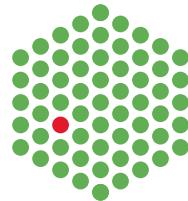


Detection of SV using single-cell approach and Mosaicatcher

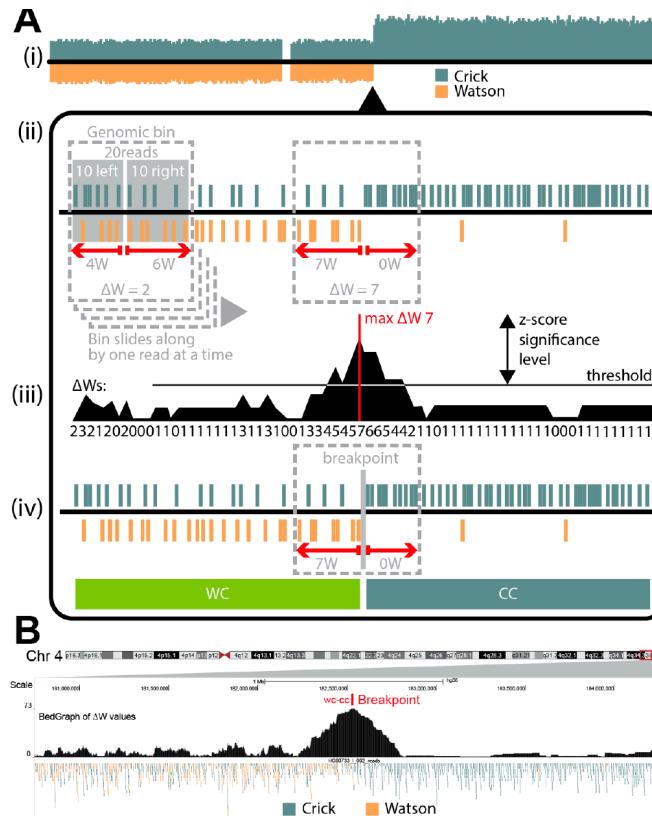


Recap of Day3 : BreakpointR

Model system

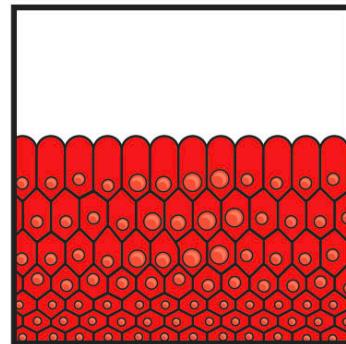


RPE1 WT



Recap of Day3 : Subclonal SVs revisited!! (WT BM510 mixture VAF 20% vs WT)

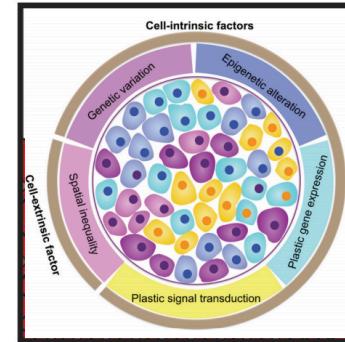
Real
world



Normal cells

Model
system

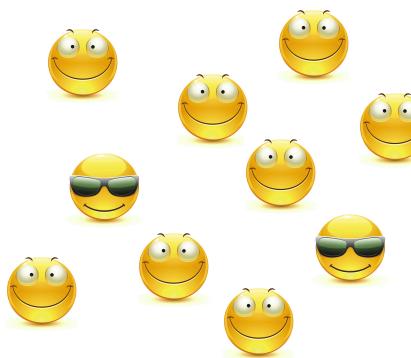
WT



Cells forming a tumour

VAF20% BM510

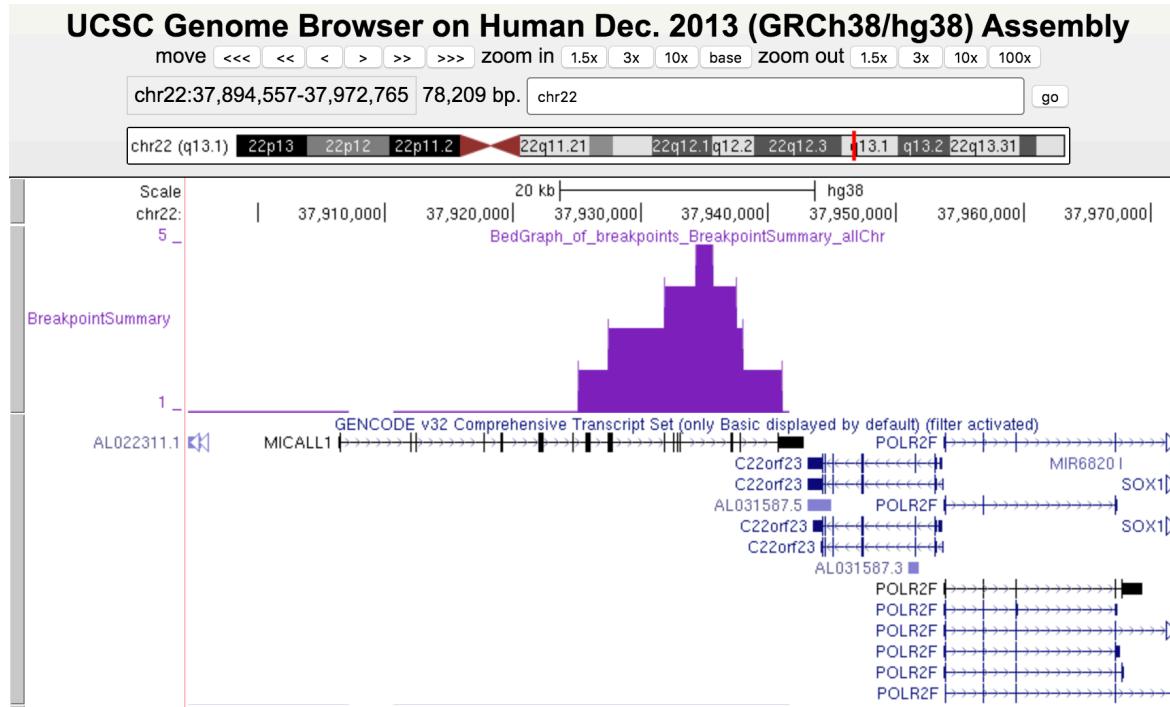
vs



Recap of Day3 : Find subclonal recurrent breakpoint in chr22!!

- STEP1 : Run breakpointR using following input folder and chr22
 - inputfolder = “/scratch/jeong/pipeline_20190625_RPE_mixture/bam/RPE/selected/”
 - outputfolder = “BreakpointR_results_100cells”
 - chromosomes = ‘chr22’
- STEP2: Compare breakpoint with the Delly based breakpoint from WGS data (hint: /breakpoints/breakPointSummary.txt)
- STEP3: Upload bedgraph file to the UCSC browser track and see which gene is there around translocation breakpoint
- STEP4: Open question: Using overview plot and breakpointR result of 100 cells (mixture), can you guess which cells are come from BM510 (20 cells) and WT (80 cells)?

BreakpointR detects translocation breakpoint in chr22 (VAF 20%)



Comparison of Delly and BreakpointR based translocation breakpoint in chr22!!

