Link to manuscript google doc:

https://docs.google.com/document/d/1uXVsyQsi46pPwH0wOXjkDKrOgnShCzdfXadnraanGJs/edit?usp=sharing

Chromatin structure and function changes in re-arranged genomes

Gilbert lab, Dekker Lab, Yue Lab, Ay Lab

Vishnu Dileep (Gilbert lab)

Rationale

- Cancer genomes and other re-arranged genomes provide a set of random structural changes

- Use 2 methods to find high confidence breakpoints and CNVs. HiC, Irys. Some of them can be validated with RT.

- TADs (Structure) changes due to these perturbations

- Replication timing and gene expression changes (function) due to the structural changes

Data generation

Hi-C (Dekker Lab)

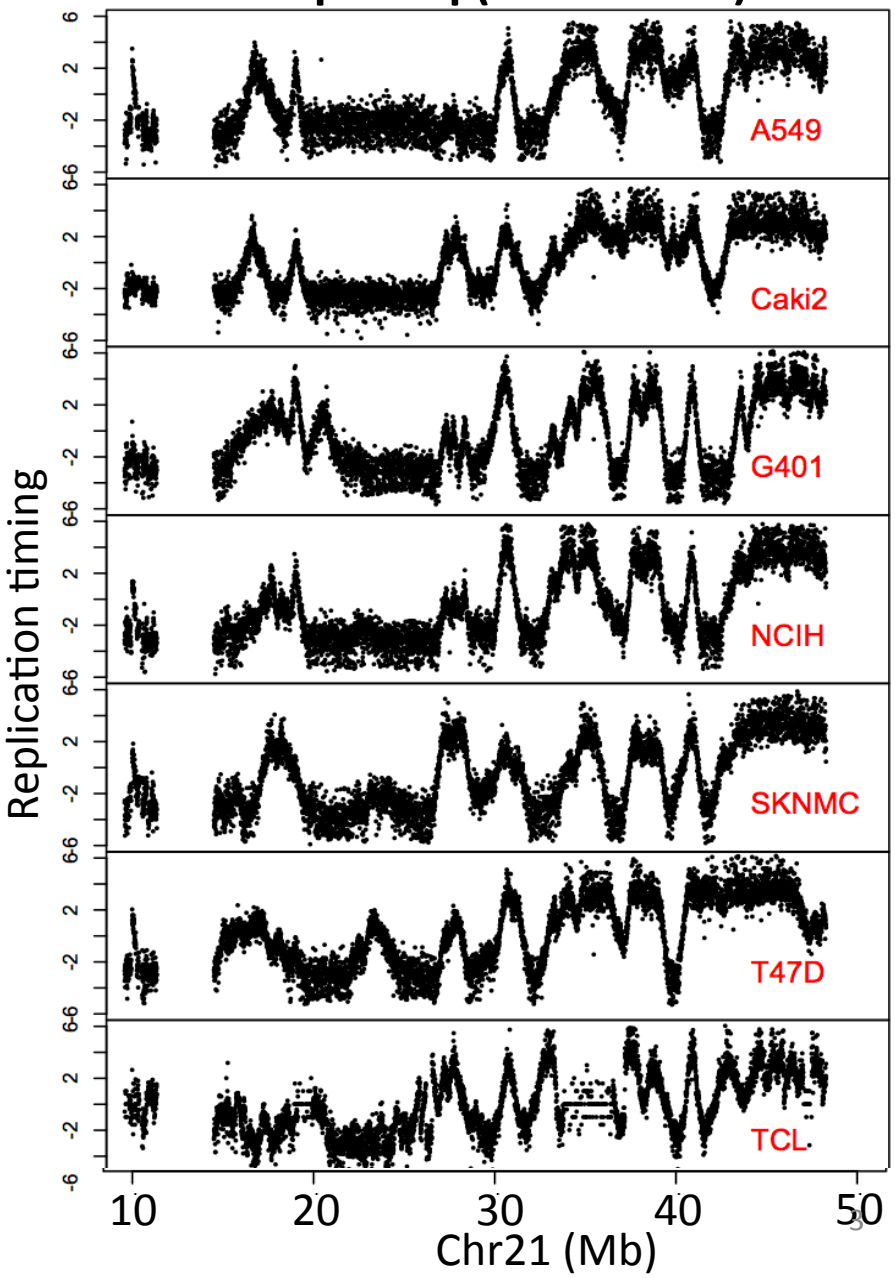
Cell line with HiC	HiC reads
CAKI2	168540398
T47D	137395622
NCIH460	155078021
SKNMC	175820923
G401	155413318
A549	136384017
TC1(re-arraged Chr21)	In progress (Hadjur/Odom lab)
PANC1	168012696
RPMI7951	189823508
LNCaP	149387648
K562	published

Irys mapping (Yue Lab)

-T47D

-Caki2

Repli-seq (Gilbert Lab)



