



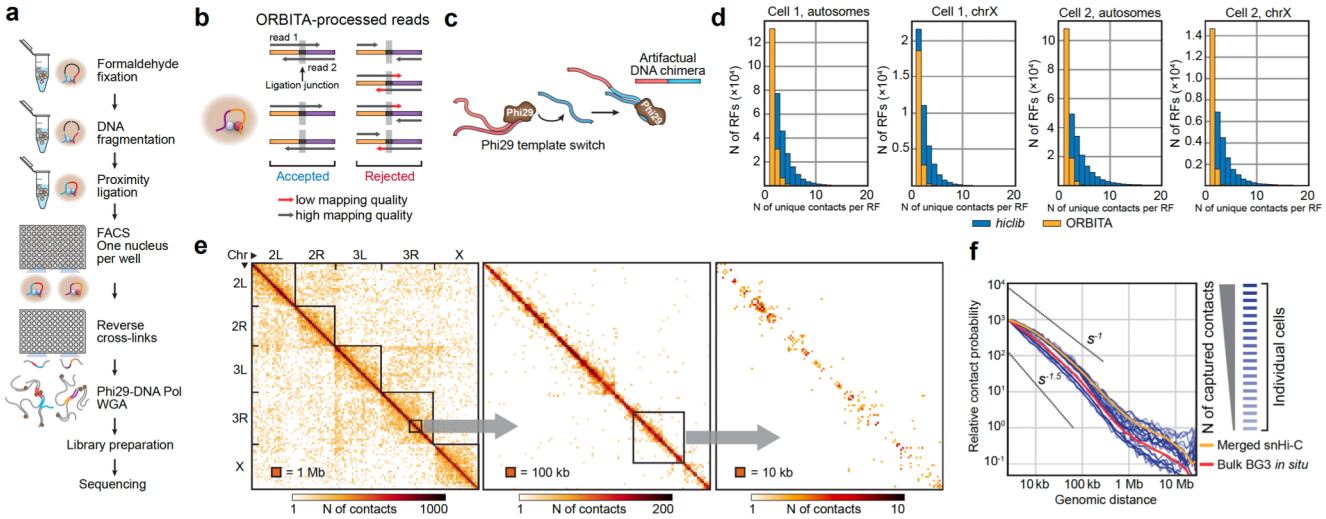
Ministry of Science and Higher Education of the Russian Federation
Institute of Gene Biology Russian Academy of Sciences

# Single-nucleus Hi-C analysis of the Drosophila genome folding

Sergey V. Ulianov, PhD

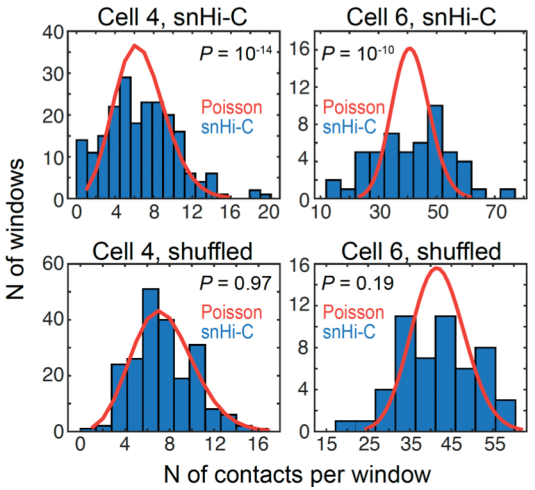
Institute of Gene Biology, Moscow, Russia
Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

# ORBITA-processed Hi-C data from single *Drosophila* nuclei



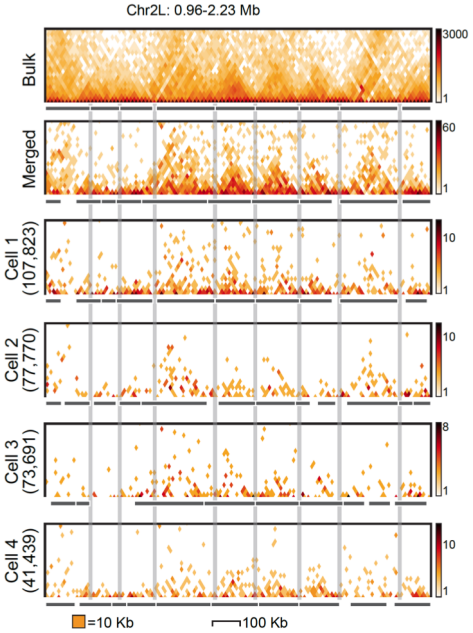

**a** Single-nucleus Hi-C protocol scheme.
**b** Workflow of ORBITA function for detection of unique Hi-C contacts. ORBITA processes only chimeric reads with good mapping quality containing ligation junction marked by the cleavage site for restriction enzyme used for the snHi-C map construction.
**c** Scheme of an artefactual DNA chimera formation by Phi29-DNA-polymerase.
**d** Number of unique contacts per restriction fragment (RF) captured by ORBITA (orange) and *hiclib* (blue) for autosomes and the X chromosome.
**e** Visualization of a single-nucleus Hi-C map at 1-Mb, 100-kb, and 10-kb resolution for the cell with 107,823 captured unique contacts.
**f** Dependence of the contact probability $P_c(s)$ on the genomic distance s for single nuclei (shades of blue reflect the number of unique contacts captured in individual nuclei), merged snHi-C data (orange), and bulk *in situ* BG3 Hi-C data (red). Black lines show slopes for $P_c(s) = s^{-1.5}$ and $P_c(s) = s^{-1}$.