Delly

- CHROM, POS : [chr22, 37934425], [chr13, 20412250]
- MAPQ : mapping quality [60]
- GT : genotype [0/1] [0/0]
- GL : log10-scaled genotype likelihoods for RR, RA, AA genotypes [-23.6655,0,-90.8635] [0,-7.5218,-88.7961]
- GQ : Genotype Quality [10000] [75]
- FT : Per-sample genotype filter [PASS] [PASS]
- RC : Raw high-quality read counts for the SV [0] [0]
- RCL : Raw high-quality read counts for the left control region [0] [0]
- RCR : Raw high-quality read counts for the right control region [0] [0]
- CN : Read-depth based copy-number estimate for autosomal sites [-1] [-1]
- DR : # high-quality reference pairs [25] [23]
- DV : # high-quality variant pairs [5] [0]
- RR : # high-quality reference junction reads [27] [25]
- RV : # high-quality variant junction reads [9] [0]
- PRECISE/IMPRECISE :SVs refined using split-reads [PRECISE]
- PASS/LowQual : genotype quality [PASS]



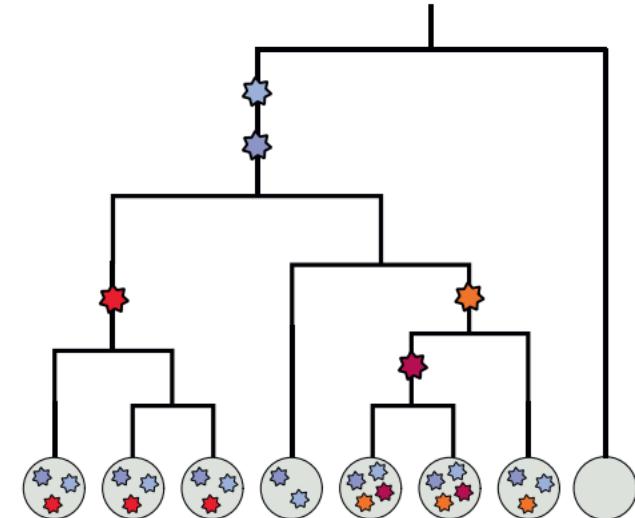
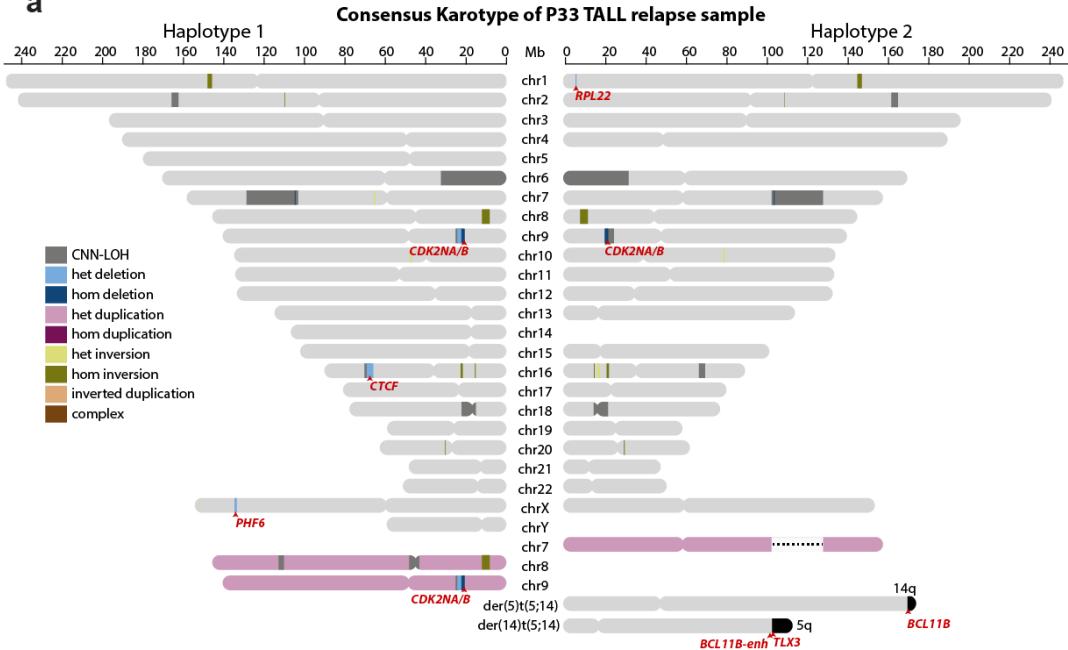
BreakpointR

```
seqnames start end CI.start CI.end genoT filenames
chr22 17553352 17567000 12693467 17642952 wc-cc Cell001.sort.mdu
chr22 37925136 37935646 37899037 37944092 cc-wc Cell001.sort.mdu
chr22 17876897 17876936 11974156 17949682 wc-cc Cell003.sort.mdu
chr22 38043988 38071100 38010266 38159426 cc-wc Cell003.sort.mdu
chr22 17544084 17554974 12293238 17583405 wc-ww Cell004.sort.mdu
chr22 17579347 17580826 12341586 17604242 wc-ww Cell006.sort.mdu
chr22 37934301 37938044 37901540 37952181 ww-wc Cell006.sort.mdu
chr22 17690760 17696506 12176341 17738540 wc-ww Cell008.sort.mdu
chr22 17524264 17525992 12402384 17564272 wc-ww Cell009.sort.mdu
chr22 37931913 37937498 37862362 37940415 ww-wc Cell009.sort.mdu
chr22 17523641 17532812 12411496 17567615 wc-ww Cell010.sort.mdu
chr22 37927507 37941108 37897587 37950947 ww-wc Cell010.sort.mdu
chr22 17068018 17104535 11975700 17278662 wc-cc Cell015.sort.mdu
chr22 18046125 18057965 12331299 18091681 wc-cc Cell018.sort.mdu
chr22 37910696 37941561 37837157 37943494 cc-wc Cell018.sort.mdu
chr22 17594004 17652660 12040556 17721340 wc-ww Cell020.sort.mdu
chr22 37855356 37907122 37814089 37985502 ww-wc Cell020.sort.mdu
chr22 17724174 17742638 11855691 17788145 wc-cc Cell021.sort.mdu
chr22 17286594 17295368 12363227 17315410 wc-ww Cell024.sort.mdu
chr22 17455607 17497839 12195010 17541676 wc-ww Cell026.sort.mdu
chr22 17169240 17172328 16037693 17192657 wc-cc Cell027.sort.mdu
chr22 48286474 48305477 48268423 48337762 wc-ww Cell028.sort.mdu
```

*Strand-seq can catch low variant allele frequency SV breakpoint,
but it gives confidence interval rather than precise breakpoint*

What if the genome has bunch of different SVs and complex subclonality?

a



Sanders et al. 2019

Genome Biology, 2016

Can we find all potential SVs and subclonal structures from the sample?

Mosaicatcher towards the automatic single-cell SV calling and clustering

<https://github.com/friendsofstrandseq/pipeline>

The screenshot shows the GitHub repository for Mosaicatcher. The logo features the word "MOSAIC" in large, bold, black letters, with "CATCHER" below it. A stylized, colorful DNA double helix graphic is integrated into the letter "I". Below the logo, a sub-headline reads: "Structural variant calling from single-cell Strand-seq data - summarized in a single [Snakemake](#) pipeline." A section titled "Overview of this workflow" provides a brief description of the process, mentioning the use of Snakemake and listing seven steps. The GitHub interface includes standard navigation elements like a back arrow and a search bar.

Structural variant calling from single-cell Strand-seq data - summarized in a single [Snakemake](#) pipeline.

