Cancer genomic rearrangements

Marcin Imielinski M.D. Ph.D.

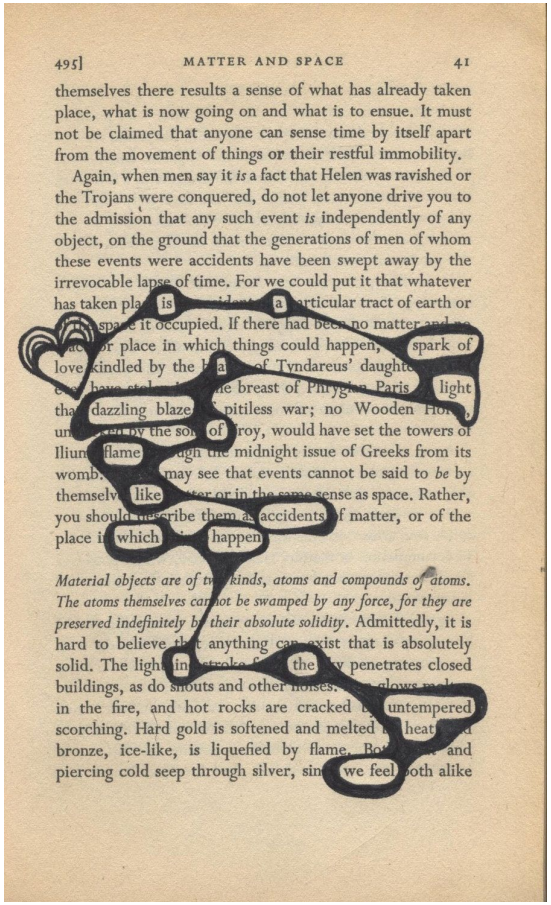
Assistant Professor (Weill Cornell)

Core Member (NYGC)

Clinical and Research Genomics

April 4 2018

If the human genome is a book ...



495] MATTER AND SPACE 41

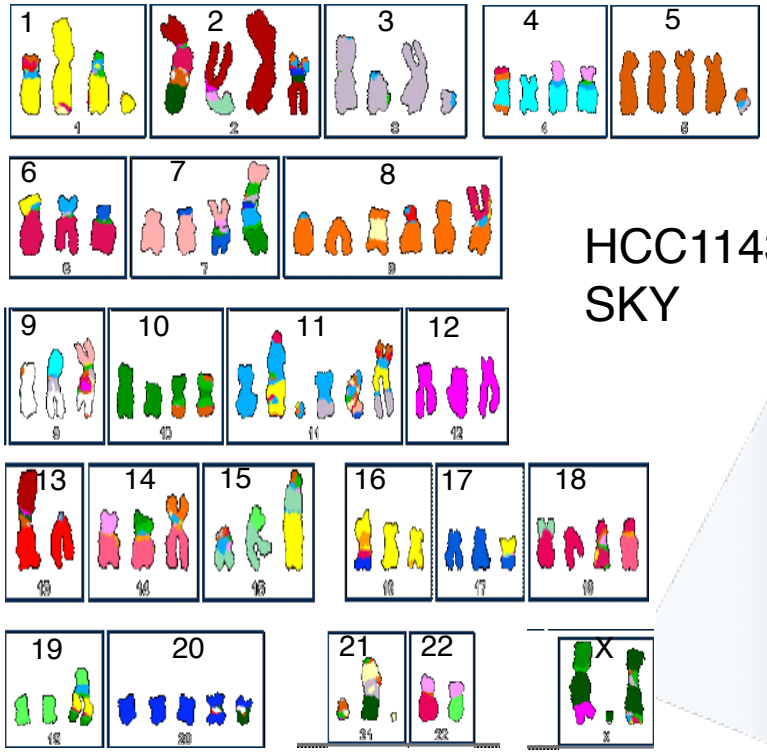
themselves there results a sense of what has already taken place, what is now going on and what is to ensue. It must not be claimed that anyone can sense time by itself apart from the movement of things or their restful immobility.

Again, when men say it *is* a fact that Helen was ravished or the Trojans were conquered, do not let anyone drive you to the admission that any such event *is* independently of any object, on the ground that the generations of men of whom these events were accidents have been swept away by the irrevocable lapse of time. For we could put it that whatever has taken pla is a rticular tract of earth or space it occupied. If there had be no matter and place in which things could happen, spark of love kindled by the of Tyndareus' daughte breast of Phrygian Paris light dazzling blaze pitiless war; no Wooden Ho by the so of Troy, would have set the towers of Ilium flame ugh the midnight issue of Greeks from its womb. may see that events cannot be said to *be* by themselv like ter or in the same sense as space. Rather, you should scribe them as accidents f matter, or of the place i which happen.

Material objects are of tw kinds, atoms and compounds of atoms. The atoms themselves cannot be swamped by any force, for they are preserved indefinitely by their absolute solidity. Admittedly, it is hard to believe th anything can exist that is absolutely solid. The ligh in stroke the penetrates closed buildings, as do shouts and other noises. in the fire, and hot rocks are cracked untempered scorching. Hard gold is softened and melted heat bronze, ice-like, is liquefied by flame. Bo and piercing cold seep through silver, sin we feel both alike

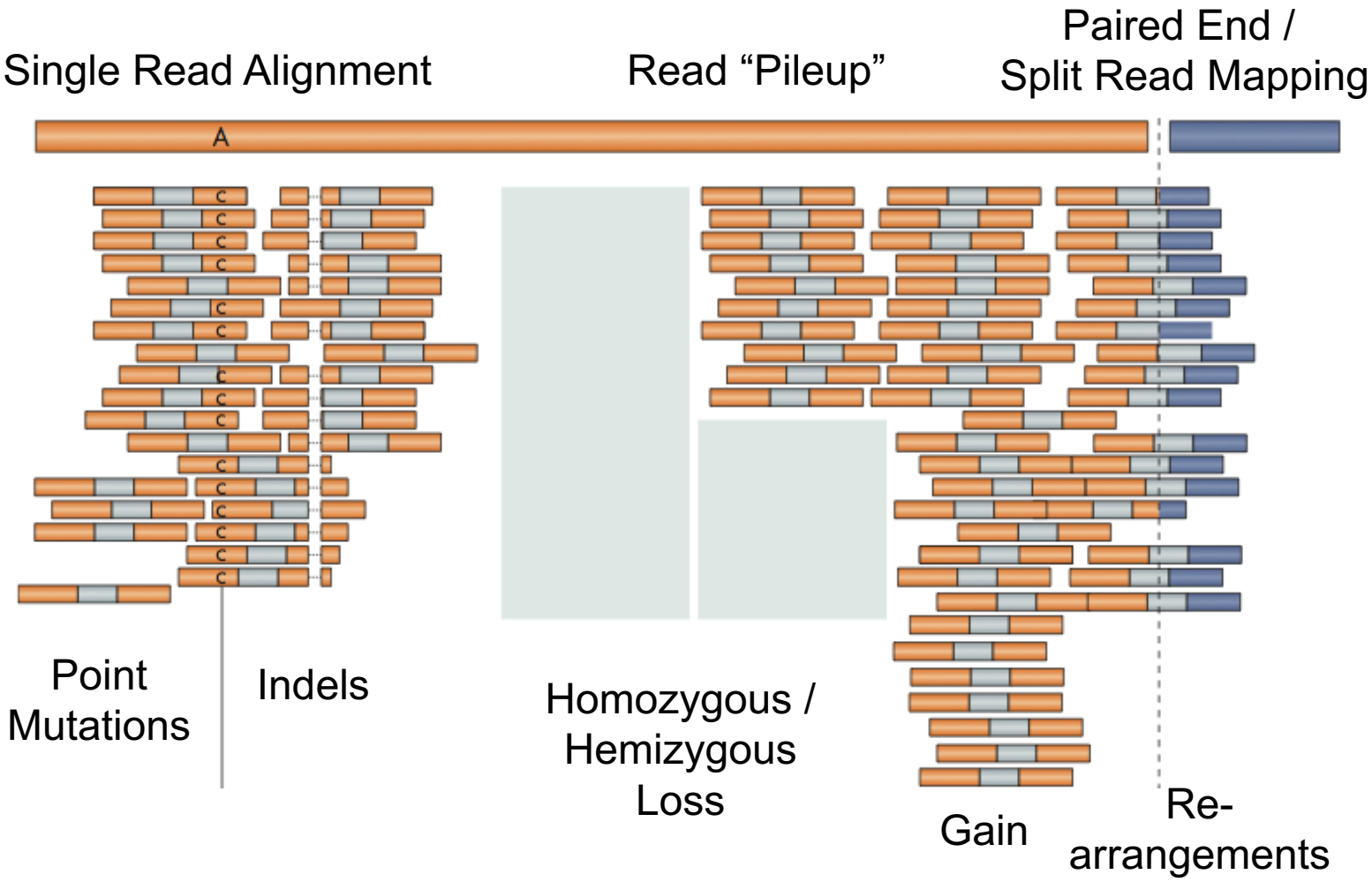
... cancer is collage poetry

Chromosomal collages



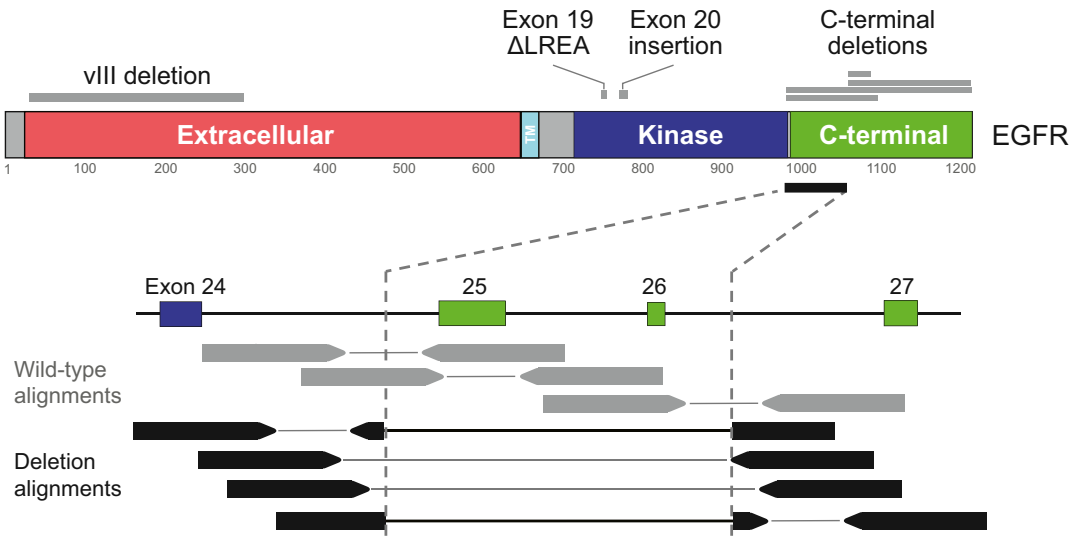
HCC1143
SKY

Chromosomal shreds

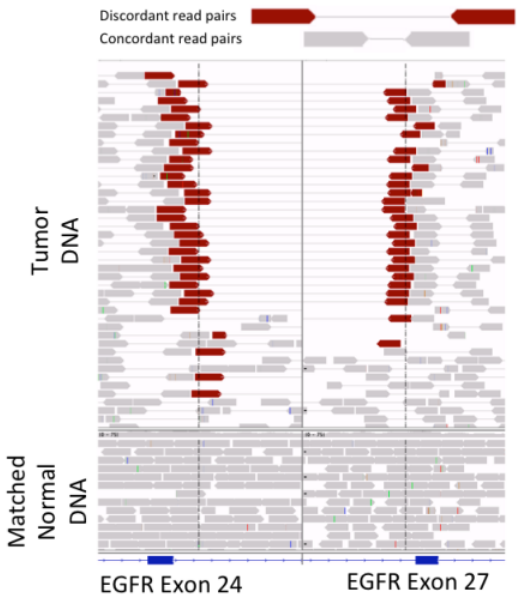


Adapted from Meyerson, Getz, Gabriel (Nature Reviews Genetics 2010)

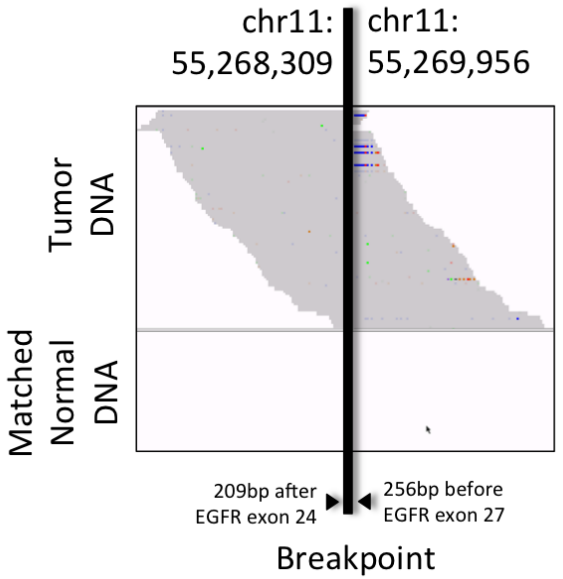
Junction detection from tiny eads



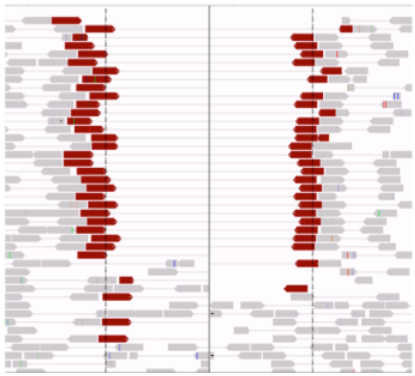
EGFR deletion paired end mapping

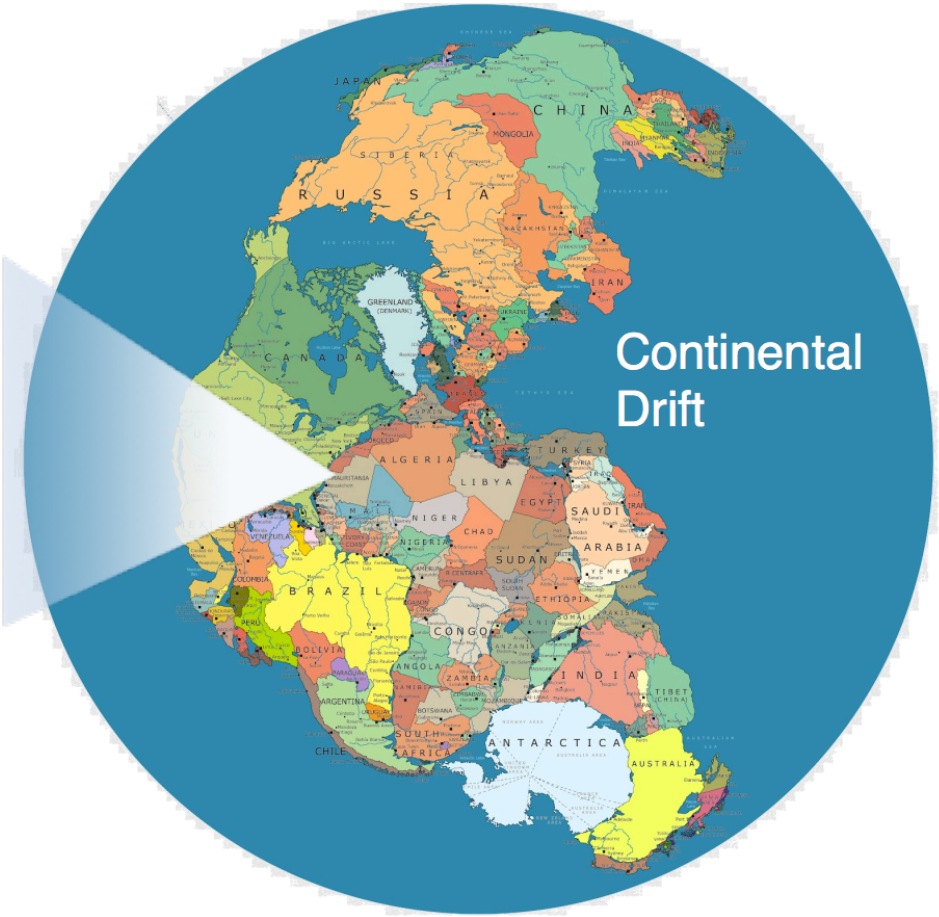


EGFR deletion Split read mapping



Chromosomal Pangaea

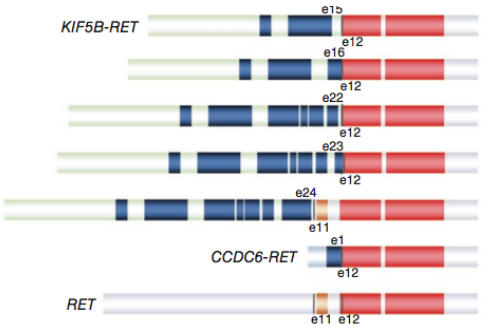




The Earth
(500 million years ago)

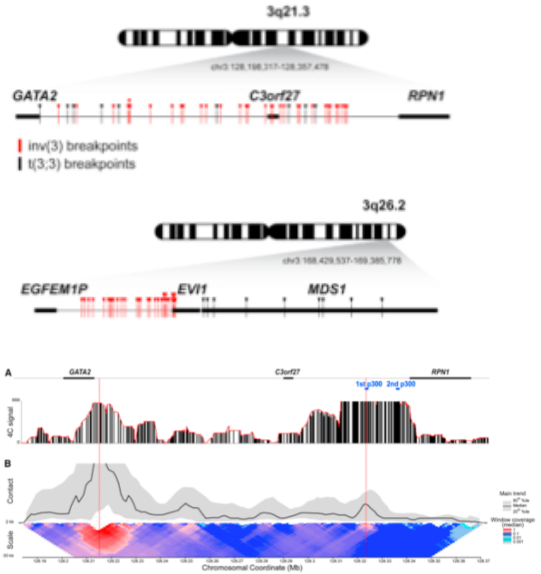
Studying cancer genome structure: motivation

Gene Fusions



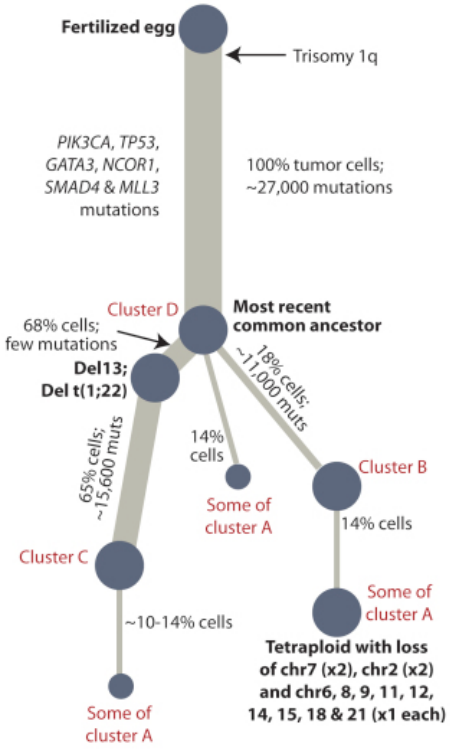
Takeuchi .. Ishikawa
(Nature Medicine 2012)