Available Data

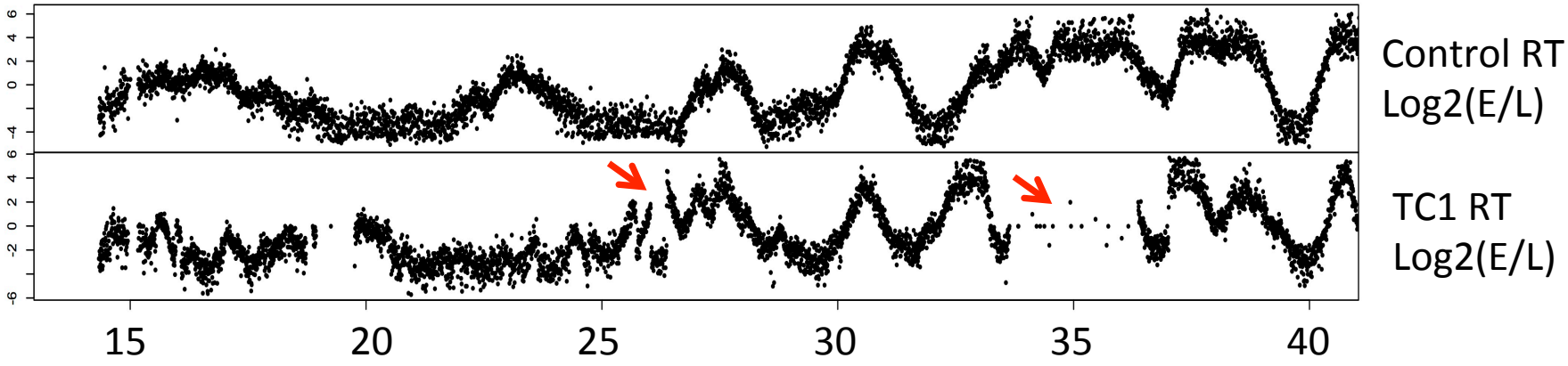
ENCODE

Cell Line	Description	Tissue	Age	Sex	DNAse I	RNA-seq
Caki-2	Carcinoma, clear cell	Kidney	69 yo	M	STAM	ENCODE3_Gingeras
T-47D	ductal carcinoma	Breast	54 yo	F	STAM	ENCODE2/Myers
NCIH460	large cell lung carcinoma	Lung	N.D.	M	STAM	ENCODE3_Gingeras
SK-N-MC					STAM	
G-401	Tumor, rhabdoid	Kidney	3 months	M	STAM	ENCODE3_Gingeras
A549	Lung carcinoma	Lung	58 yo	M	STAM	ENCODE_Rdmap
PANC-1	epithelioid carcinoma	Pancreas	56 yo	M	STAM	ENCODE_Rdmap/Myers
RPMI-7951	melanoma	Skin	18 yo	F	STAM	ENCODE3_Gingeras
LnCAP	adenocarcinoma				STAM	

Alex Urban shared structural variations in K562 identified deep paired end sequencing with large and small insert sizes

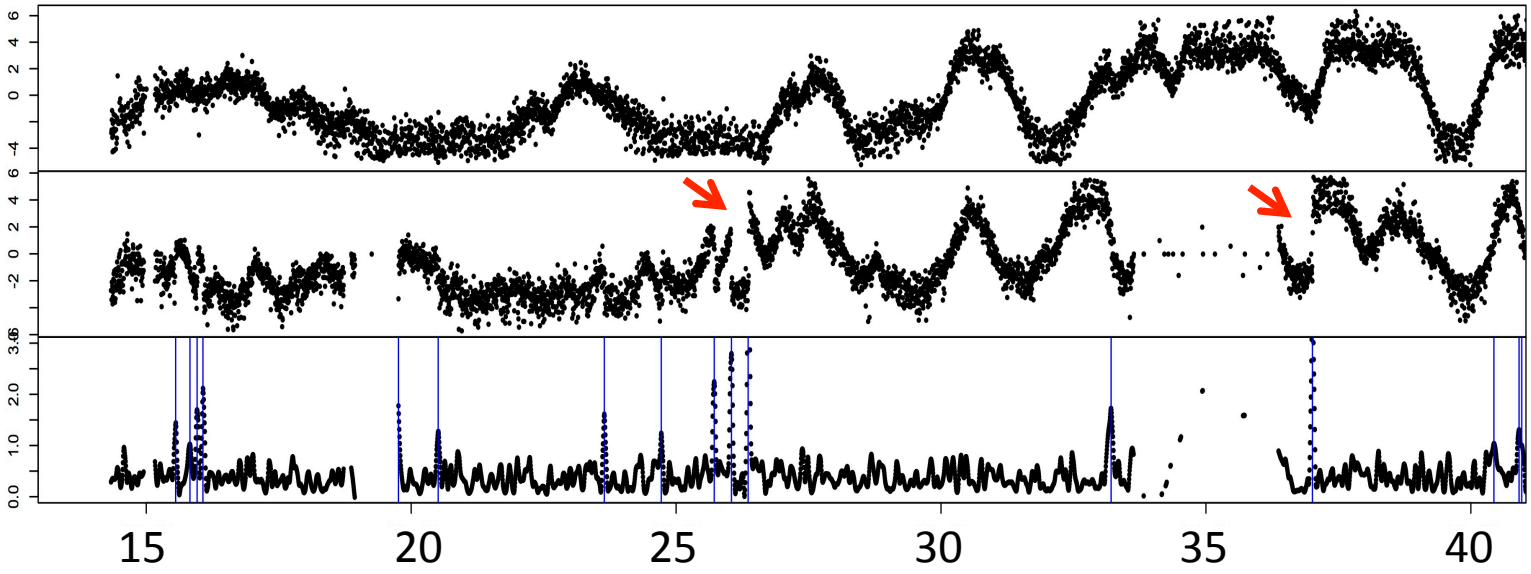
Calling breakpoints from Replication Timing Data
TC1 re-arranged human chr21 in mouse

Translocations connect DNA of different replication timing, show up as abrupt changes in RT, when mapped back to normal genome assembly