Overview of this workflow

This workflow uses [Snakemake](#) to execute all steps of MosaiCatcher in order. The starting point are single-cell BAM files from Strand-seq experiments and the final output are SV predictions in a tabular format as well as in a graphical representation. To get to this point, the workflow goes through the following steps:

1. Binning of sequencing reads in genomic windows of 100kb via [mosaicatcher](#)
2. Normalization of coverage with respect to a reference sample (included)
3. Strand state detection (included)
4. Haplotype resolution via [StrandPhaseR](#)
5. Multi-variate segmentation of cells ([mosaicatcher](#))
6. Bayesian classification of segmentation to find SVs using [mosaiClassifier](#) (included)
7. Visualization of results using custom R plots (included)

*Korbel group,
EMBL*



Ashley
Sanders



Ashley
Sanders

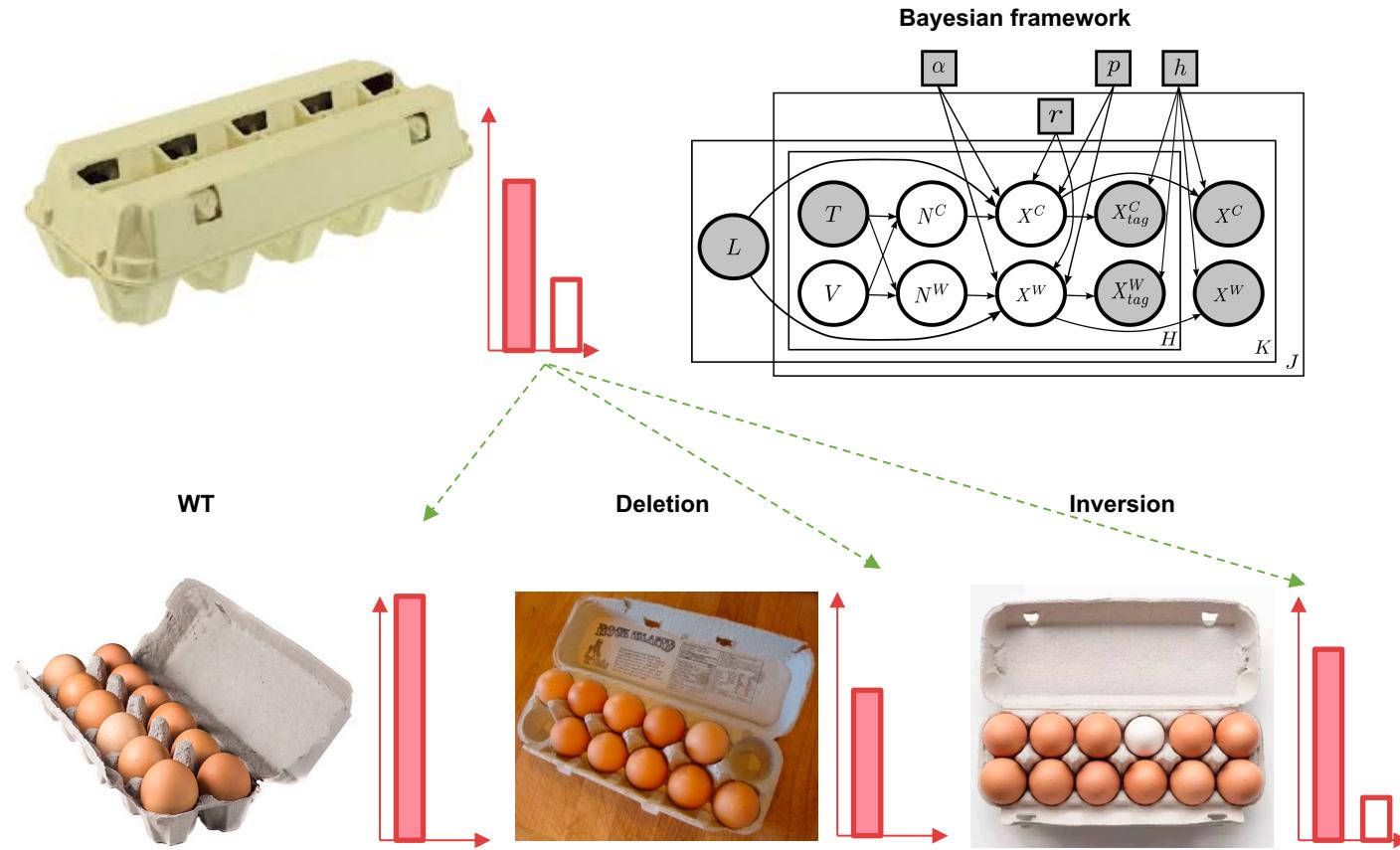
*Marschall group,
MPI informatics*



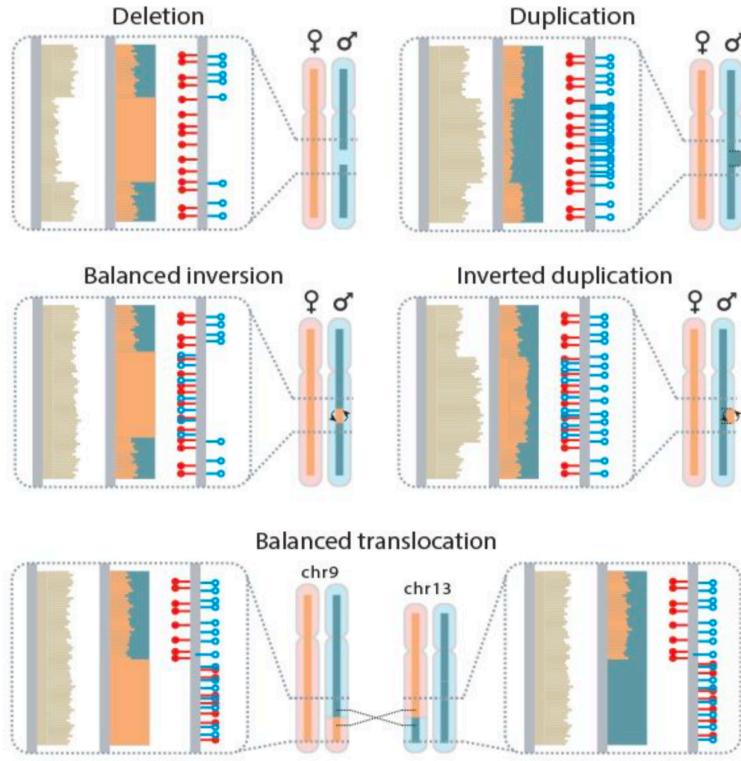
David
Porubsky

Maryam
Ghareghani

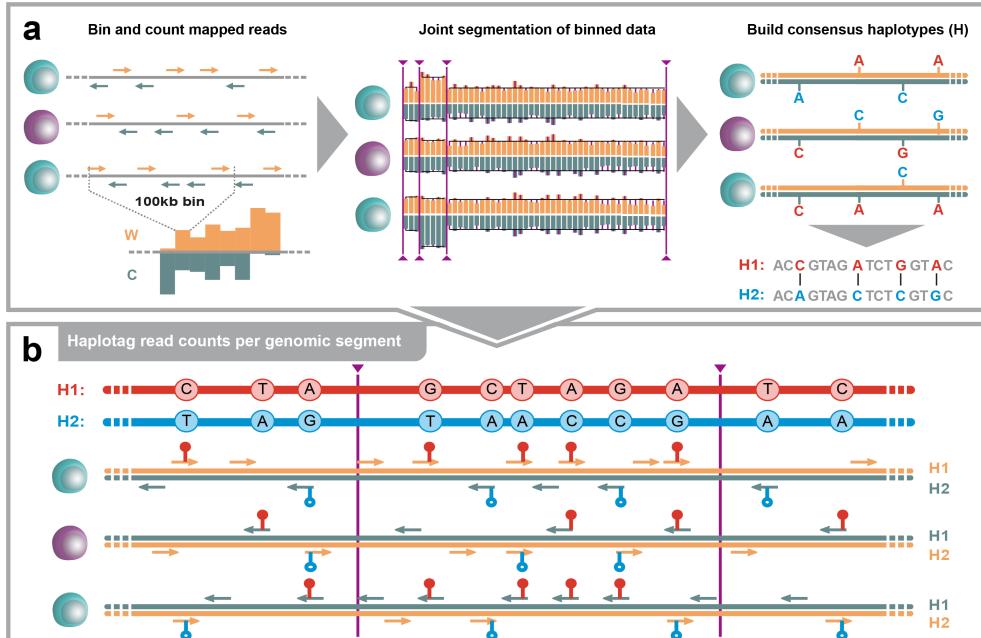
MosaiCatcher to automatically detect abnormal pattern in the genome



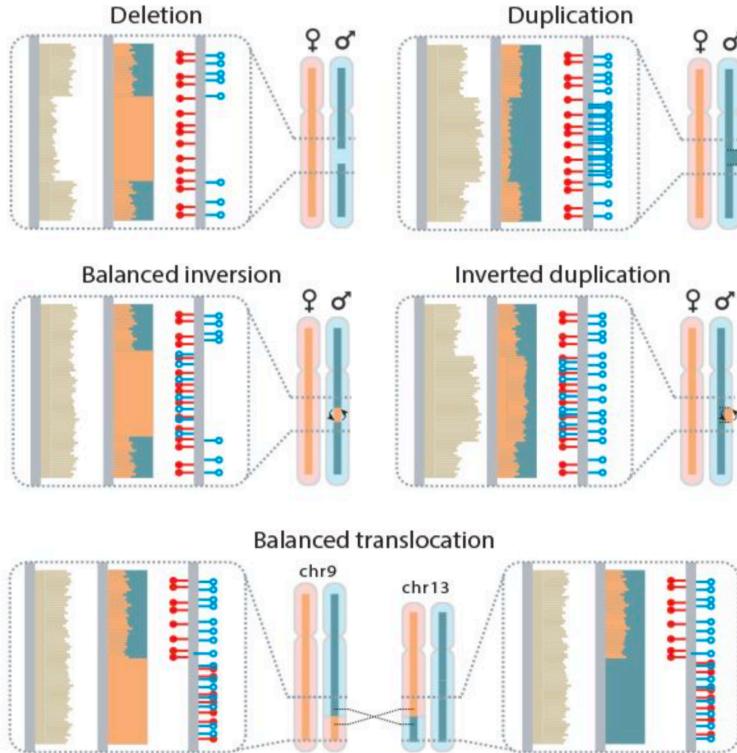
MosaiCatcher to detect characteristic footprints of SVs from Strand-seq data



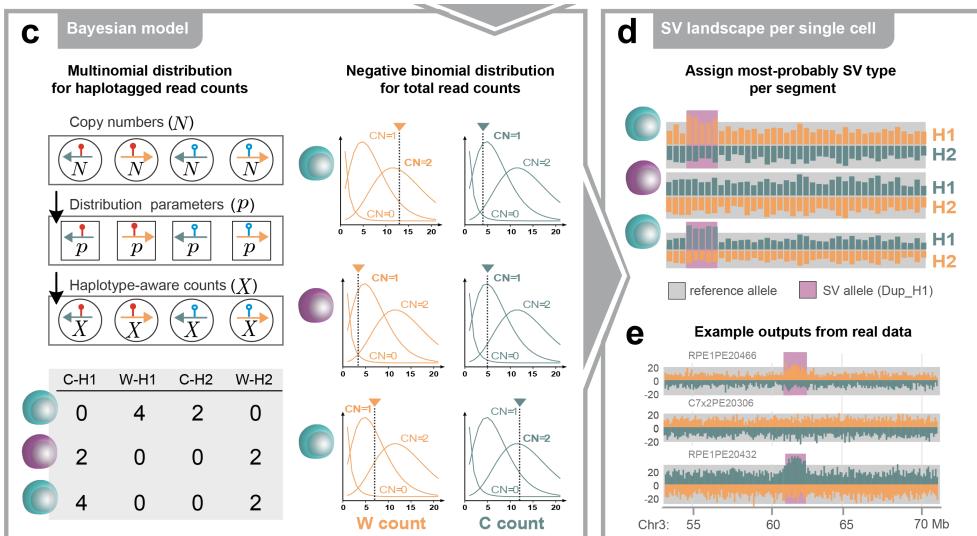
- Input: single-cell BAM files
- Workflow management: Snakemake



