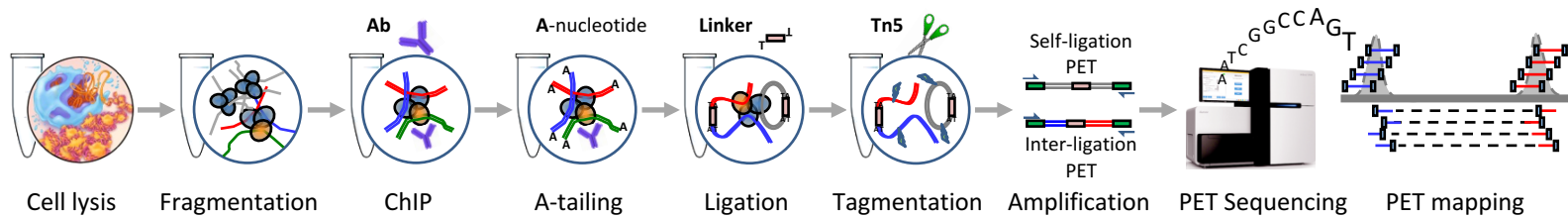


The ChIA-PET data and analysis pipeline



Sheng Li

The Jackson Laboratory for Genomic Medicine

ENCODE NA call, May 26th, 2017

The goals and scopes of the ChIA-PET mapping center in ENCODE4

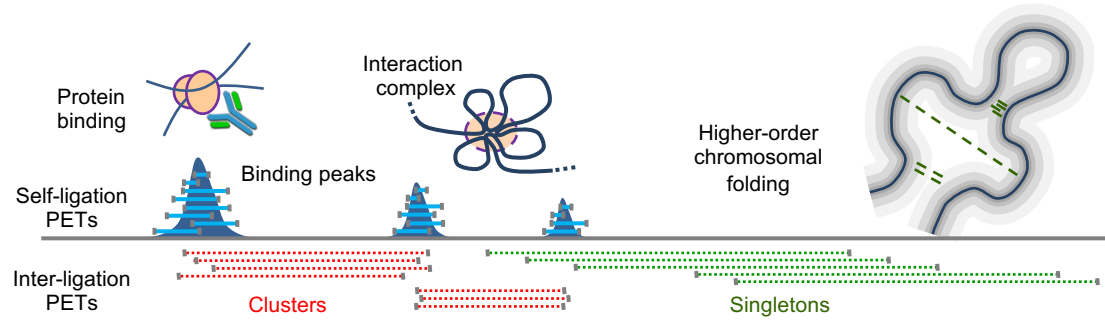
***Our major goal** is to produce high-quality, high-resolution, and comprehensive maps of long-range chromatin interactions between the structural and functional elements in human and mouse genomes.*

Our scopes of data production:

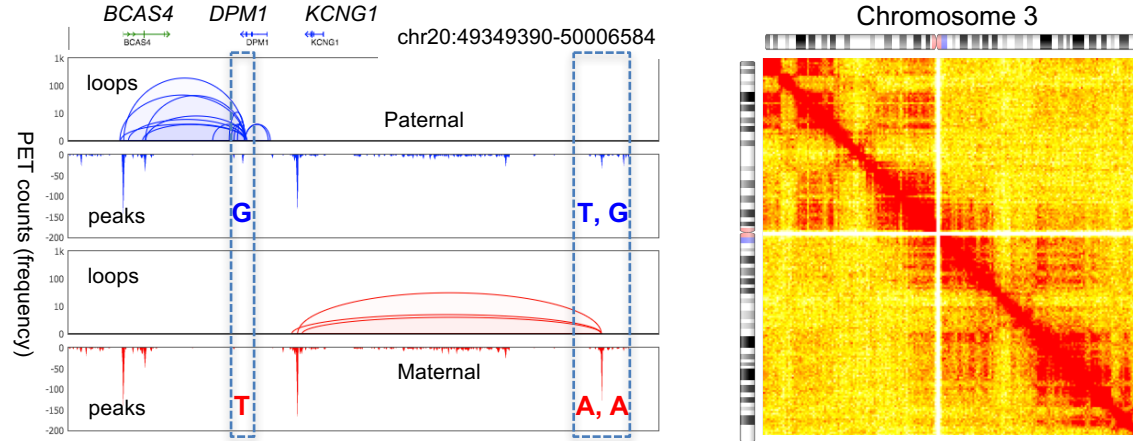
- CTCF ChIA-PET for structural genome organization
- RNAPII ChIA-PET for transcriptional chromatin architecture
- 125 different biological samples focusing on
 - ENCODE cell lines
 - Subtypes of immune cells from human blood donors
 - Patient-derived iPSCs and iPSC-derived neurons
 - Mouse cell lines and primary tissues