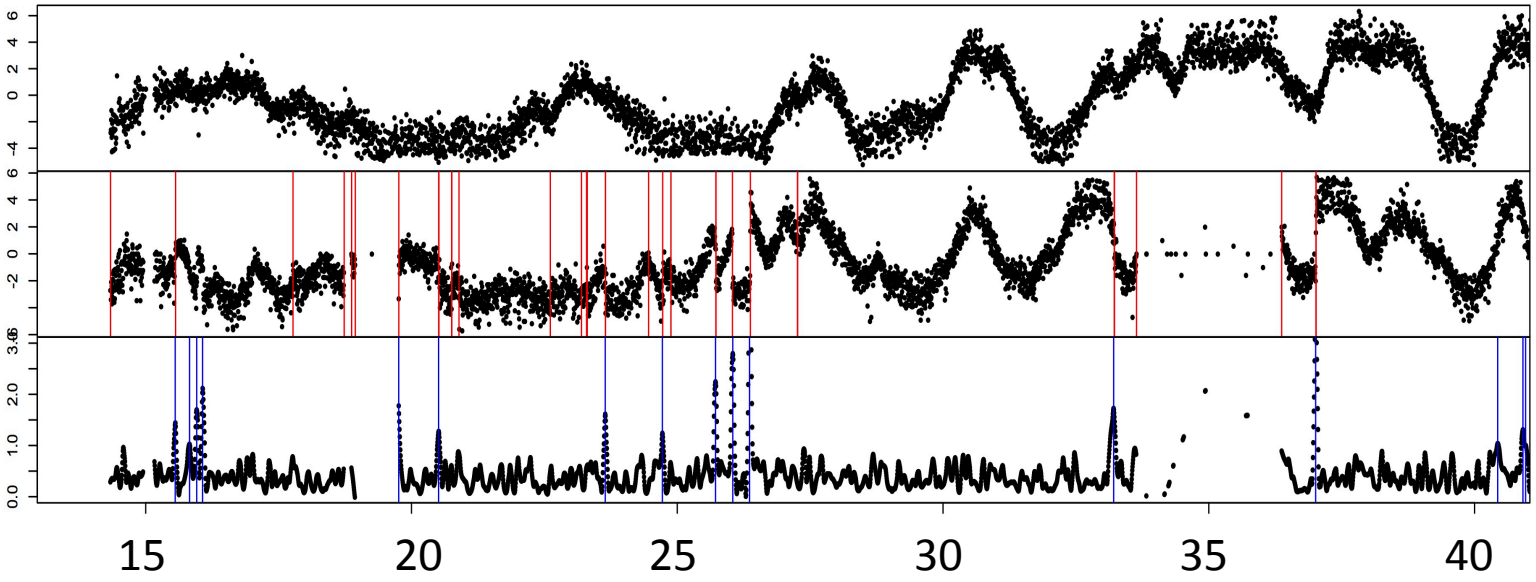
For every 5Kb bin:
RT difference= mean (upstream 10 windows)-mean (downstream 10 windows)

Looking for abrupt shifts in RT



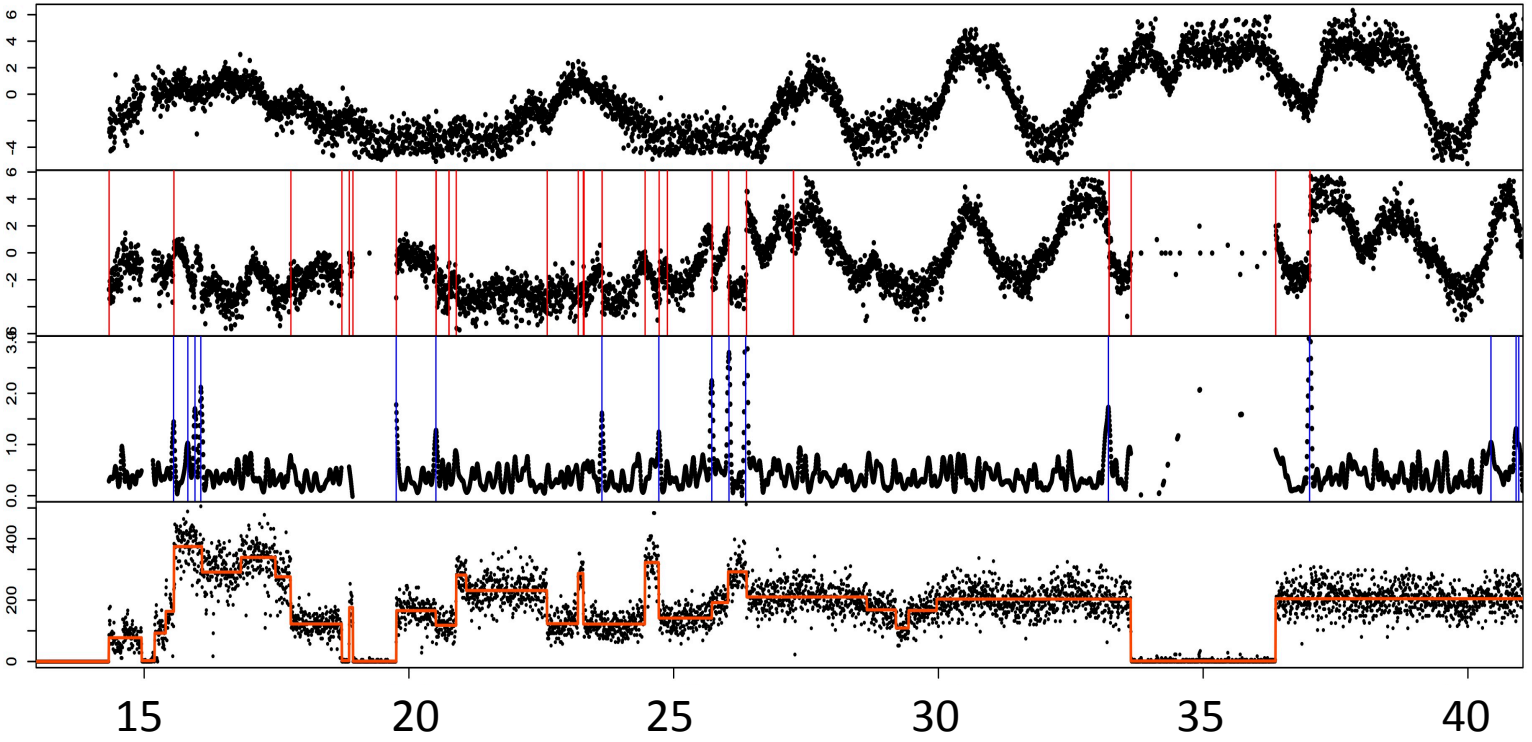
Call local maximas above certain thresholds (blue vertical lines)

