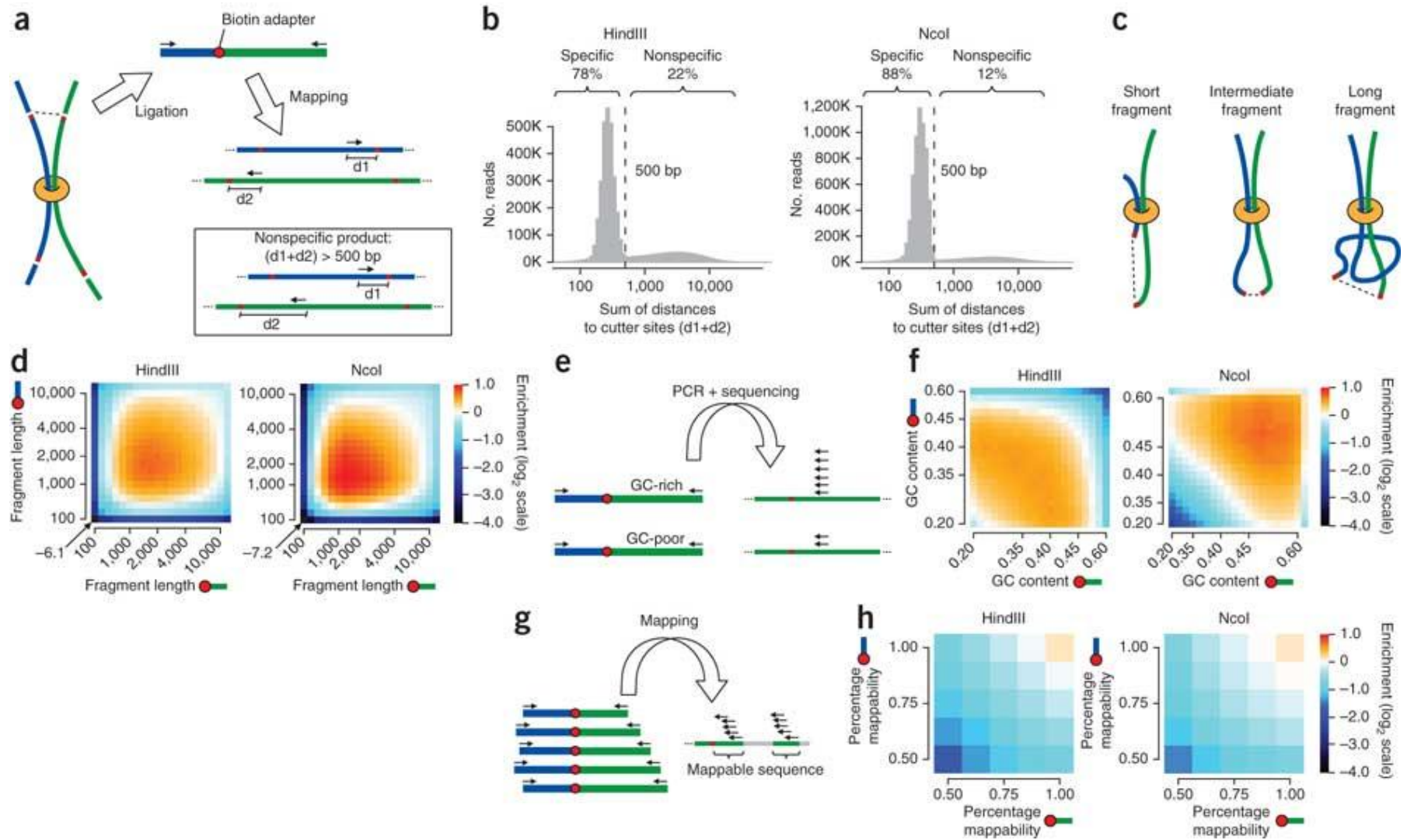
# HiC bias removal

Chromosomal Conformation Course 2017

CRG, Barcelona

[enrique.vidal@crg.eu](mailto:enrique.vidal@crg.eu)

# Intro



Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture.

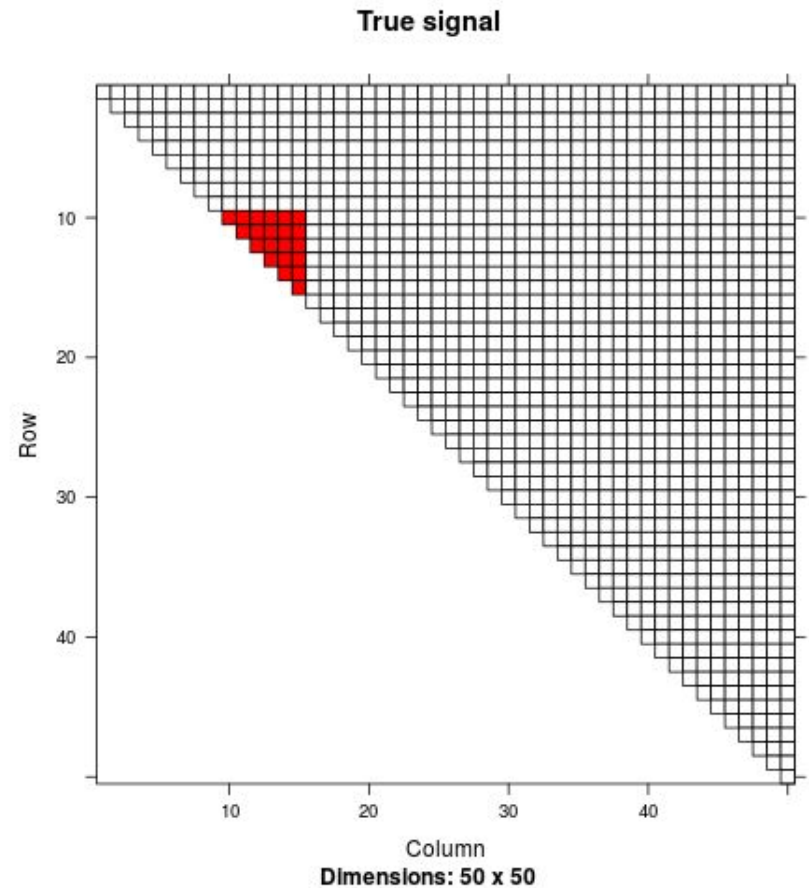
Nat Genet. 2011 Oct 16;43(11):1059-65. doi: 10.1038/ng.947. PubMed PMID: 22001755.

# A toy simulation

## Base

Ultra-simplistic example

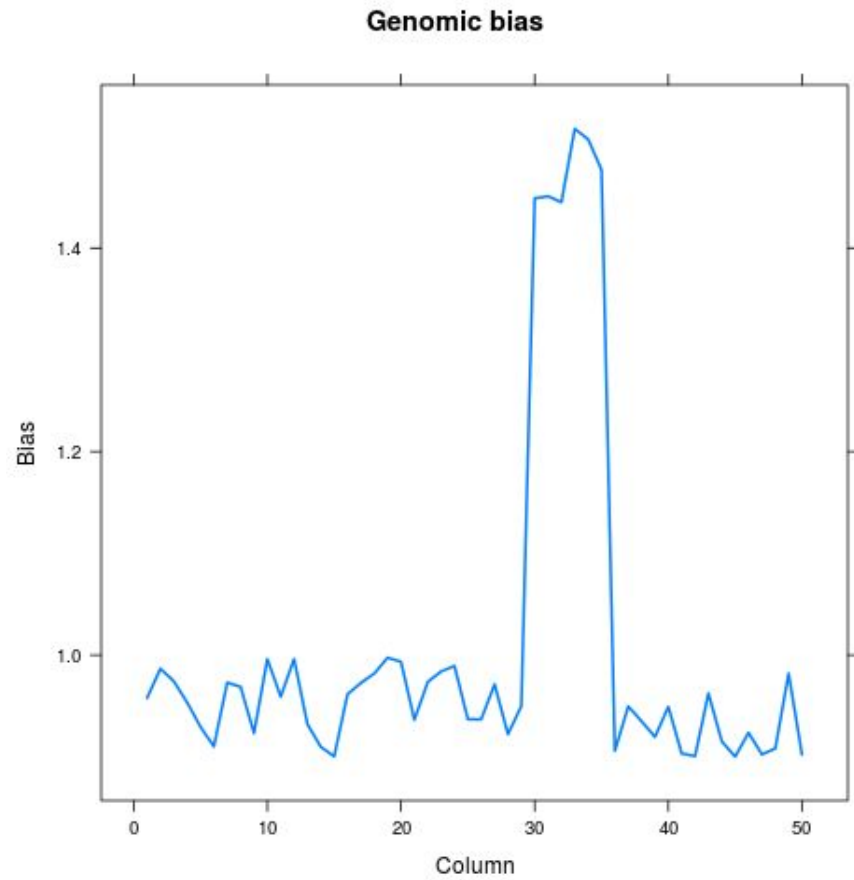
One domain of interaction



# A toy simulation

## Bias

Genomic bias  
(different location)