Mapping and visualization

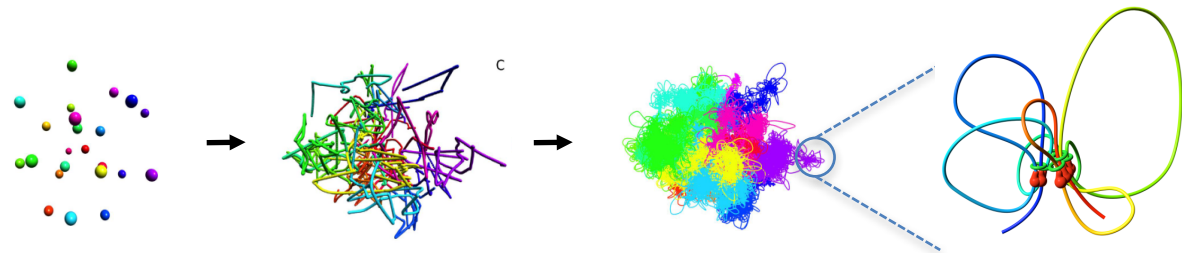
1D-Mapping



2D-Visualization



3D-Simulation

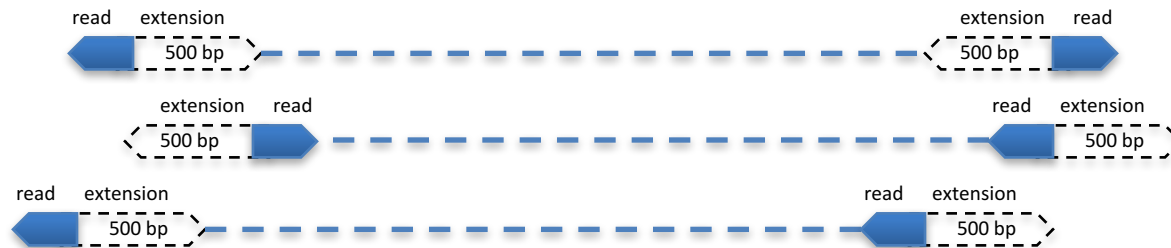


