Chromatin structure and function changes in re-arranged genomes

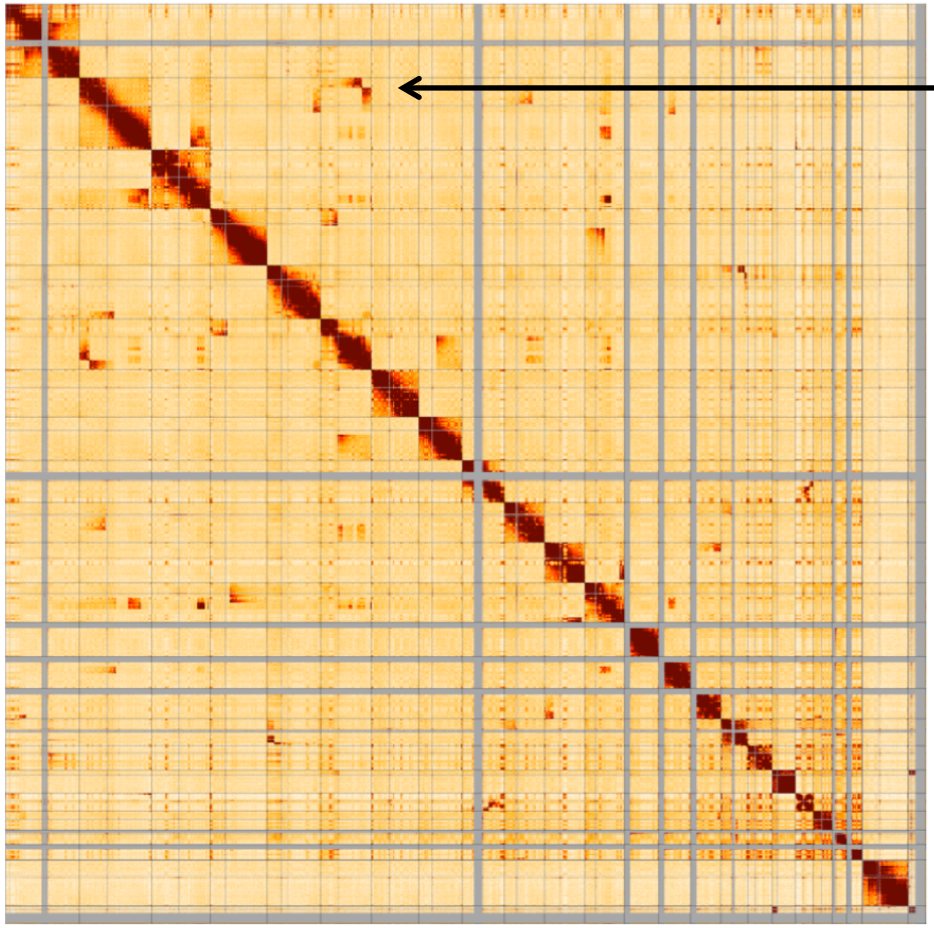
Dekker Lab, Gilbert lab, Yue Lab, Ferhat Ay

Vishnu Dileep (Graduate student, Gilbert lab)

Goals

> Pipeline to identify translocations from Hi-C datasets: Using Hi-C from cancer cell lines

Caki2 (Kidney epithelial carcinoma)



Strong off-diagonal interaction indicates translocation

Ferhat Ay and

Noam Kaplan (Dekker lab)

Validate the translocations identified from Hi-C

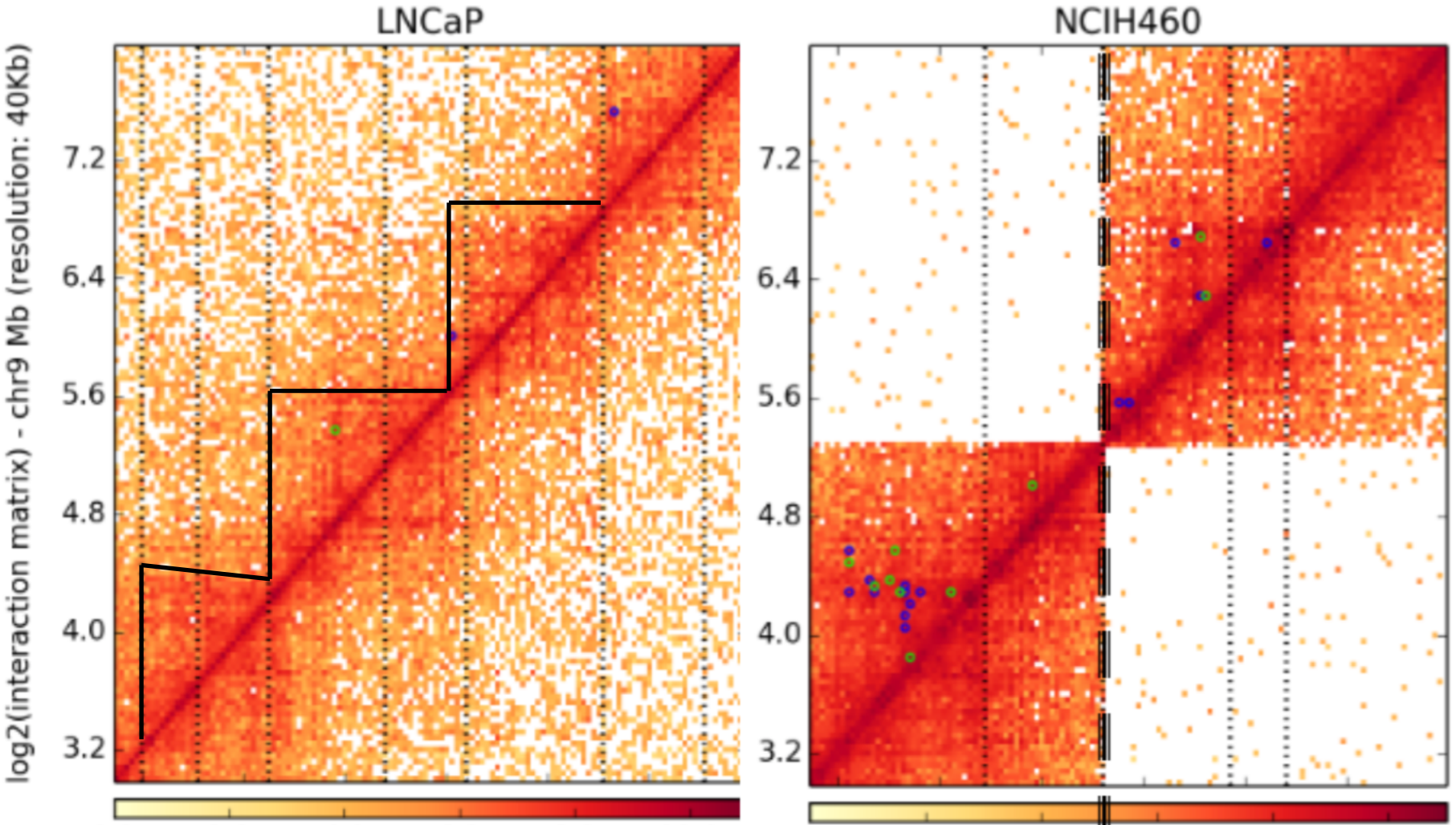
>Irys system to make physical genome maps

>Selected 2 cell lines with maximum number of strong off-diagonal interactions (Caki2, T47D)

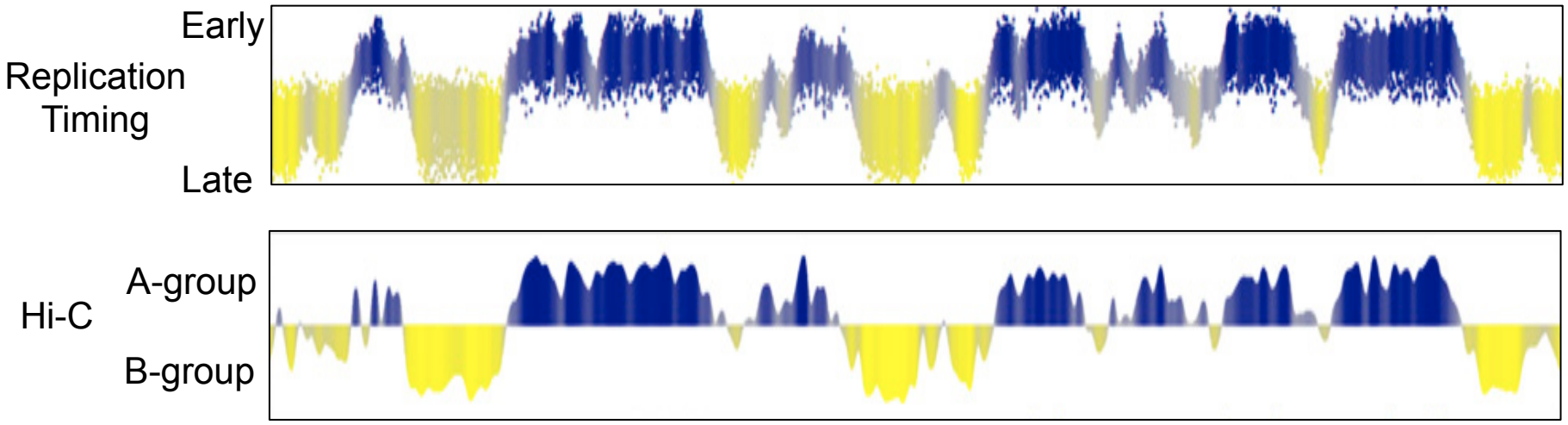
>Jie Xu (Yue Lab) will talk about the technology and the data we have so far

Structural and functional changes due to the breakpoints

- What happens to the TAD structure when it is split by a translocation?