Noncoding Rearrangements



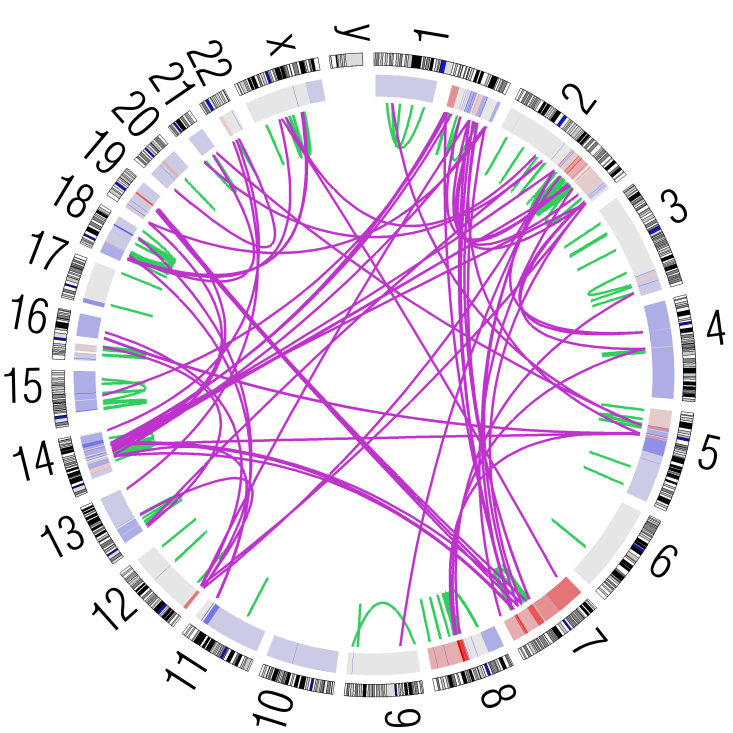
Groschel .. Delwel
(Cell 2014)

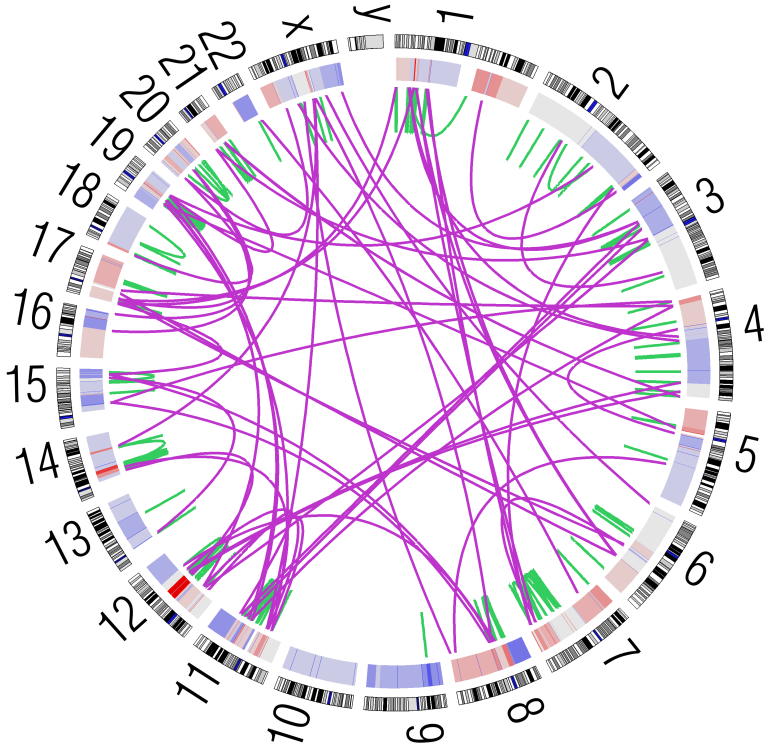
Mutational processes



Nik-Zainal .. Stratton
(Cell 2012)

Which is the smoker?



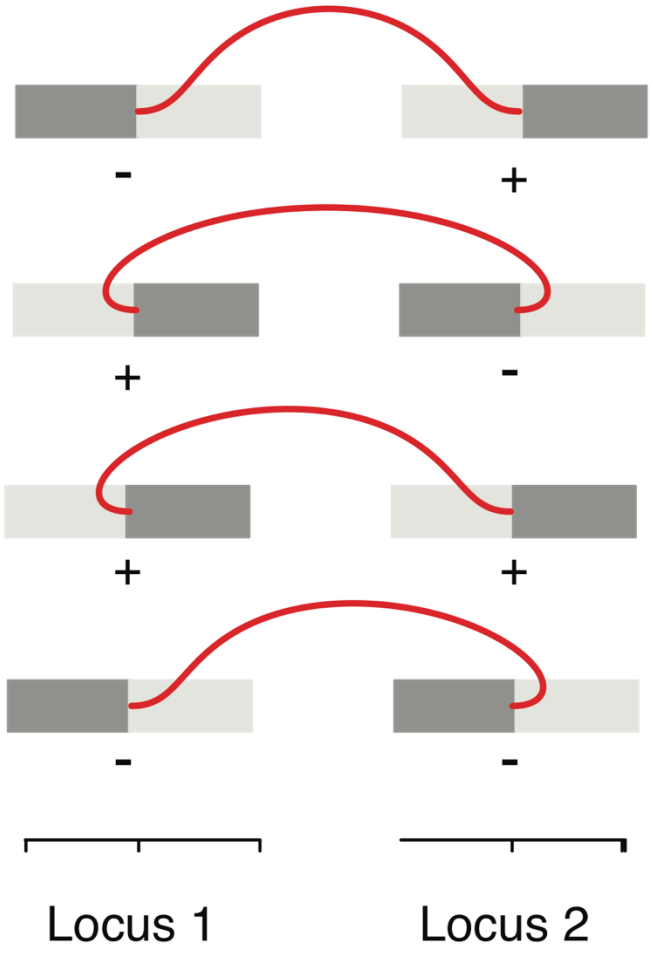


Imielinski et al *Cell* 2012

Circos .. so beautiful

http://www.circos.ca/

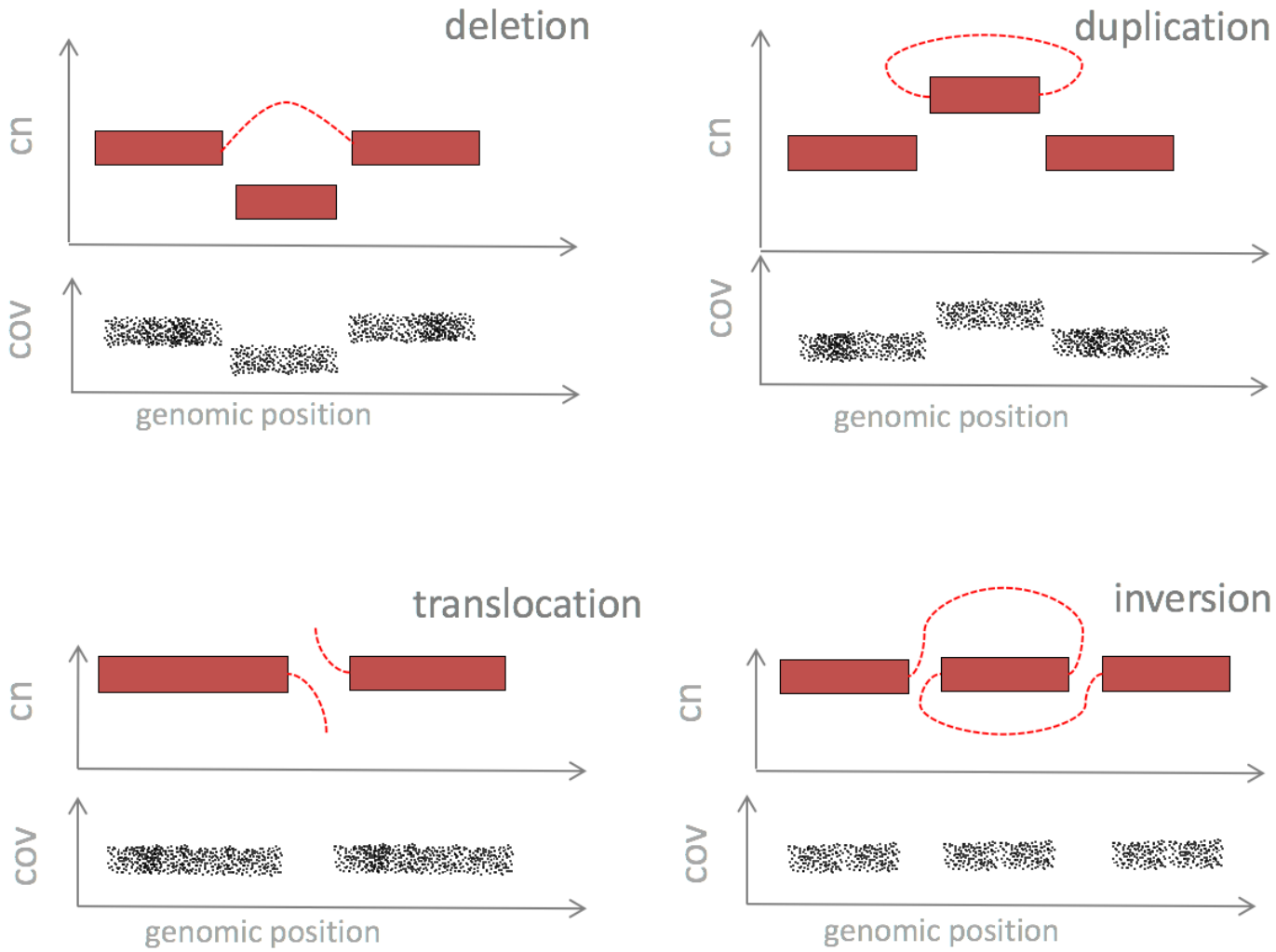
Junction: the “atomic unit” of a genomic rearrangement



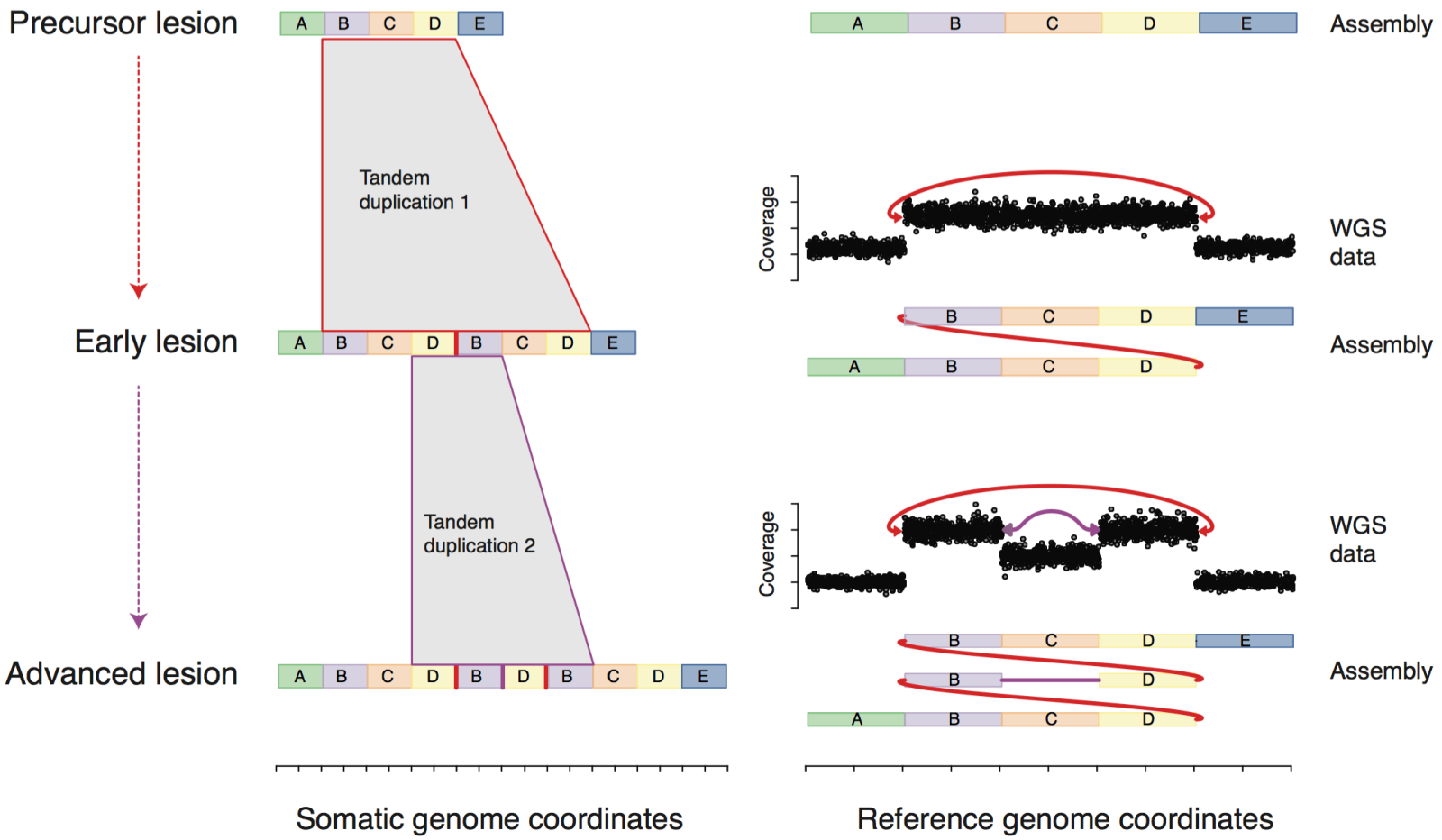
Junction = pair of locations AND orientations

Maciejowski and Imielinski 2017

Can it be all so simple?



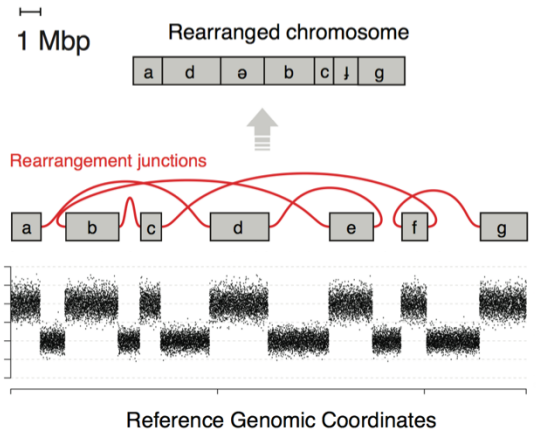
What is an “event”?



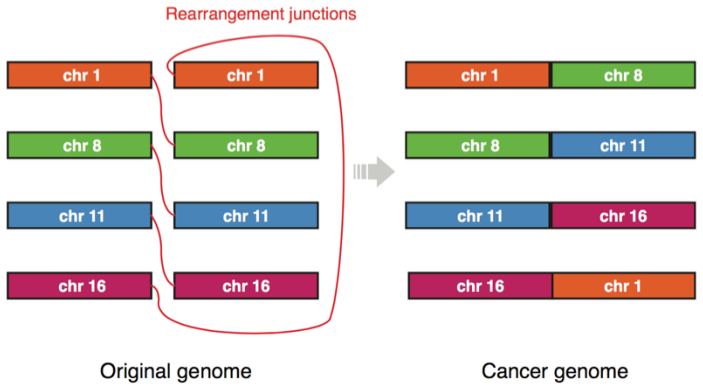
Maciejowski and Imielinski (2017)