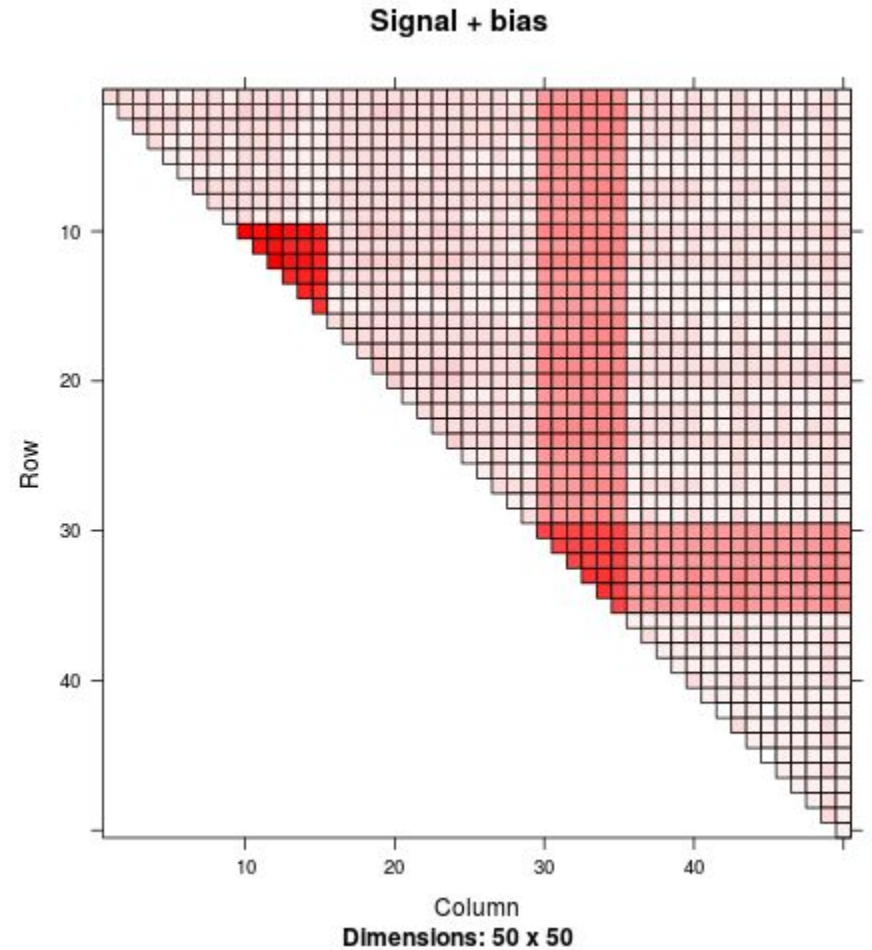


# A toy simulation

**Include bias**

Fake domain

Fake interaction bands

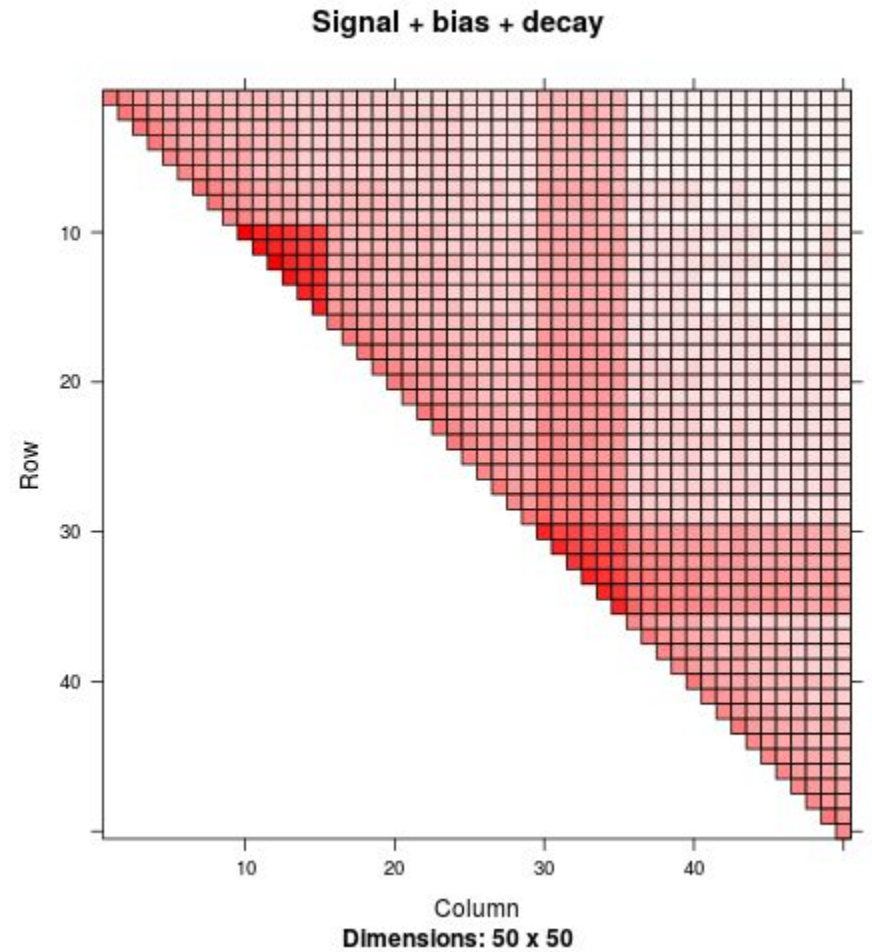


# A toy simulation

**Include decay**

$\log(n) \sim 1 / \log(\text{distance})$

Fading at long distances

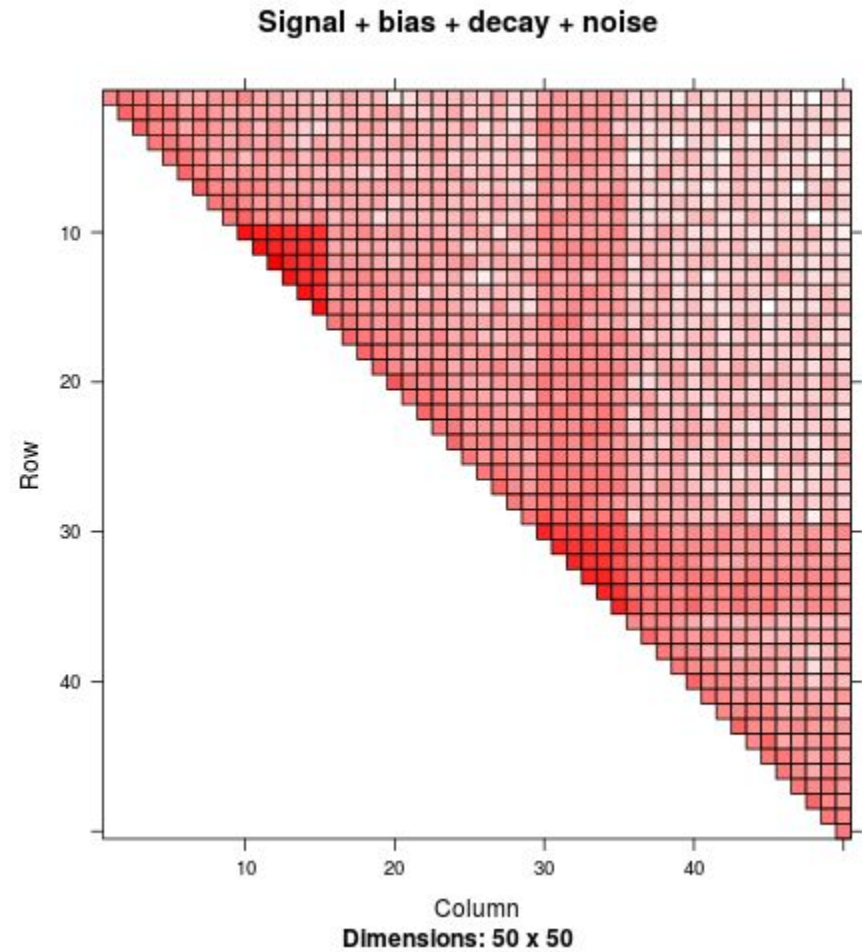




# A toy simulation

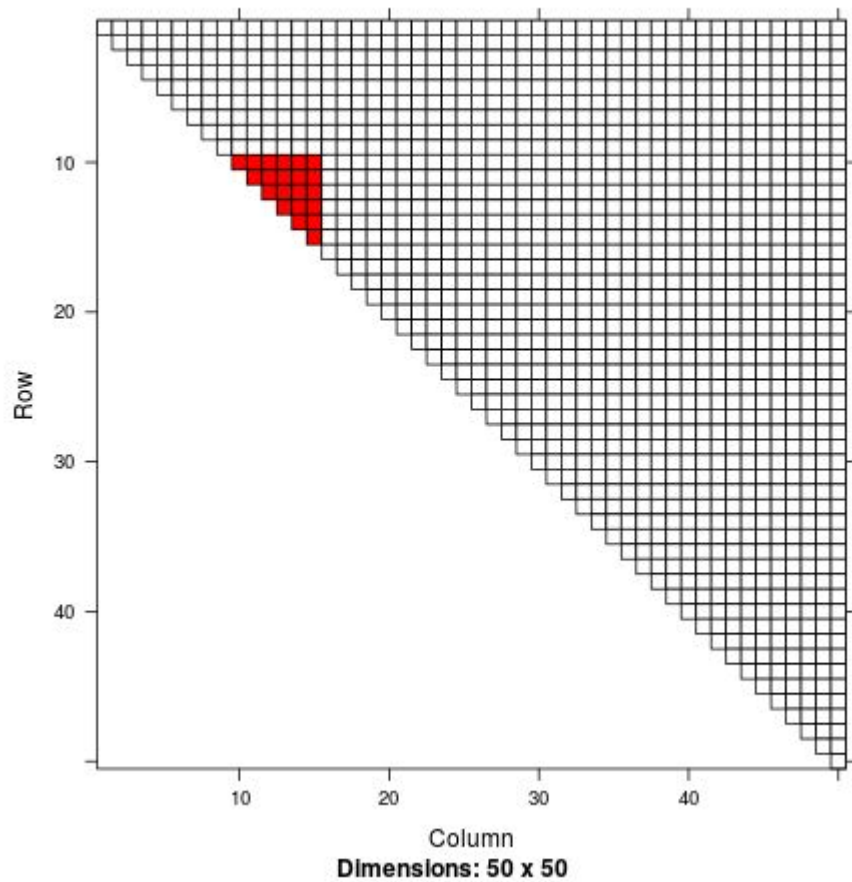
Include noise

Poisson

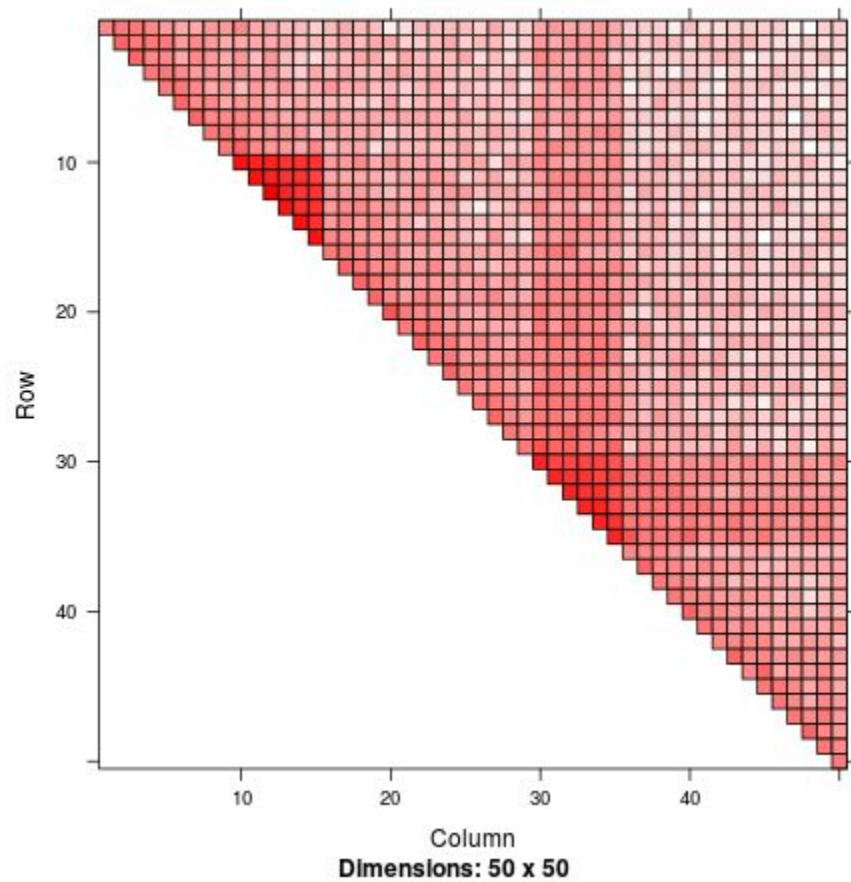


# A toy simulation

**True signal**



**Signal + bias + decay + noise**



Usual  
approaches

# Usual approaches

LGF

## BIOINFORMATICS APPLICATIONS NOTE

Vol. 28 no. 23 2012, pages 3131–3133  
doi:10.1093/bioinformatics/bts570

Genome analysis

Advance Access publication September 27, 2012

### HiCNorm: removing biases in Hi-C data via Poisson regression

Ming Hu<sup>1</sup>, Ke Deng<sup>1</sup>, Siddarth Selvaraj<sup>2,3</sup>, Zhaohui Qin<sup>4</sup>, Bing Ren<sup>2</sup> and Jun S. Liu<sup>1,\*</sup>