How does translocations affect function? Replication Timing Program (RT)

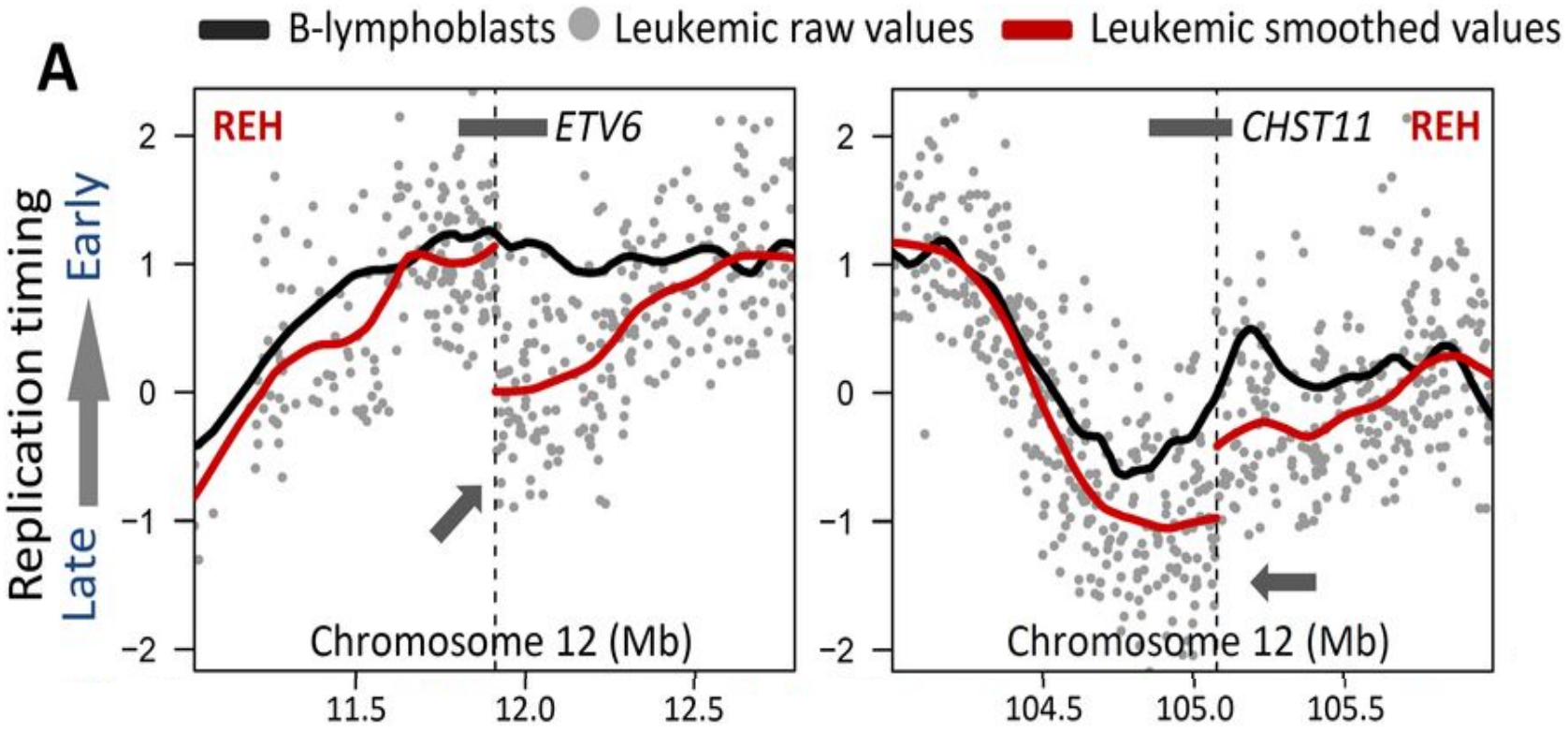


- **DNA replicates in a spatio-temporal order**
- **Distinct domains that replicate earlier or later in S-phase**

- **Early/Late replicating DNA corresponds to A/B compartment.**
- **50% of genome changes RT during mouse and human development**
- **Changes in RT during differentiation occur in units of 400-800kb which have a one-to-one correspondence with TADs.**
- **A/B compartments and TADs are established along with the establishment of RT in early G1**

Ryba et al. 2010; Naumova et al. 2013 ; Pope et al. 2014; Dileep et al. 2015

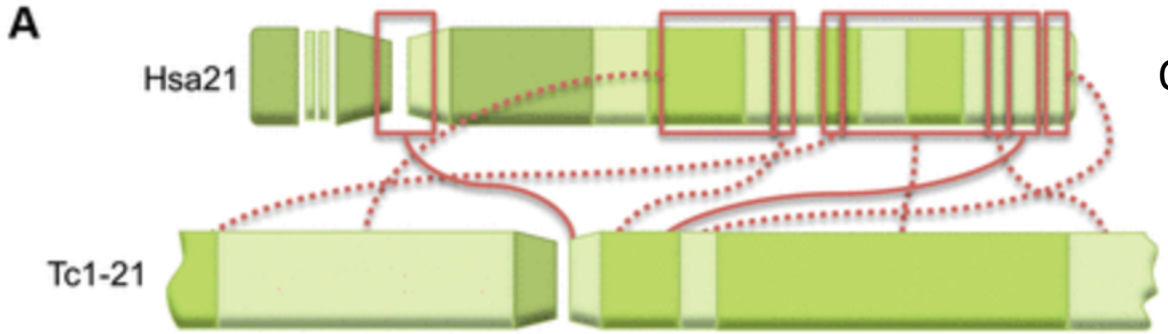
Replication timing at translocation in leukemia cell lines