# snHi-C datasets in *Drosophila* are not random matrices

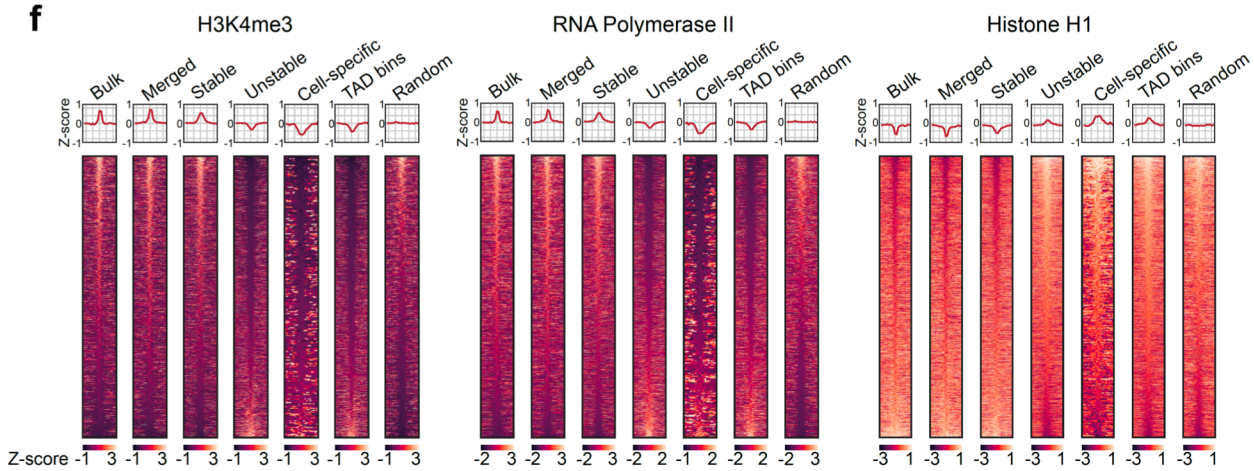



Distributions of the number of contacts in windows of fixed size (100 kb for the Cell 4, and 400 kb for the Cell 6; chr2R) in snHi-C data and shuffled maps for two individual cells (blue bars). The red curve shows the Poisson distribution expected for an entirely random matrix with the same number of contacts.

# Stable TAD boundaries are defined by high level of active epigenetic marks

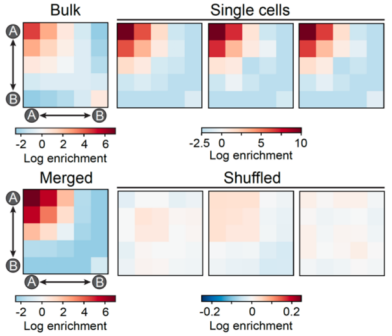



Example of a genomic region on Chromosome 2L with a high similarity of TAD profiles (black rectangles) in individual cells and bulk BG3 *in situ* Hi-C data. Number of unique captured contacts is shown in brackets. Positions of TAD boundaries identified in bulk BG3 *in situ* Hi-C data (top panel) are highlighted with grey lines




Heatmaps of active (H3K4me3, RNA Polymerase II) and inactive (H1 histone) chromatin marks centered at single-cell TAD boundaries from different groups (+/- 100 kb). Bulk – conventional BG3 *in situ* Hi-C; merged – aggregated snHi-C data from all individual cells; stable and unstable – boundaries found in more and in less than 50% of cells, respectively; cell-specific – boundaries identified in any one individual cell; TAD bins – genomic bins from TAD interior; random – randomly selected genomic bins.

## Compartments in individual cells



# Our team


Vlada Zakharova


Sergey Ulianov

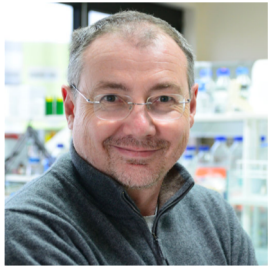

Aleksandra Galitsyna

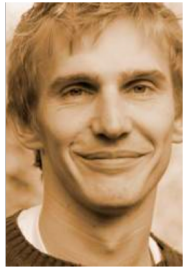

Pavel Kos


Kirill Polovnikov


Ilya Flyamer


Ekaterina Khrameeva


Yegor Vassetzky


Alexander Chertovich


Sergey Nechaev


Mikhail Gelfand


Yuri Shevelyov


Sergey Razin

Skoltech

Skolkovo Institute of Science and Technology

Faculty of Physics

M.V.Lomonosov Moscow State University

RUSSIAN ACADEMY OF SCIENCES

INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS

(Kharkevich Institute)