MosaiCatcher to detect characteristic footprints of SVs from Strand-seq data

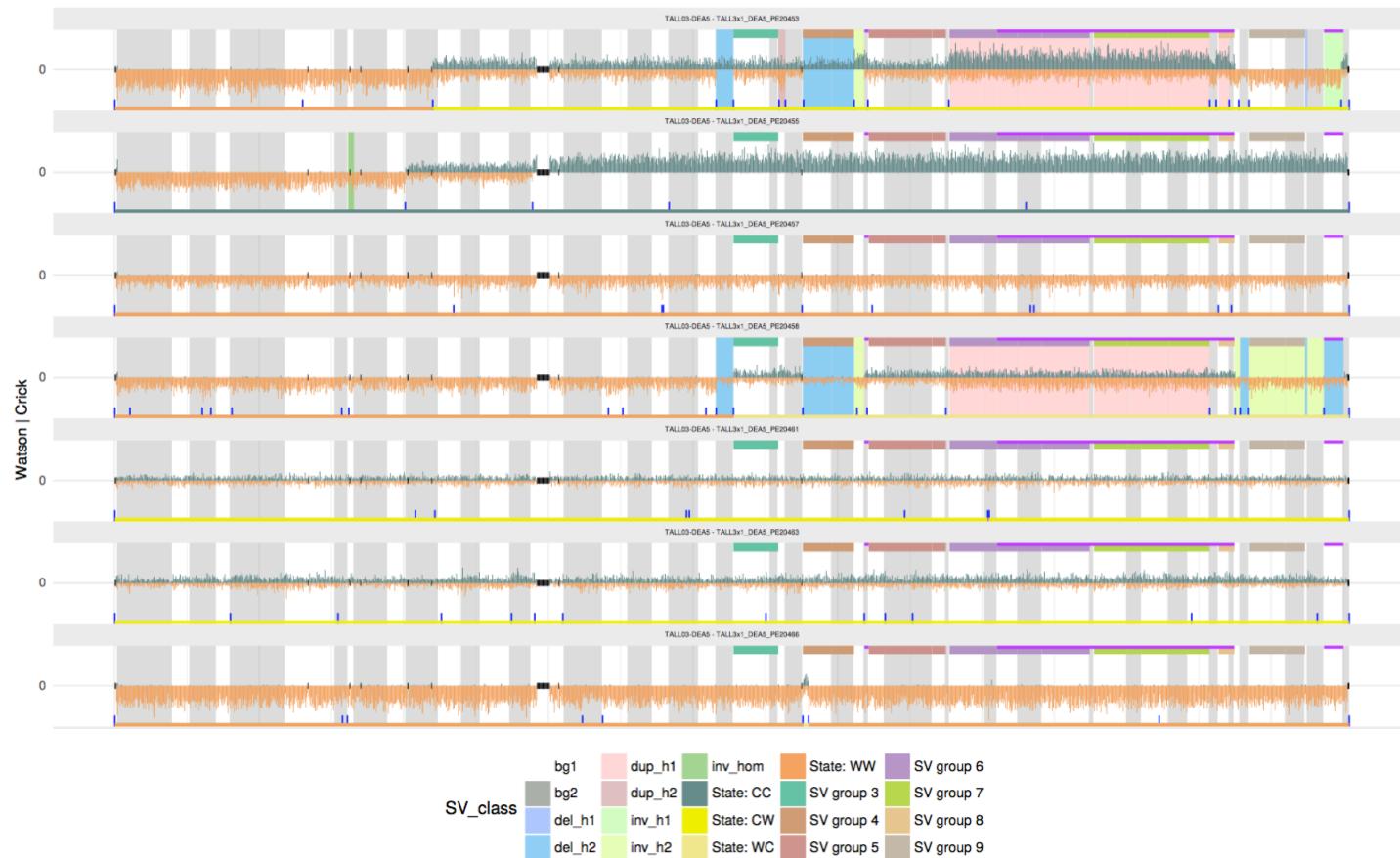


- Input: single-cell BAM files
- Workflow management: Snakemake

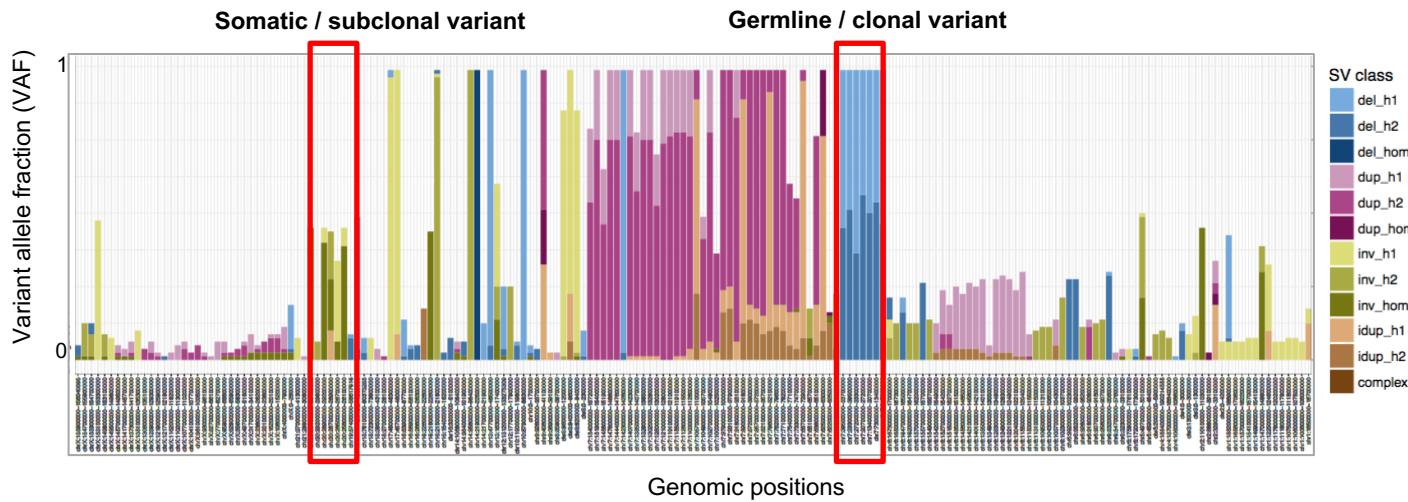


Chromosome plot with SVs called by MosaiCatcher framework

Strand-seq from T-ALL PDX (P1) Chromosome6

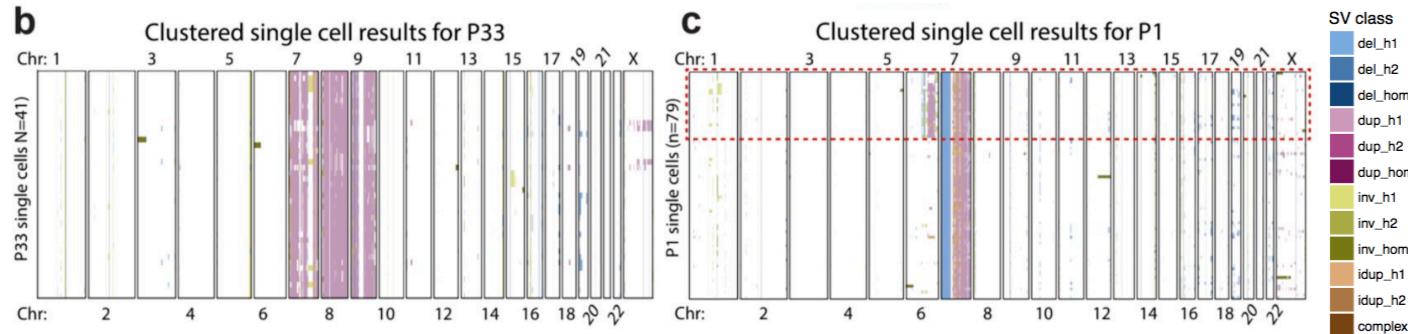


Variant allele fraction of SVs called by MosaiCatcher framework



- If the VAF is close to 1, the SVs are expected to be germline variant
- If the VAF is below 1, the SVs are expected to be somatic variant
- If the SVs only detected by one cell, it can be rare SV event, or an SCE (sister chromatid exchange) event
- SCEs happen independently in each single cell, and unlike SVs, SCEs are not transmitted clonally to daughter cells. Hence, changepoints resulting from SCEs are very unlikely to recur at the same position in >1 cell of a sample

Heatmap of single-cells based on SVs called by MosaiCatcher framework

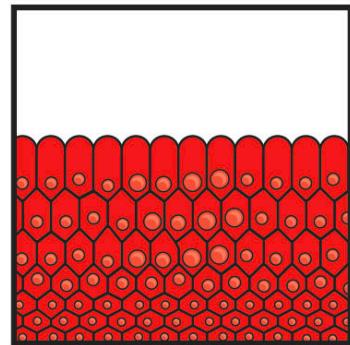


Sanders et al. 2019

- This heatmap was arranged using Ward's method for hierarchical clustering of SVs genotype likelihoods in two PDX samples
- P33 shows single dominant clone but P1 shows subclonal population in the sample represented by 23 cells

Let's try Mosaicatcher ourselves! (WT BM510 mixture VAF 20% vs WT)