Ryba et al. 2012

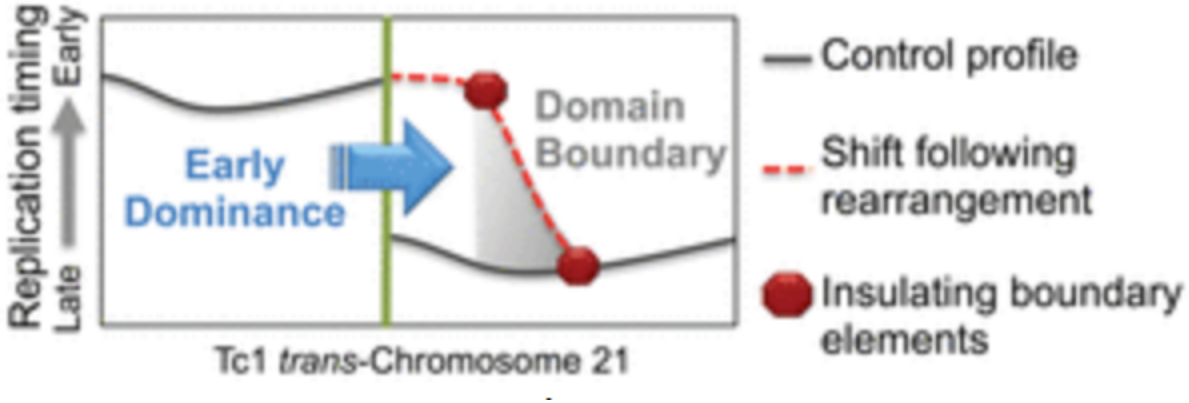
Re-arranged human Chr21 in mouse

- Mouse model for down syndrome
- Chr21 got re-arranged in mouse cells



Original Human Chr21

Re-arranged human Chr21 in mouse

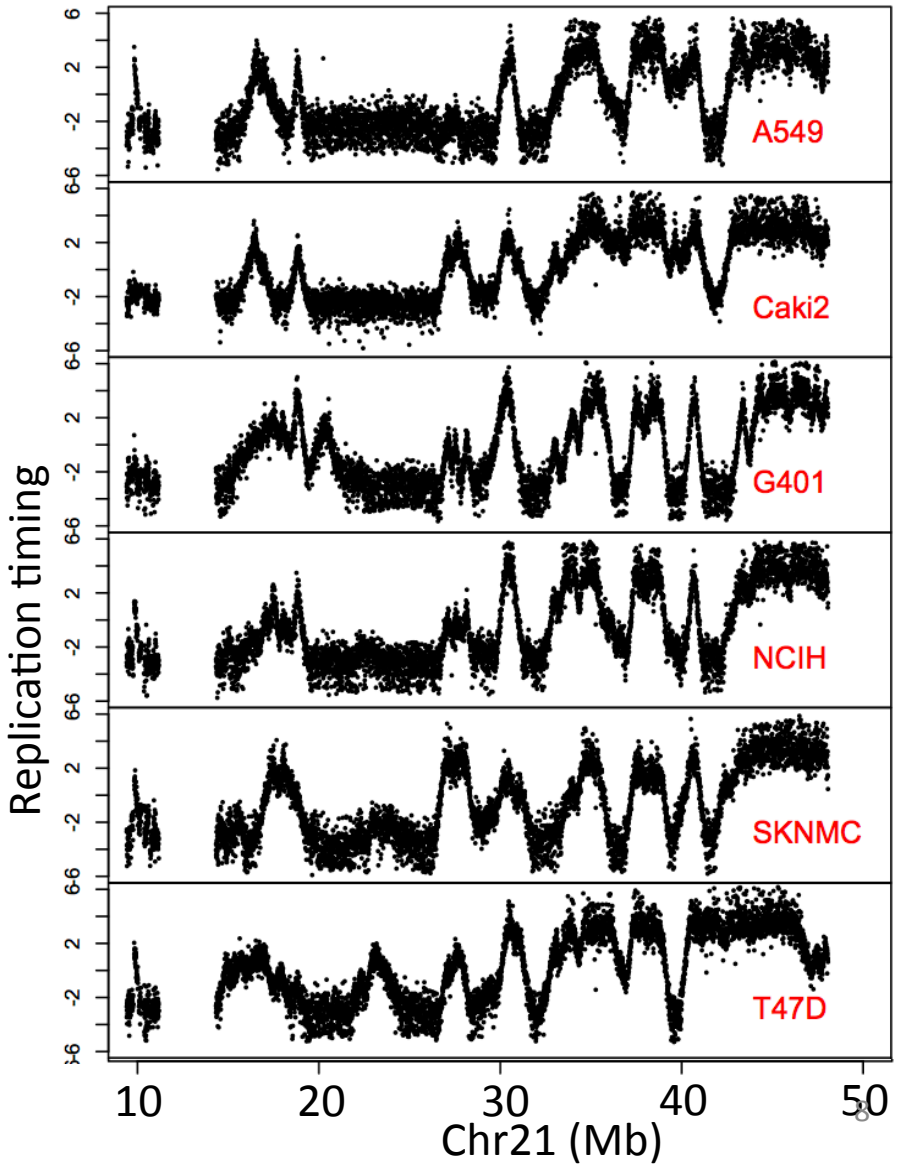


When early and late replicating DNA is juxtaposed, early replication spreads into late until it meets a TAD/RD boundary

Pope et al. 2012

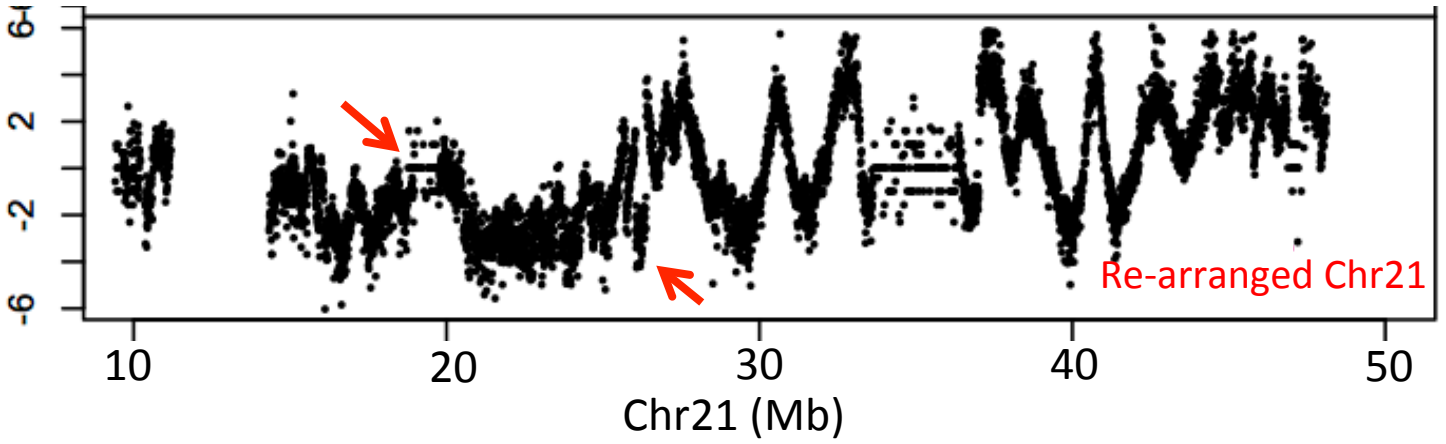
Data generation

Cell line	Toatl HiC reads	Translocations
CAKI2	168540398	6
T47D	137395622	6
NCIH460	155078021	5
PANC1	168012696	3
G401	155413318	3
A549	136384017	1
RPMI7951	189823508	0
SKNMC	175820923	0
LNCaP	149387648	0



HiC on Re-arranged human Chr21 in mouse
Suzana Hadjur (University of London)

- TAD structure at translocations
- Will Human Chr21 maintain human TADs in mouse cells?
- Extra validation for the pipeline for calling breakpoints from HiC data.