<sup>1</sup>Department of Statistics, Harvard University, Cambridge, MA 02138, USA, <sup>2</sup>Department of Cellular and Molecular Medicine, UCSD School of Medicine, La Jolla, CA 92093, USA, <sup>3</sup>Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA and <sup>4</sup>Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322 USA

## ARTICLES

### Iterative correction of Hi-C data reveals hallmarks of chromosome organization

Maxim Imakaev<sup>1,3</sup>, Geoffrey Fudenberg<sup>1,3</sup>, Rachel Patton McCord<sup>3</sup>, Natalia Naumova<sup>3</sup>, Anton Goloborodko<sup>1</sup>, Bryan R Lajoie<sup>3</sup>, Job Dekker<sup>3</sup> & Leonid A Mirny<sup>1,2,4</sup>

ICE

# Usual approaches

## LGF

Regression on all cells

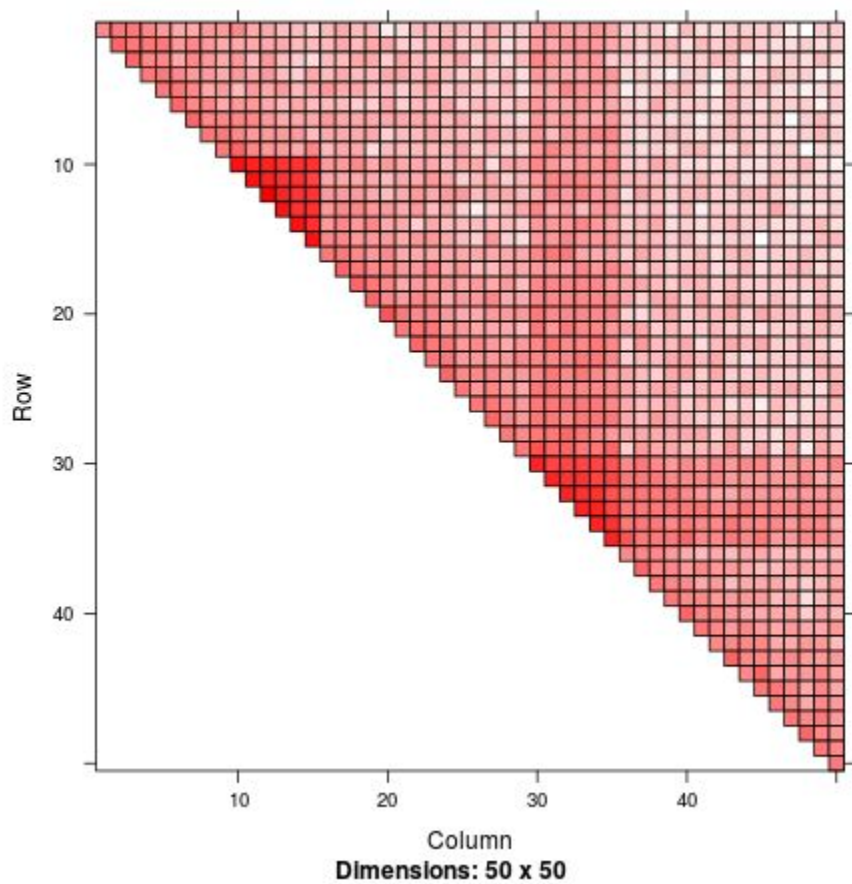
$$x_{ij} \sim NB(map_i map_j \lambda_{ij}, \theta)$$

$$\log(\lambda_{ij}) = \beta_{cg} (cg_i cg_j) + \beta_{res} (res_i res_j)$$

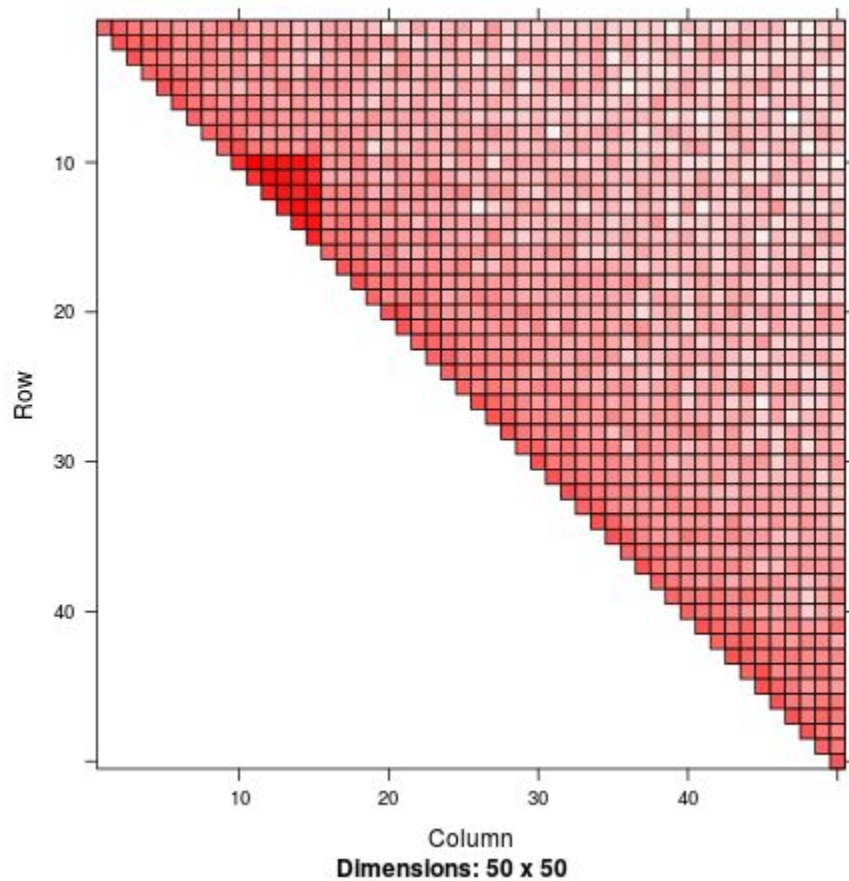
$$n_{ij} = \frac{x_{ij}}{map_i map_j \lambda_{ij}}$$

# Usual approaches

**Signal + bias + decay + noise**



**LGF**



# Usual approaches

## ICE

Matrix balancing

Equal visibility

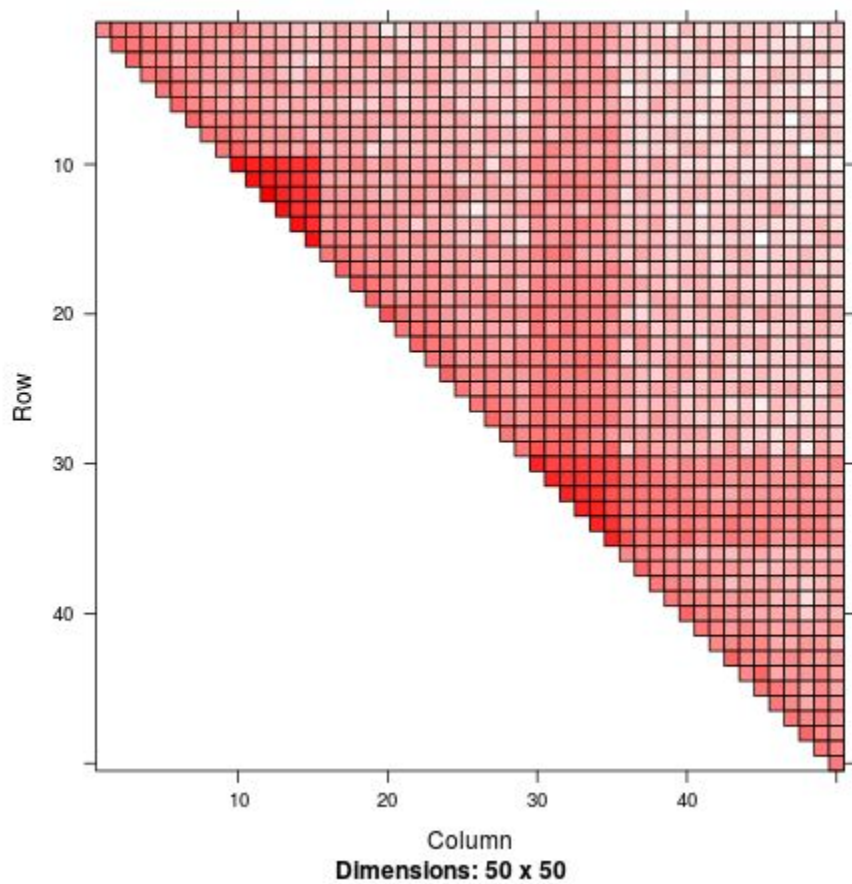
$$s_i^n = \frac{\sum_j x_{ij}^n}{\frac{\sum_i \sum_j x_{ij}^n}{N}}$$