Complex structural variation in cancer

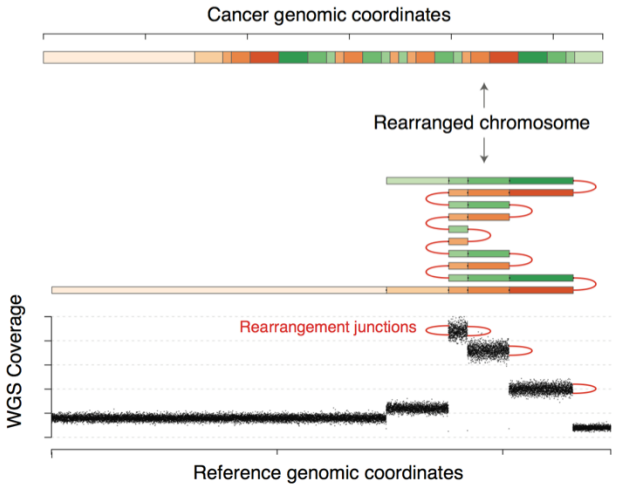
Chromothripsis



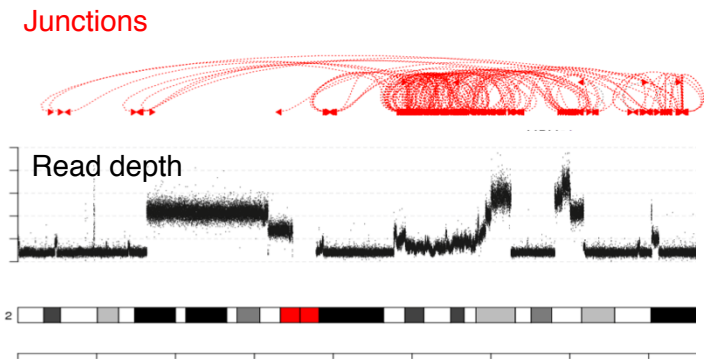
Chromoplexy



Breakage-fusion-bridge

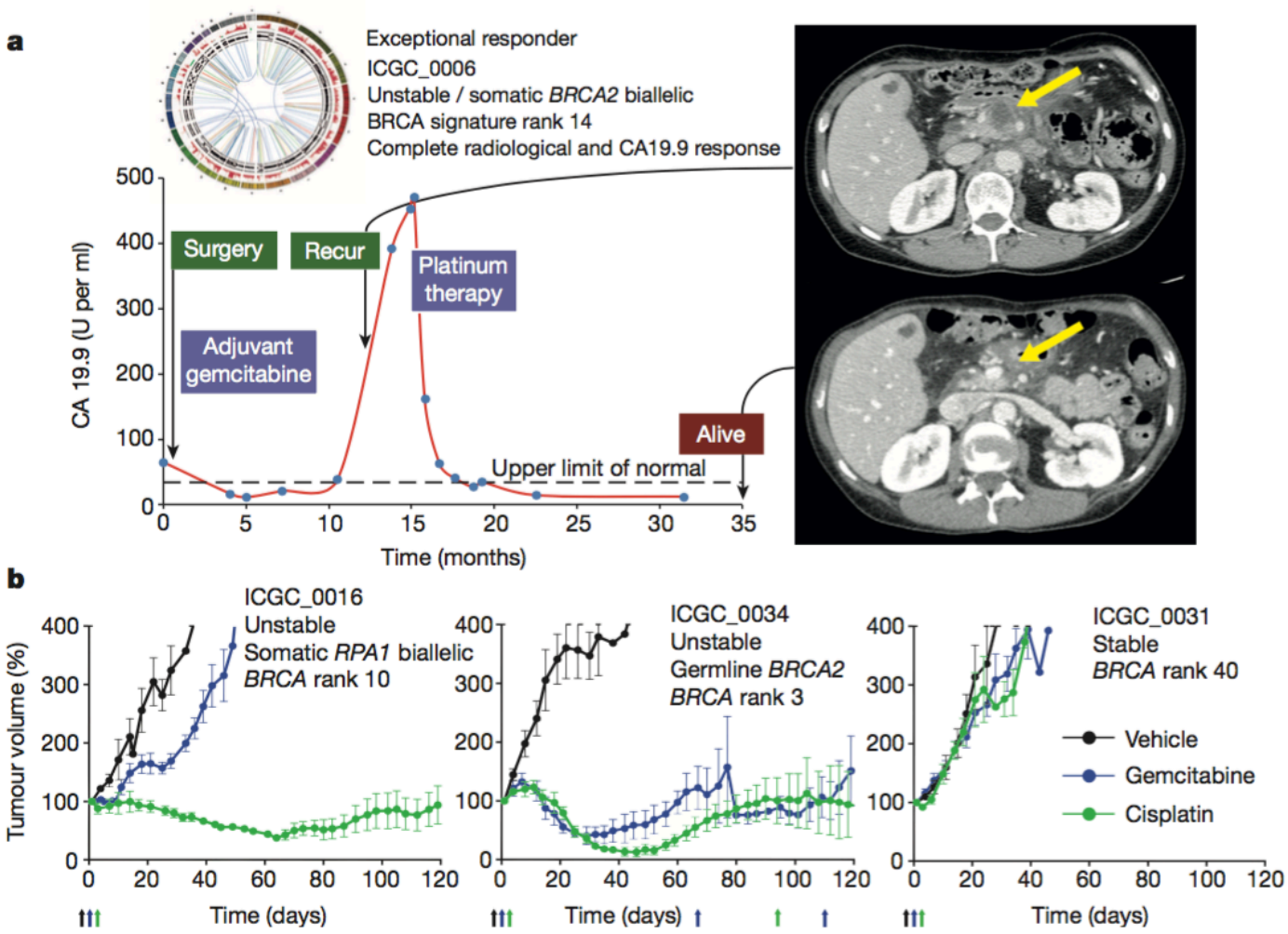


??????



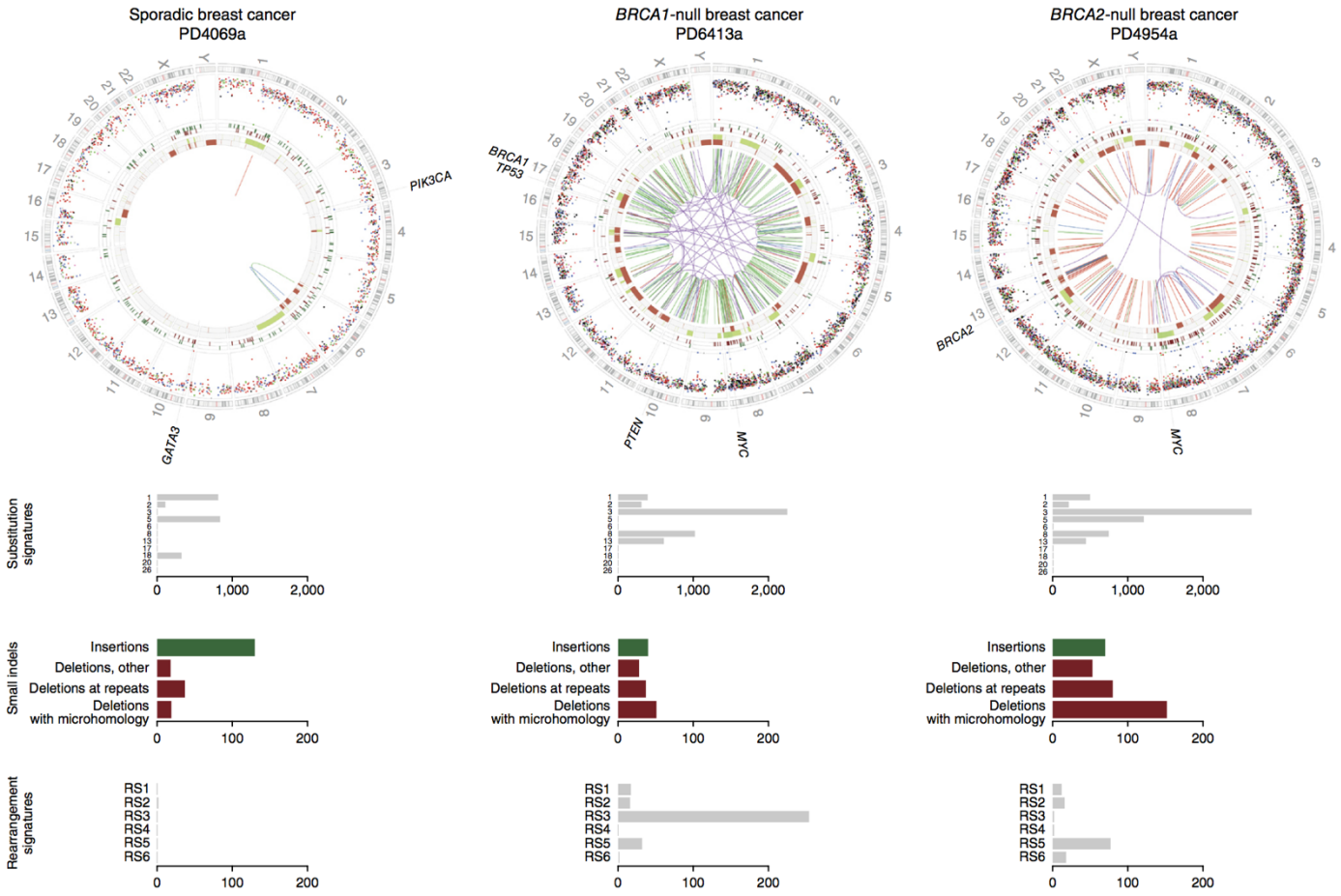
Imielinski and Rubin (Nat Clin Oncology 2017)

Clinical consequences of rearrangement signatures: exceptional chemotherapy response



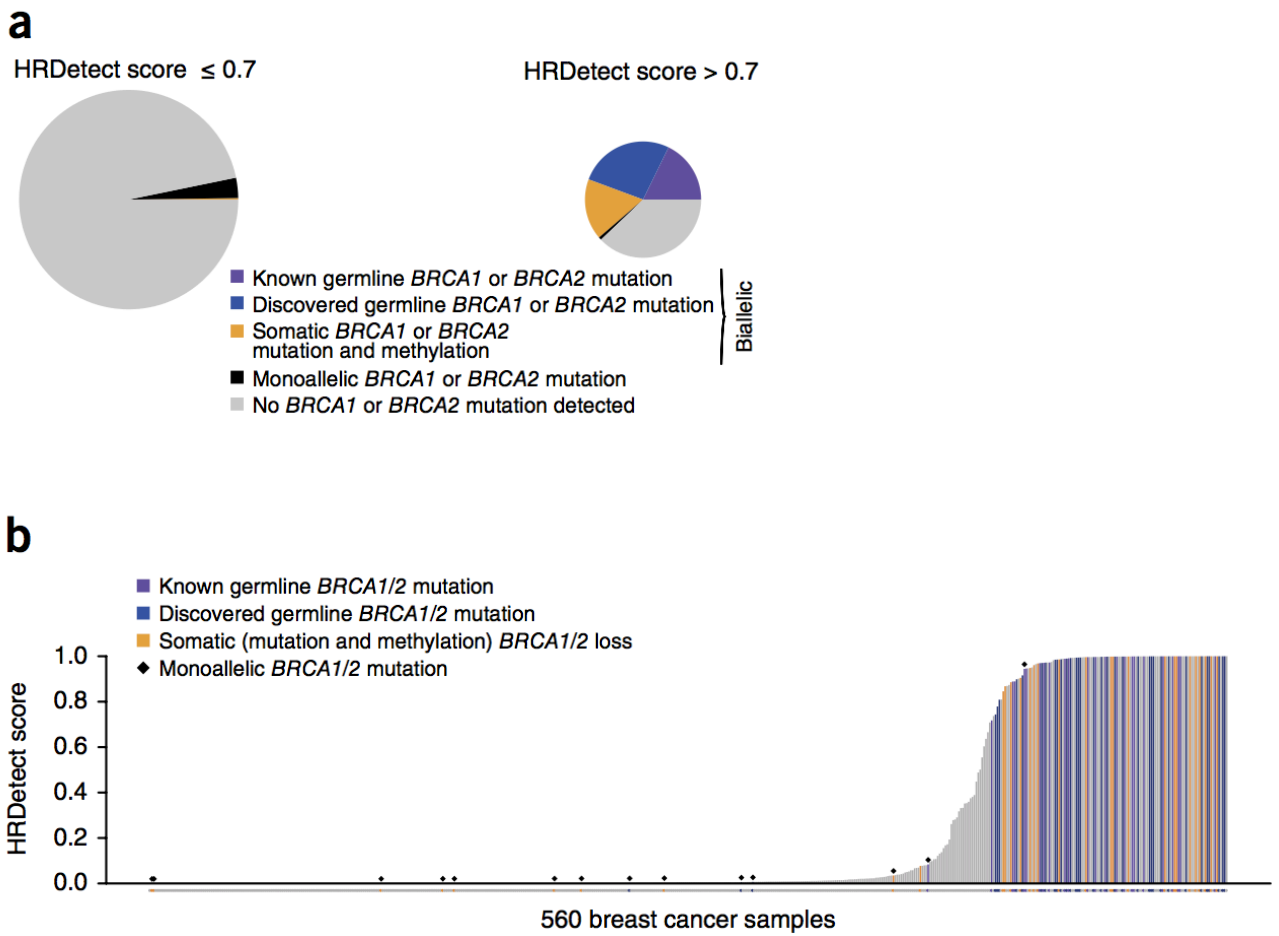
Waddell et al *Nature* 2015

Clinical consequences of rearrangement signatures: “BRCAness phenotype”



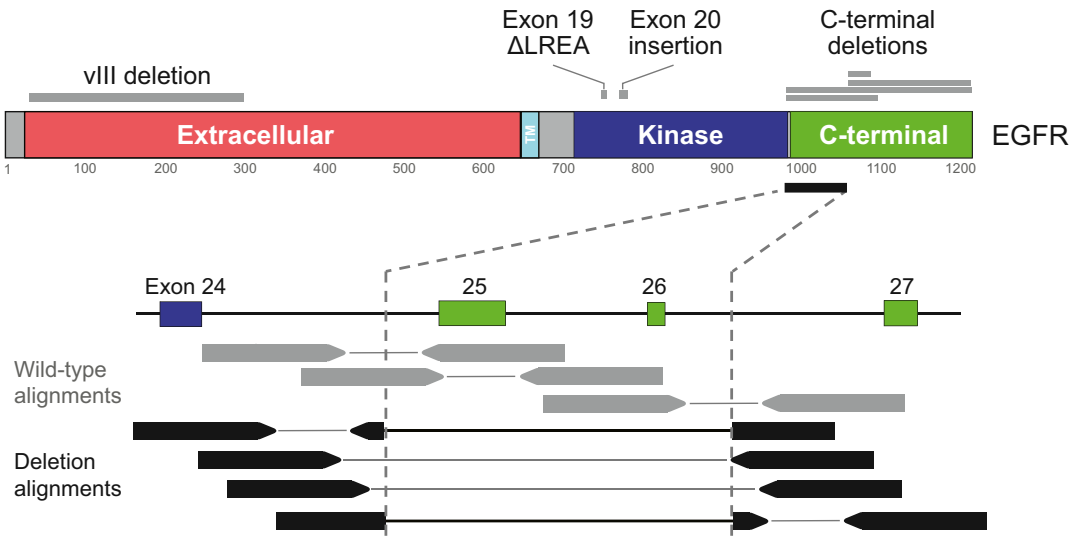
Davies et al *Nature Medicine* 2017

Clinical consequences of rearrangement signatures: “BRCAness phenotype”

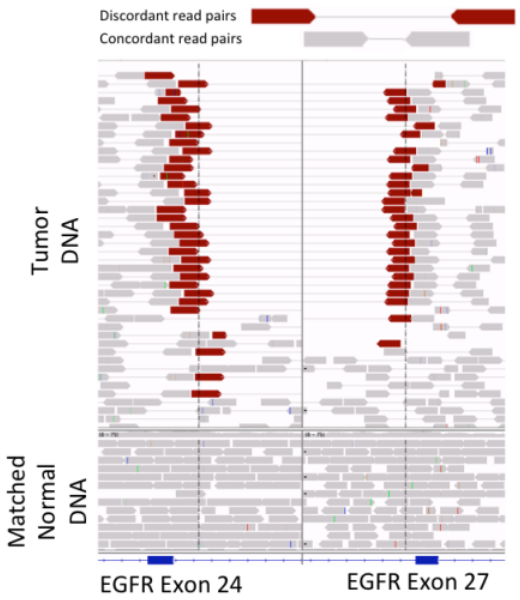


Davies et al *Nature Medicine* 2017

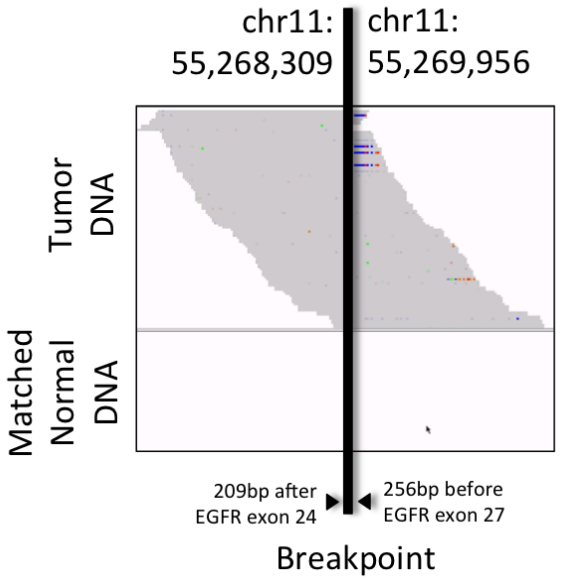
Standard WGS
Paired-end rearrangement mapping