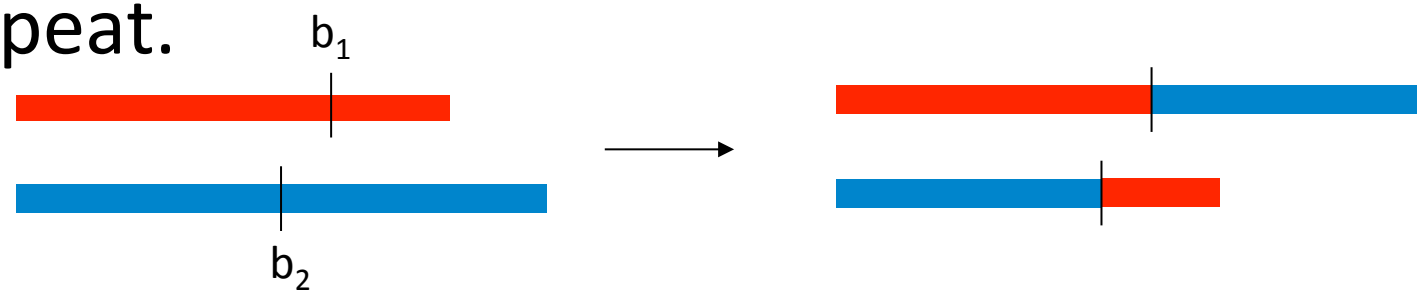
Noam Kaplan

Dekker Lab

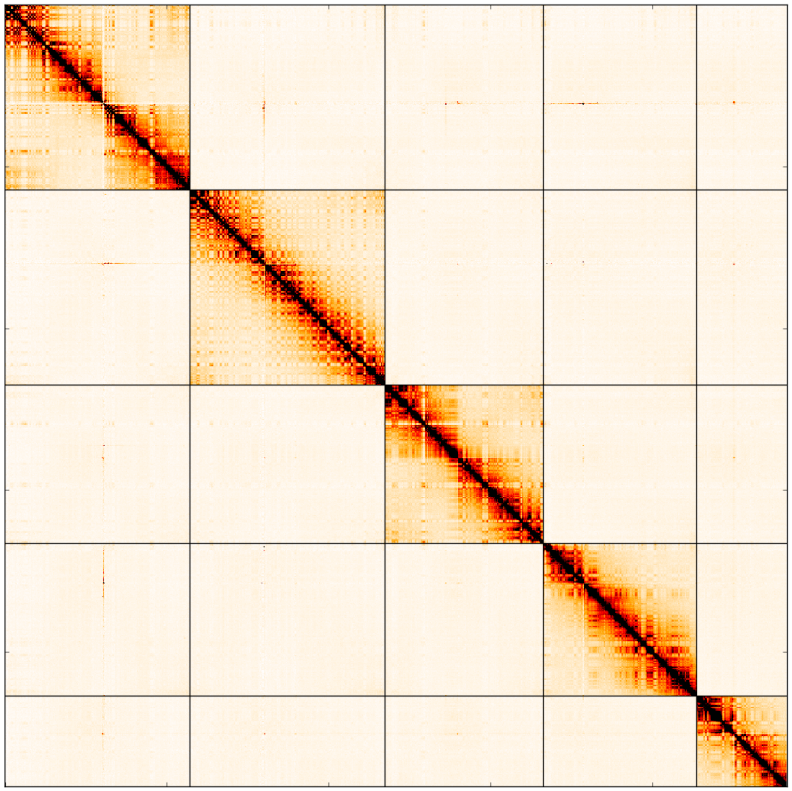
Finding breakpoints from simulated translocations

Translocation simulation

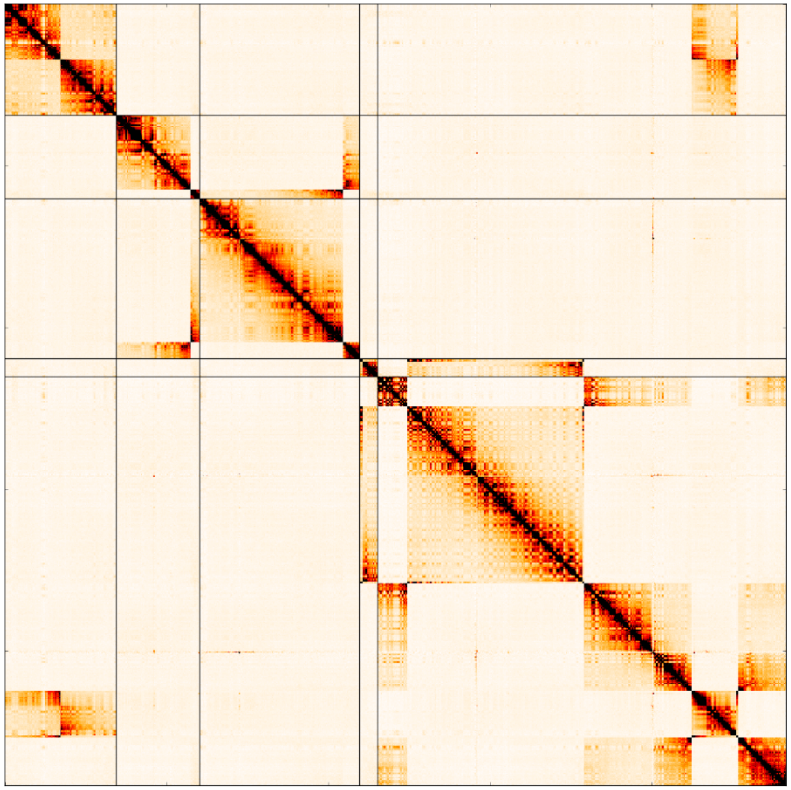
- Pick random bins b_1,b_2 on different chromosomes.
- Remove chromosomes chr(b_1),chr(b_2) from data.
- Add 2 new chromosomes with translocations.
- Repeat.



Original data



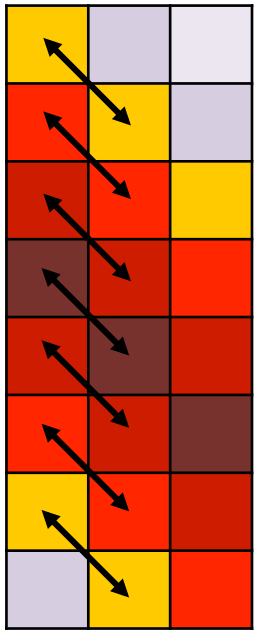
Data with translocations



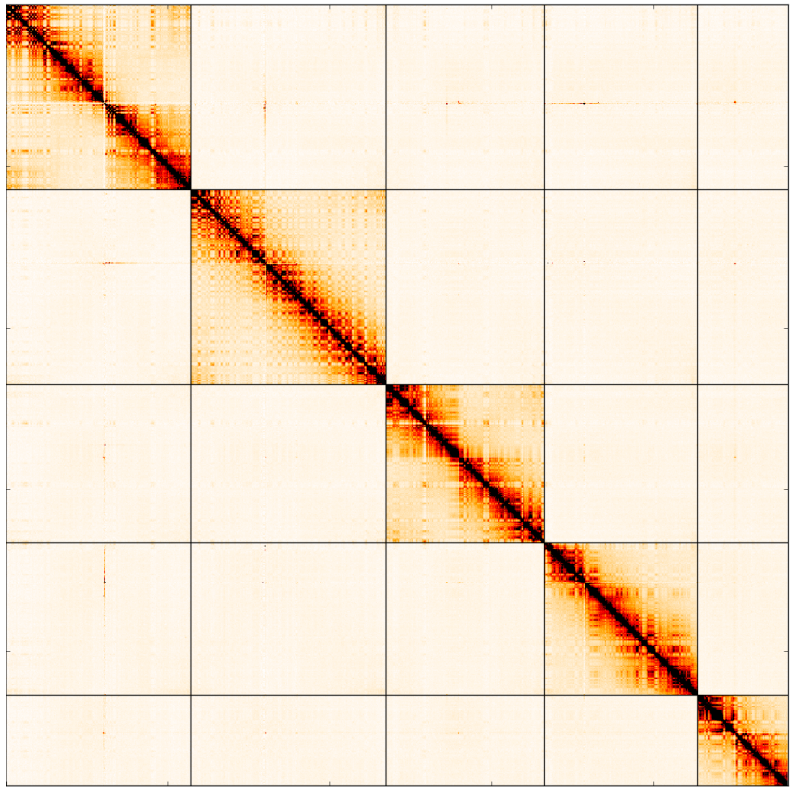
Finding breakpoints

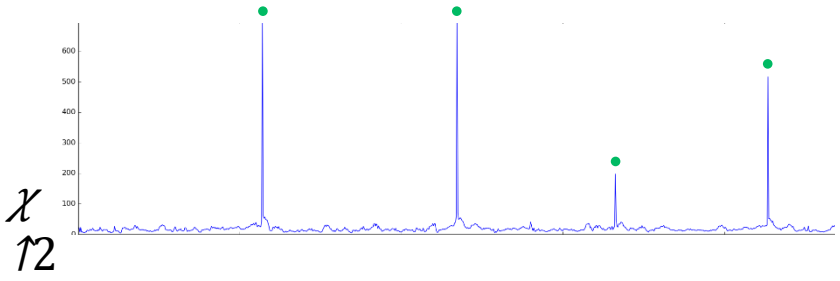
- Strategy: identify adjacent rows/columns that are different from each other.