ChIA-PET process pipeline

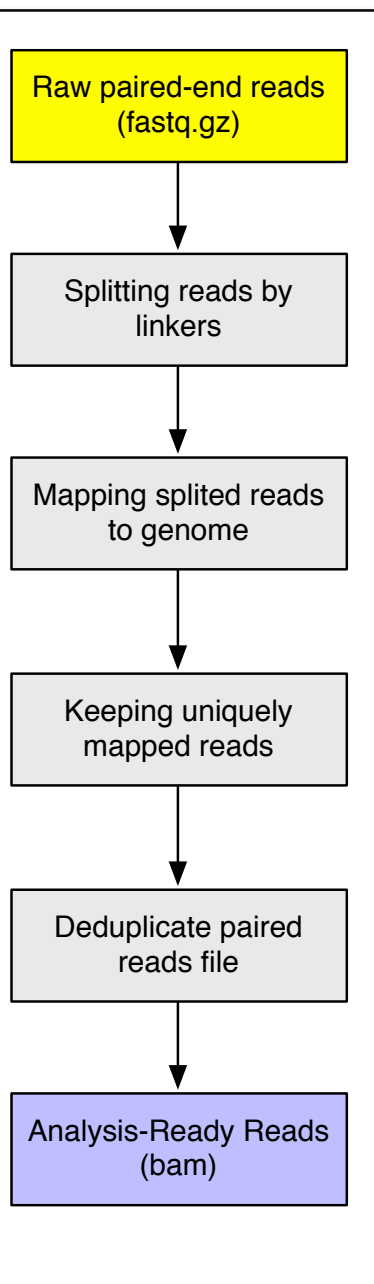
Binding peaks



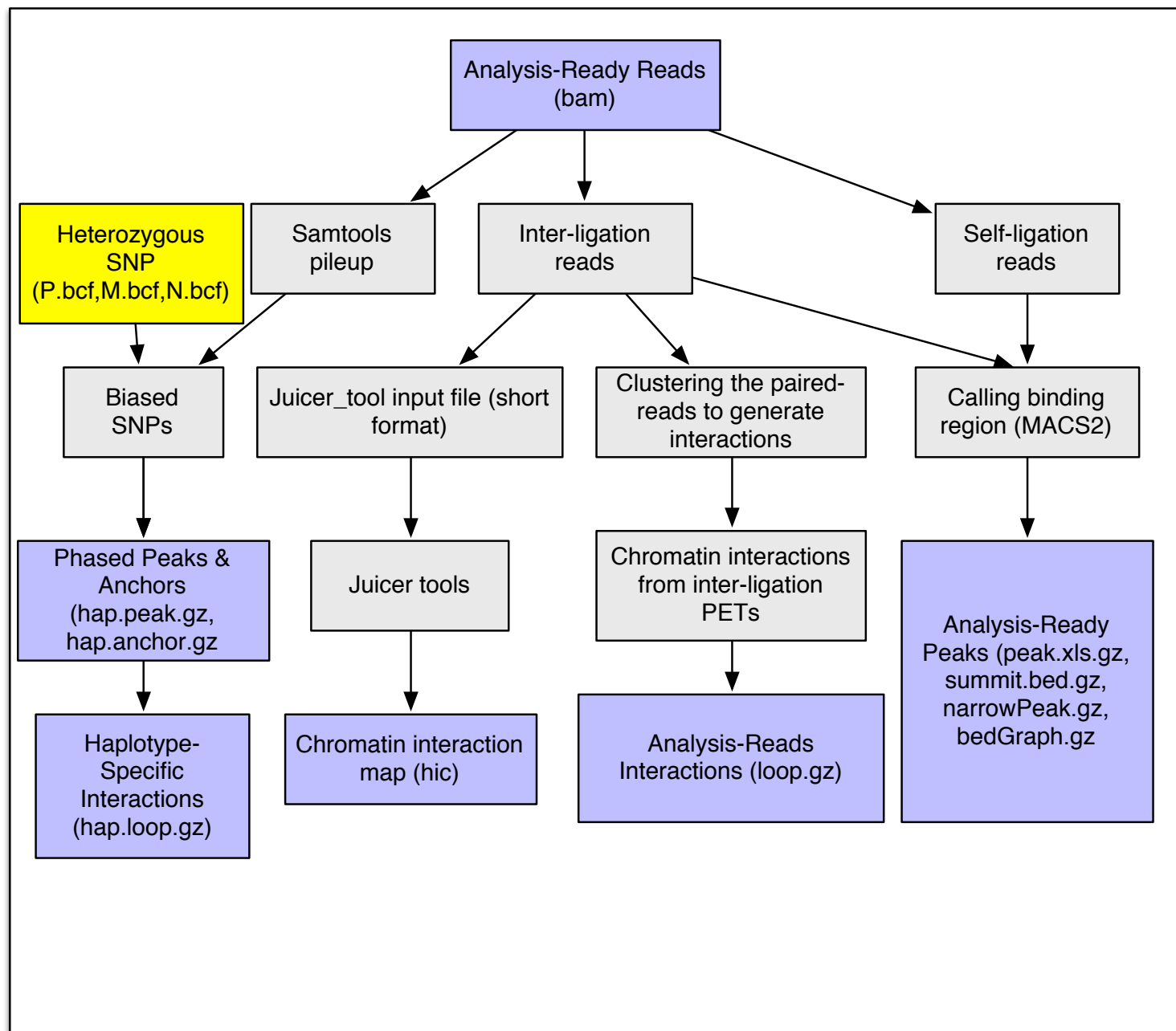
PET clusters



