

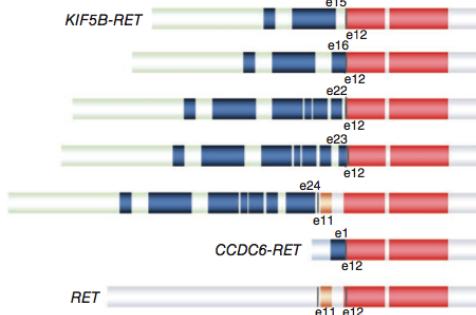
# **Distinct Classes of Complex Structural Variation across Thousands of Cancer Whole Genomes**

**Marcin Imielinski MD PhD**

**Clin Gen Talk**

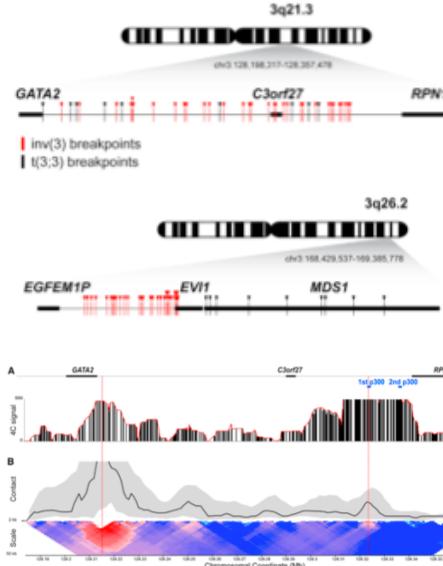
# Studying cancer genome structure: motivation

## Gene Fusions



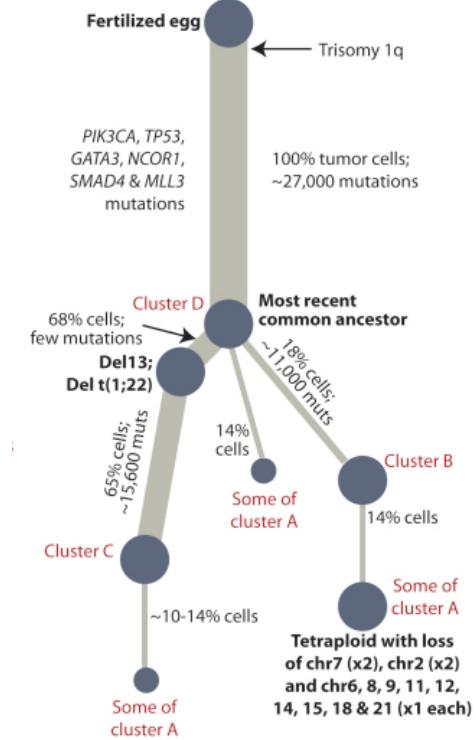
Takeuchi .. Ishikawa  
(Nature Medicine 2012)

## Noncoding Variants



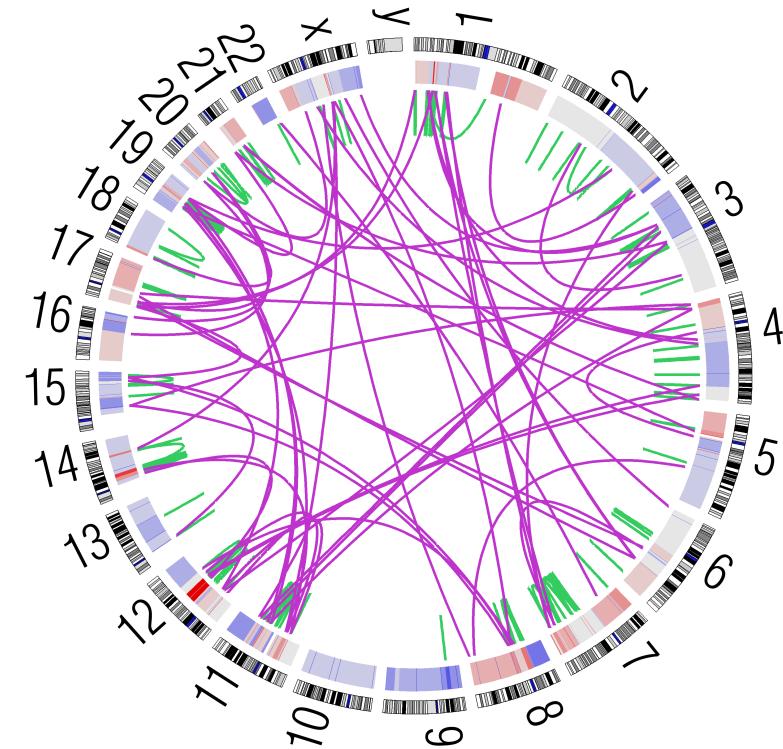
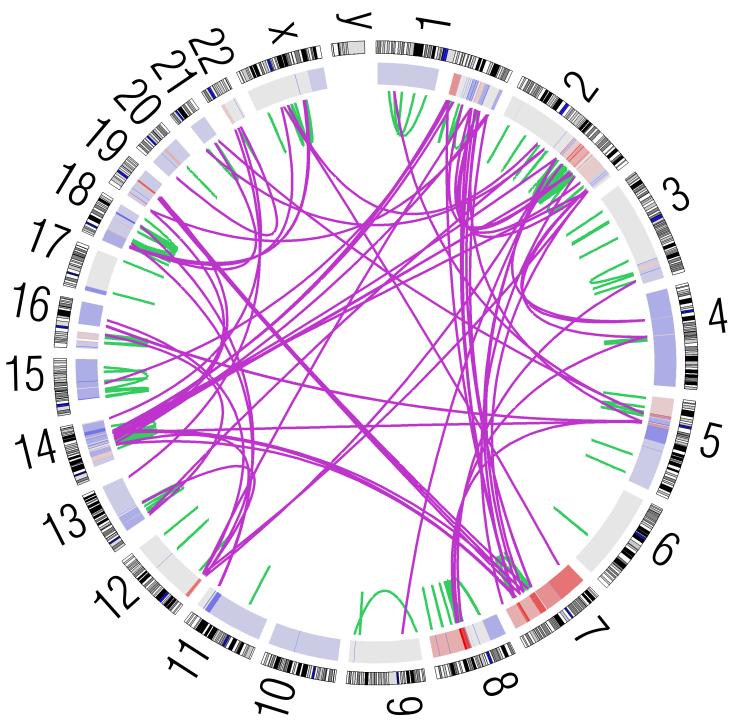
Groschel .. Delwel  
(Cell 2014)