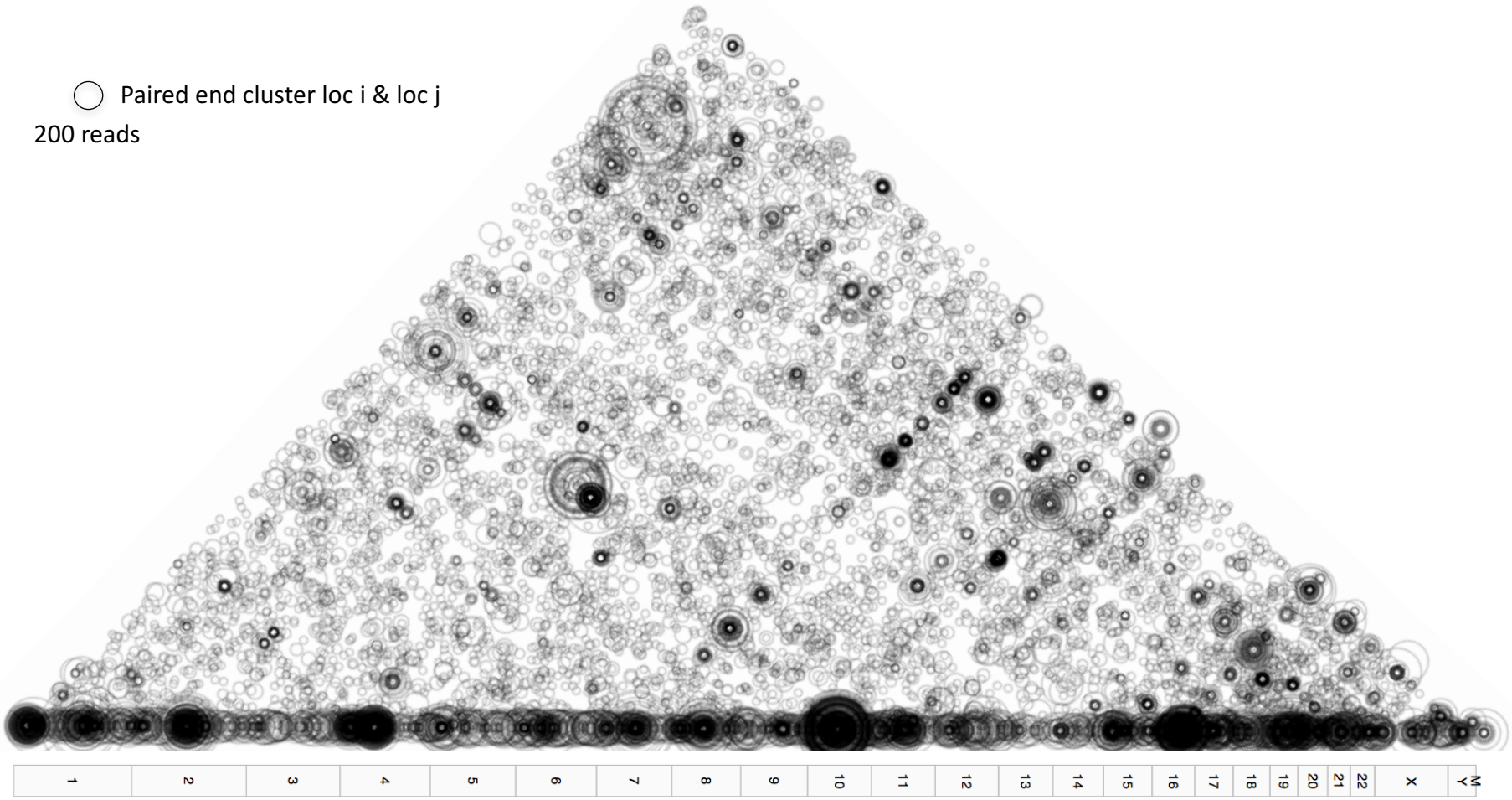
EGFR deletion paired end mapping



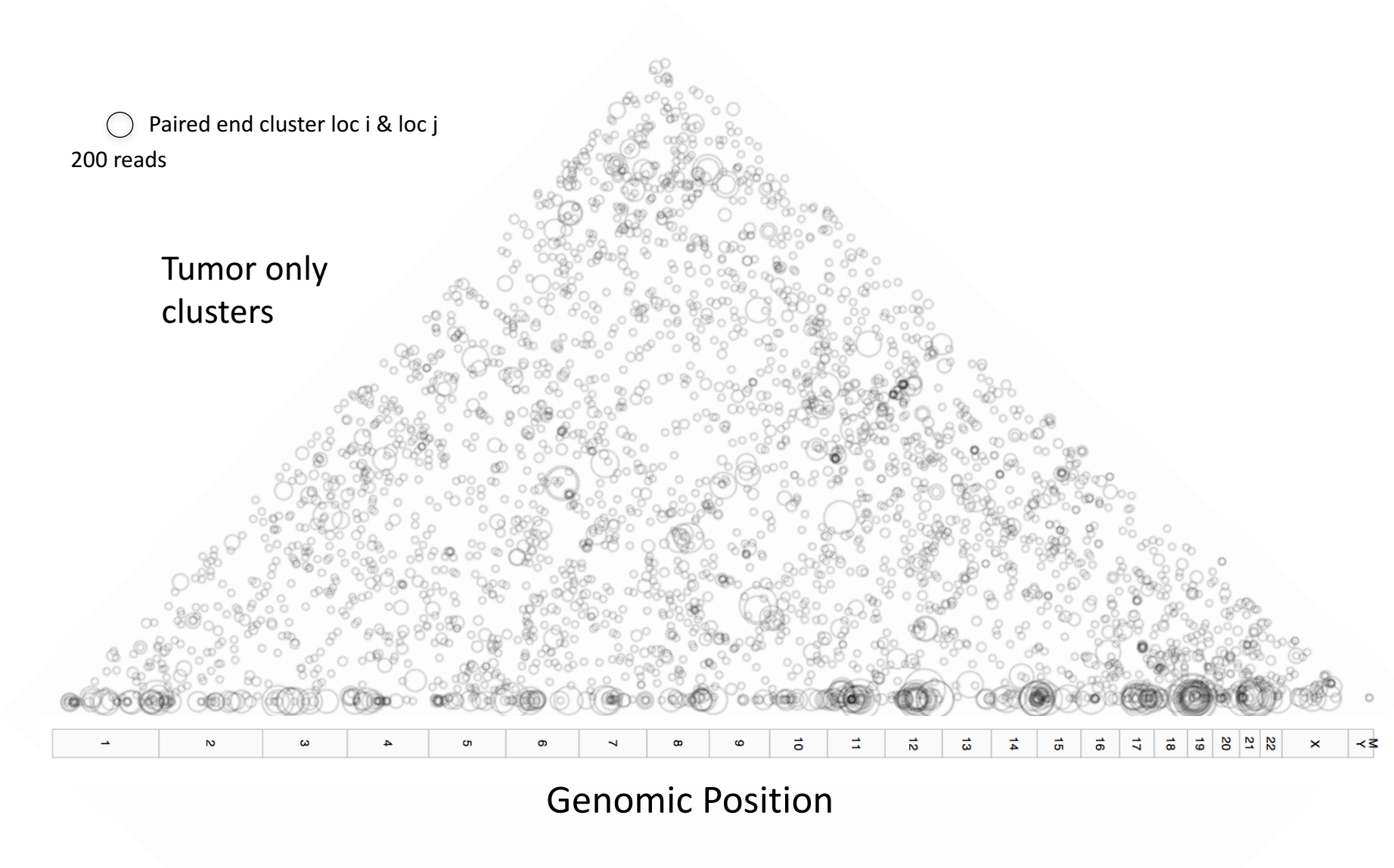
EGFR deletion Split read mapping



Challenges: signal vs noise in rearrangement analysis



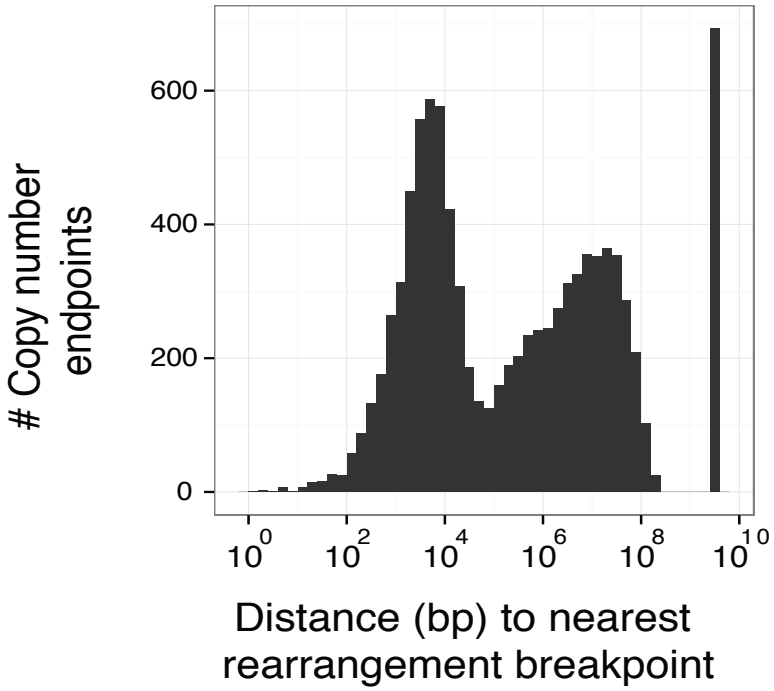
Challenges: signal vs noise in rearrangement analysis

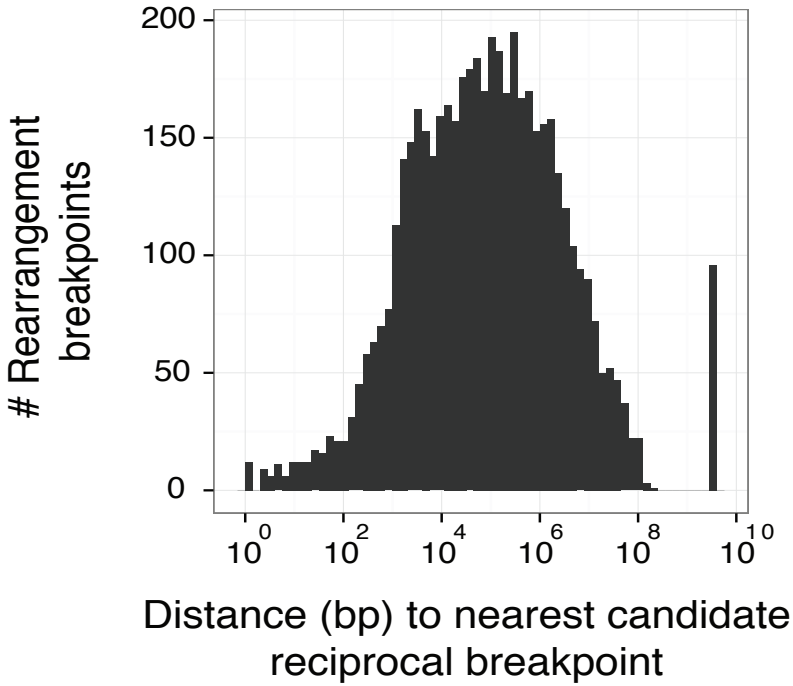


ALERT:

Copy number and rearrangement data … don't agree!

Data from 80 lung cancer whole genome T/N pairs

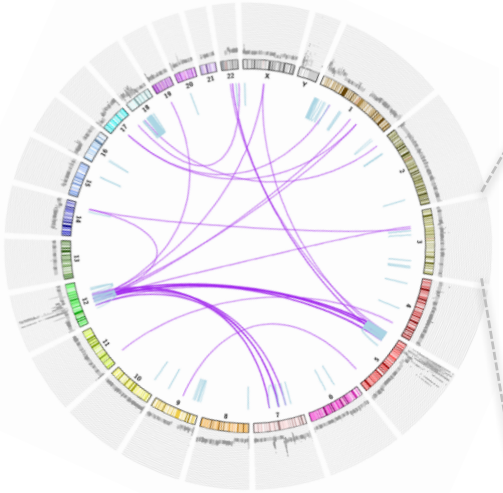




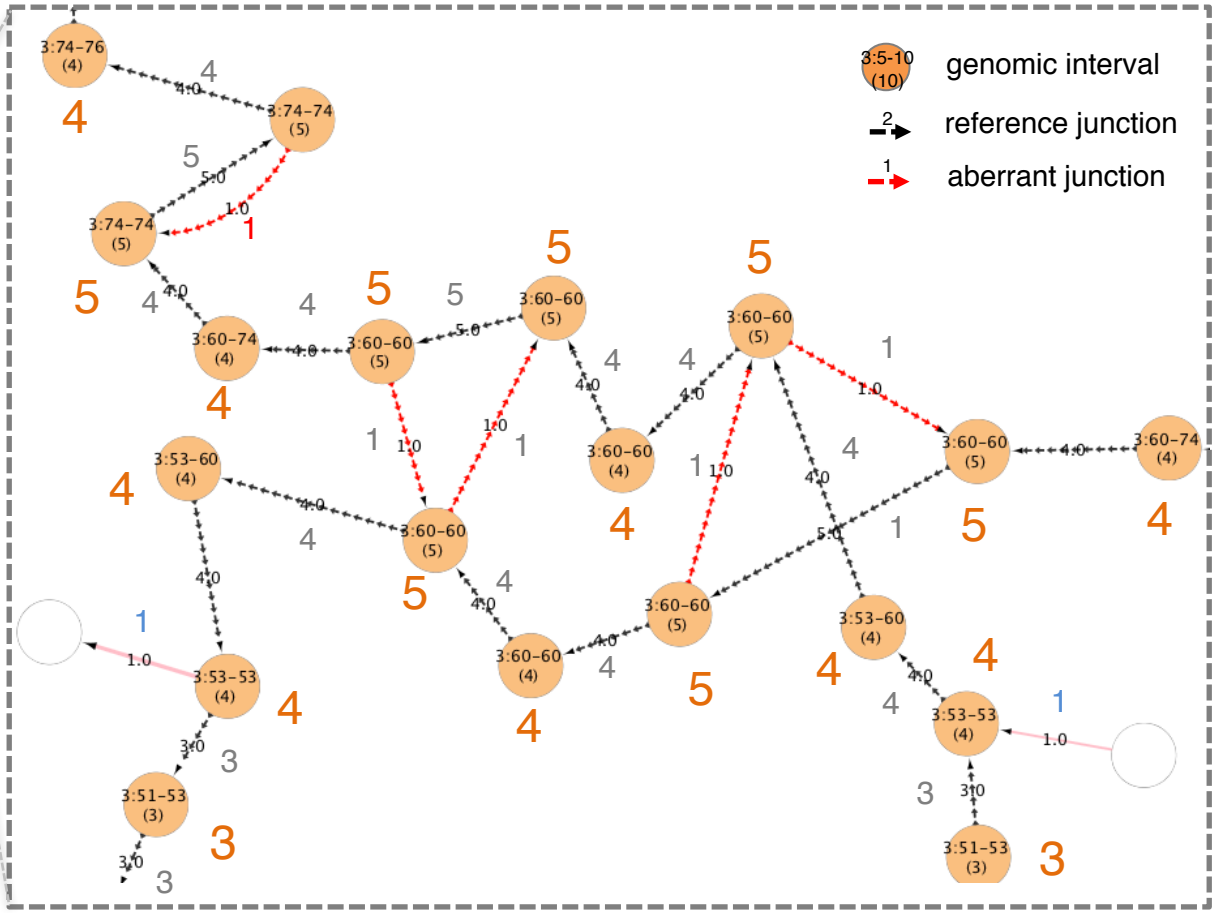
- Separately inferred → Inconsistent
- Over-segmentation of WGS data
- Unmapped and false positive junctions

JaBbA: From junctions to balanced assembly graphs

Whole genome sequence



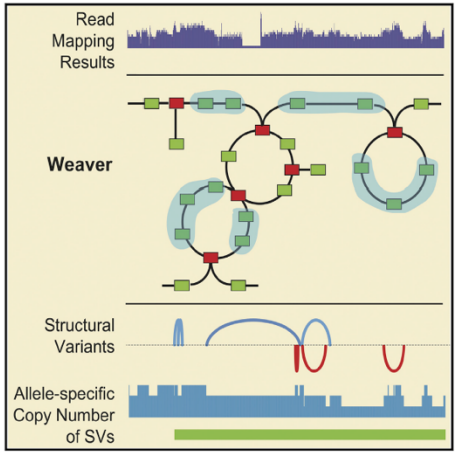
$$
\begin{aligned}
&\textit{minimize} \\
&\epsilon^T \Sigma^{-1} \epsilon \; + \; \lambda_e \left(\|\delta\|_2 + \|\eta\|_2 \right) \\
&\textit{subject to} \\
&\mathrm{v_i} + \gamma = \mu_i \beta \; + \epsilon_i \\
&\frac{\mathrm{v^T L}}{\|\mathrm{L}\|_1} + \gamma = \mu_0 \beta \\
&(\mathrm{B_{ev}})^{\mathrm{T}} \mathrm{e} + \delta \; - \; \mathrm{v} = 0 \\
&\mathrm{B_{ve}} \mathrm{e} + \eta \; - \; \mathrm{v} \; = 0 \\
&\tau_{min} \leq \mathrm{v^T L} / ||\mathrm{L}||_1 \leq \tau_{max} \\
&\mathrm{v}, \mathrm{e}, \gamma, \beta, \delta, \eta \; \geq \; 0
\end{aligned}
$$



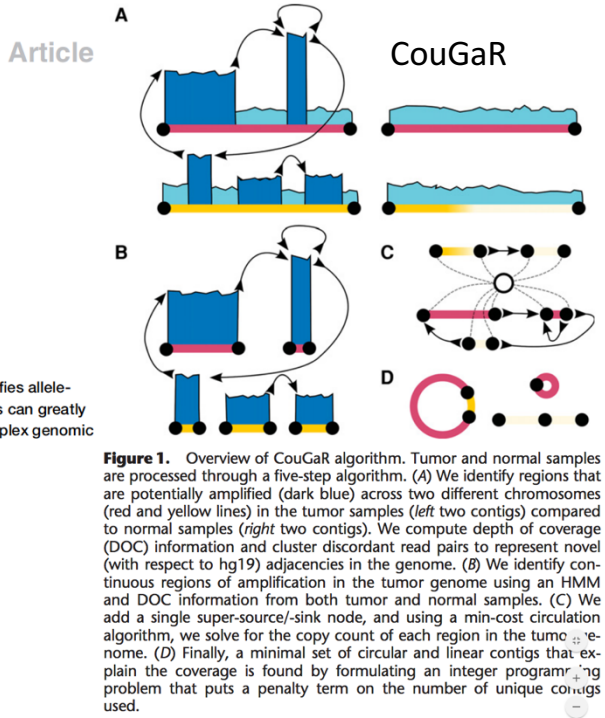
JaBbA graph

Yao et al (in preparation)

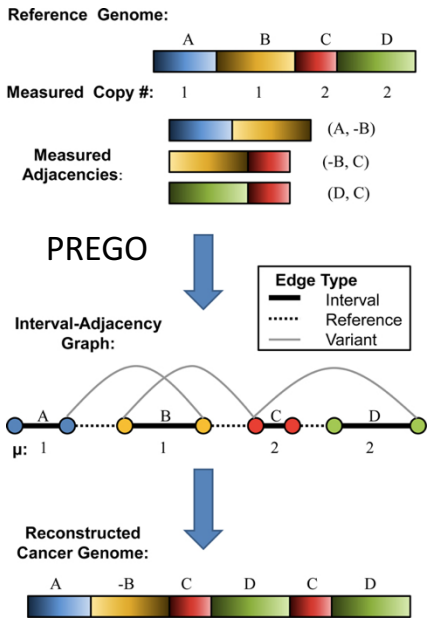
Other “graph callers”



Li .. Ma
Cell Systems 2016



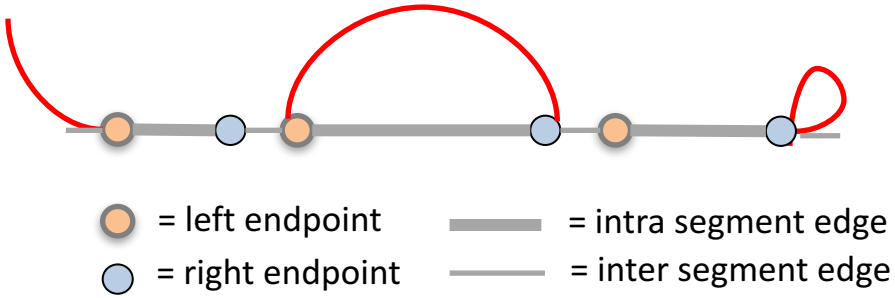
Dzamba .. Brudno
Genome Research 2016



Osper .. Raphael
BMC Bioinf 2012

Graph representation of whole genomes

Bidirected Graph



- Nodes represent left and right **sides** of intervals
- Undirected edges of two flavors (intra and inter segment)
- **Paths must be**

https://github.com/mskilab/gGnome

anti-path.