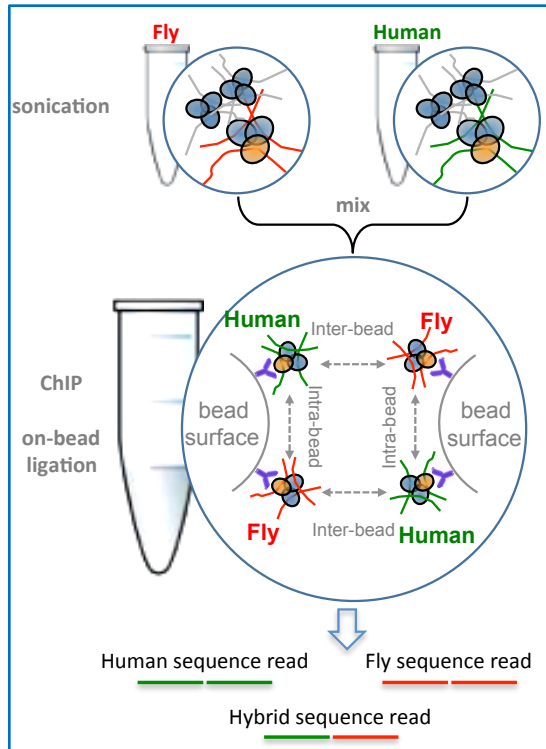
PRE-PROCESSING



INTERACTION DISCOVERY

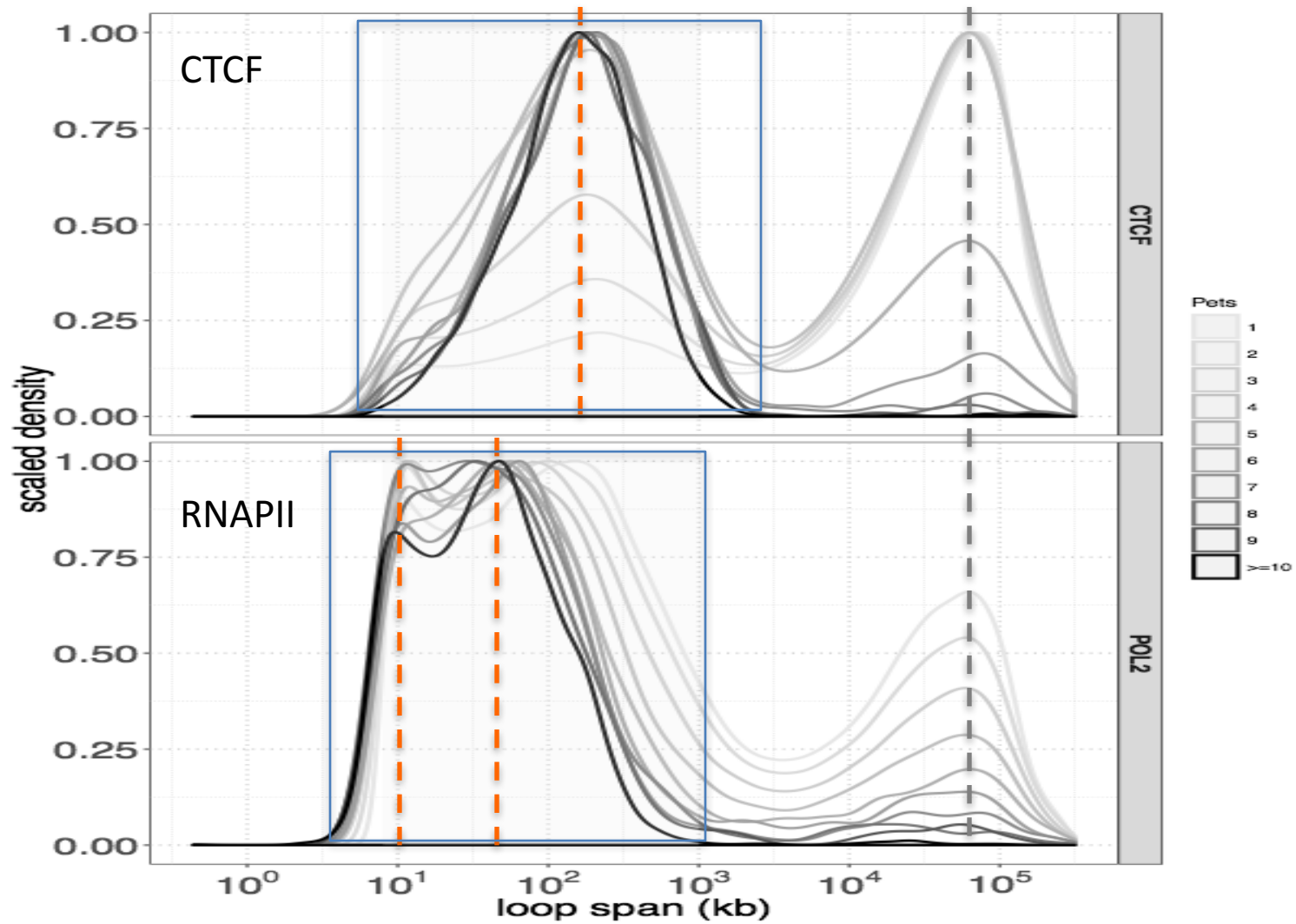


Signal vs noise assessment (2)