  1. Edge-preserving smoothing: median filter (9x9 bins)
  2. Perform test between every 2 adjacent rows, shifted
  3. Use test statistic as metric



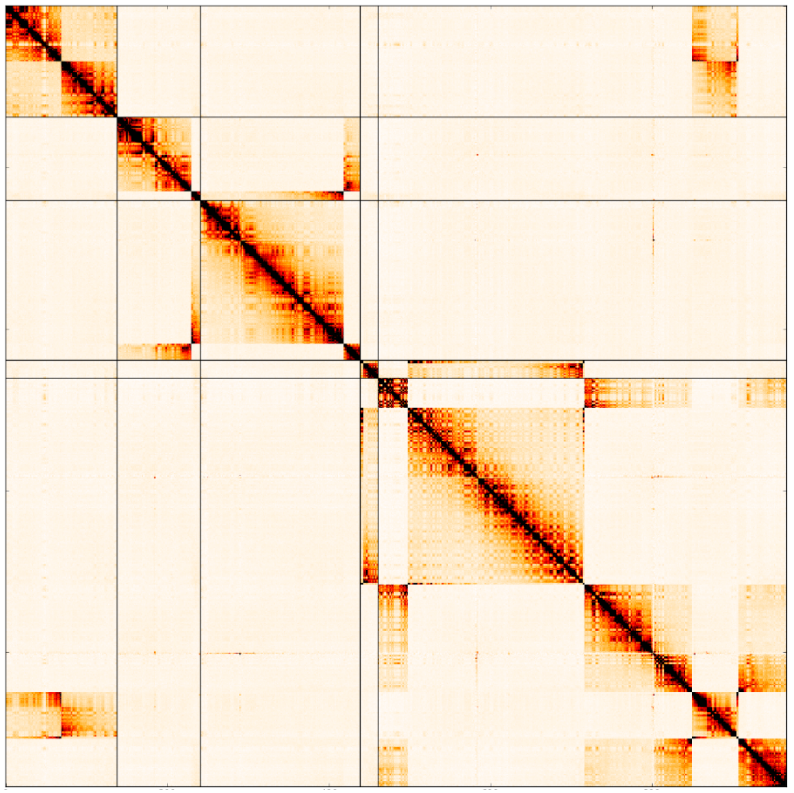
Original data

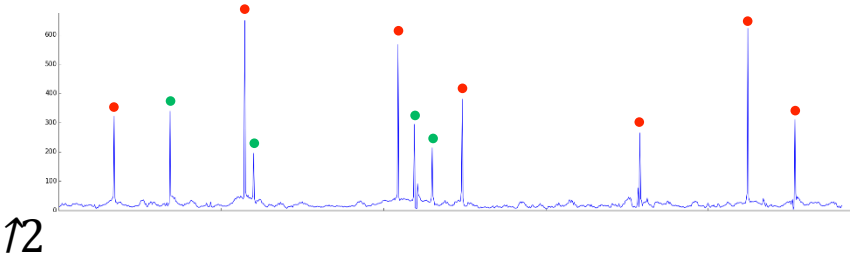




χ^2

Data with translocations





χ^2

Re-scaffolding

- Breakpoints partition the genome into large contigs.
- Contigs can be scaffolded (rearranged) using Hi-C.
- We search for an arrangement that maximizes the likelihood of the data, given an expected interaction model.

Kaplan N, Dekker J. (2013).
High-throughput genome scaffolding from in vivo DNA interaction frequency.
Nature Biotechnology, 31:1143-1147.

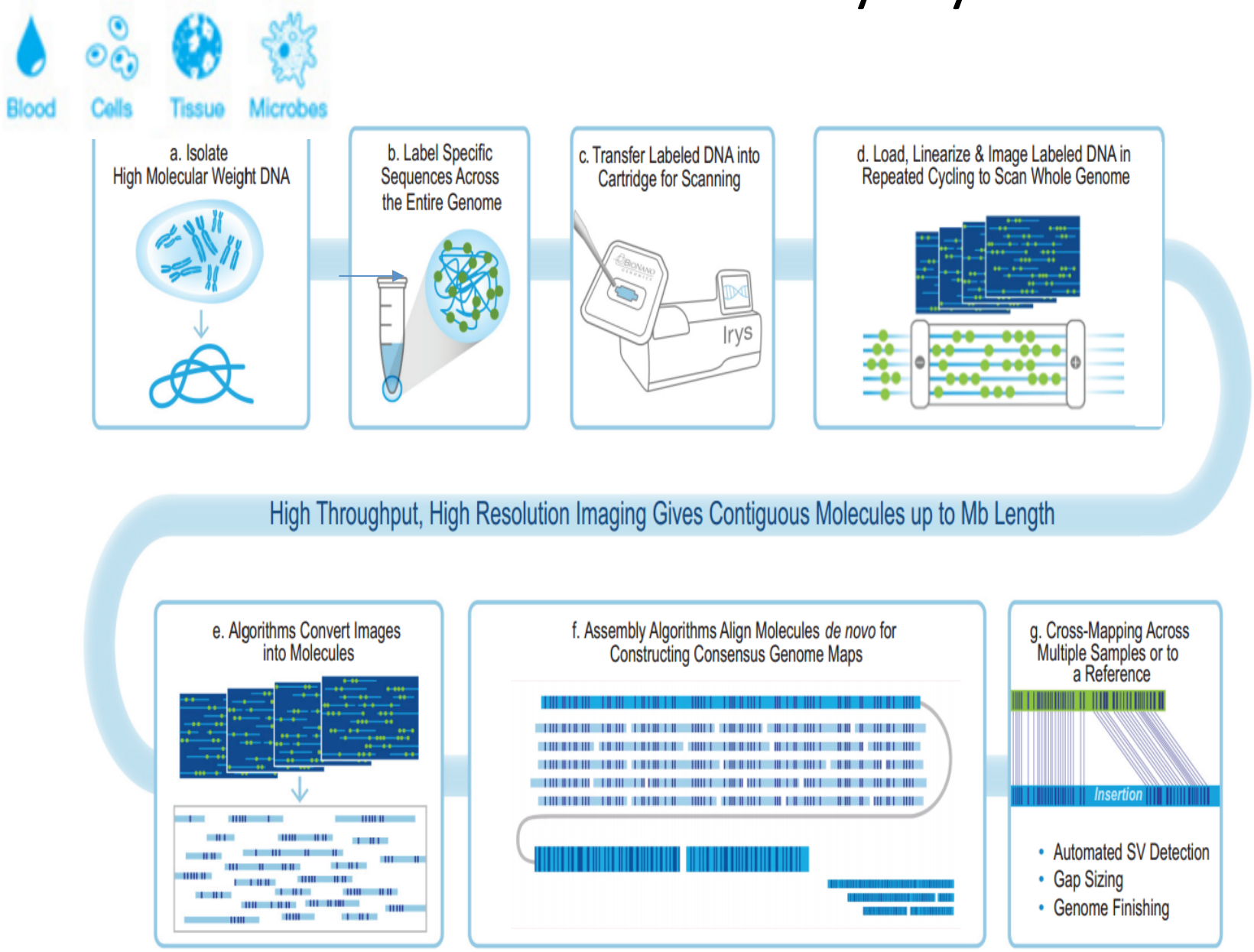
Detecting structural variation by Bionano Irys

Jie Xu (PhD student)