Looking for abrupt shifts in RT and comparing to know breakpoints



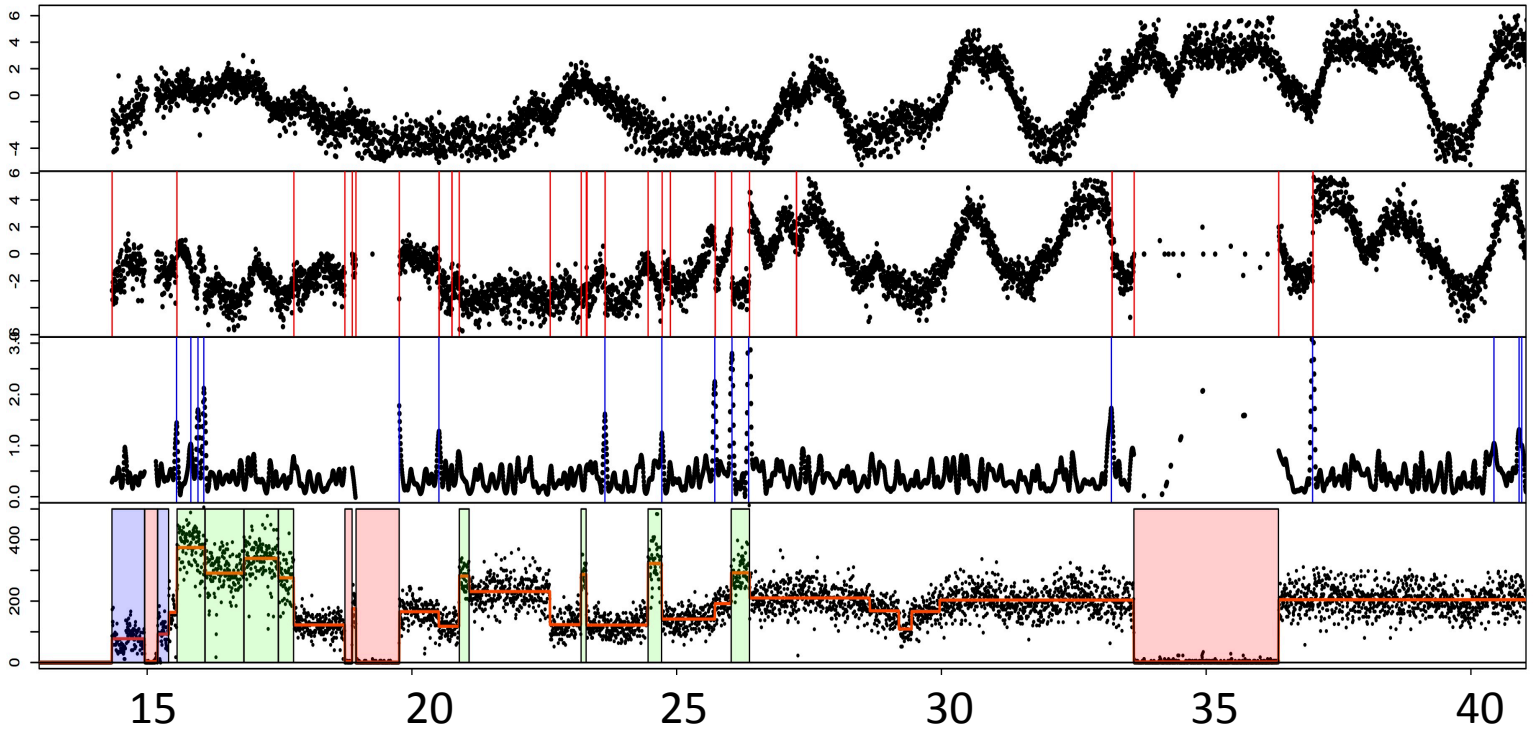
Breakpoints predicted by RT(blue vertical lines)
Known breakpoints (red vertical lines)

Calling CNVs from coverage



Breakpoints predicted by RT (blue vertical lines)
Known breakpoints (red vertical lines)