Real
world



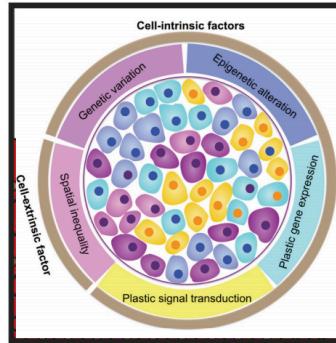
Normal cells

Model
system

WT

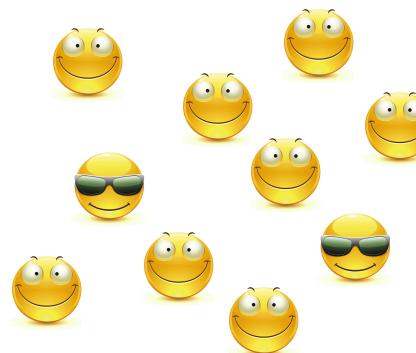


vs

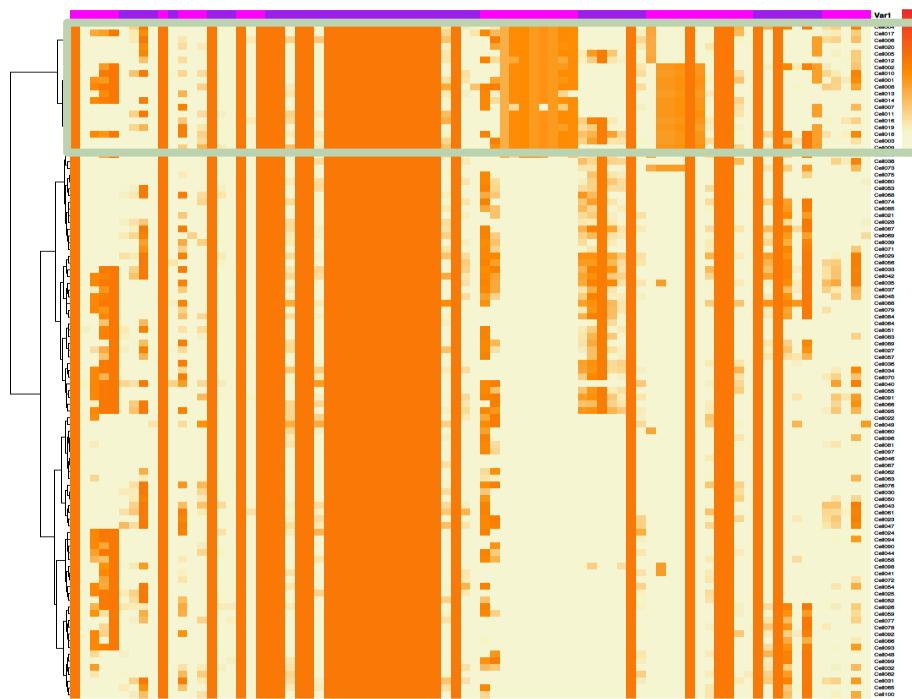


Cells forming a tumour

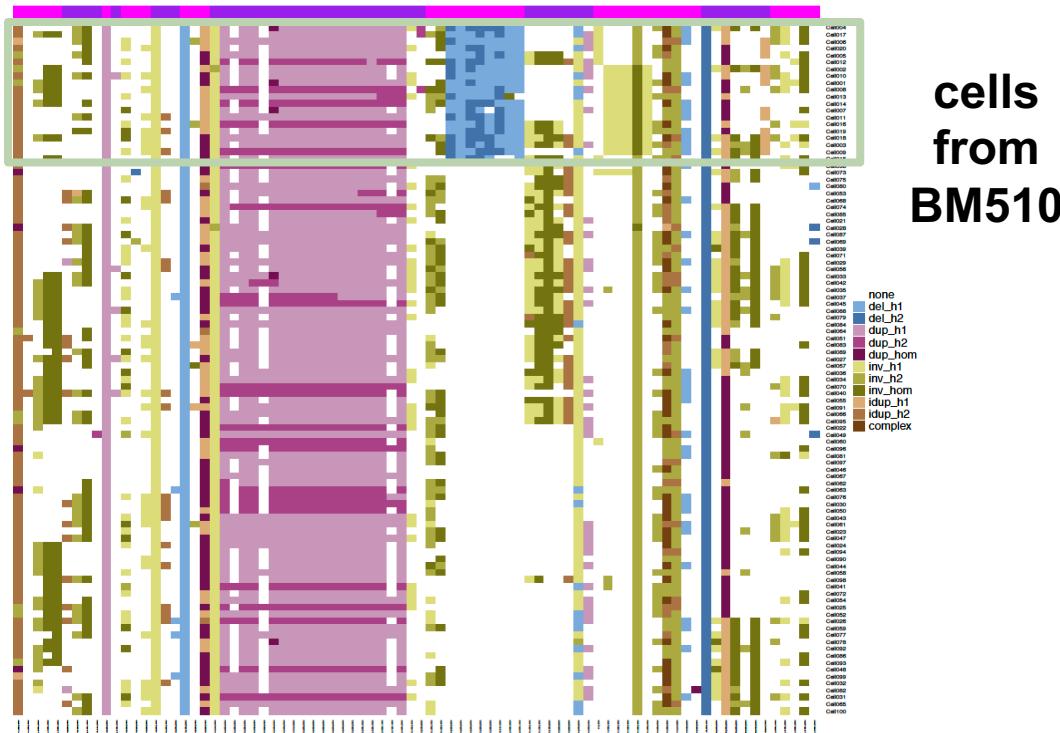
VAF20% BM510



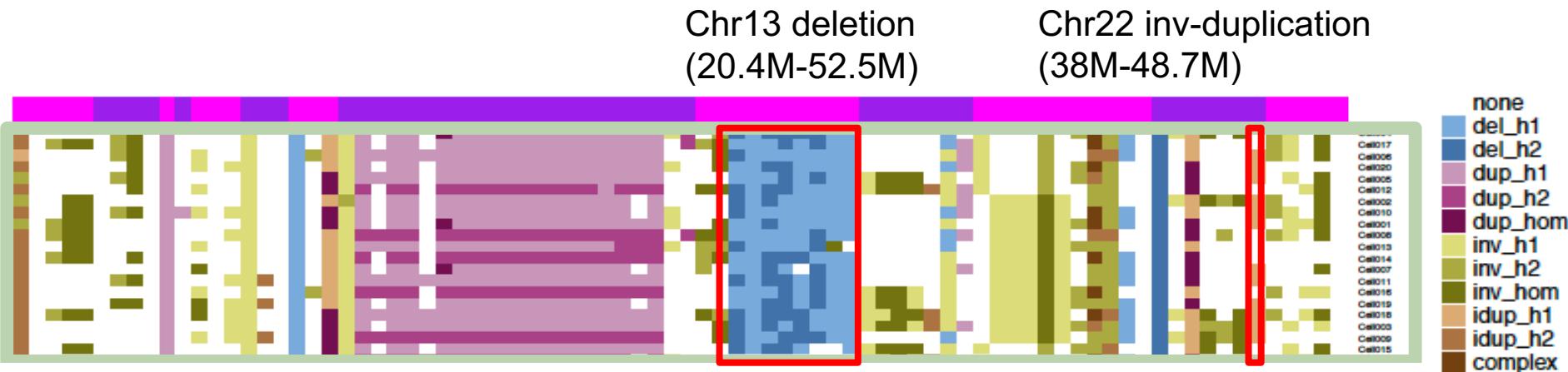
Single-cell SV calls and clusters (WT BM510 mixture VAF 20% vs WT)



Single-cell SV calls and clusters (WT BM510 mixture VAF 20% vs WT)