Feng Yue Lab

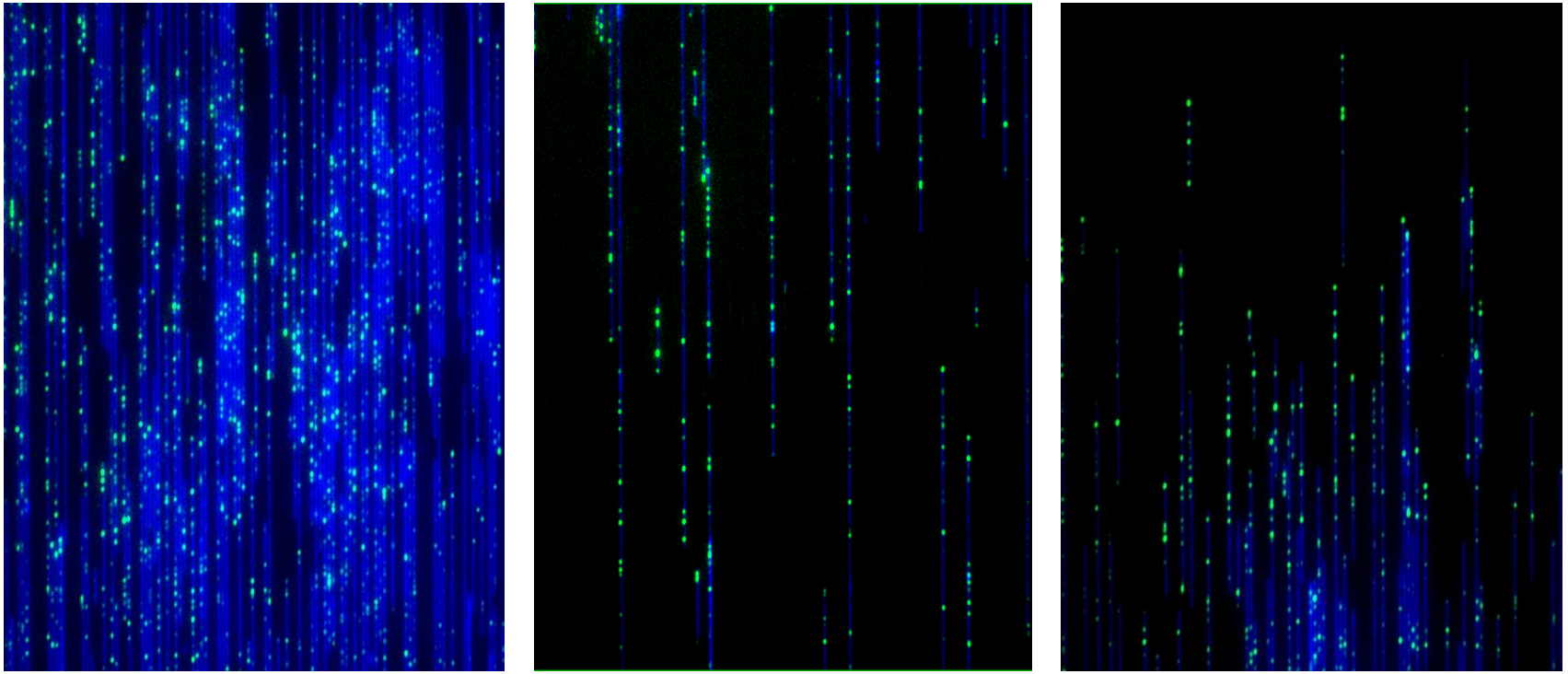
Penn State University

Flowchart of how Bionano Irys system works

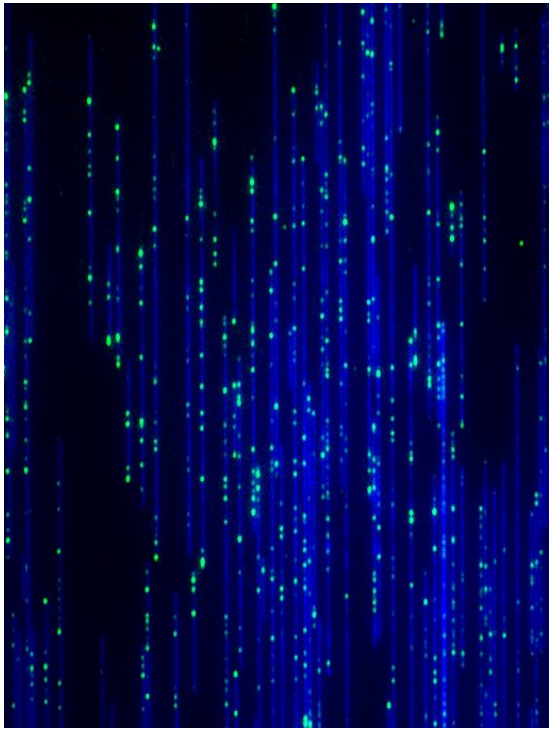
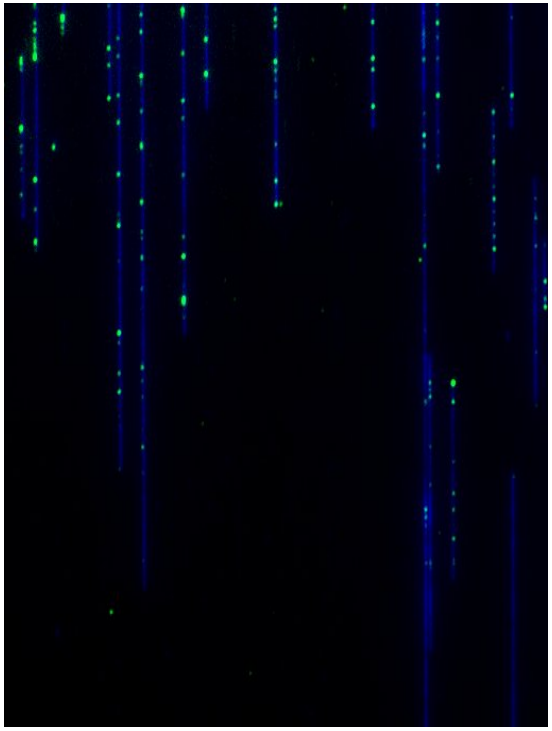
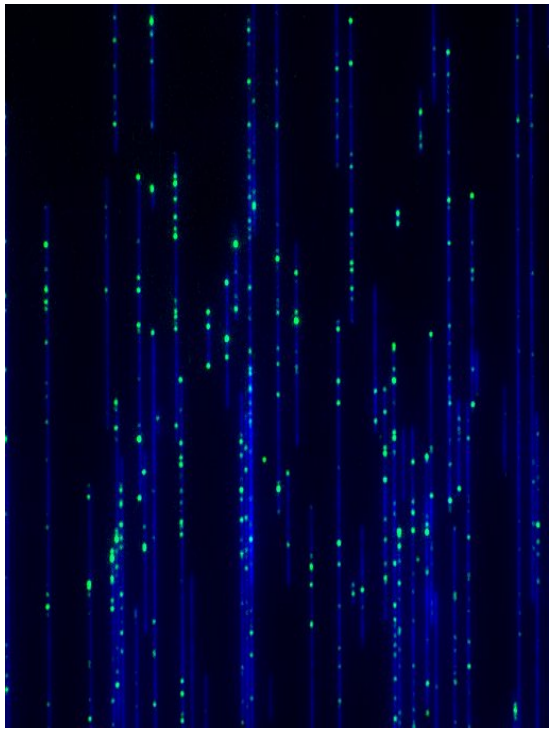