$$x_{ij}^{n+1} = \frac{x_{ij}^n}{s_i^n s_j^n}$$

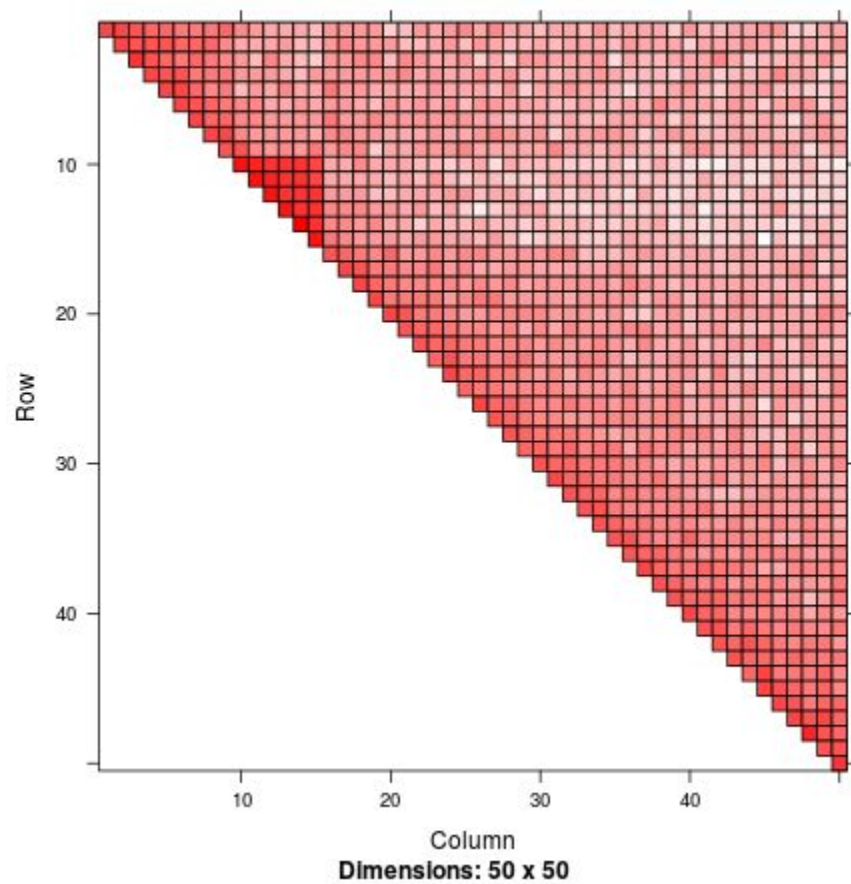


# Usual approaches

**Signal + bias + decay + noise**



**ICE**





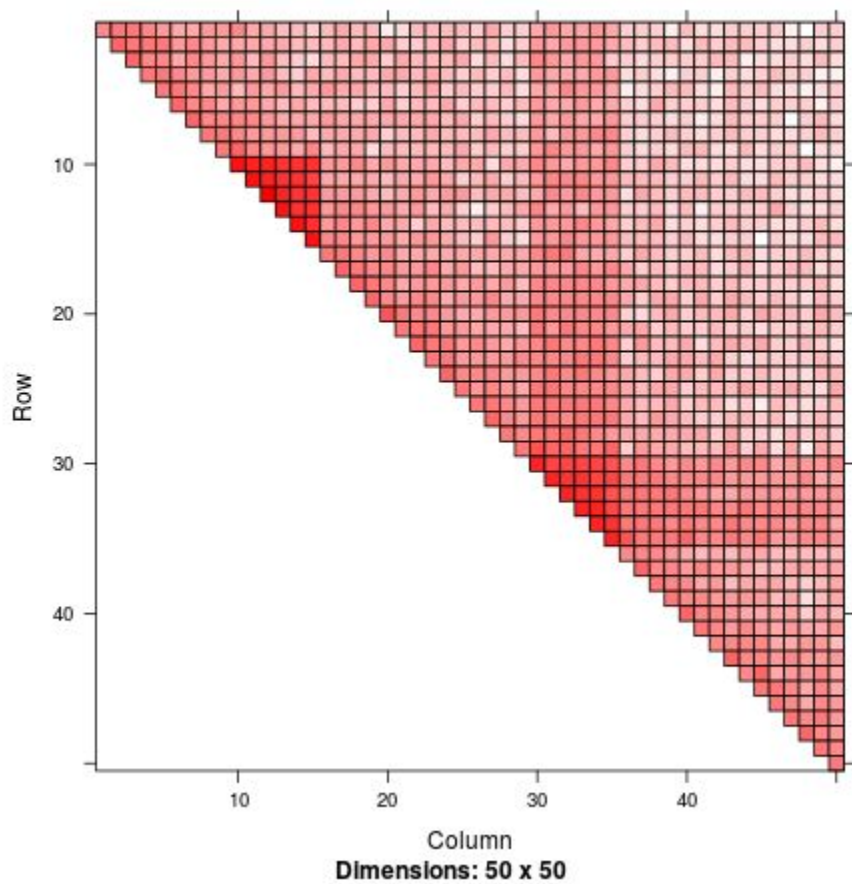
# Usual approaches

**Vanilla**

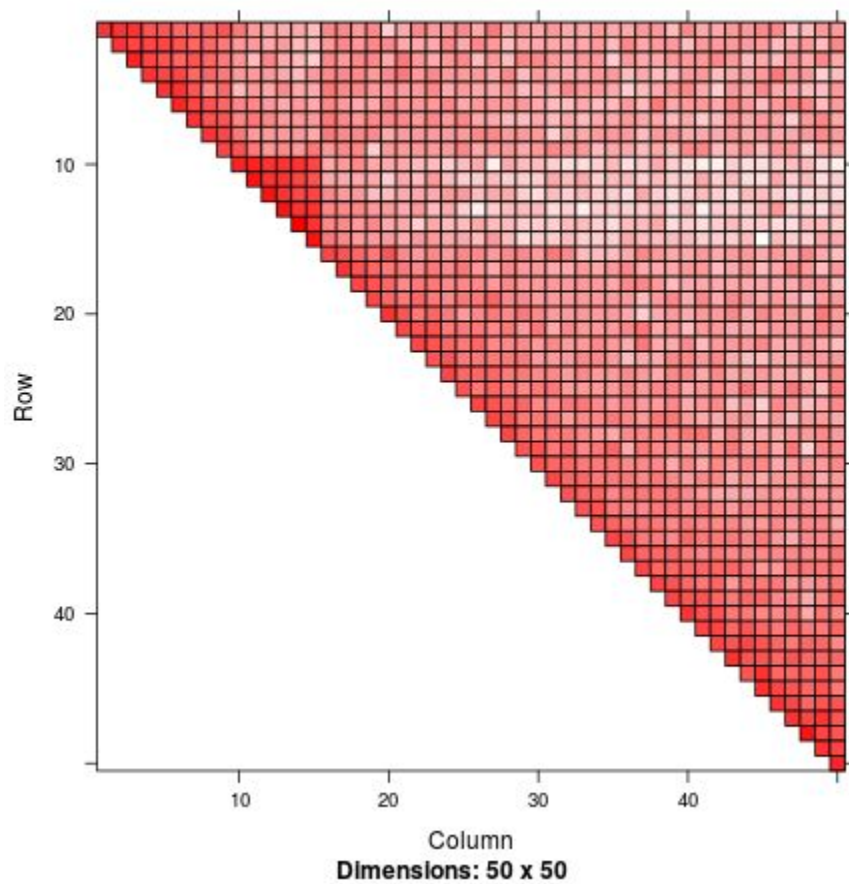
One ICE iteration

# Usual approaches

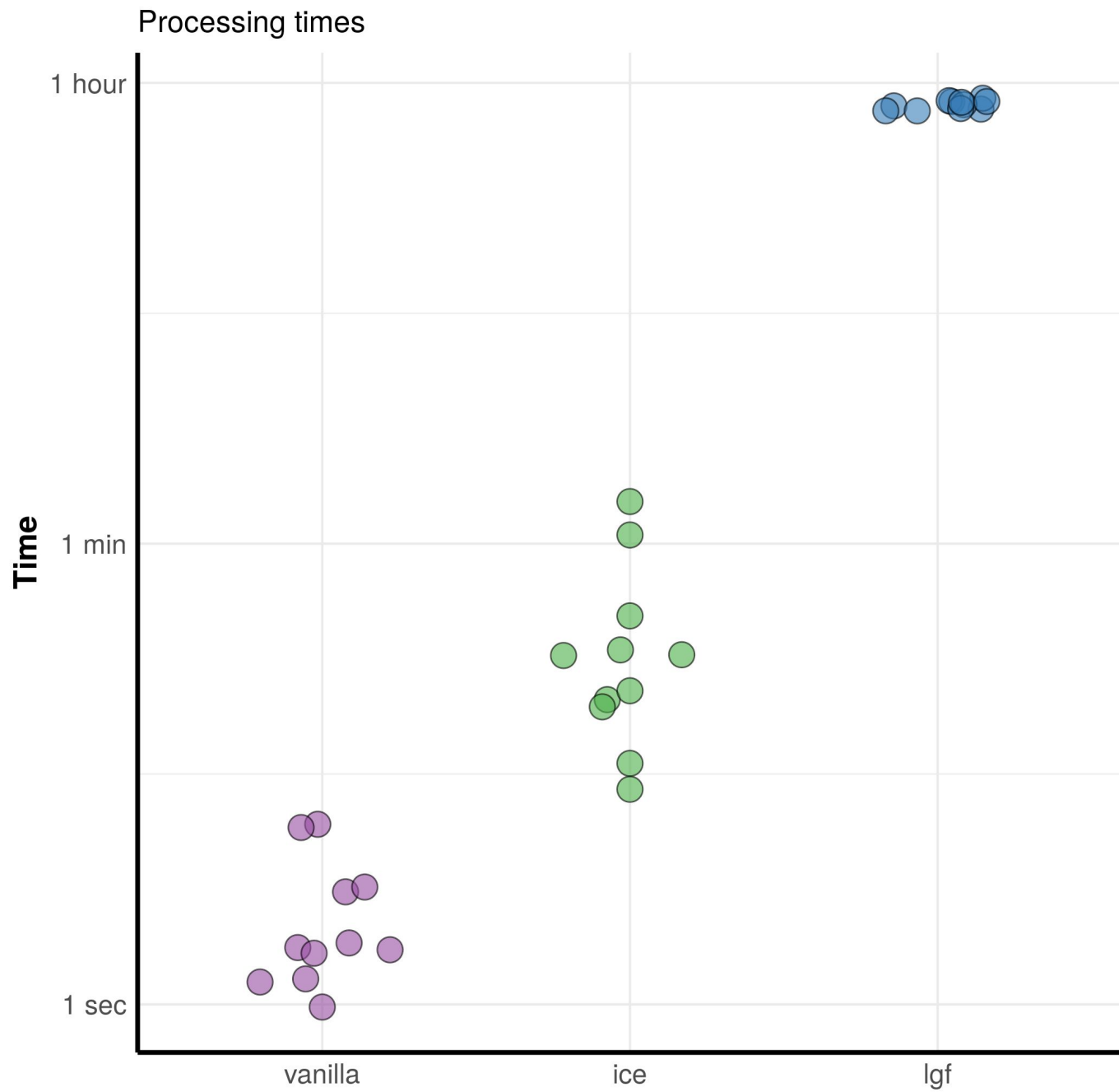