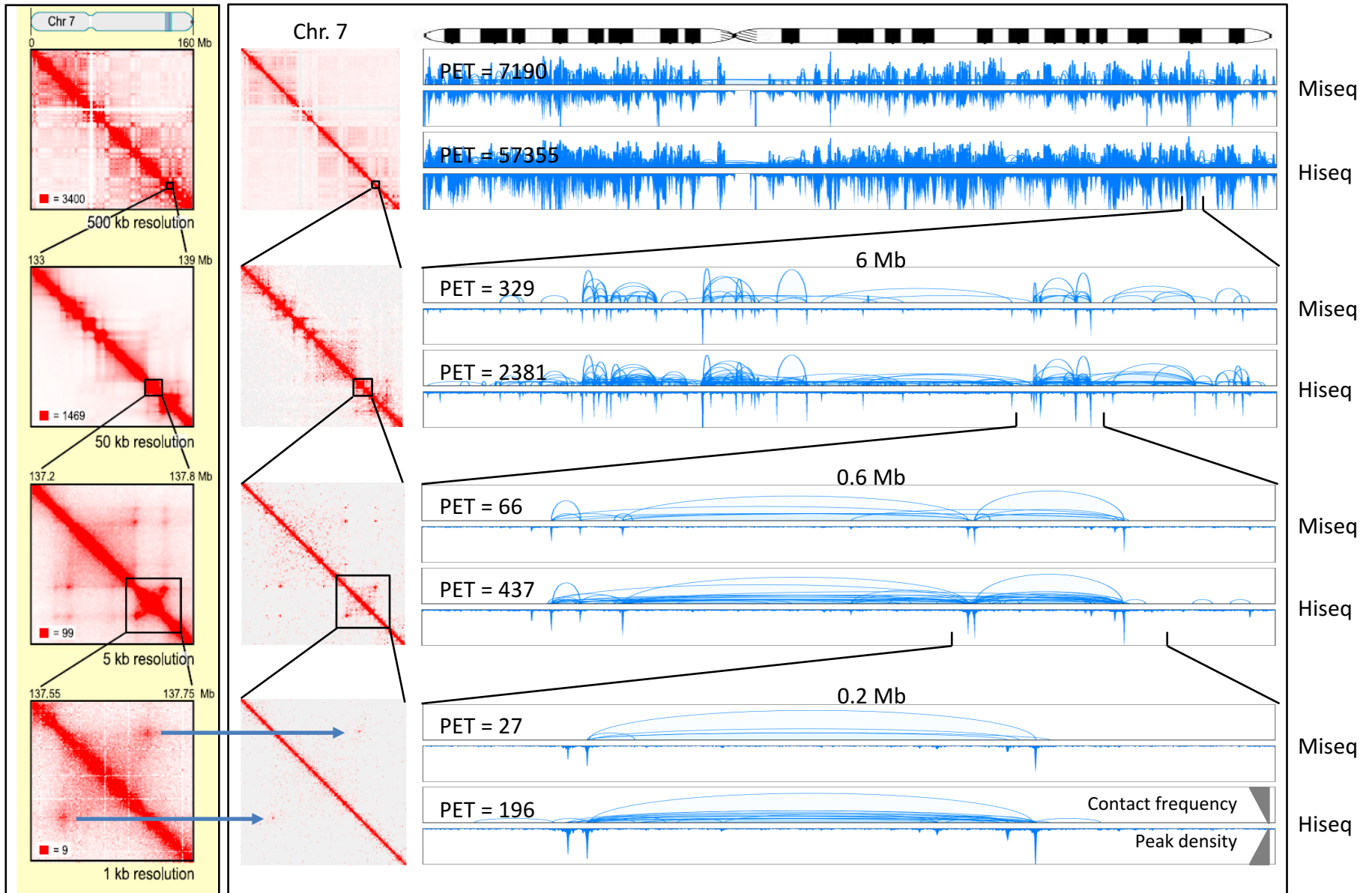
| Lib ID | Seq method | Bead dilution | Uniquely mapped PETs | Inter-species PETs | Noise rate | |
|--------------|------------|---------------|----------------------|--------------------|--------------|-----------------------|
| LDS0221 | Miseq | Low | 5,746,593 | 608,618 | 10.59% | low dilution library |
| LDS0224 | Miseq | Low | 8,667,508 | 721,158 | 8.32% | |
| LDS0225 | Miseq | Low | 8,337,772 | 833,831 | 10.00% | |
| LDS0226 | Miseq | Low | 9,137,320 | 852,414 | 9.33% | |
| Total | | | 31,889,193 | 3,016,021 | 9.46% | |
| LDS0227 | Miseq | High | 8,128,874 | 387,685 | 4.77% | high dilution library |
| LDS0228 | Miseq | High | 5,768,369 | 259,150 | 4.49% | |
| LDS0229 | Miseq | High | 5,745,104 | 226,540 | 3.94% | |
| LDS0230 | Miseq | High | 8,302,149 | 400,521 | 4.82% | |
| LDS0231 | Miseq | High | 5,192,920 | 83,834 | 1.61% | |
| LDS0232 | Miseq | High | 5,148,641 | 135,162 | 2.63% | |
| LDS0230N | Nextseq | High | 141,314,800 | 6,801,514 | 4.81% | |
| LDS0231N | Nextseq | High | 190,200,755 | 3,194,298 | 1.68% | |
| LDS0232N | Nextseq | High | 157,849,191 | 4,167,384 | 2.64% | |
| Total | | | 527,650,803 | 15,656,088 | 2.97% | |