http://www.bionanogenomics.com/technology/irys-technology/

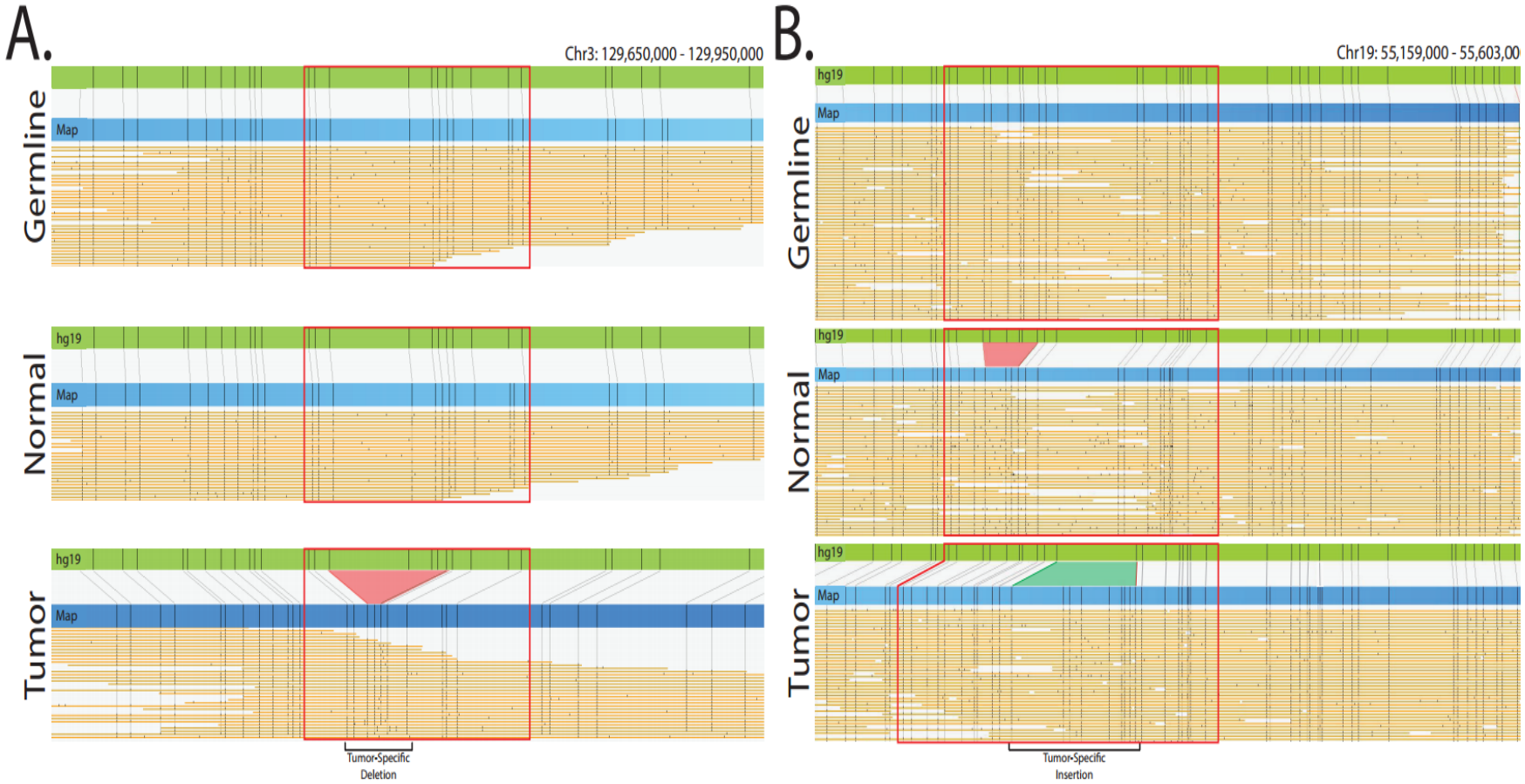
Running Scan of Caki2 DNA



Running Scan of T47D DNA

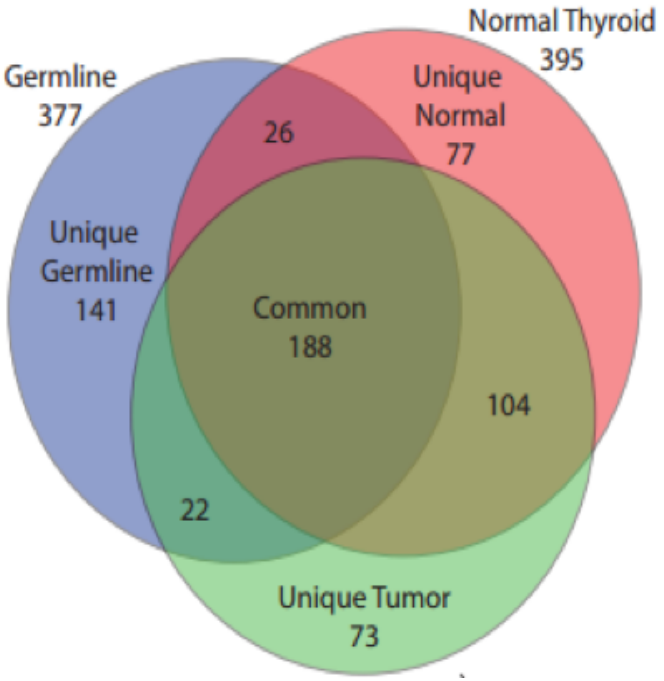




Cancer-specific deletions and insertions detected by Irys



* Note: above results is from our recent run in thyroid cancer samples that have 30X coverage. For T47D and Caki2 cell lines, we achieved 10X coverage and need to sequence deeper to generate the genome assembly.

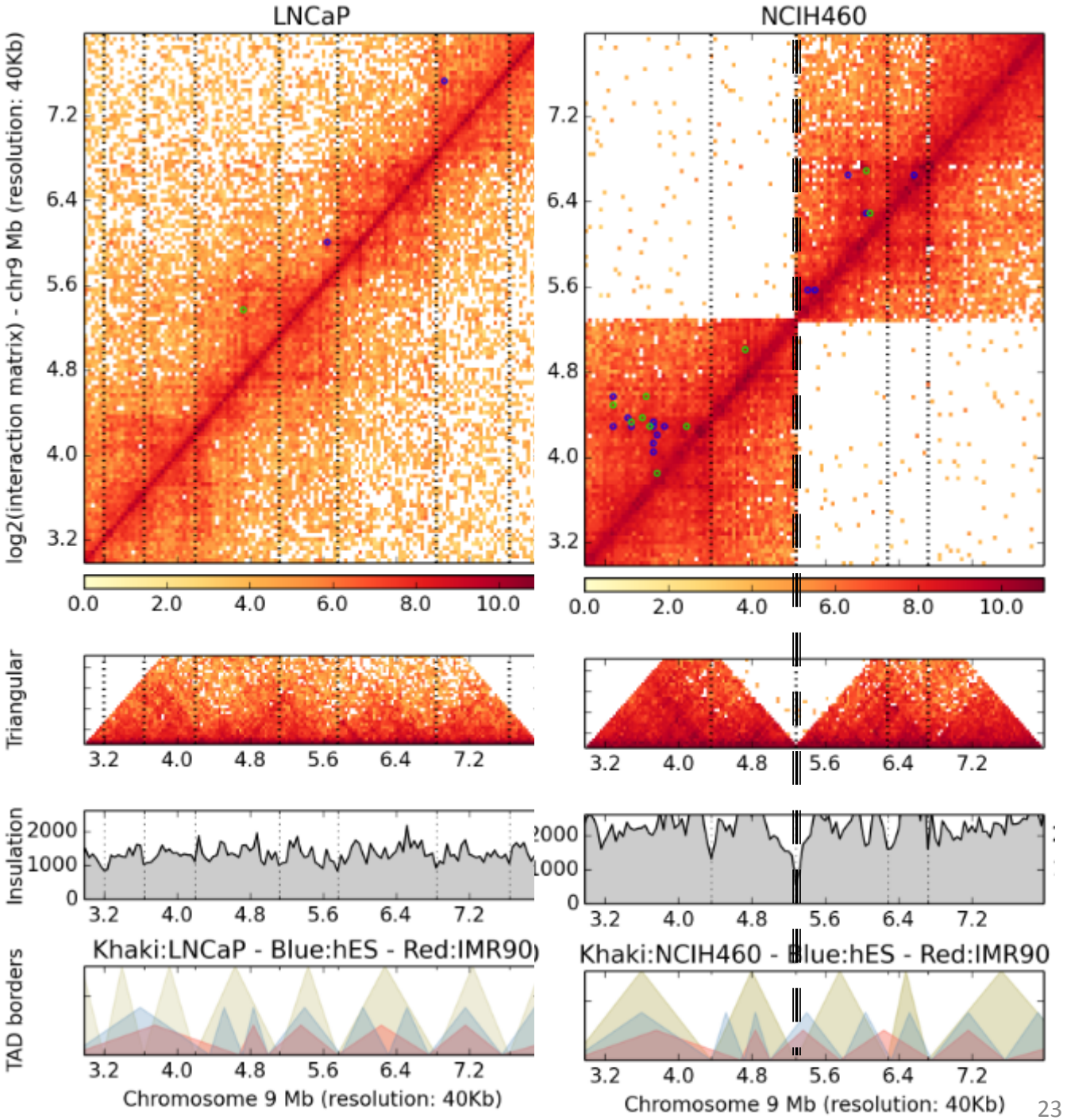
Summery of structure variations detected in thyroid cancer by Irys



Ferhat Ay

Identifying breakpoints from Hi-C data

A TAD splitting translocation on chr9



Translocations in NCIH460

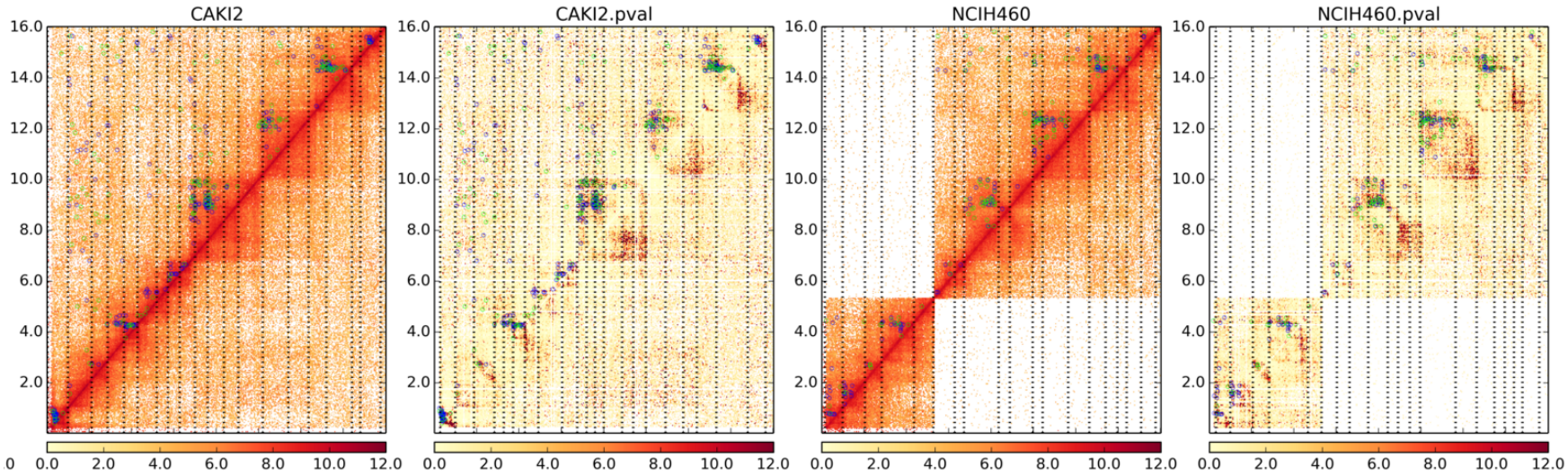
NCI-H460 [H460] (ATCC® HTB-177™)