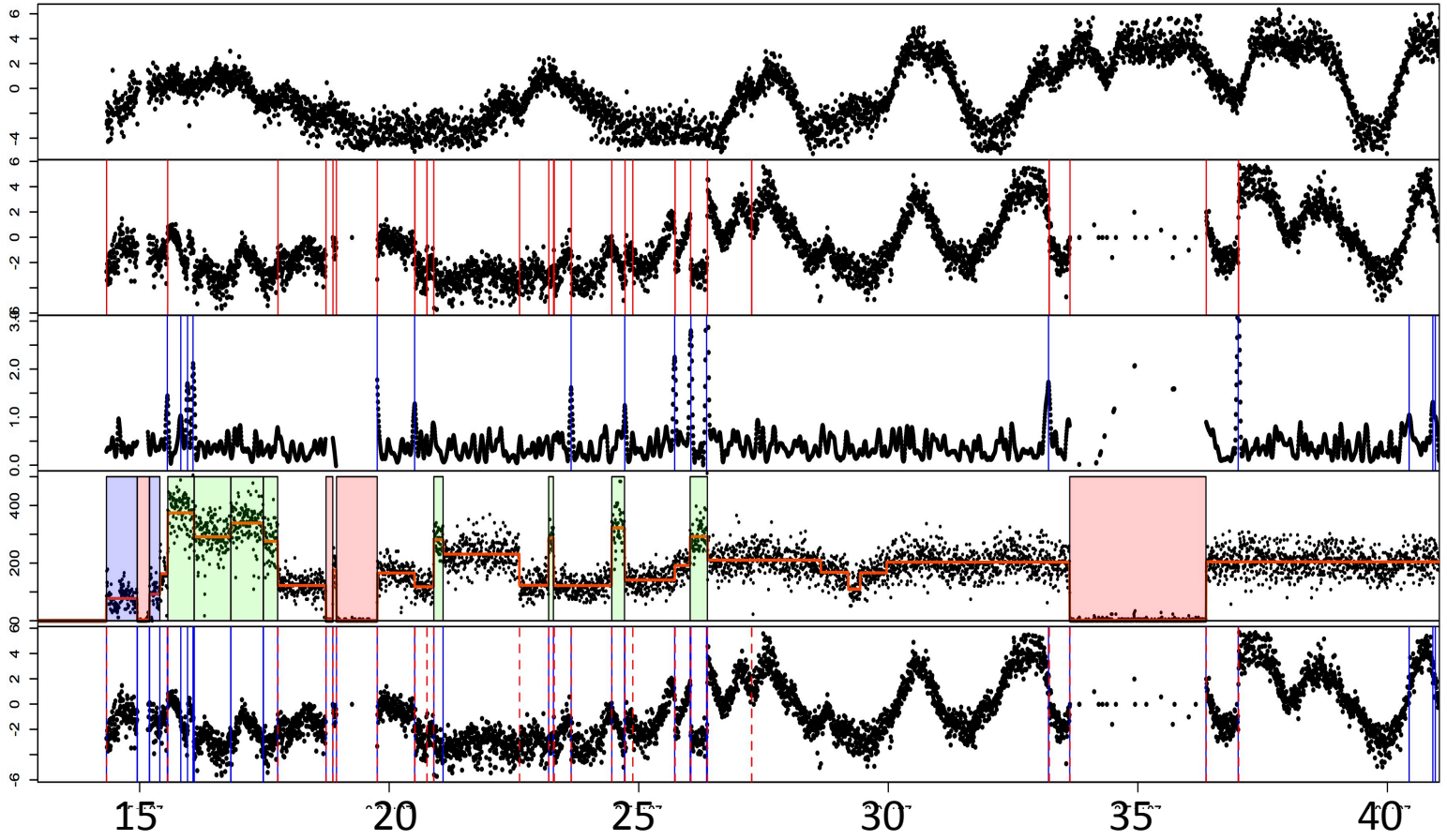
Calling CNVs from coverage



Breakpoints predicted by RT (blue vertical lines)
Known breakpoints (red vertical lines)

Deletions (Red boxes), Duplications (Green boxes), Loss(Blue boxes)

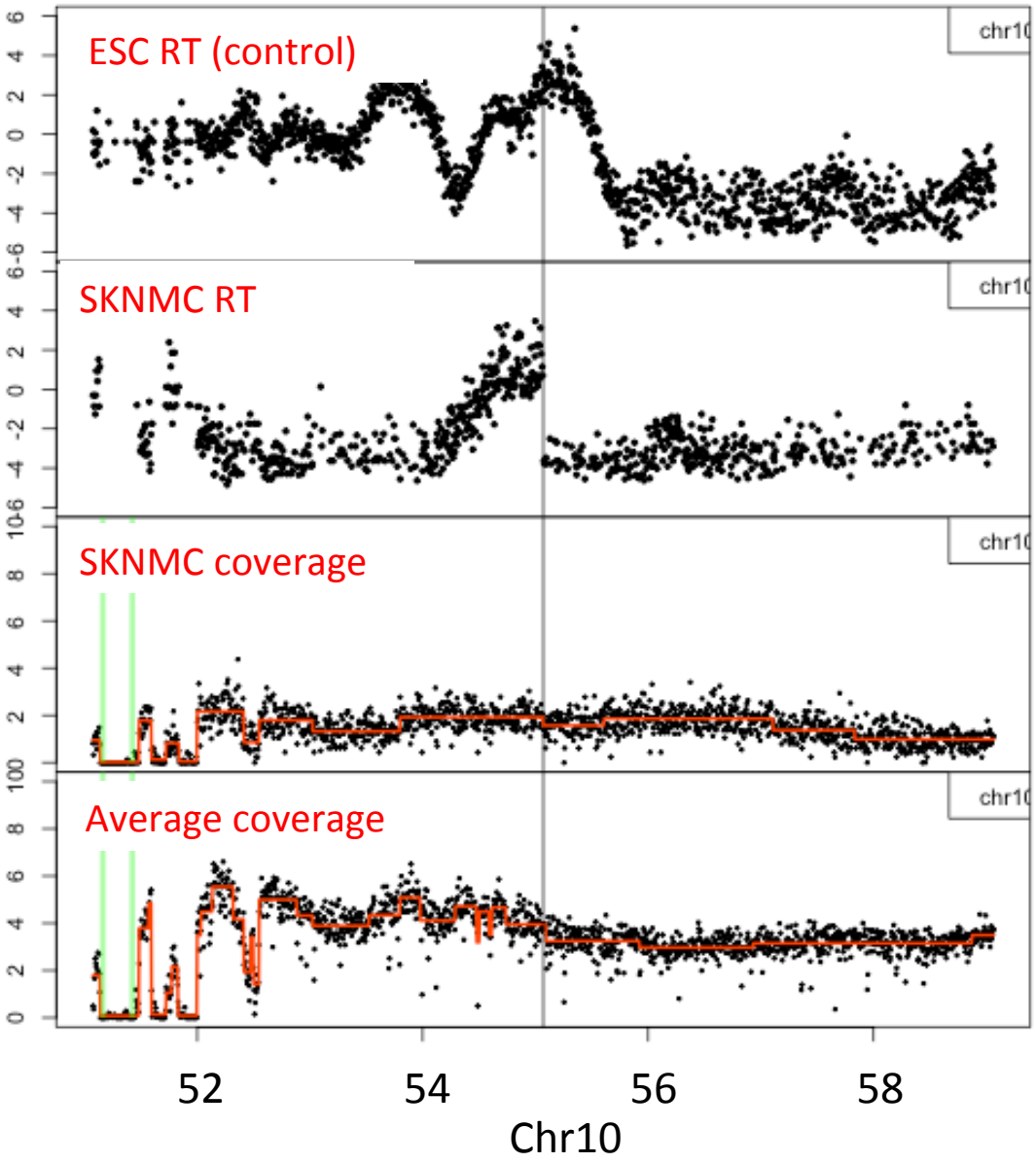
Combining abrupt shift in RT and CNV to predict most known breakpoints