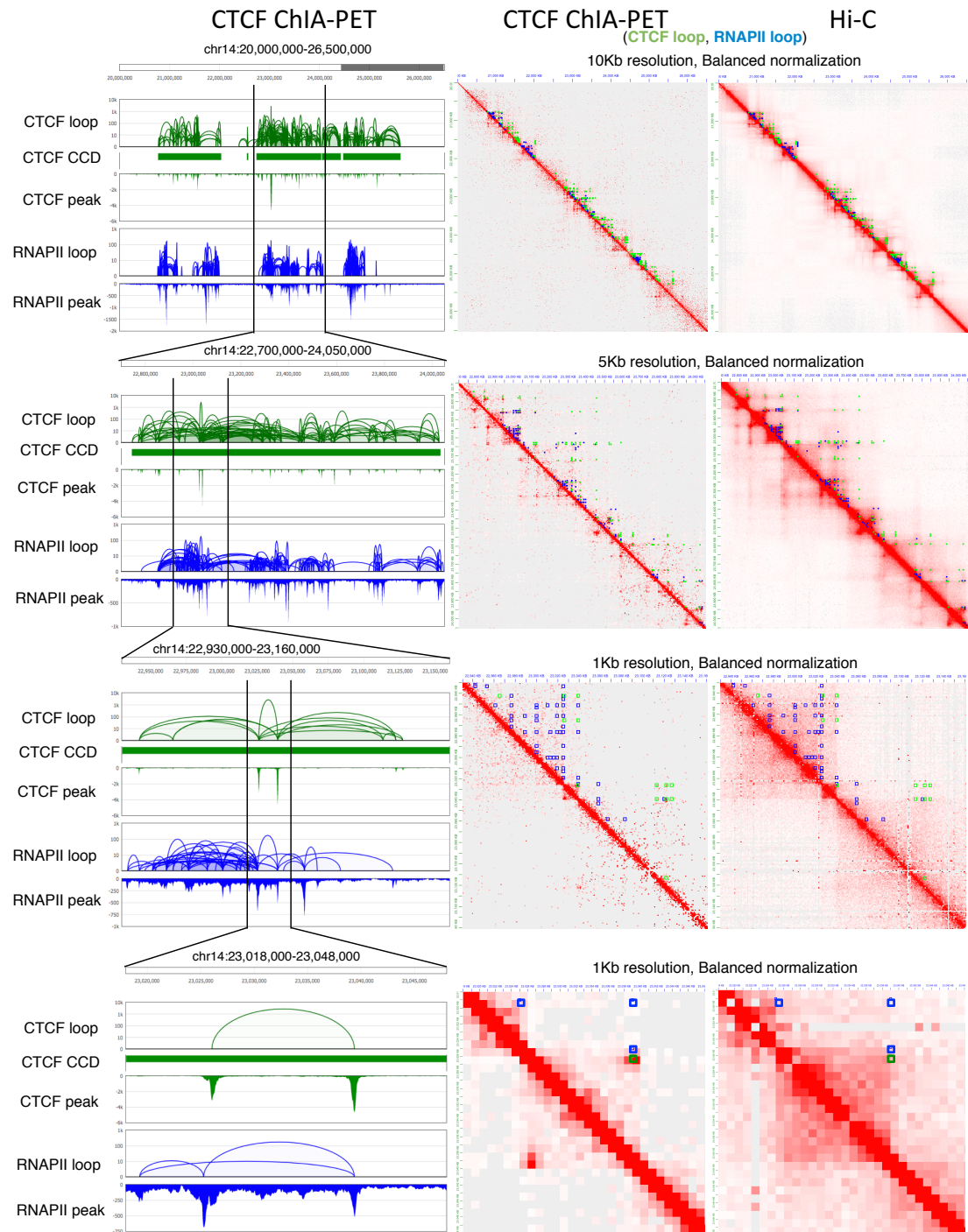
Signal vs noise assessment (3)