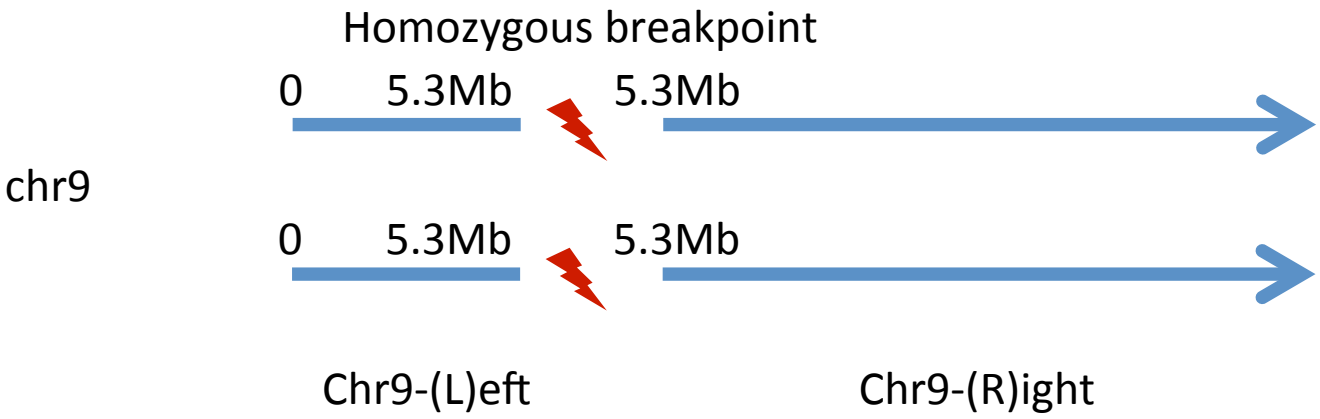
Karyotype

modal numbr = 57; range = 53 to 65.This is a hypotriploid human cell line. The modal chromosome number is 57 although cells with 58 chromosomes occurred with a comparable frequency. The frequency of higher ploidies was 1.7%. Seven marker chromosomes, der(9)t(1;9)(q21;p24), der(9)t(7;9)(p11;p22), t(10q14q), der(16)t(7;16)(q11.23;q22), a small ring (about 1/2 the size of a G chromosome) and two others, were common to all cells. Three other markers were found in some cells only. The markers, t(7;9) and t(7;16) were mostly paired. Normal N9 was absent, and N7 and N16 had only a single copy per cell. Two copies each of the X and the Y were present in all cells.

- "Normal chr9 was absent" →Translocations will show up in each and every copy of chr9
- "Normal chr7 had only single copy per cell" → No translocations for one copy of chr7

Translocations in NCIH460 – chr9





Breakpoint is very clear visually and will be easy to identify with any method

What happens to each piece? Chr9-R

ChrA	BinMidA	ChrB	BinMidB	ContactCount	Pvalue
chr7	49660000	chr9	5340000	10	2.18E-26
chr7	49700000	chr9	5340000	5	5.02E-19
chr7	49740000	chr9	5340000	12	1.78E-31
chr7	49820000	chr9	5340000	16	6.47E-29
chr7	49860000	chr9	5340000	12	6.47E-29
chr7	49900000	chr9	5340000	15	1.78E-31
chr7	49940000	chr9	5340000	15	1.78E-31

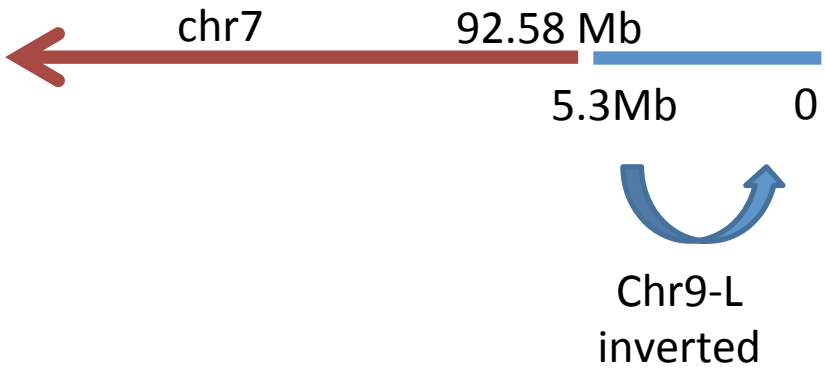
Significant inter-chr contacts from chr9-R leftmost bin

All copies of chr9-R join with chr7:49.94Mb

What happens to each piece? Chr9-L

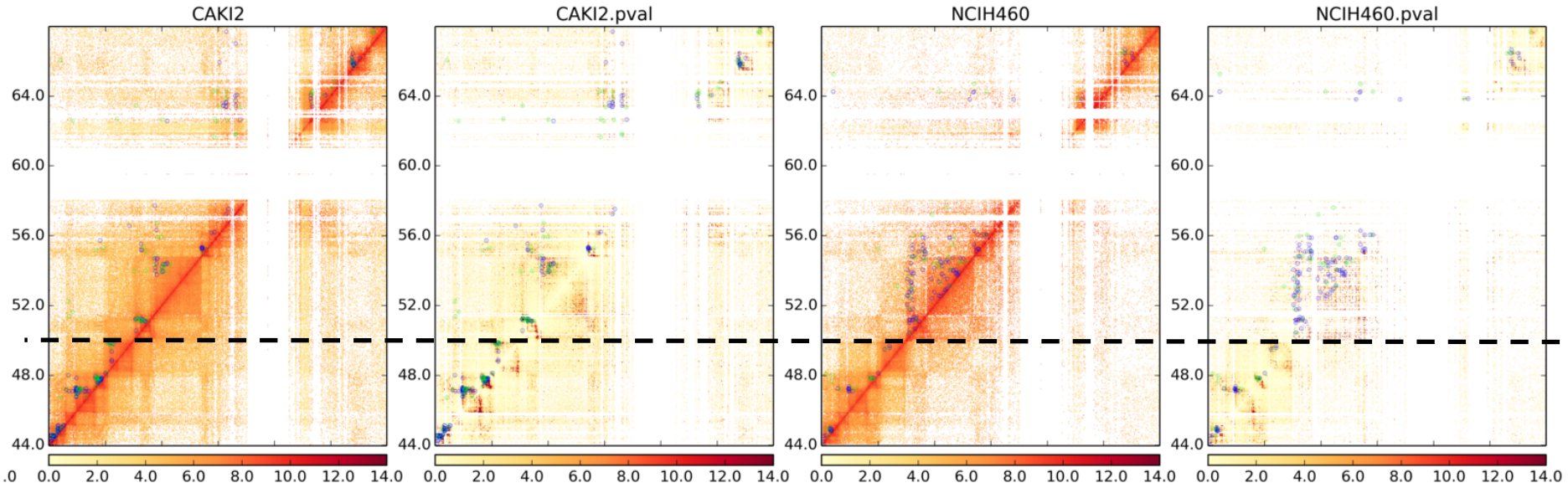
ChrA	BinMidA	ChrB	BinMidB	ContactCount	Pvalue
chr7	92300000	chr9	5260000	12	4.58E-34
chr7	92340000	chr9	5260000	15	1.78E-31
chr7	92380000	chr9	5260000	10	1.10E-36
chr7	92420000	chr9	5260000	13	6.78E-24
chr7	92460000	chr9	5260000	10	6.78E-24
chr7	92500000	chr9	5260000	22	4.16E-64
chr7	92540000	chr9	5260000	18	6.75E-53
chr7	92580000	chr9	5260000	24	2.10E-81

Significant inter-chr contacts from chr9-L rightmost bin



All copies of chr9-L join with chr7:92.58 Mb

What if we look from chr7?



chr7:49.94Mb region that joins with chr9-R

Breakpoint is not very clear visually because of the one normal chr7 that does not have it.
It will not be as easy to identify.

Challenges ahead

- Copy number differences of translocation partners either at the whole chromosome level or smaller
- Normalization carried out before copy numbers are accounted for is likely to create biases
- Heterogeneity of translocations and breakpoints among the cell population
- Identification of intra-chromosomal translocations will require taking genomic distance into effect (Are they abundant?)
- Other routine challenges:
  - Non-mappable regions (e.g., centromeres)
  - Partially mappable regions (e.g., only one half of a bin)