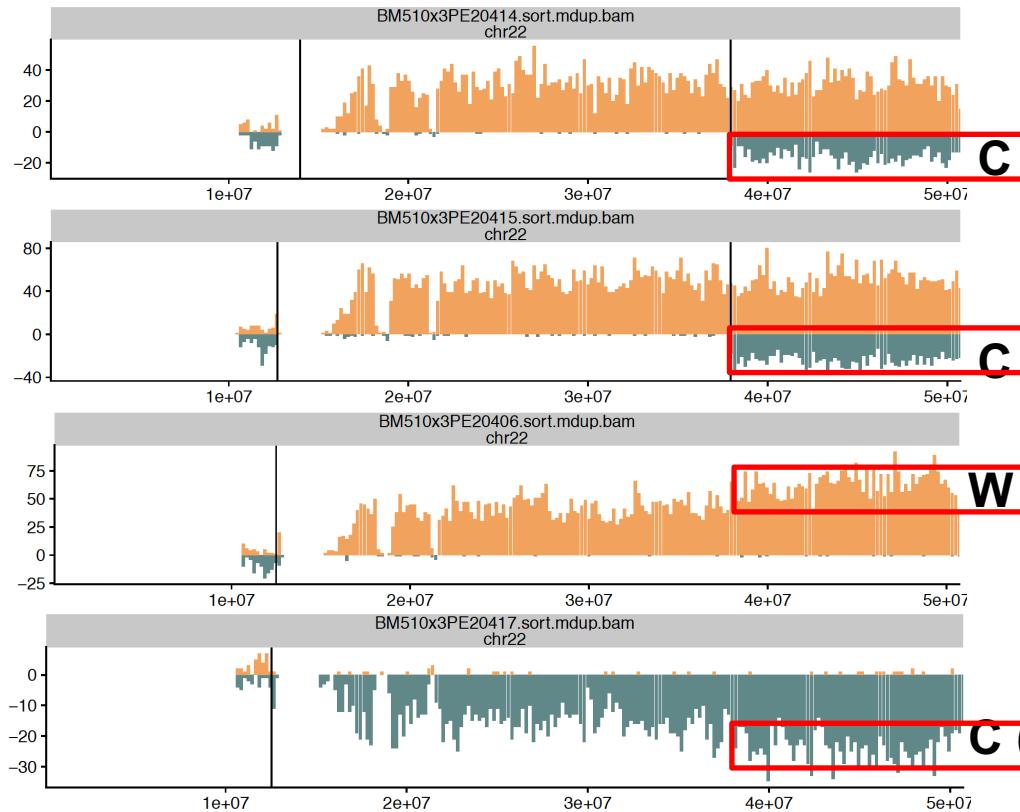
Identified SVs specific to BM510 rather than WT



- Two potential situations
 - 1) Inverted duplication of subset of 20 cells?
 - 2) Clonal translocation of 20 cells?

Why should we suspect that inverted duplication in subset of the cells can be a translocation?

Translocation
partner
chromosome N



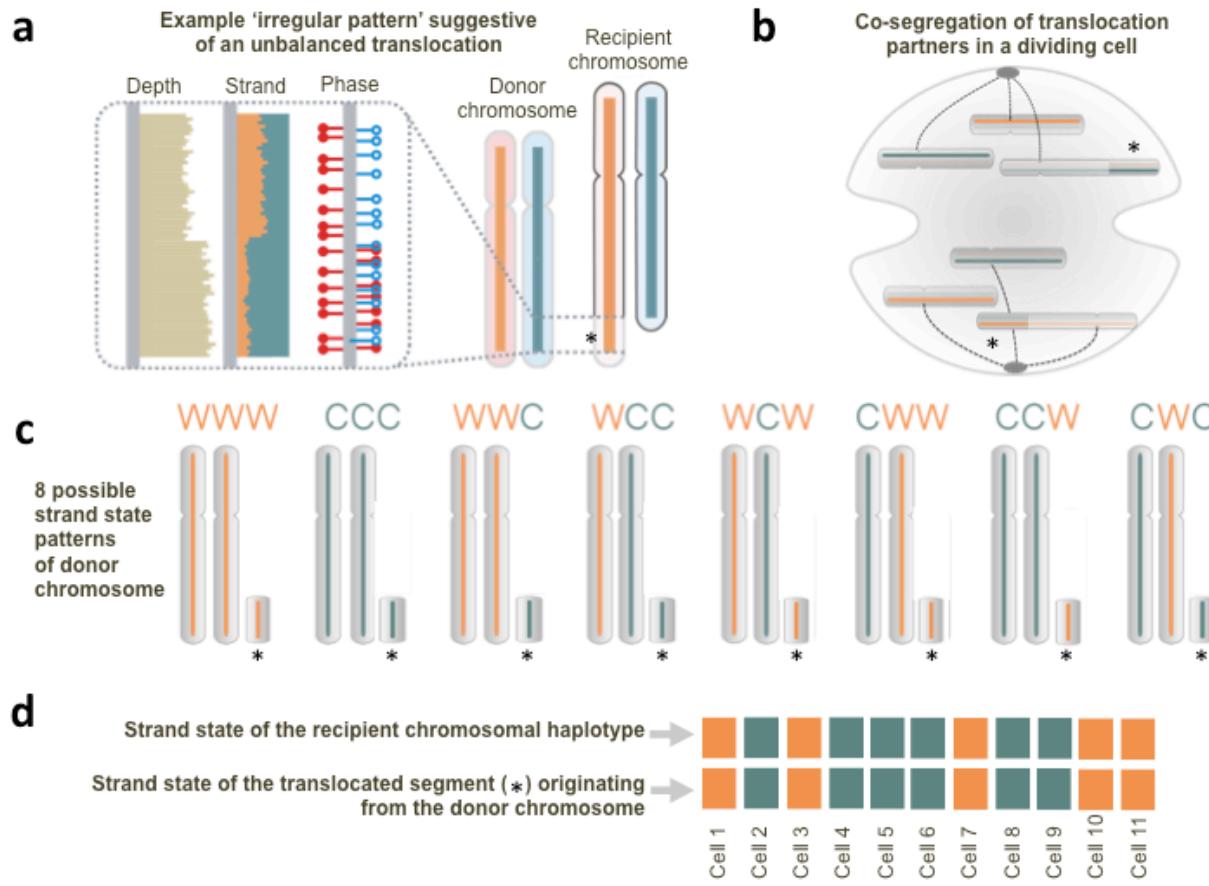
C

C

W

C

Translocated part will not follow strand state of donor chromosome, but follow the pattern of recipient chromosome