**Signal + bias + decay + noise**



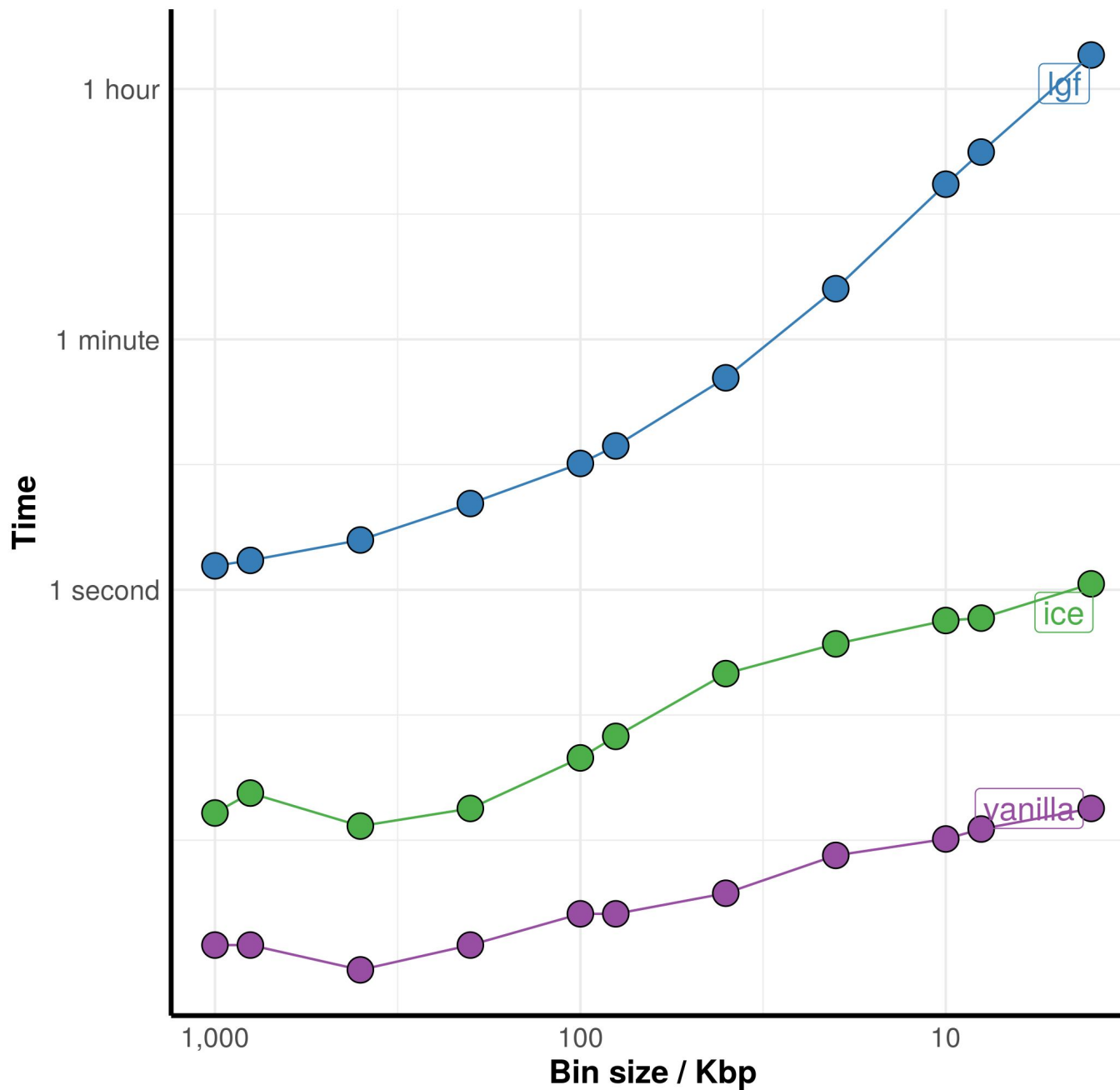
**Vanilla**



# Timing

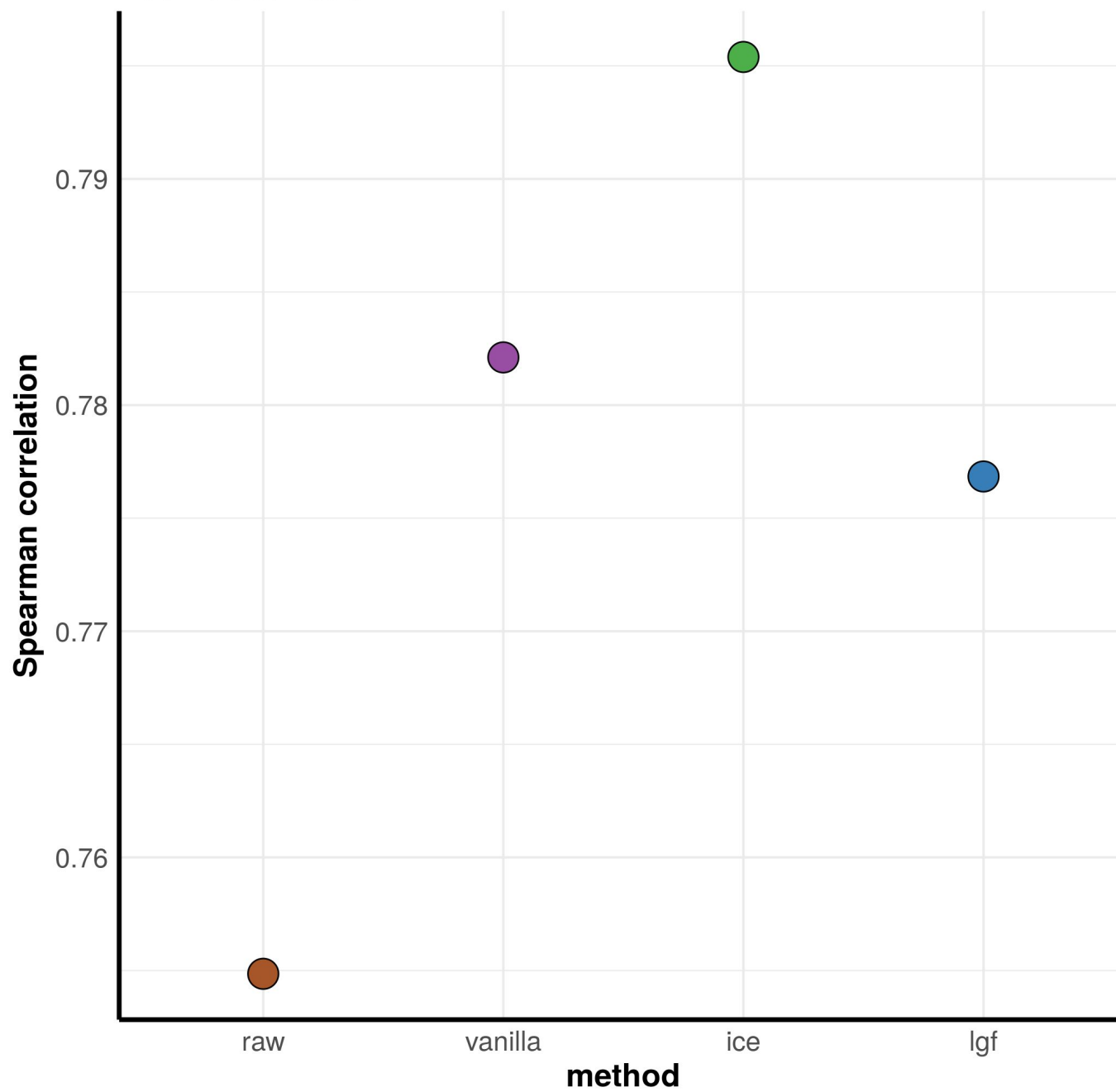


Computing time required to normalize chr22

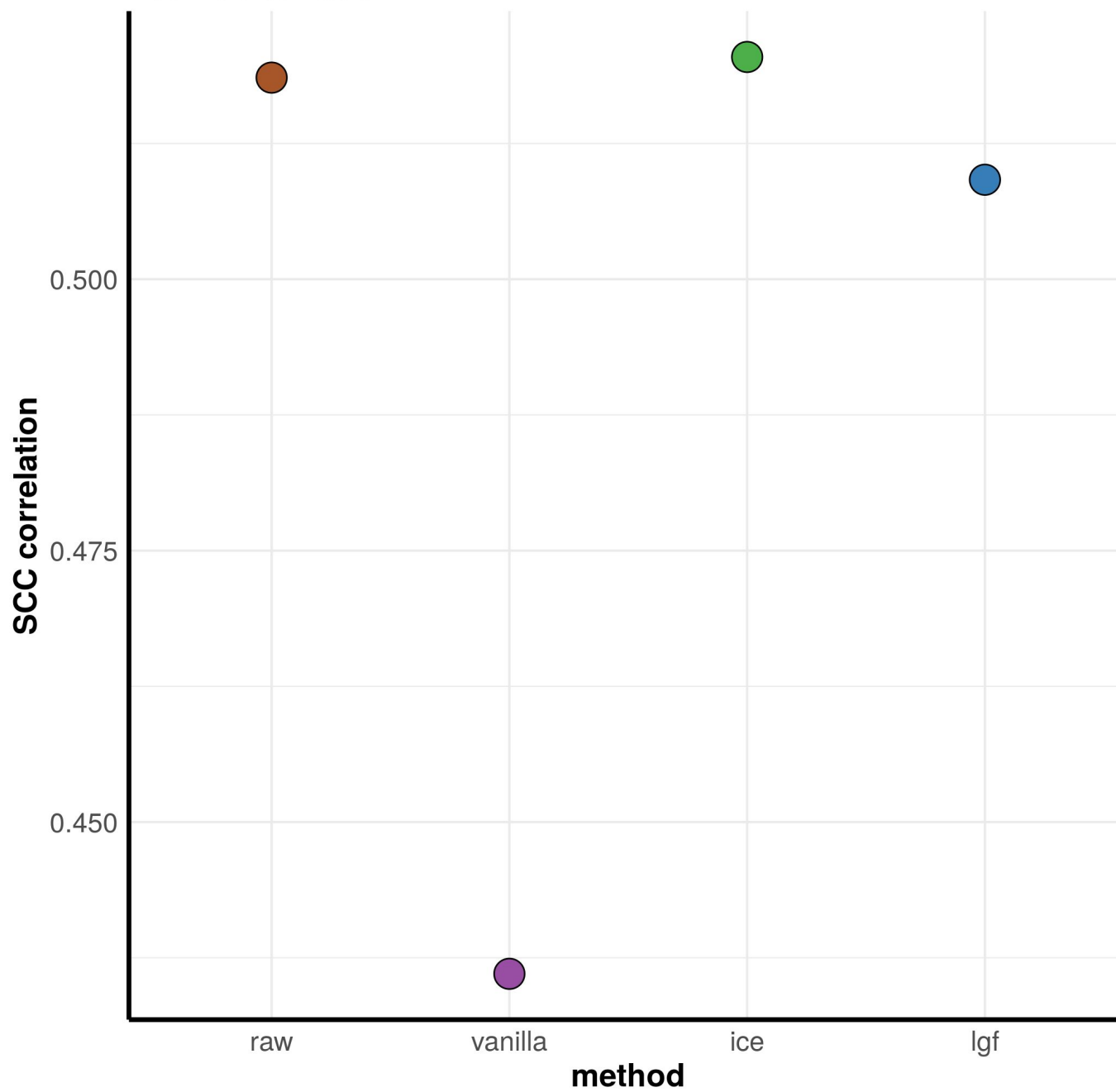


# Performance

K562: cis contacts



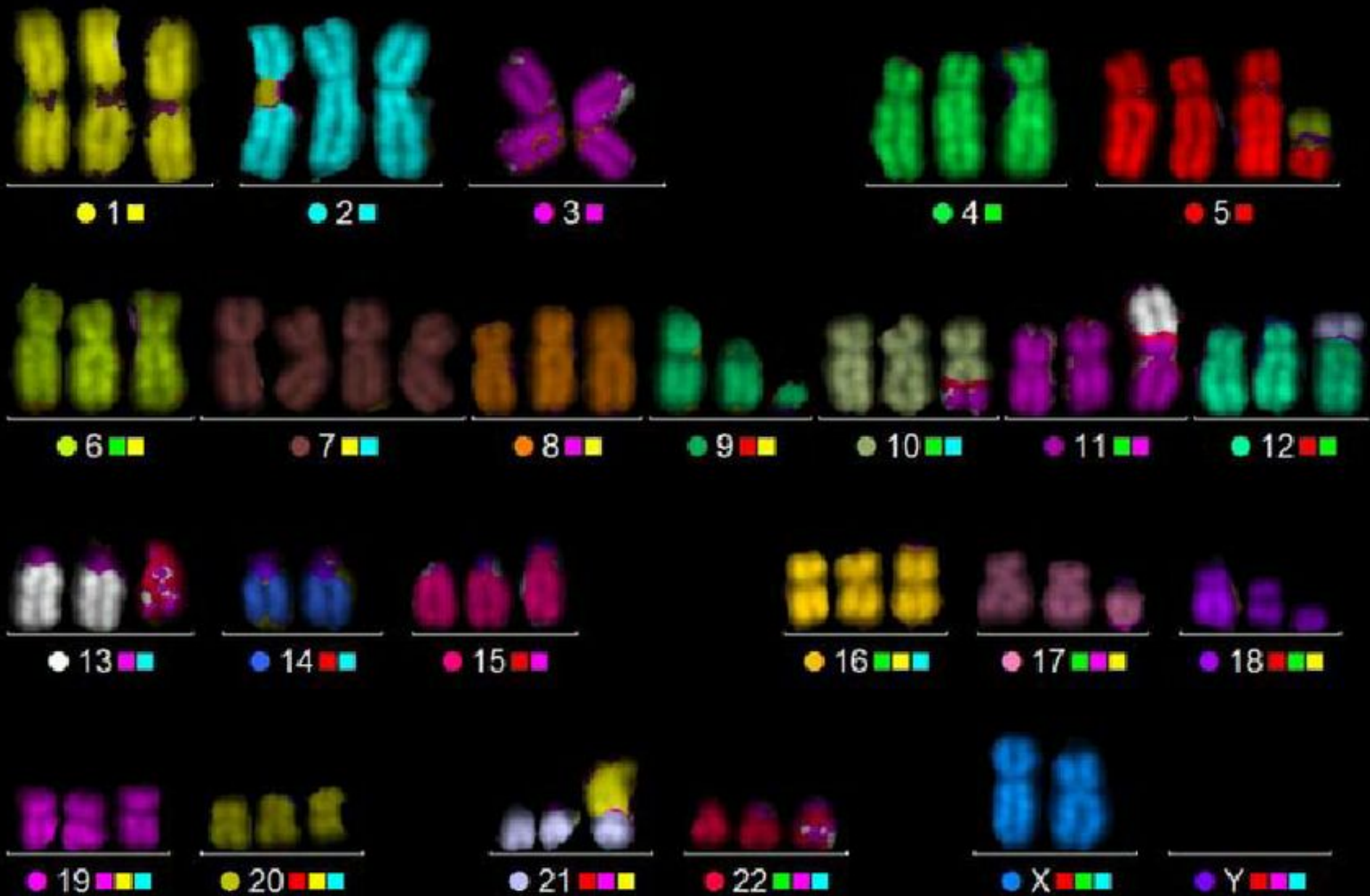
K562: cis contacts



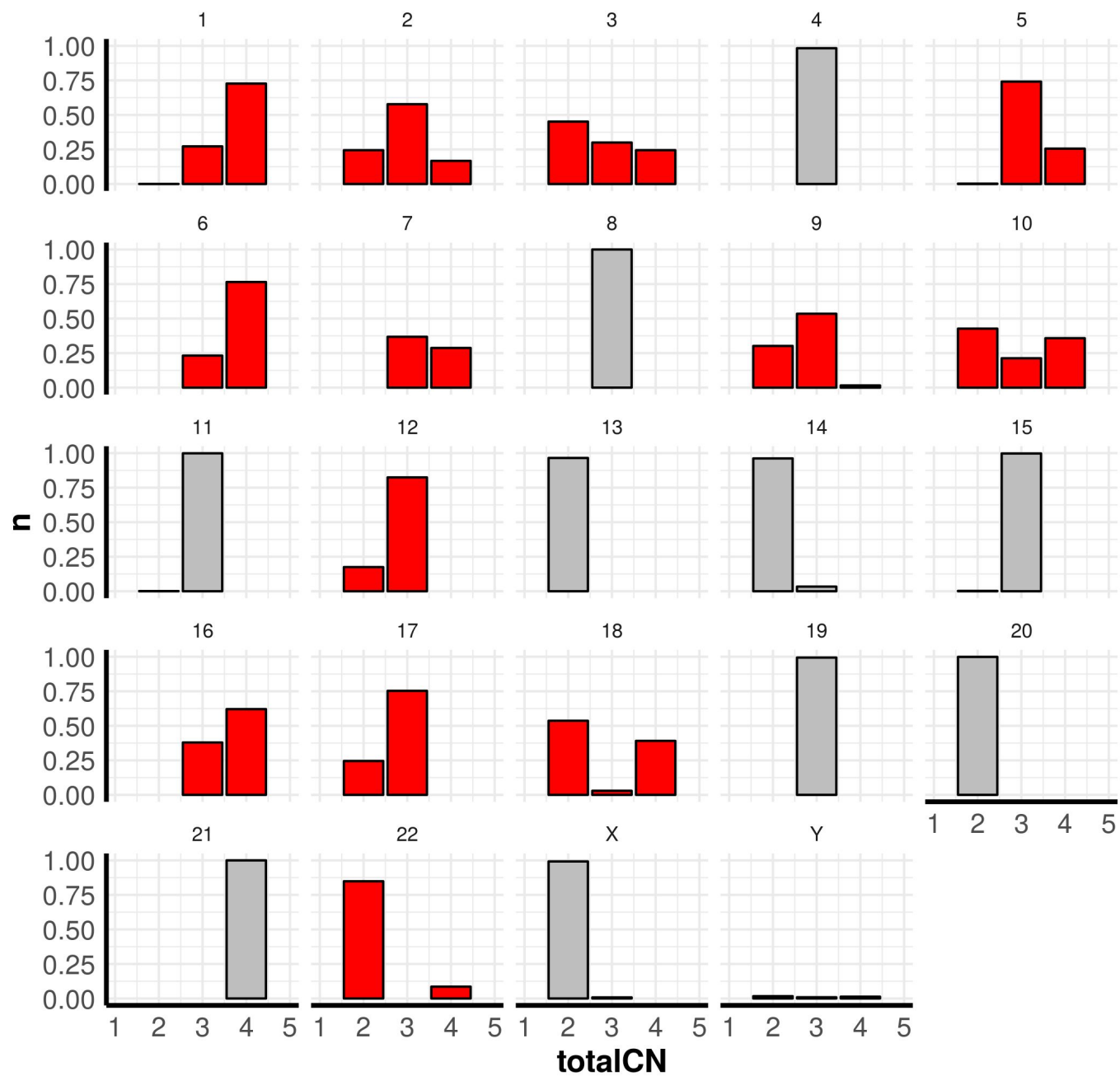