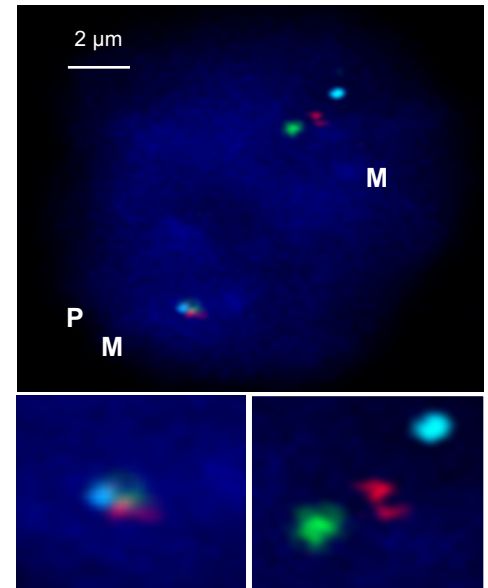
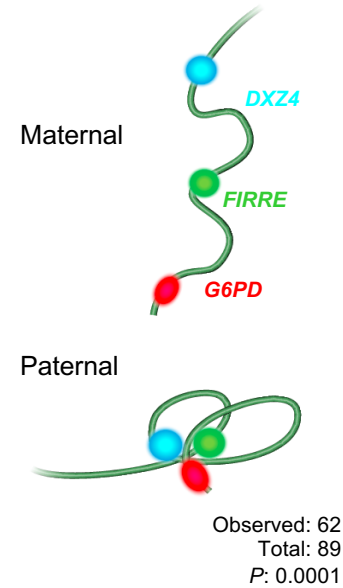
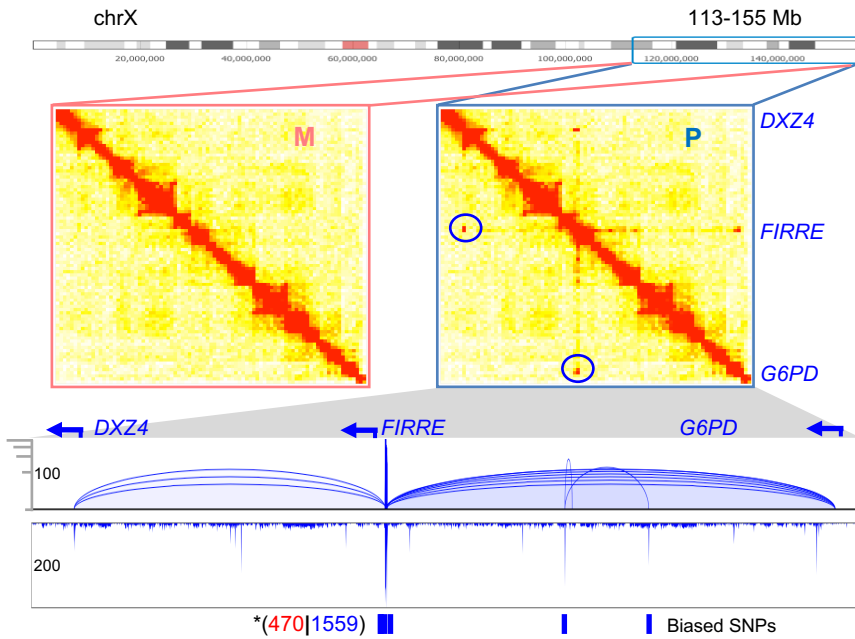
Hi-C