Stranded adjacency matrix A

$$A = \begin{array}{c|cc} & + & - \\ + & A_1 & A_2 \\ - & A_3 & A_1^T \end{array}$$

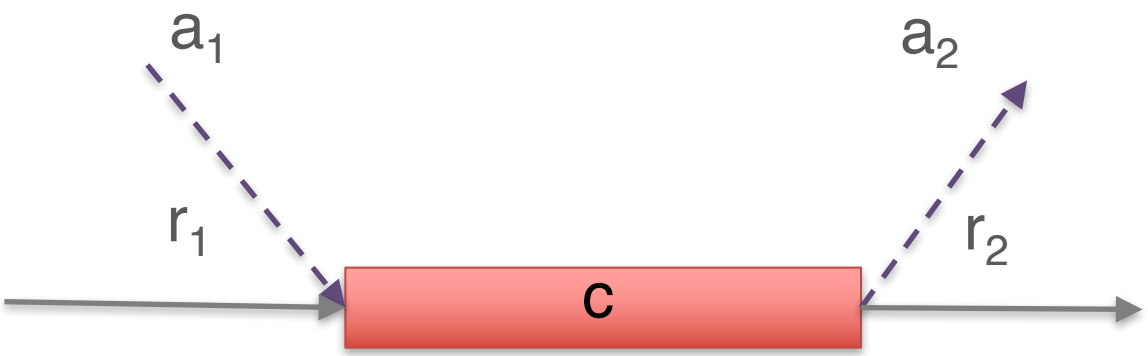
$$a_{ij} = a_{\bar{j}\,\bar{\imath}}$$

= number of copies of junctions joining intervals i and j

$$A_2 = A_2^T \quad A_3 = A_3^T$$

JaBbA (Junction Balance Analysis):

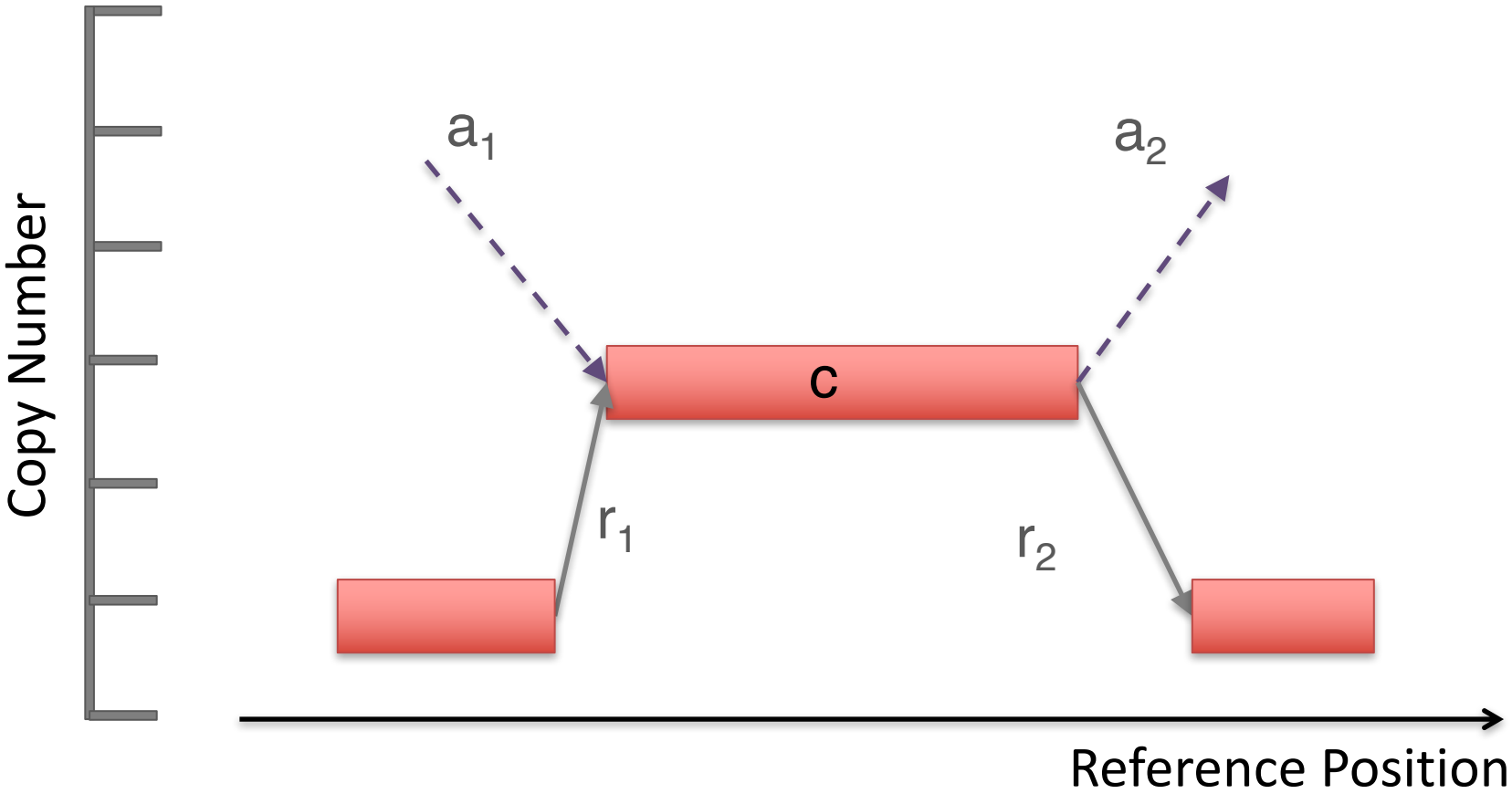
Integrating rearrangements and copy state



$$a_1 + r_1 = c = a_2 + r_2$$

JaBbA (Junction Balance Analysis):

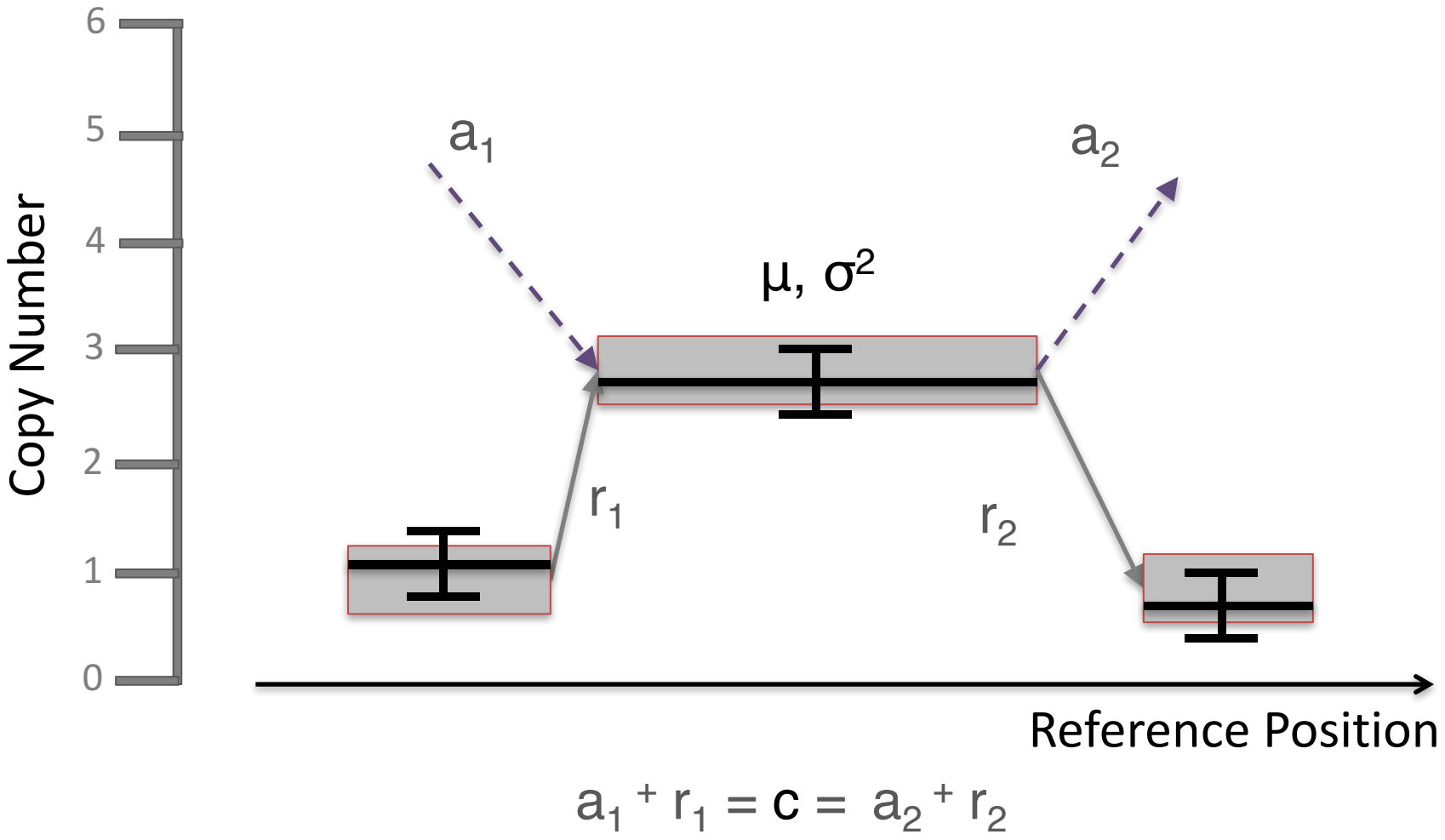
Integrating rearrangements and copy state



$$a_1 + r_1 = c = a_2 + r_2$$

JaBbA (Junction Balance Analysis):

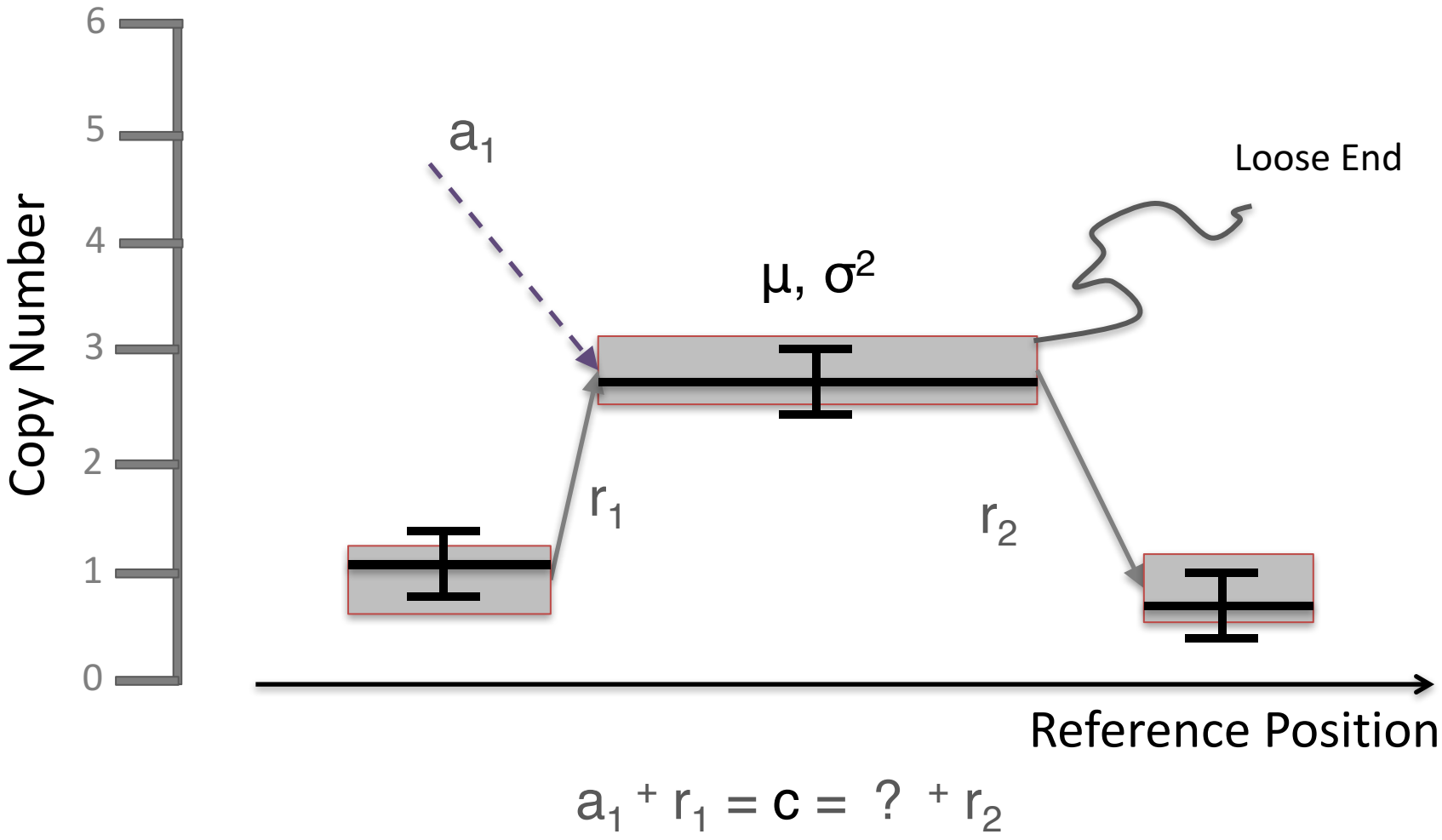
Challenge: Noisy coverage data



$$a_1 + r_1 = c = a_2 + r_2$$

JaBbA (Junction Balance Analysis):

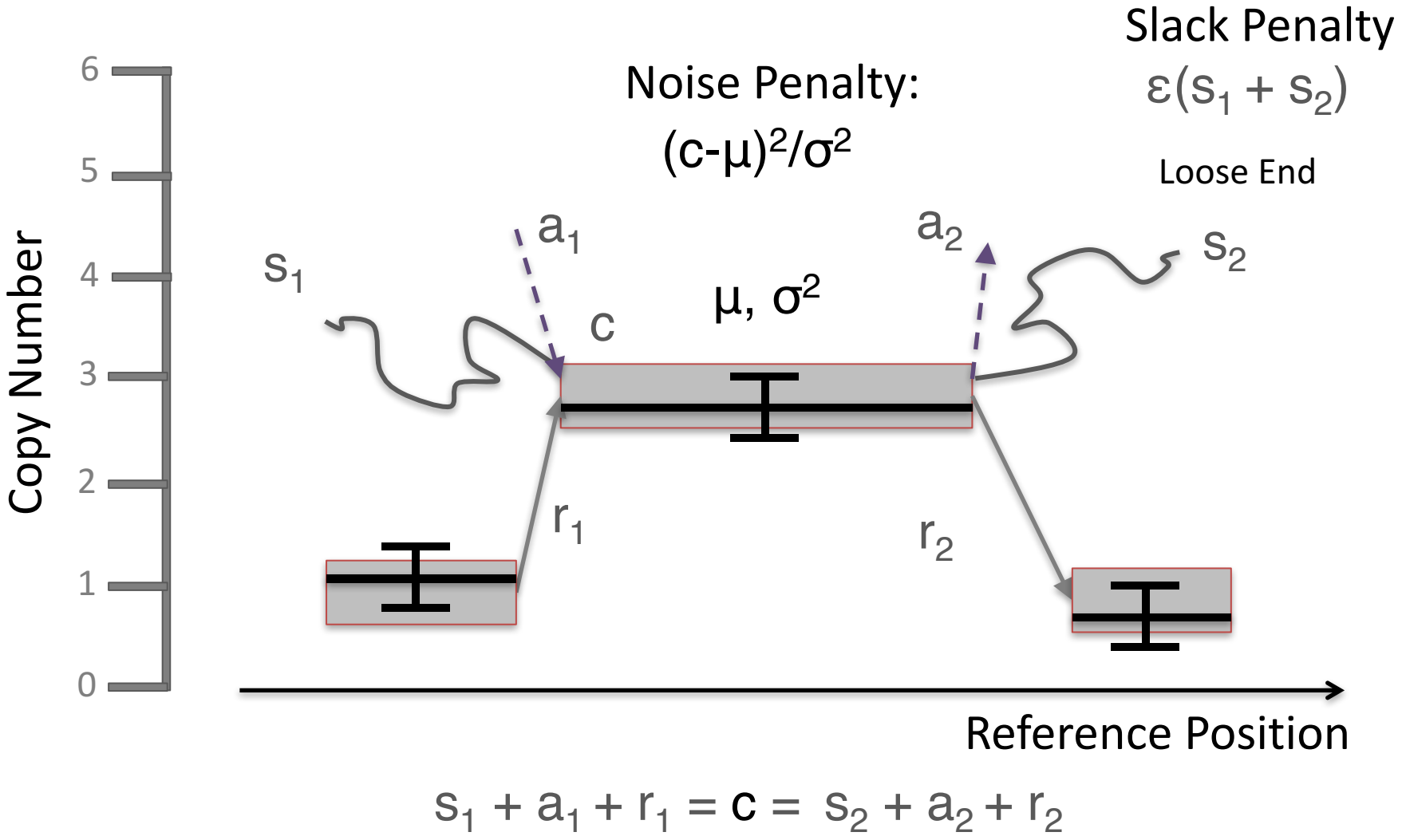
Challenge: Missing rearrangements



$$a_1 + r_1 = c = \; ? \; + r_2$$

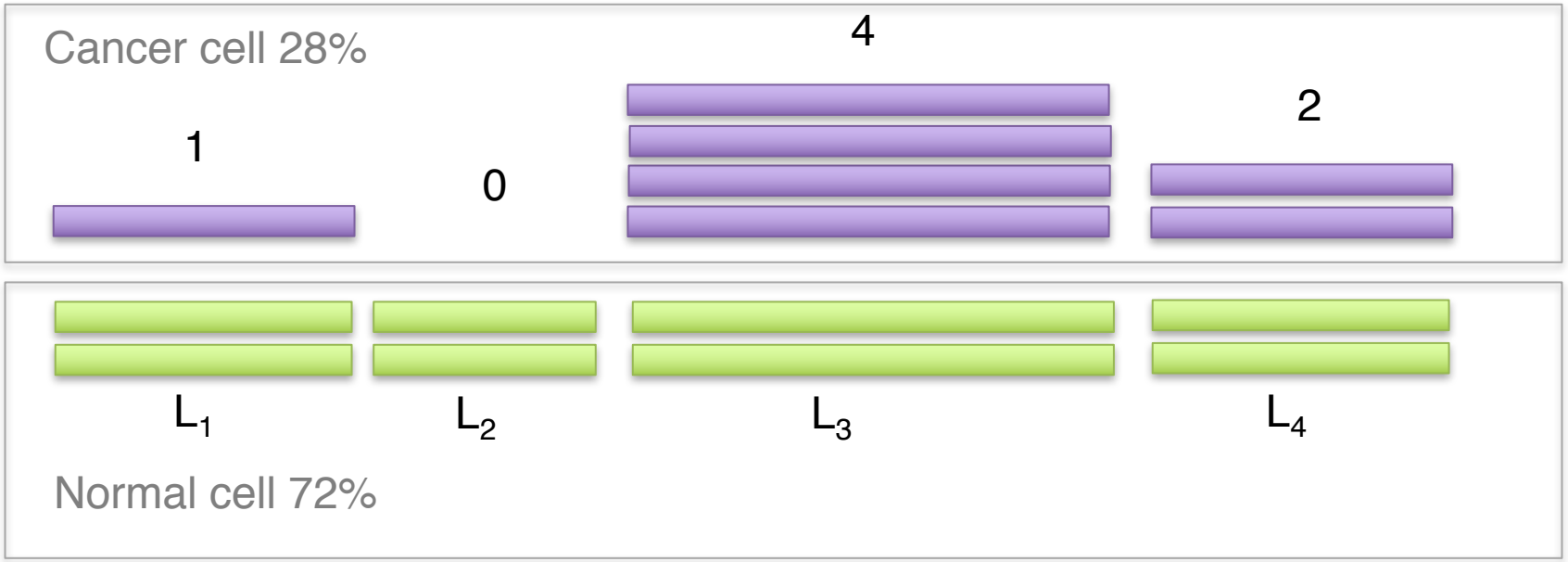
JaBbA (Junction Balance Analysis):