# Aberrant karyotypes

# K562



k562



New proposal

# New proposal

**oneD**

Regression on totals

$$t_i = \sum_j x_{ij}$$

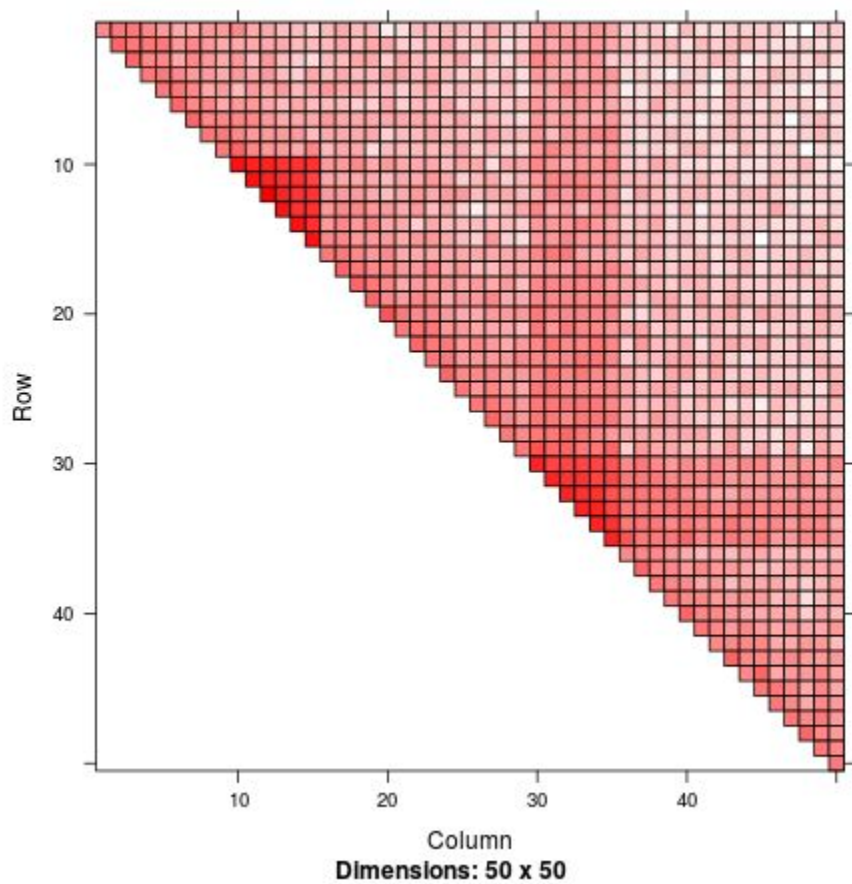
$$t_i \sim NB(\lambda_i, \theta)$$

$$\log(\lambda_i) = s(\text{map}_i) + s(\text{cg}_i) + s(\text{res}_i)$$

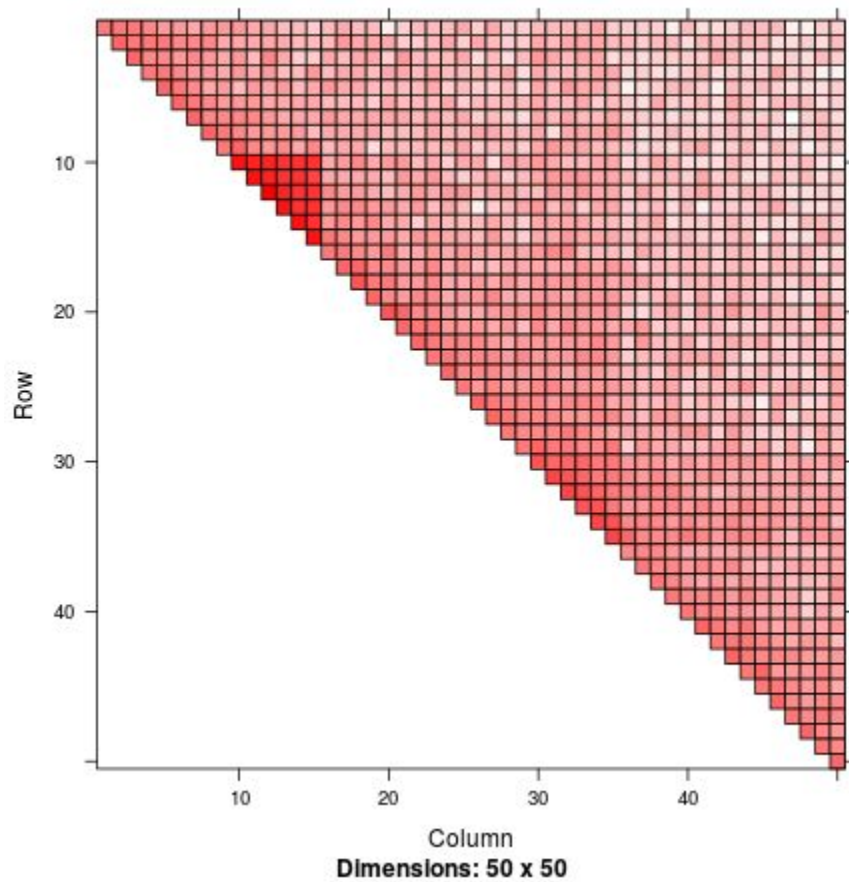
$$n_{ij} = \frac{x_{ij}}{\sqrt{\lambda_i \lambda_j}}$$