CTCF ChIA-PET

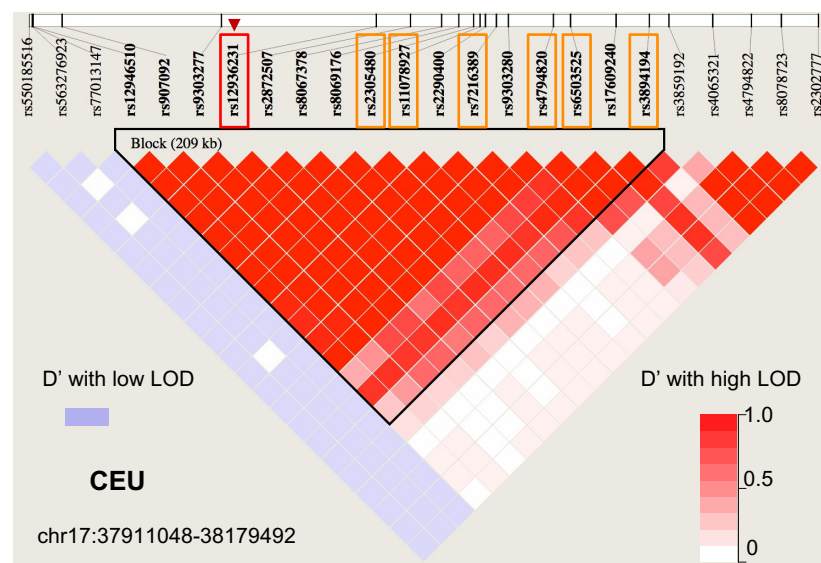
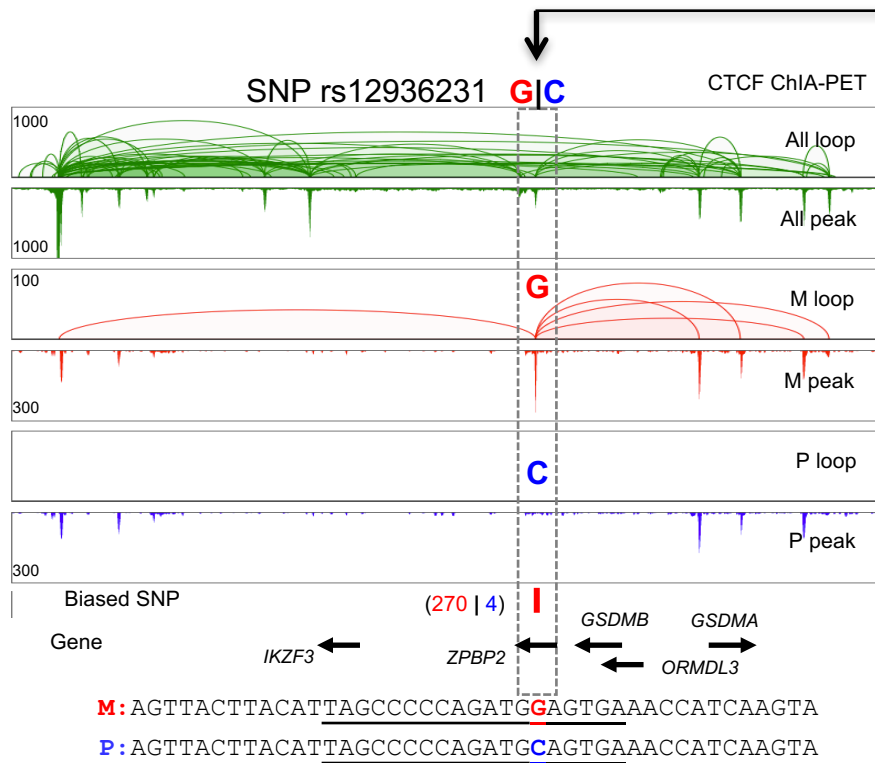




Haplotype-resolved chromatin interactions



Single nt resolution of chromatin interactions linked to genetics



High-risk SNPs for asthma and autoimmune disease alter domain-wide transcription of certain genes (Verlaan 2009)