Statistical model



$$s_1 + a_1 + r_1 = c = s_2 + a_2 + r_2$$

Transforming analog fragment density to digital copy number



Density of fragments aligning to interval 3 is

$$\frac{(4 \times 0.28 + 2 \times 0.72) \times L_3}{0.28 \times (L_1 + 4L_3 + L_4) + 0.72 \times 2 \times (L_1 + L_2 + L_3 + L_4)}$$

See Carter *et al* *Nature Biotech* 2012, Van Loo et al *PNAS* 2011

Transforming analog fragment density to digital copy number

$$\frac{\mu_i}{\mu^T L} = \frac{\alpha v_i + 2(1-\alpha)}{\alpha v^T L + 2(1-\alpha)\|L\|_1}$$

$\mu \in \mathbb{R}^n$ Vector of fragment densities across n intervals (data)

$L \in \mathbb{Z}^n$ Vector of n interval widths (data)

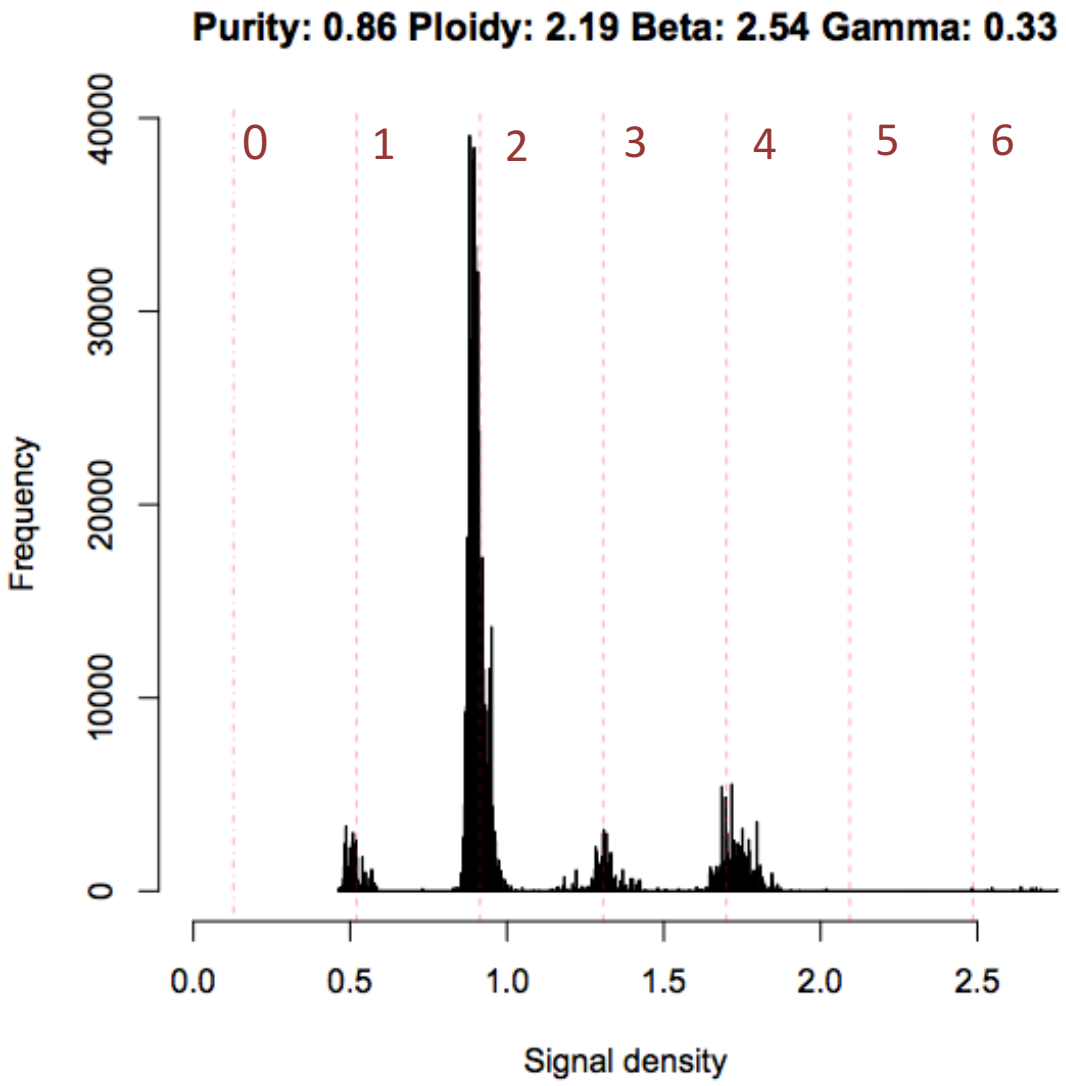
$v \in \mathbb{Z}^n$ Vector of n interval copy numbers (inferred)

$\alpha \in [0,1]$ Tumor cell fraction (purity, inferred)

Let $\gamma = \frac{2(1-\alpha)}{\alpha}$ $\beta = \frac{\alpha v^T L + 2(1-\alpha)\|L\|_1}{\alpha \mu^T L}$

$$v_i + \gamma = \beta \mu_i$$

Transforming analog fragment density to digital copy number



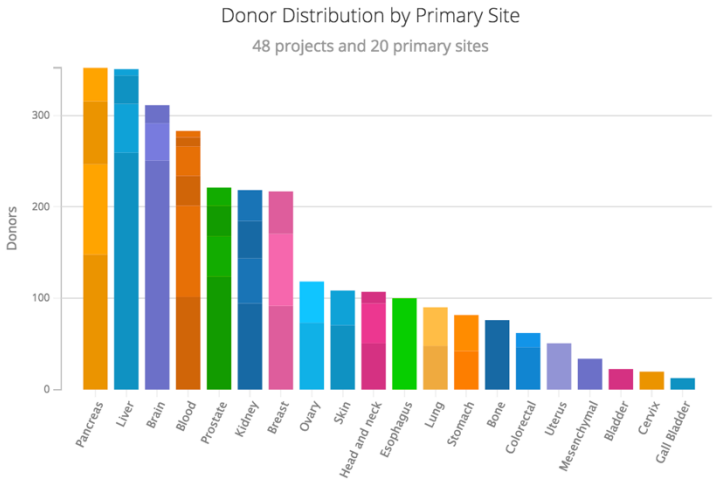
Example: lung adenocarcinoma

ICGC International Cancer Genome Consortium

in the cloud

PCAWG PanCancer Analysis OF WHOLE GENOMES

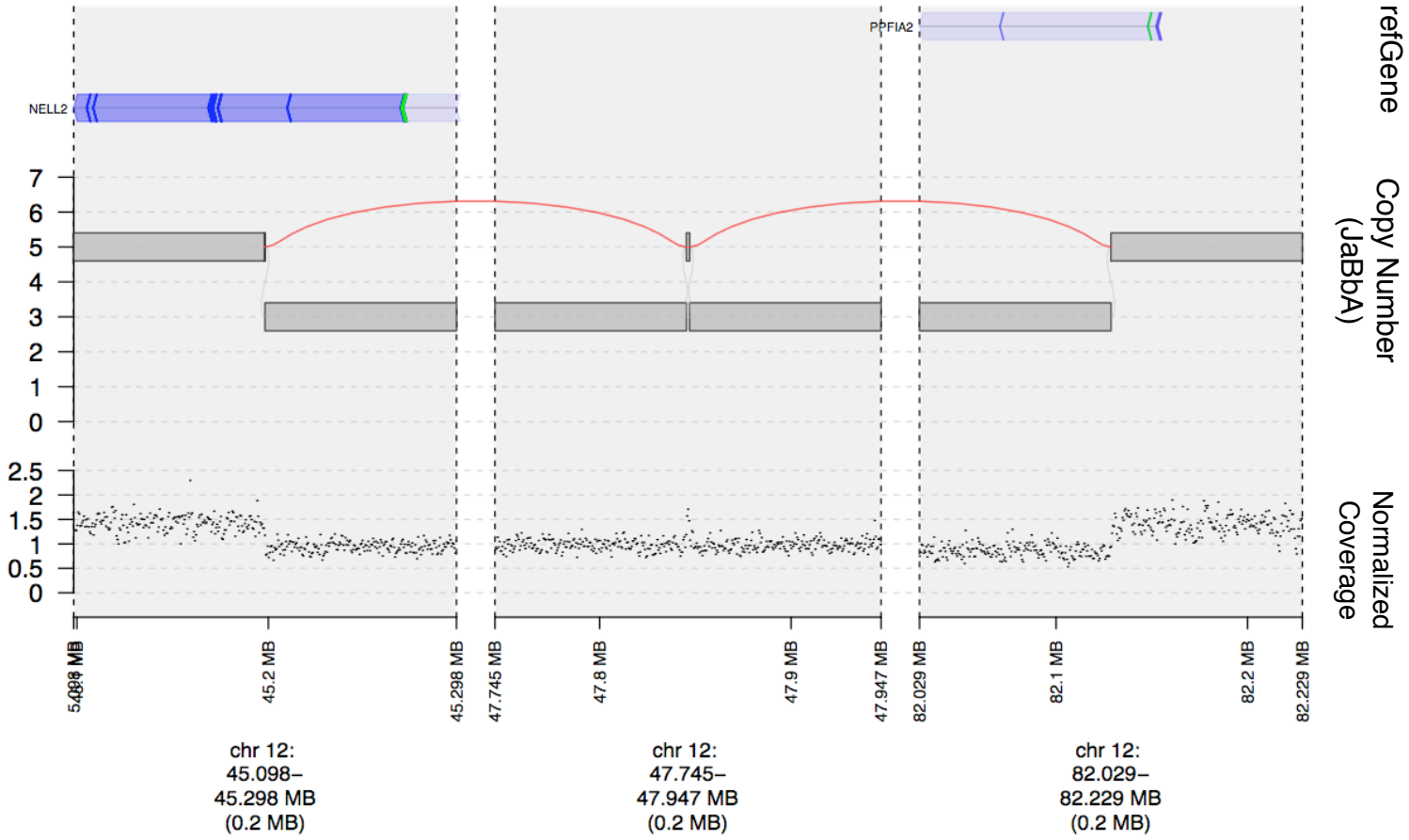
- 2834 tumor and matched normal whole genome sequences across 30 cancer types and 48 projects
- 1.5 Petabytes of raw data + downstream analytic pipelines
- 13 analysis working groups, including PCAWG-6 (structural variation dataset)



CCLE Cancer Cell Line Encyclopedia

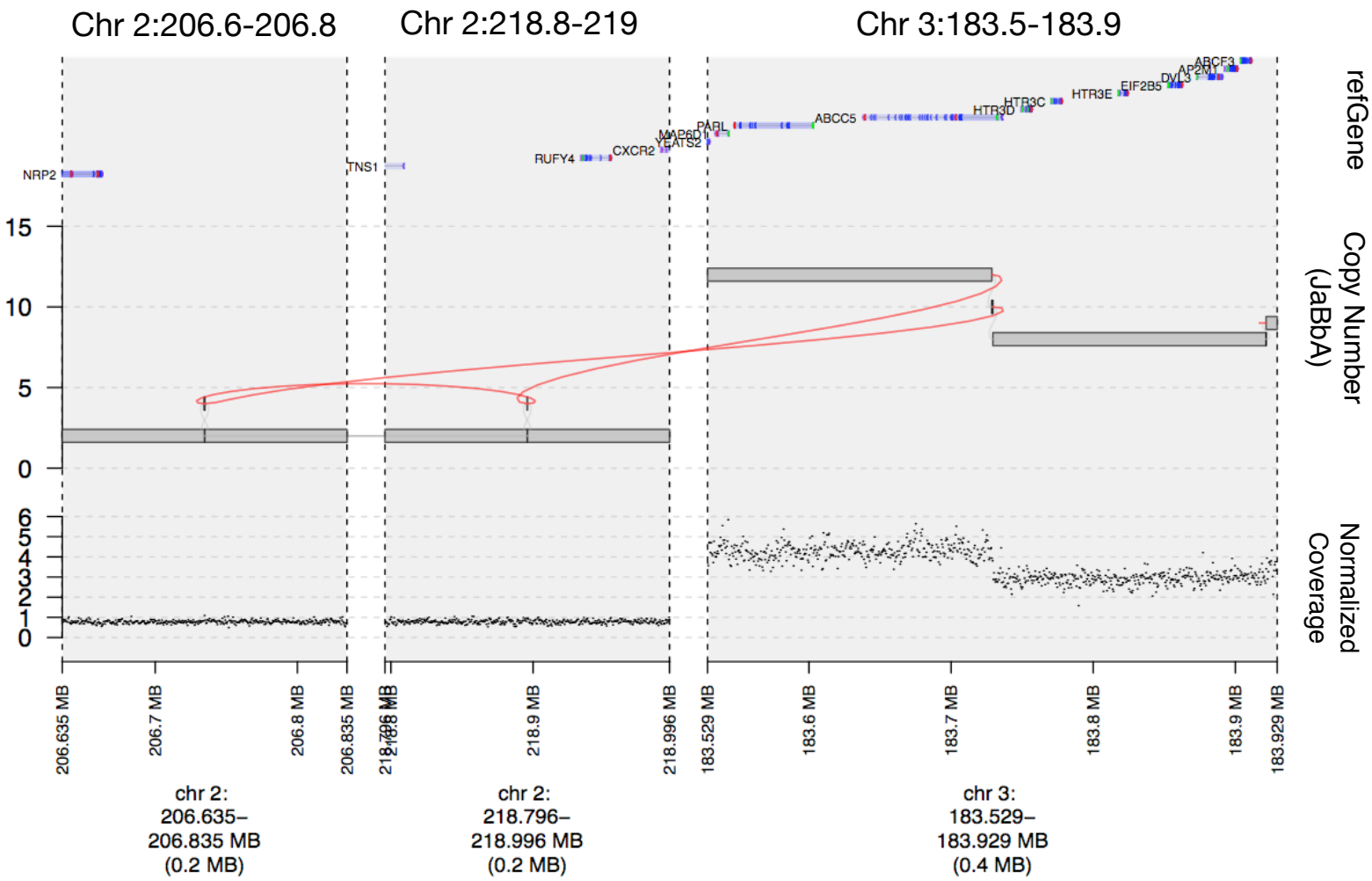
- 330 WGS cell lines across 16 cancer types
- collaboration with Mahmoud Ghandi and Jesse Boehm at Broad Institute

Long range coupling of copy changes



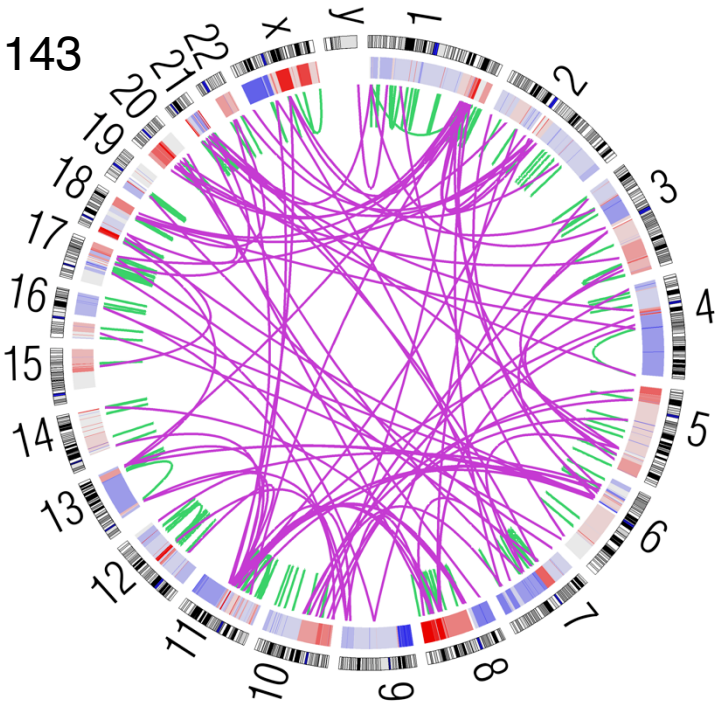
Long distance coupling of copy changes through rearrangement junctions

“Kidnapped” loci
(Lung adenocarcinoma)