# New proposal

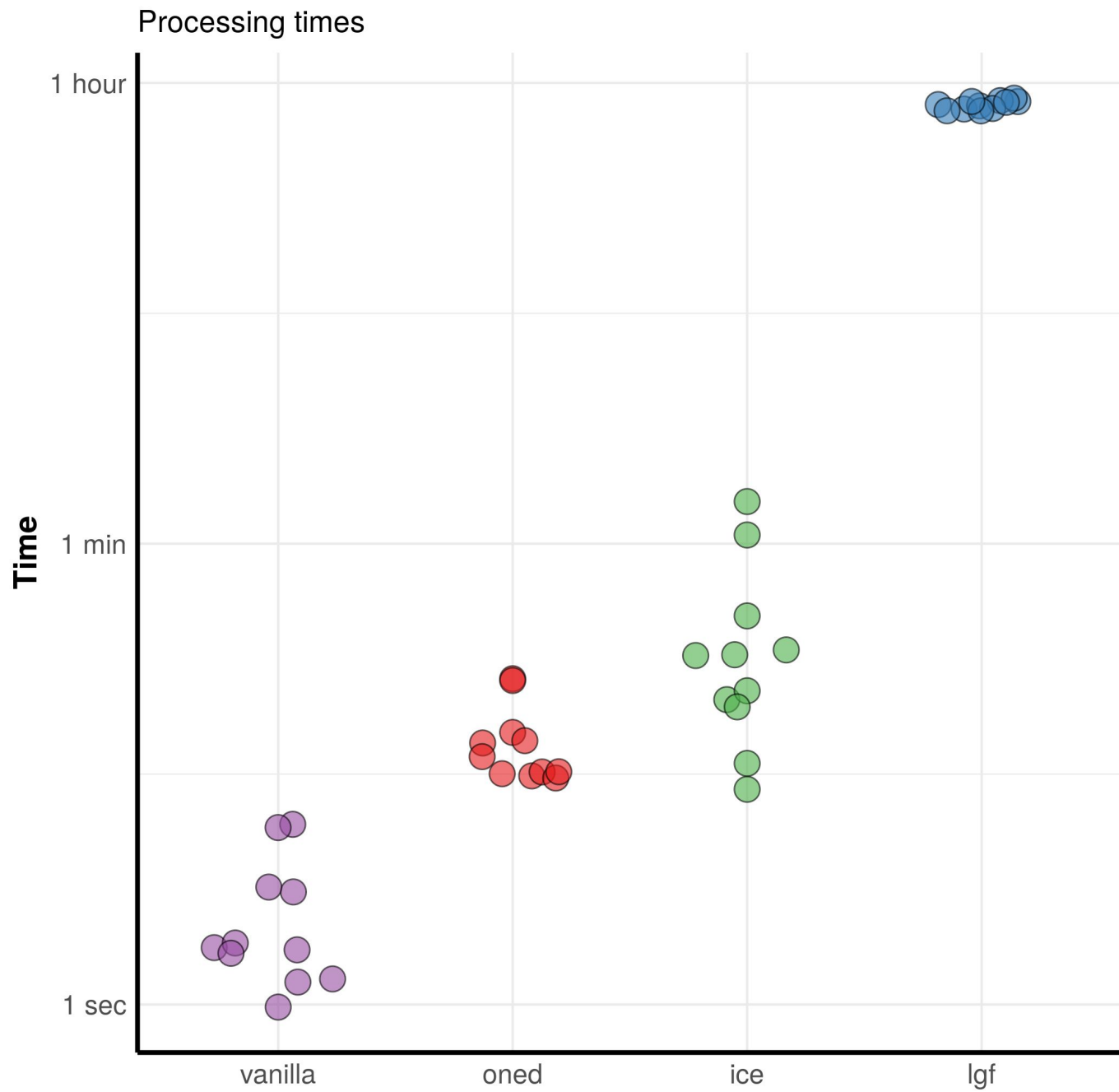
**Signal + bias + decay + noise**



**oneD**

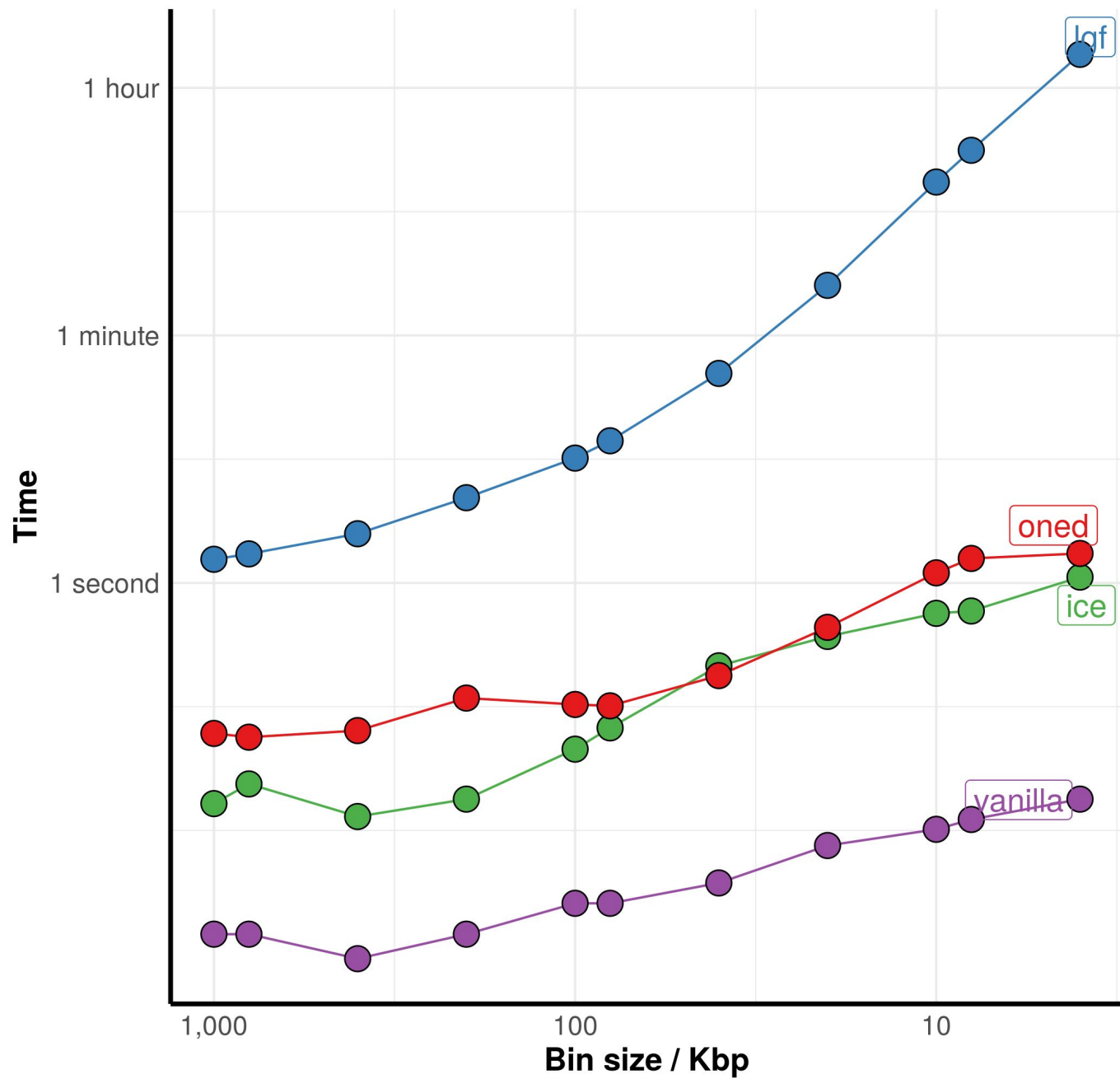


# Timing



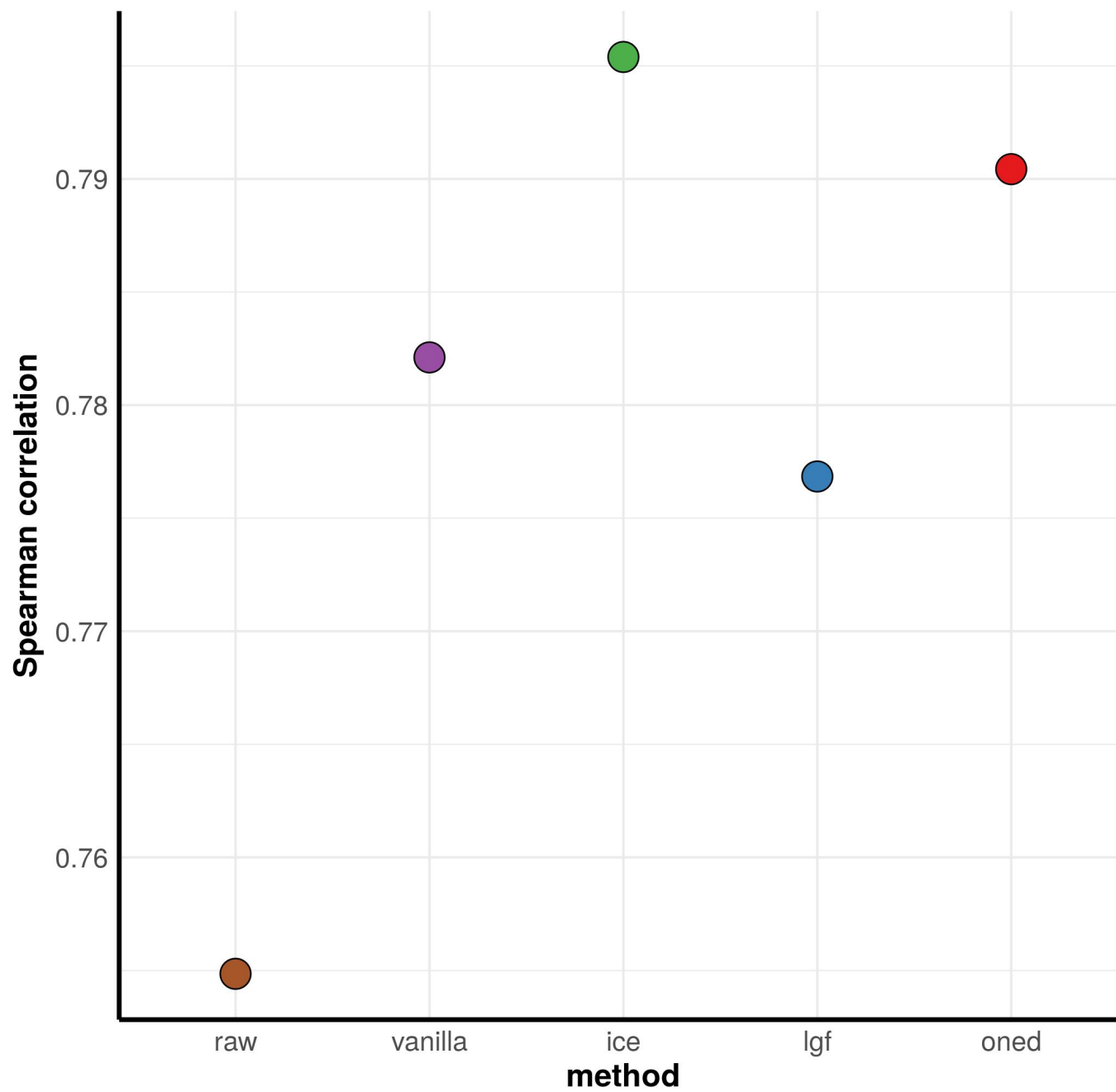