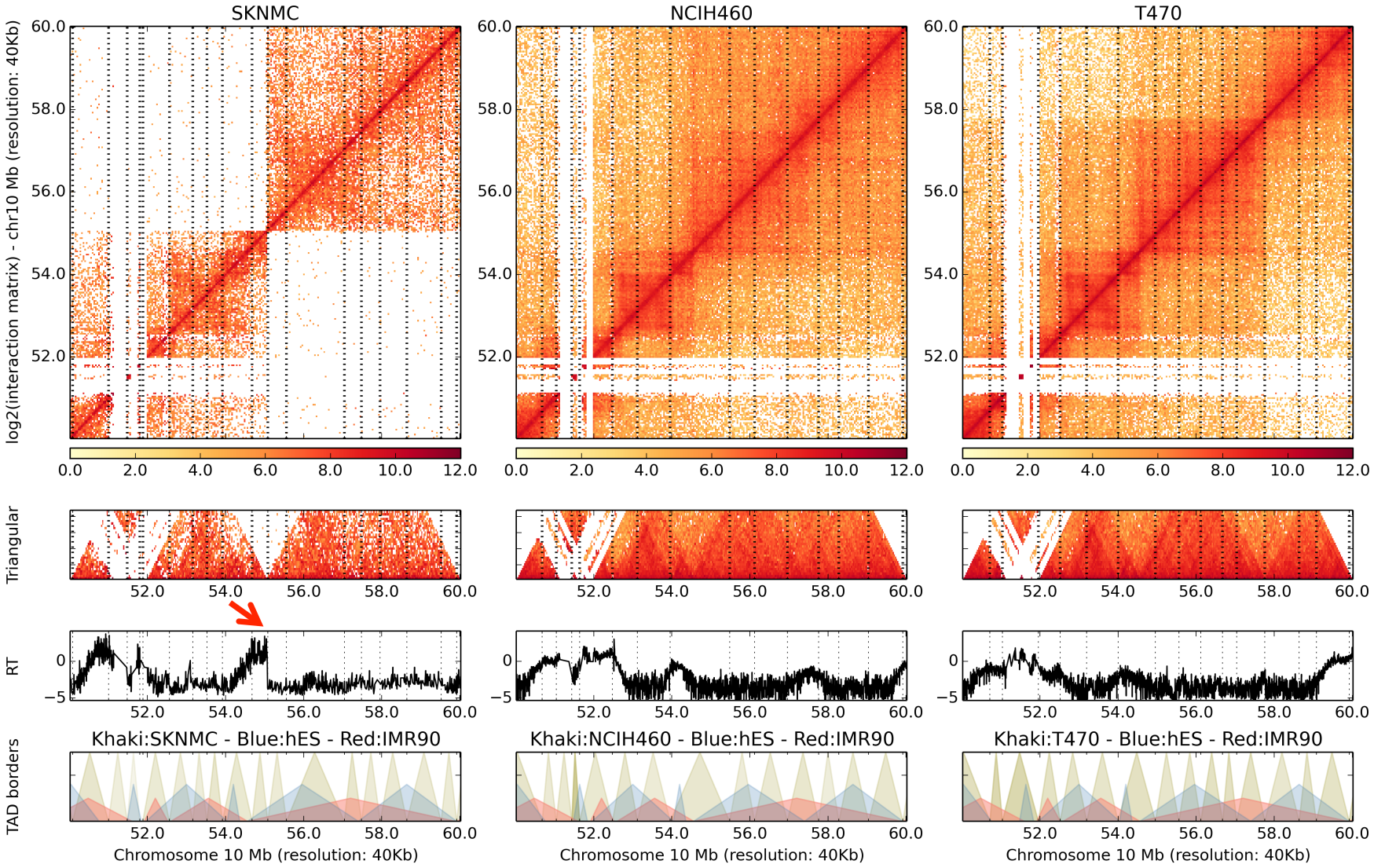
Breakpoints predicted by RT (blue vertical lines)
Known breakpoints (red vertical lines)

Deletions (Red boxes), Duplications (Green boxes), Loss(Blue boxes)

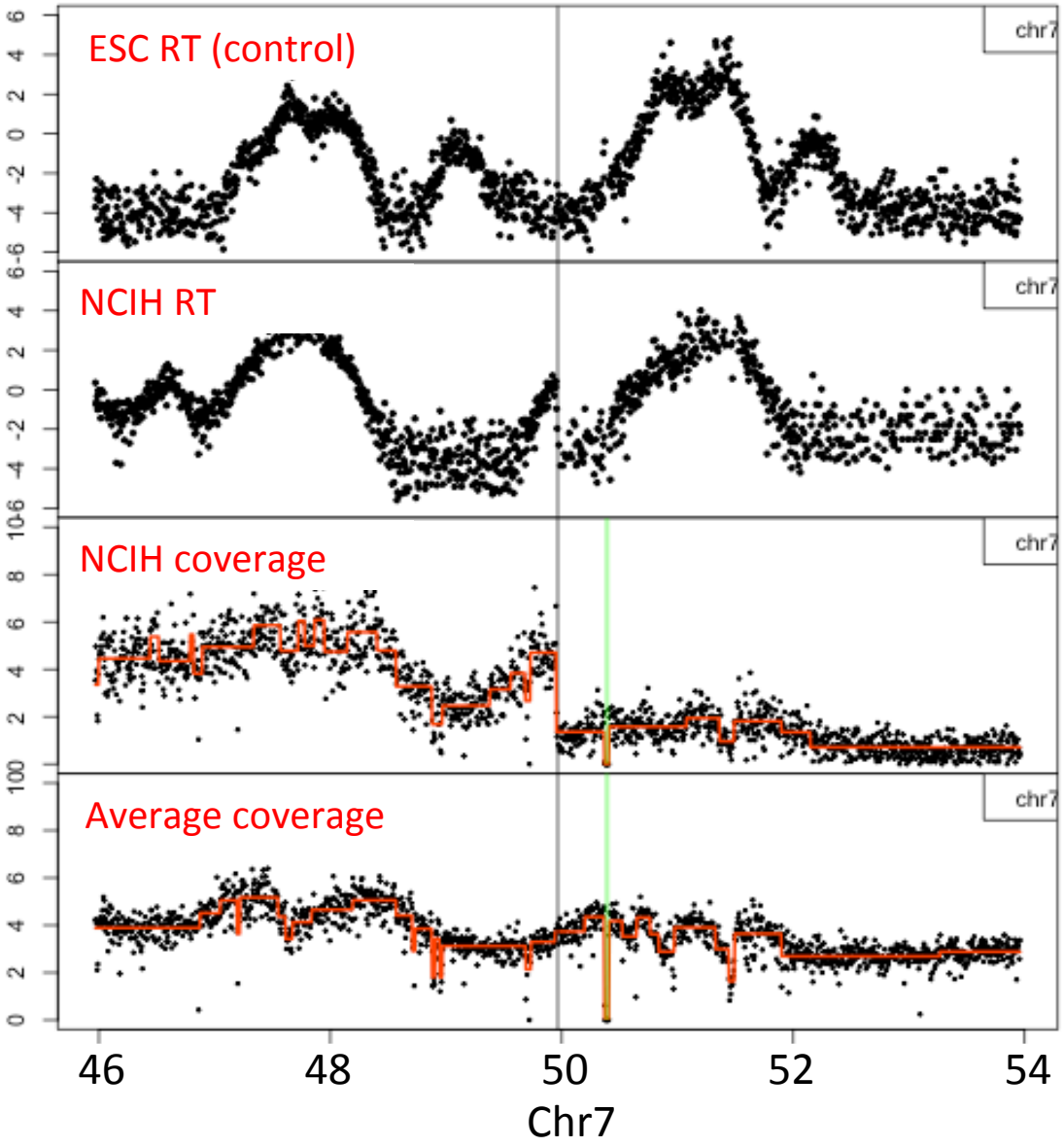
Applying above pipeline to cancer genome RT SKNMC Chr10