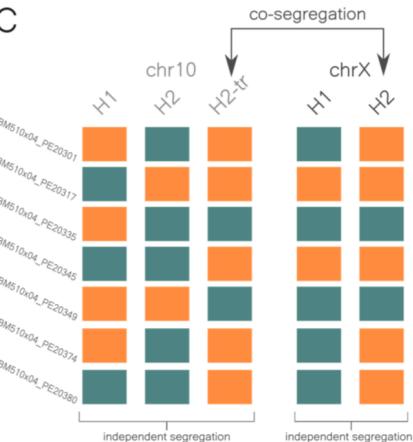


Finding translocated partners using Strand-seq data



Alex van Vliet

C



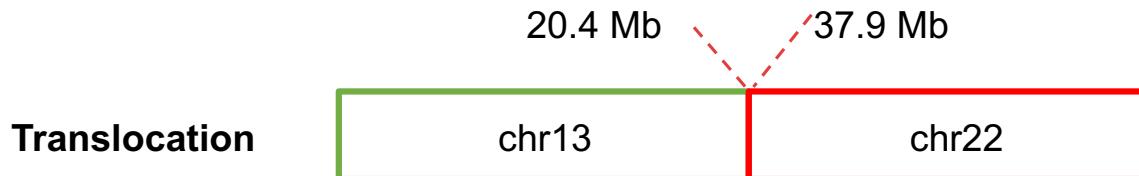
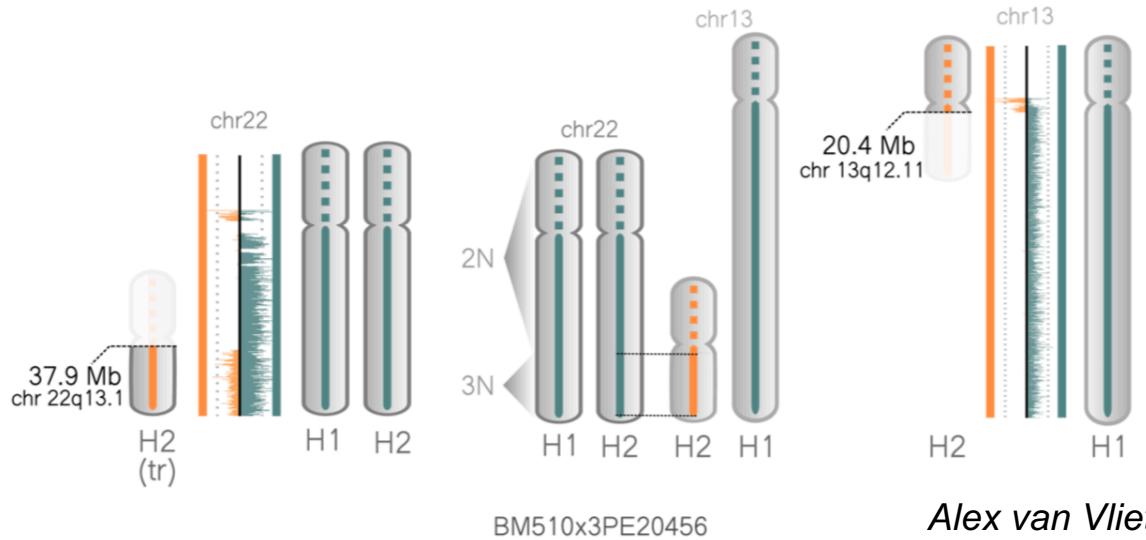
Sample	Derivative chromosome	Translocation Type	Orientation	correlation p-value (adj)	Derivative: haplotype	breakpoint	Partner: haplotype	breakpoint	Notes
C7	der(10; t(10;15)	unbalanced	D	$p=3.45e-22$	chr10p-H2	ends BFB	chr15q-H1	93.2Mb	provides telomere cap for BFB on chr10p; WGS validated
RPE-1	der(X)t(X;10)	unbalanced	I	$p=2.9e-33$	chrXq-H2	Telomere	chr10-H2	60.8Mb	cytogenetically validated
BM510	der(X)t(;10)	unbalanced	D	$p=2.26e-32$	chrXq-H2	Telomere	chr10-H2	60.8Mb	originated from RPE-1 cell line; cytogenetically validated
BM510	der(13)t(13;22)	unbalanced	D	$p=5.52e-41$	chr13p-H2	19.4Mb	chr22q-H2	37.9Mb	rest of chromosome 13 is monosomic; WGS validated
BM510	der(15)t(15;17)	reciprocal	I	$p=4.75e-29$	chr15q-H2	88.3Mb	chr17p-H2	19.6Mb	contains an inversion on chr17p; places TP53 3' of NTRK3; WGS validated
BM510	der(17)t(15;17)	reciprocal	I	$p=3.93e-30$	chr17p-H2	19.6Mb	chr15q-H2	88.3Mb	WGS validated
P33	der(5)t(5;14)	reciprocal	D	$p=4.52e-5$	chr5q-H1	171.3Mb	chr14q-H1	98.7Mb	cryptic (not seen in cytogenetic karyotype)
P33	der(14)t(5;14)	reciprocal	D	$p=1e-4$	chr14q-H1	98.7Mb	chr5q-H1	171.3Mb	places BCL11B enhancer upstream of TLX3; cryptic

*derivative assigned to chromosome containing centromere, translocation partners listed in numerical order

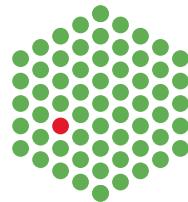
D = Direct (correlated segregation); I = indirect (anti-correlated segregation)

Sanders et al. 2019

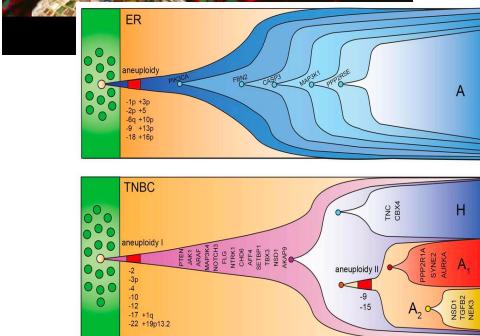
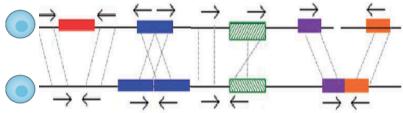
Unbalanced translocation between chr22 and chr13 in BM510, validated by Delly and WGS



Exciting biological questions to be addressed using Strand-seq



Summary



Acknowledgement

Supervisor

- Jan Korbel

Korbel group

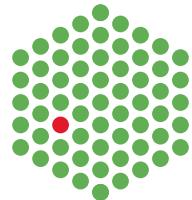
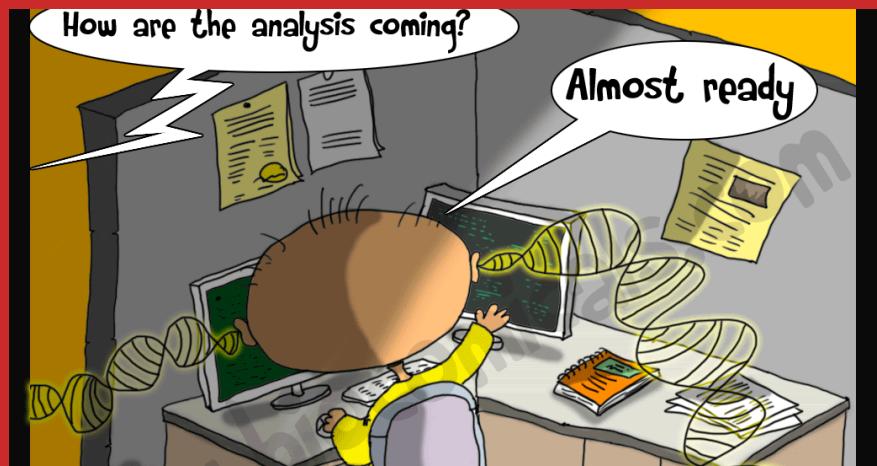
- Busra Erarslan Uysal
- ***Karen Grimes***
- Nina Habermann
- Sergei Iakhnin
- Sascha Meiers
- Esa Pitkaenen
- Benjamin Raeder
- Tobias Rausch
- Ashley Sanders
- Sebastian Waszak
- Daniel Schraivogel
- ***Wolfram Hoeps***
- Maja Starostecka
- Tania Christiansen
- Theocharis Efthymiopoulos
- Marco Raffaele Cosenze
- Alex James Ing
- Mike Smith
- Andrea Imle
- Patrick Hasenfeld
- Davide Bolognini
- Alexandra van Vliet
- Gabriel Longo
- ***Hyobin Jeong***
- Caelan Bell

MPI Informatics

- Tobias Marschall
- David Porubsky
- Maryam Ghareghani



*Thank you for
listening*



Q&A
hyobin.jeong@embl.de