Computing time required to normalize chr22

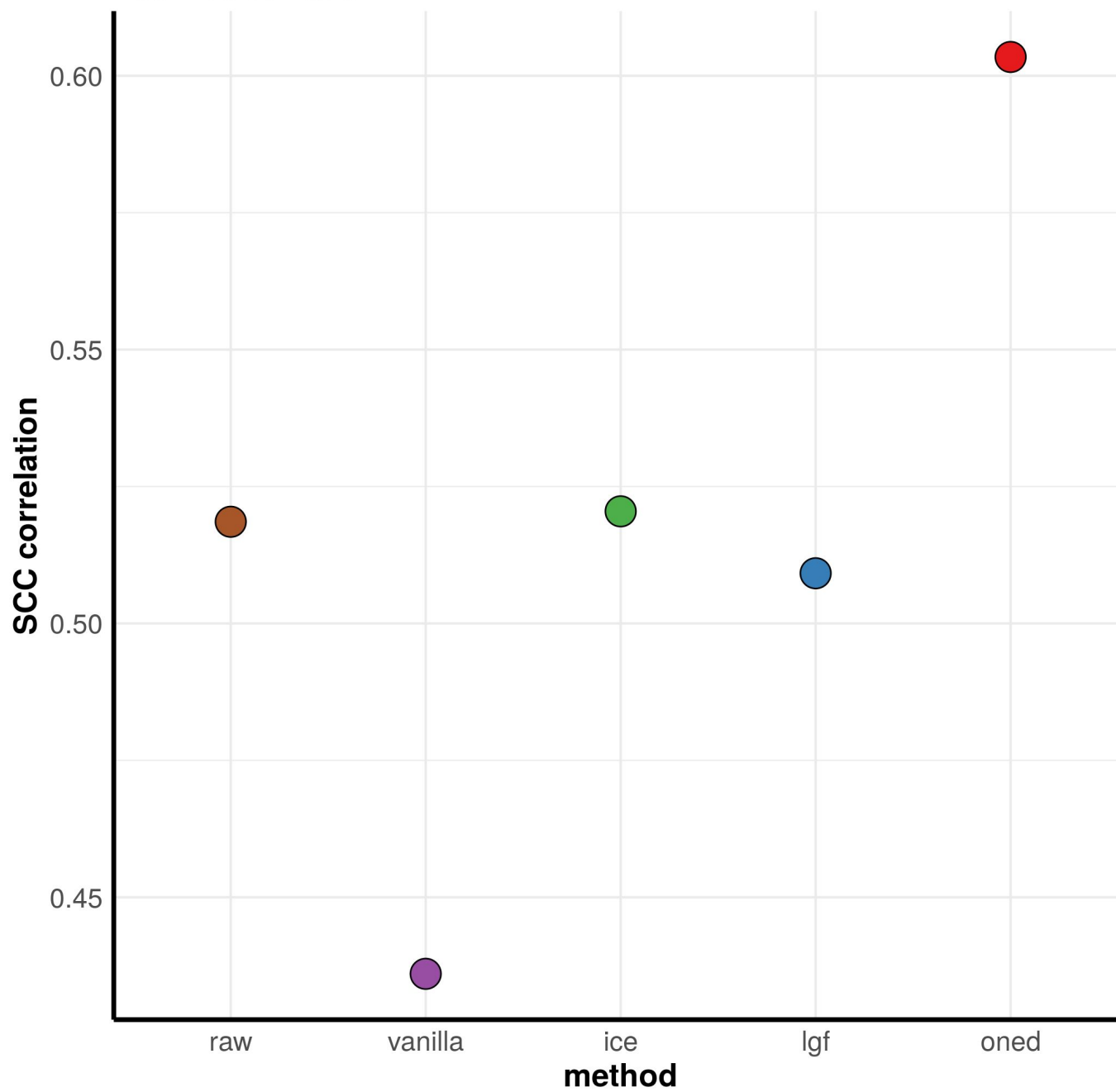


# Performance

K562: cis contacts



K562: cis contacts



**FIN**