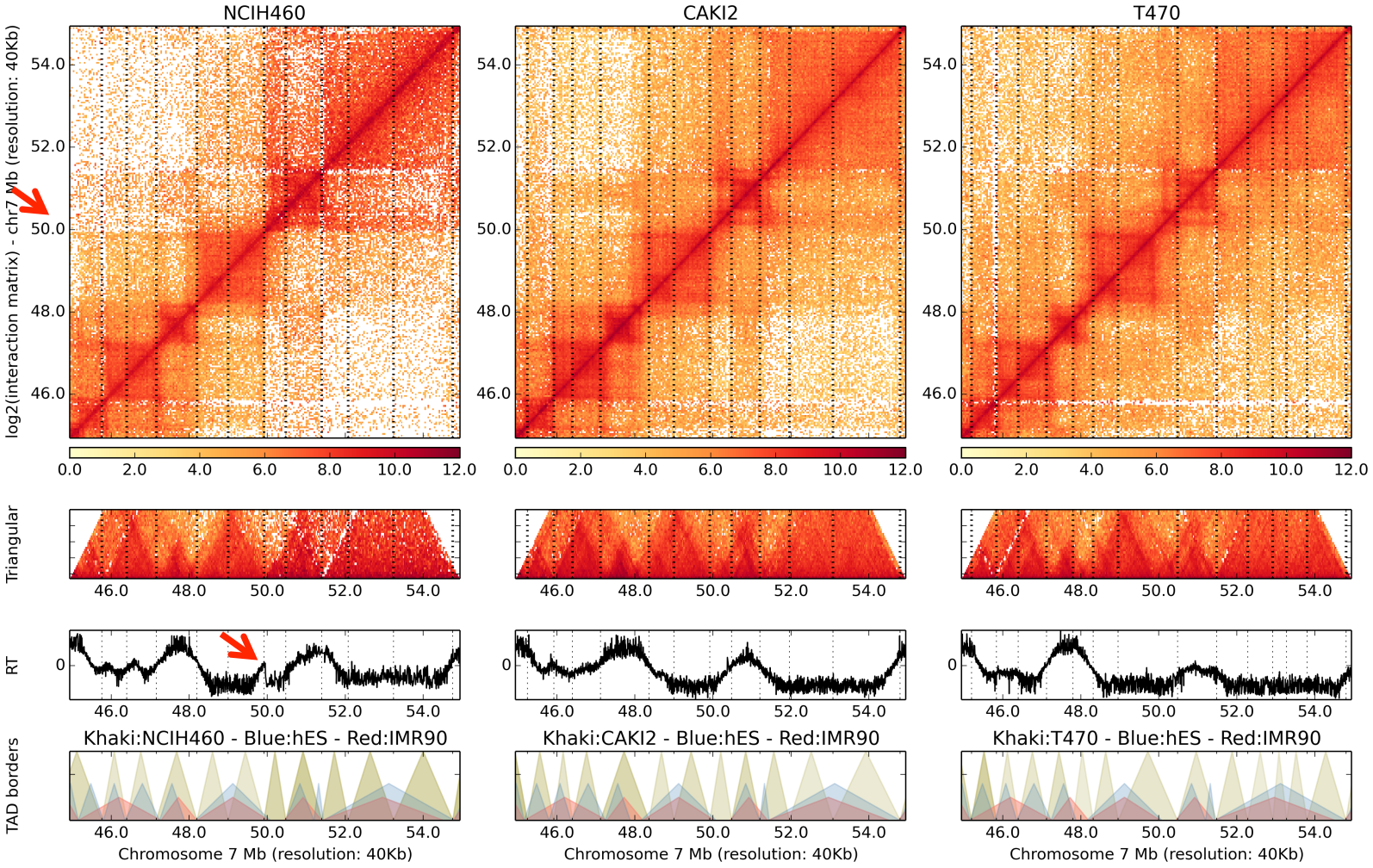


SKNMC Chr10 (HiC)