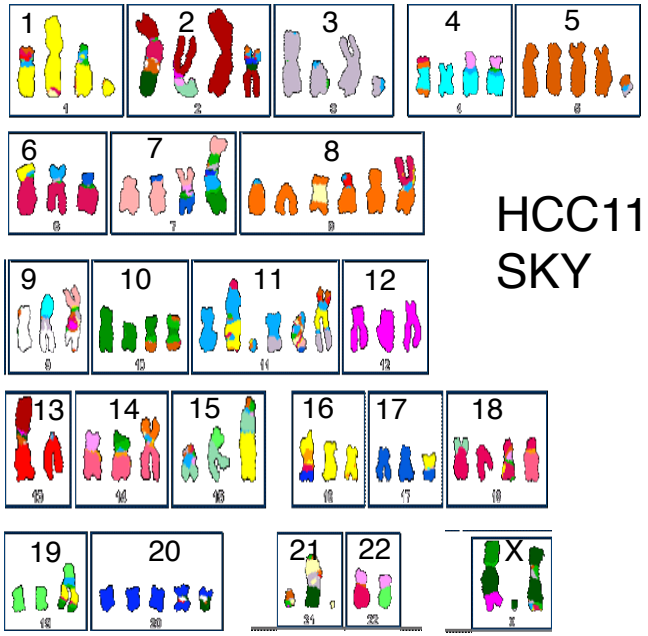


WGS vs. cytogenetics

HCC1143
WGS

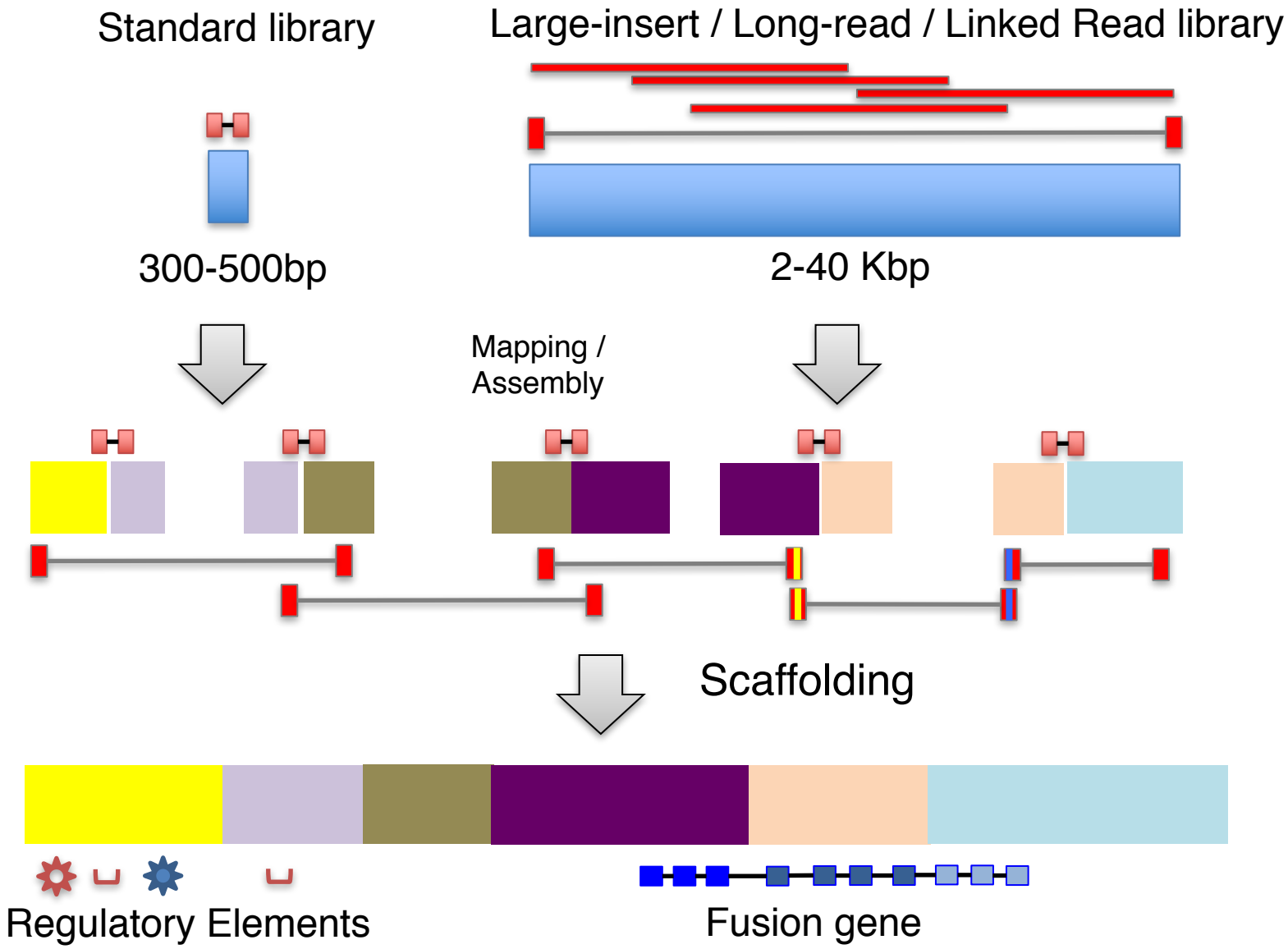


!=



HCC1143
SKY

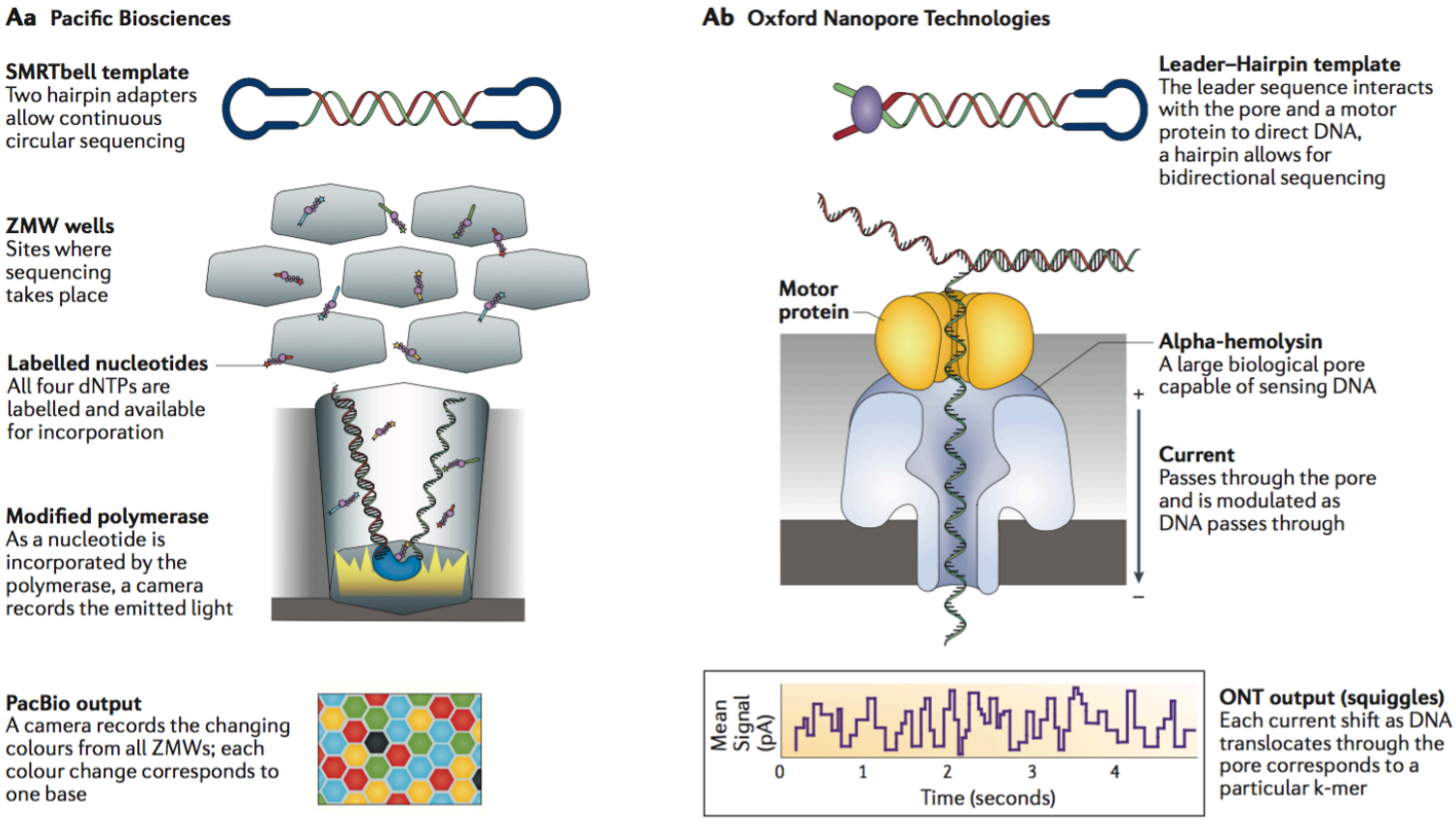
“Shattered” cancer data



Long-read sequencing

12-20 Kbp reads, $85-400/Gb**

200-900 Kbp, $100-180/Gb**

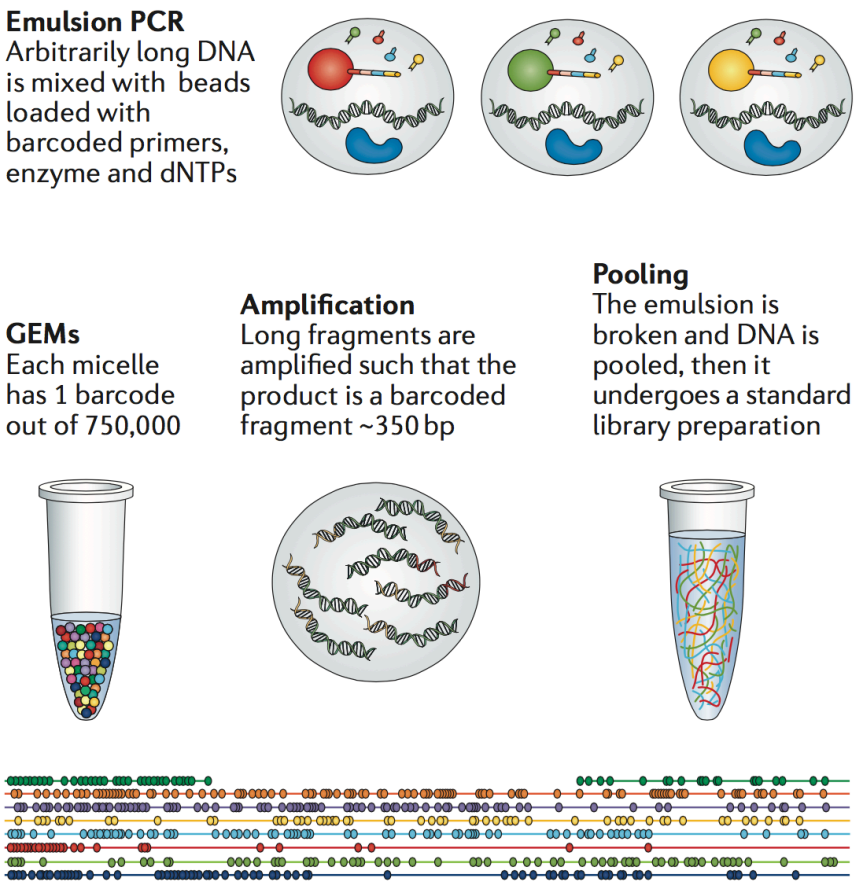


**https://blog.genohub.com/2017/06/16/pacbio-vs-oxford-nanopore-sequencing/

Goodwin, Mcpherson, McCombie *Nature Reviews Genetics* 2016

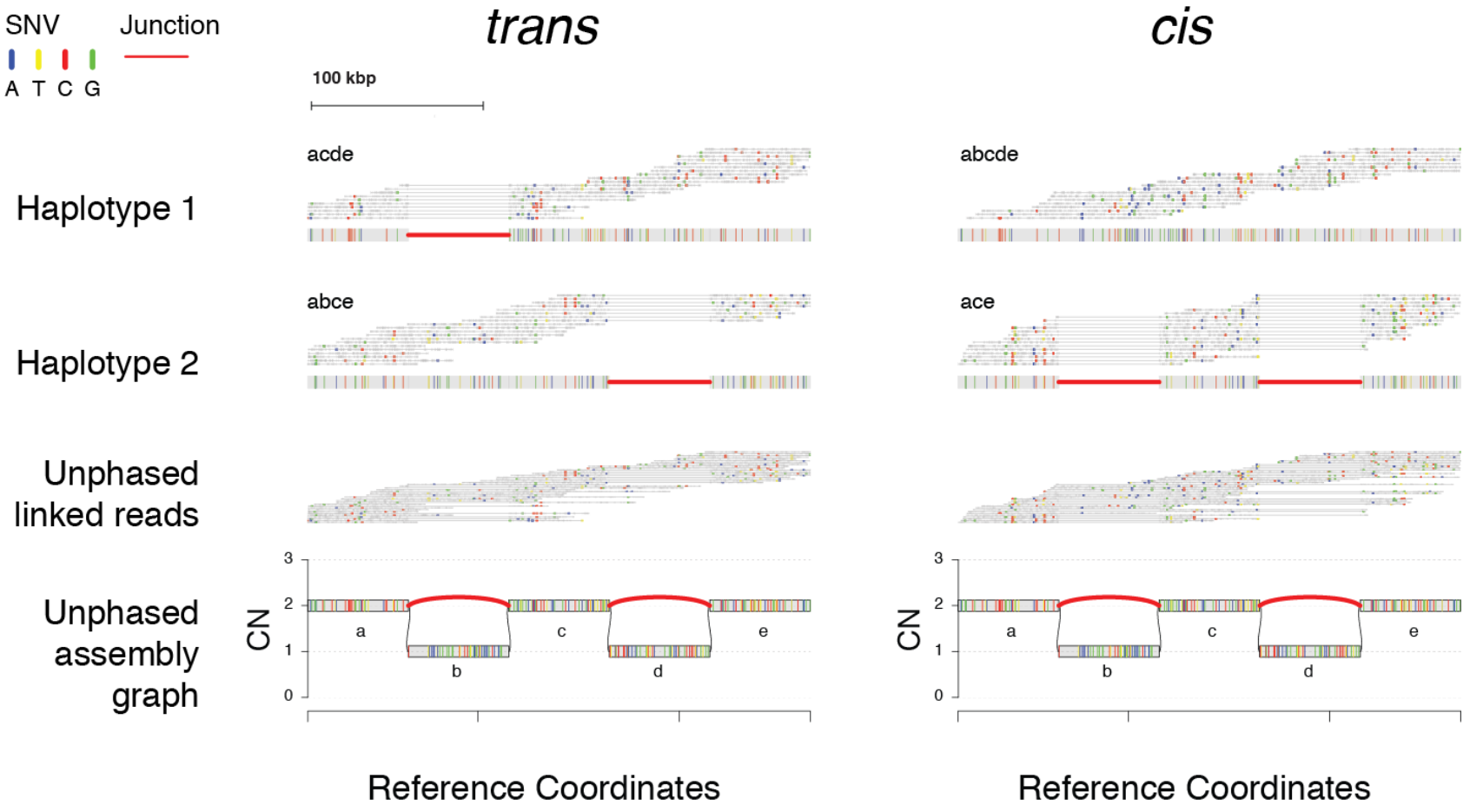
Linked-read whole genome sequencing (10x genomics)

100 Kbp “synthetic long reads” $7/Gb

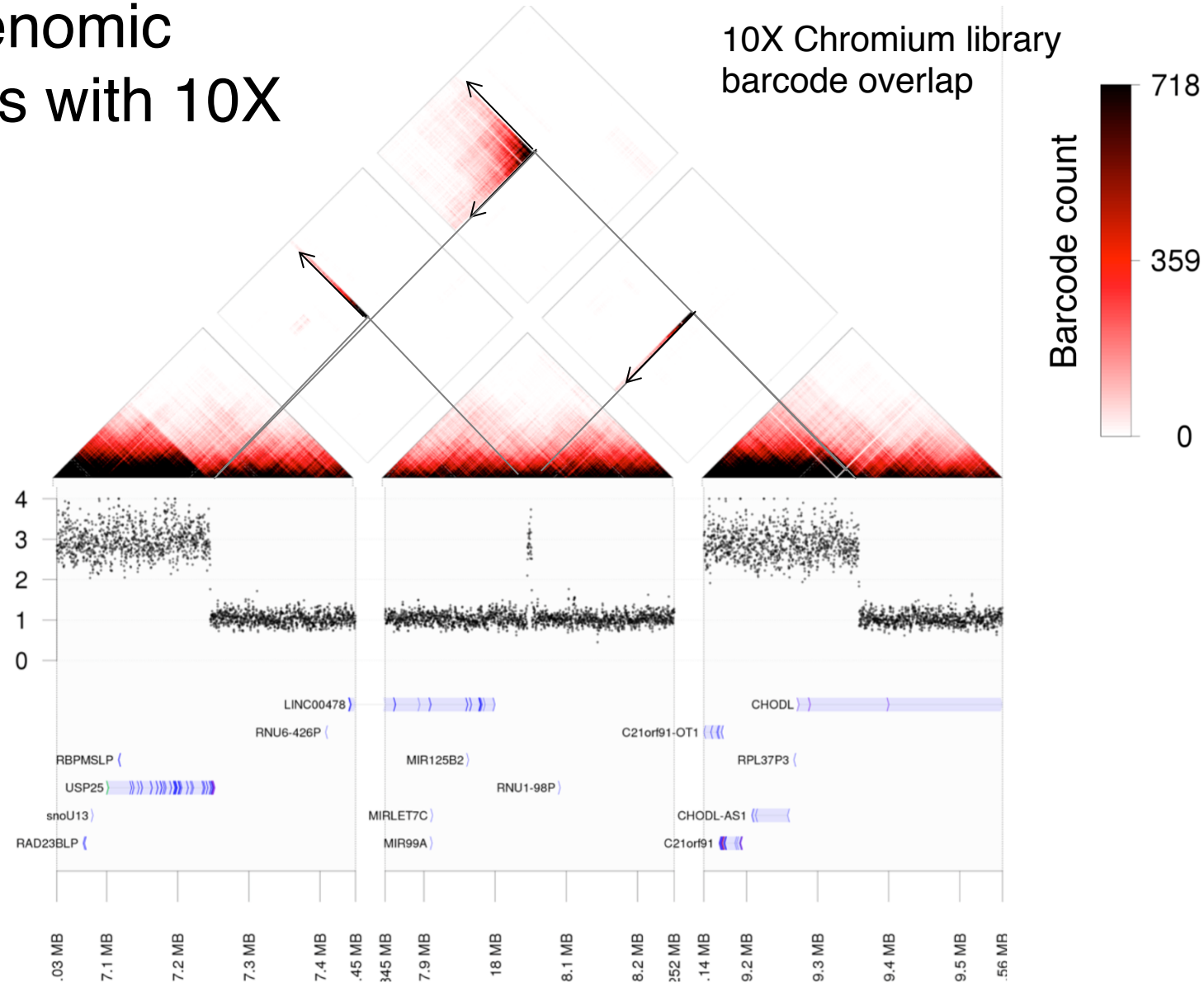


Goodwin, Mcpherson, McCombie *Nature Reviews Genetics* 2016

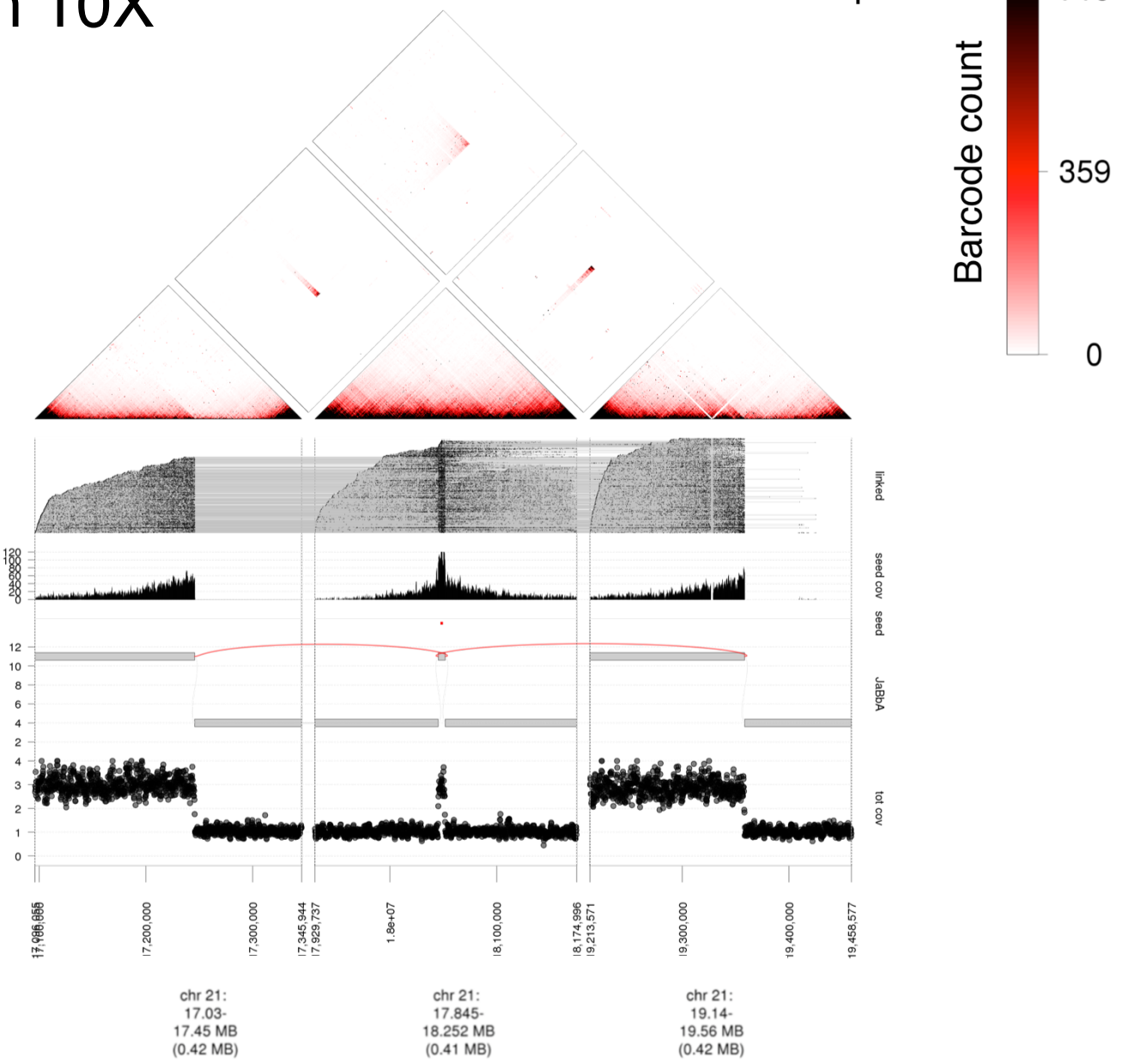
Phasing rearrangements with 10X



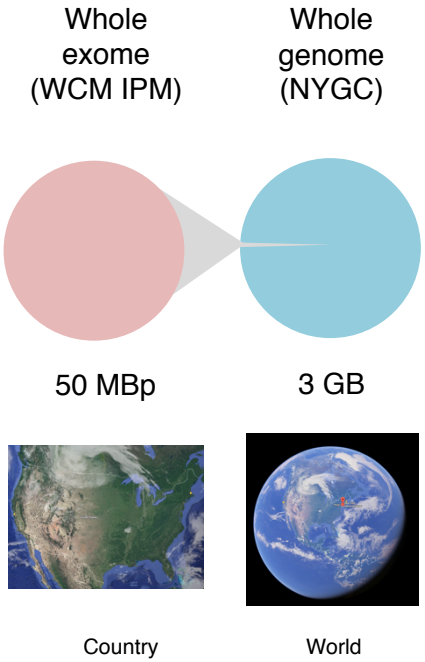
Phasing genomic kidnappings with 10X

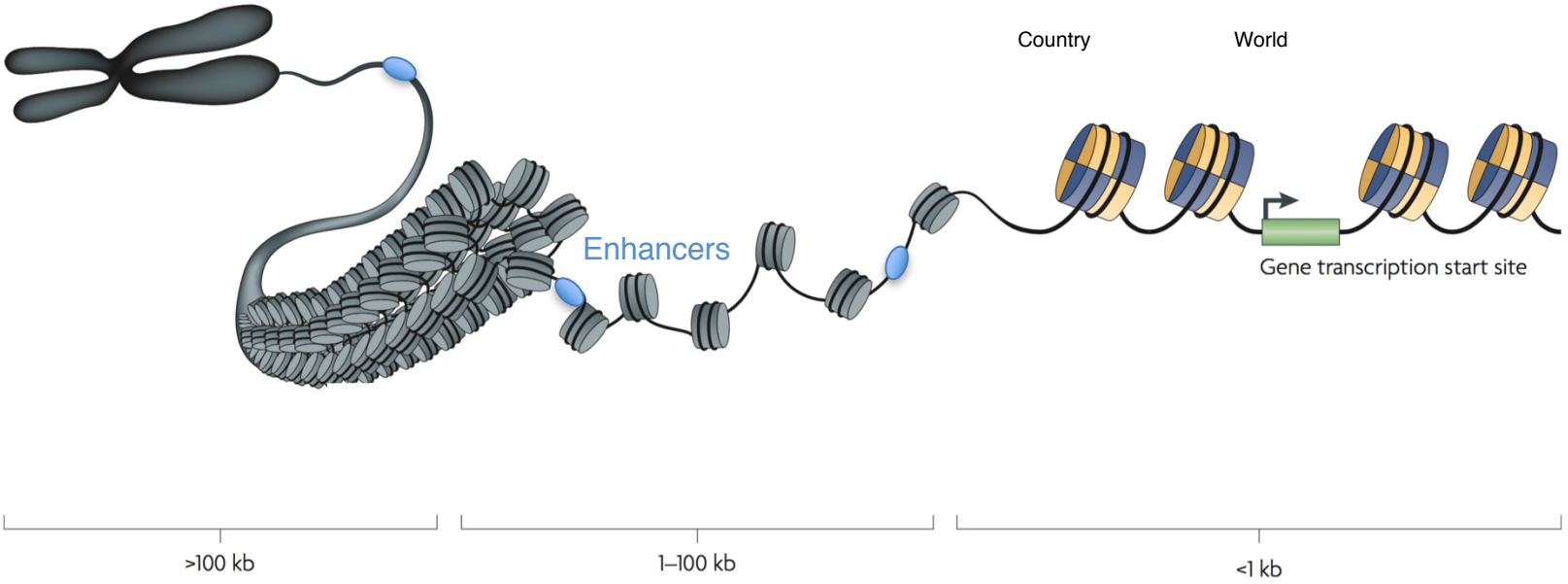


Phasing genomic kidnappings with 10X



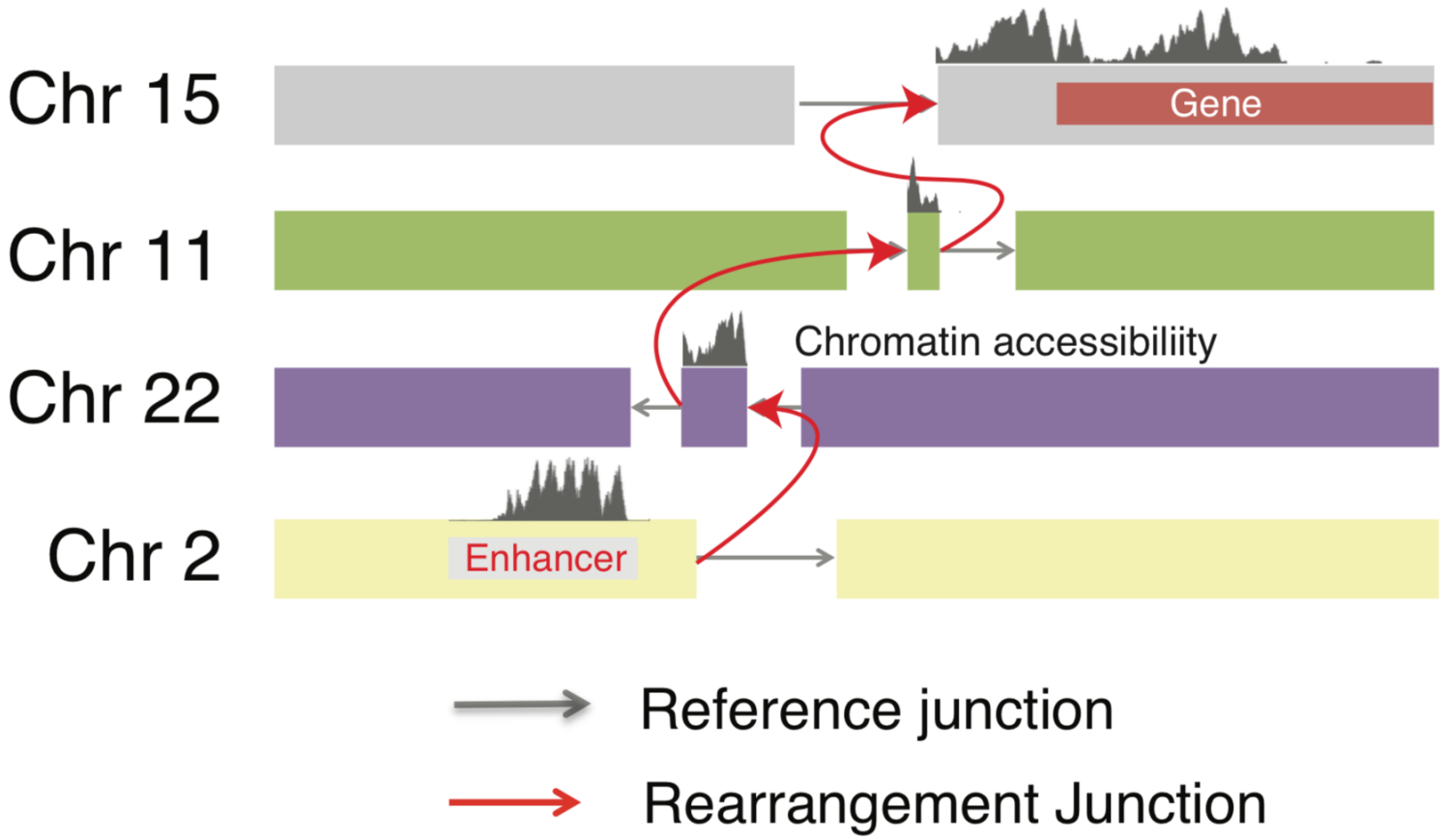
Diving into the ocean of regulatory DNA



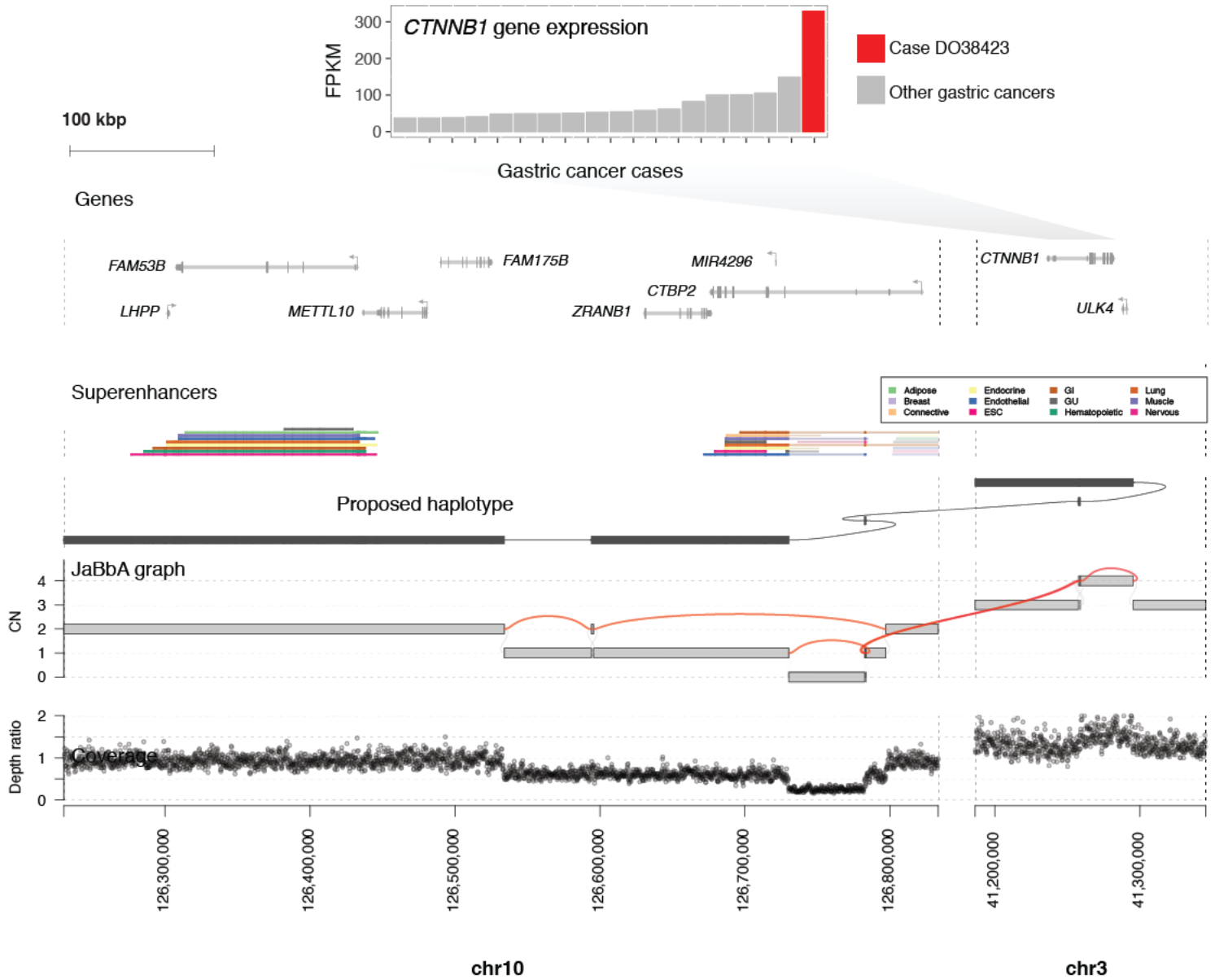


Alexander *Nature Reviews Genetics* 2011

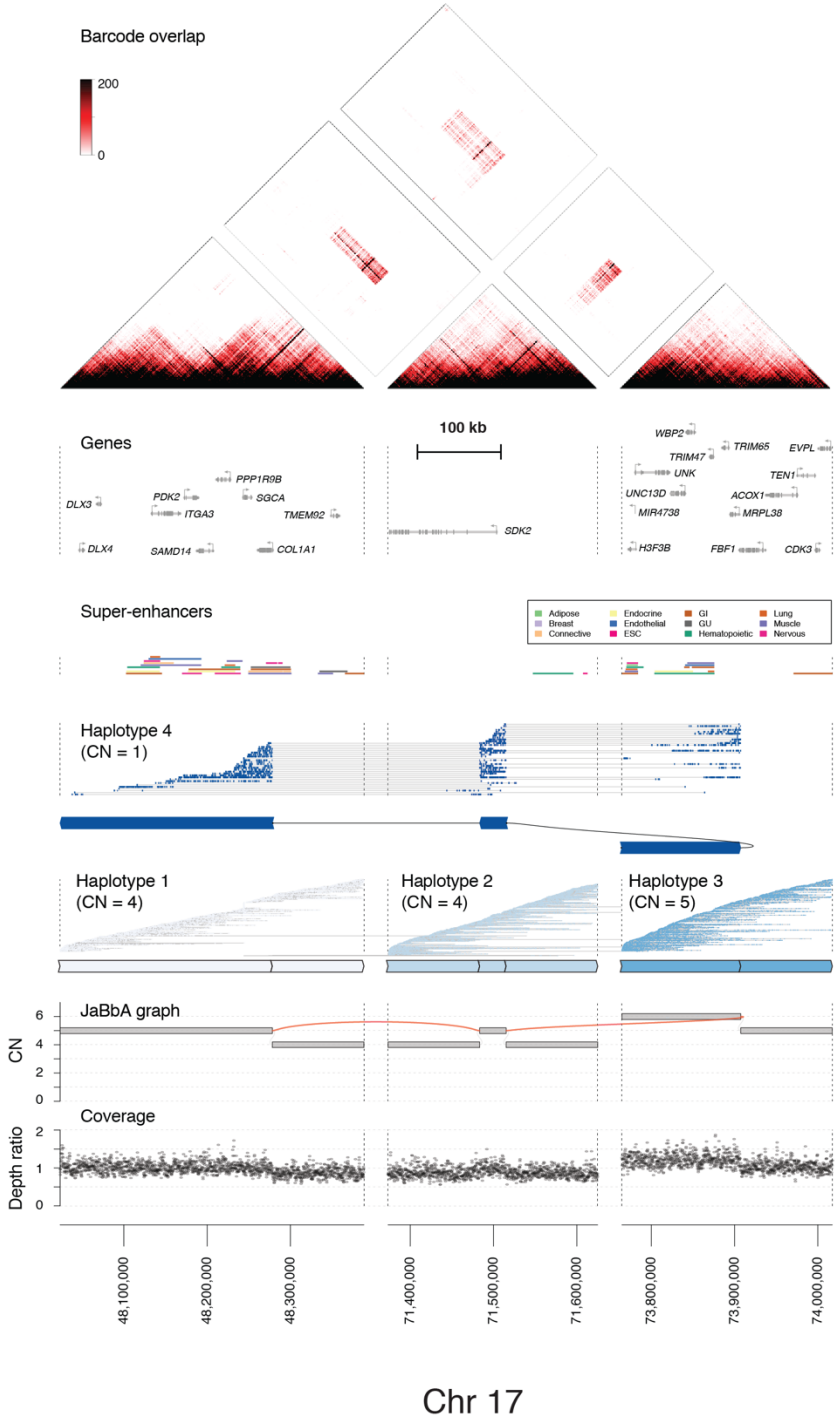
Enhancer perturbations



Enhancer hijacking in PCAWG



10X phasing of enhancer hijacking in IPM bladder cancer case



mski@mskilab.org
http://github.com/mskilab

Evan Biederstedt (research specialist)
Julie Behr (Tri-I CBM PhD student)
Aditya Deshpande (Tri-I CBM PhD student)
Zoran Gajic (undergrad)
Kofi Gyan (Tri-I CBM PhD student)
Kevin Hadi (WCM PBSB PhD student)
Khagay Nagdimov (intern)
Joel Rosiene (medical student)
Huasong Tian (research scientist)
Netha Ulahannan (postdoctoral associate)
Trent Walradt (medical student)
Charalampos Xanthopoulakis (soft eng)
Xiaotong Yao (Tri-I CBM PhD student)

Weill Cornell Medicine
Pathology & Laboratory Medicine

Weill Cornell Medicine
Englander Institute
for Precision Medicine

NEW YORK
GENOME CENTER