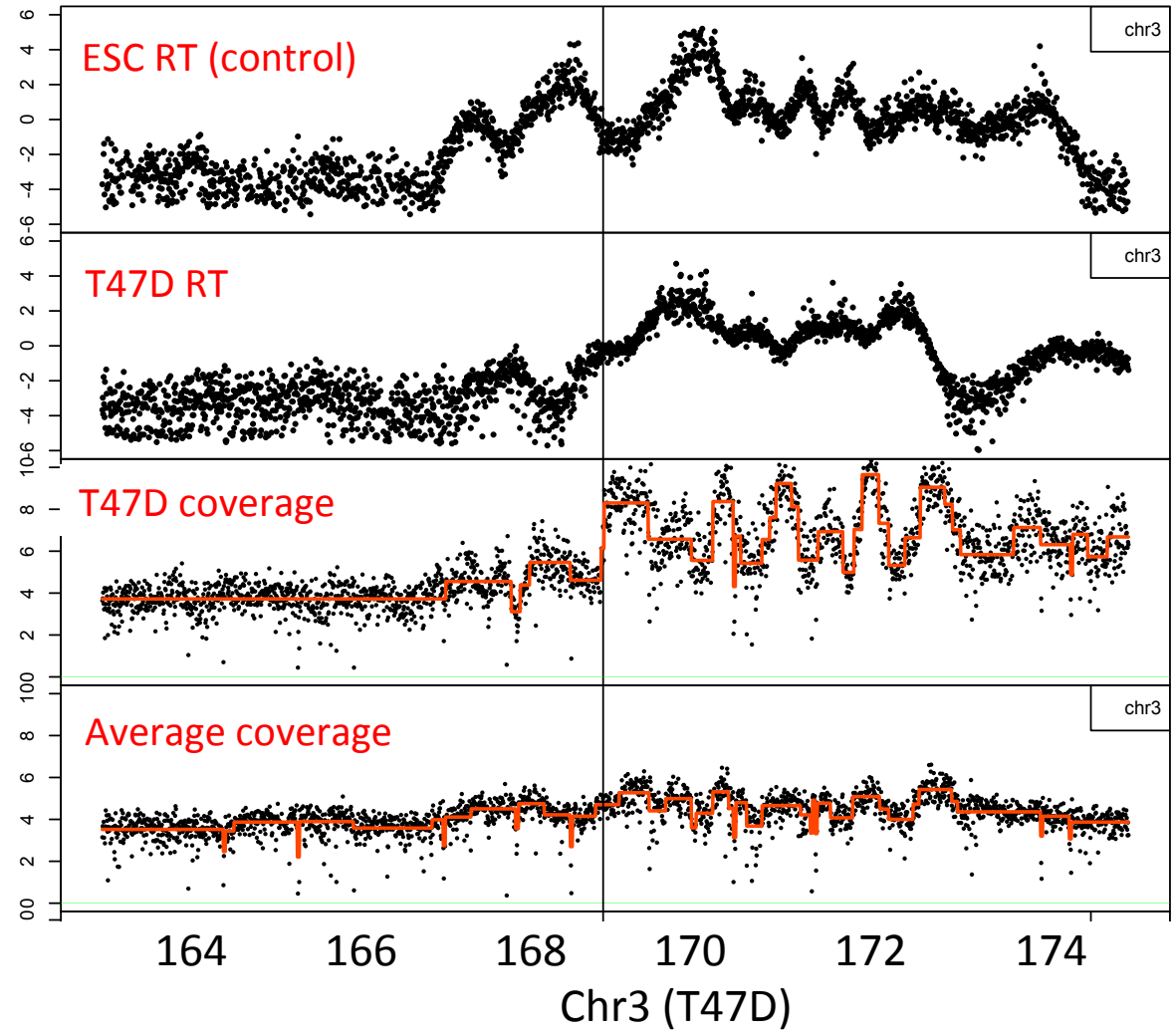


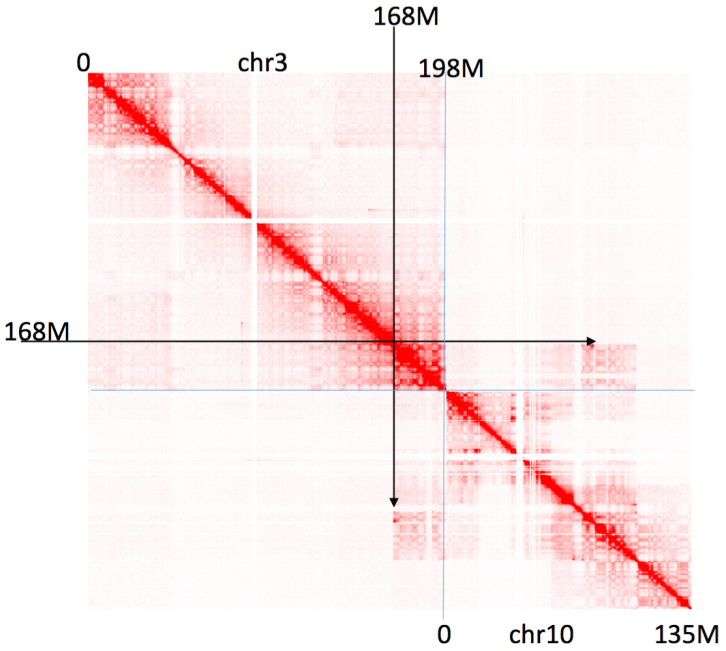
Applying above pipeline to cancer genome RT NCIH Chr7




NCIH460
CAKI2
T470
log2(interaction matrix) - chr7 Mb (resolution: 40Kb)
54.0
52.0
50.0
48.0
46.0
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Triangular
RT
0
TAD borders
Khaki:NCIH460 - Blue:hES - Red:IMR90
Khaki:CAKI2 - Blue:hES - Red:IMR90
Khaki:T470 - Blue:hES - Red:IMR90
Chromosome 7 Mb (resolution: 40Kb)

Example where breakpoint is not visible in RT T47D cell line





T47D
Chr3: 168,856,485
Chr10: 81,419,237

Identified both in Hi-C and Irys

Conclusion, Next steps

-RT can predict breakpoints when regions with different RT are connected
-Pipeline to use RT data for CNV analysis

Questions to ask:

How far does will the TAD grow in the new configuration (will it stop at the next CTCF site etc?)

How far does the RT change spread at the new location? (Pope et.al 2012 and Ryba et.al 2012 has few examples)

What happens to expression when a gene is translocated to a new location?

To do:
-Replicates for RT, RT on 3 remaining cell lines
-Can sequence some G1 cells for CNV
-Hi-C in Re-arranged chr21 (Libraries are done, sequencing soon, Hadjur and Odom Lab)
-Algorithm to predict breakpoints at HindIII resolution (Ferhat, Noam (Dekker lab)
- Possibly more cells for Irys (Jie, Yu Lab)

Acknowledgement

- Dave Gilbert
- Ferhat Ay (Hi-C analysis)
- Feng Yue, Jie Xu (Irys mapping)
- Job Dekker , Noam Kaplan (Hi-C data, analysis)
- John Stam, Rajinder Kaul (cell lines)
- Susan Hadjur, Duncan Odom, Pezic Dubravka, Christina Ernst (re-arranged chr21 Hi-C)
- Alexander Urban, Bo Zhou (K562 SVs)
- Bill Noble, Gurkan Yardimci