## Mutational processes



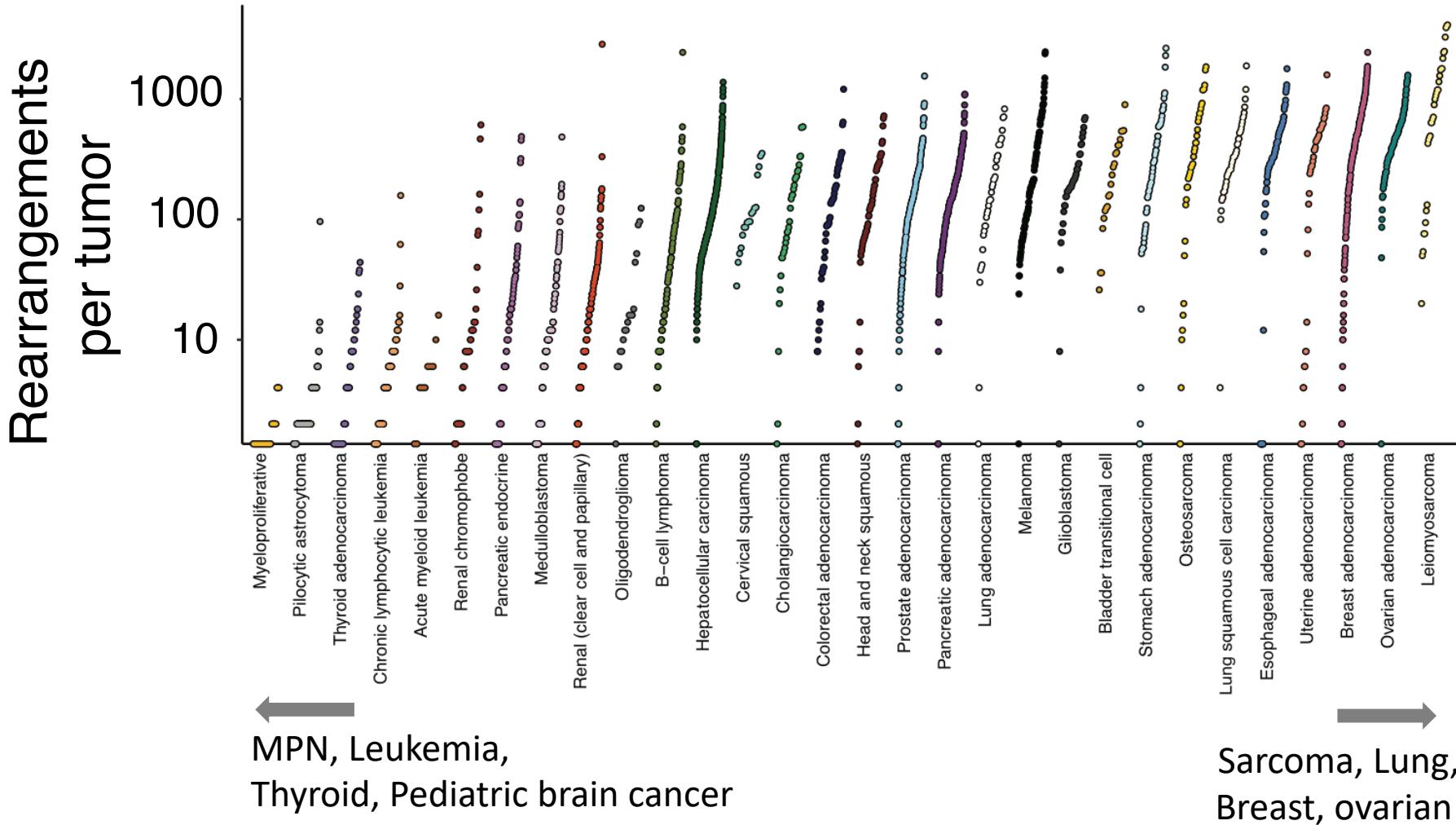
Nik-Zainal .. Stratton  
(Cell 2012)

# Which is the smoker?

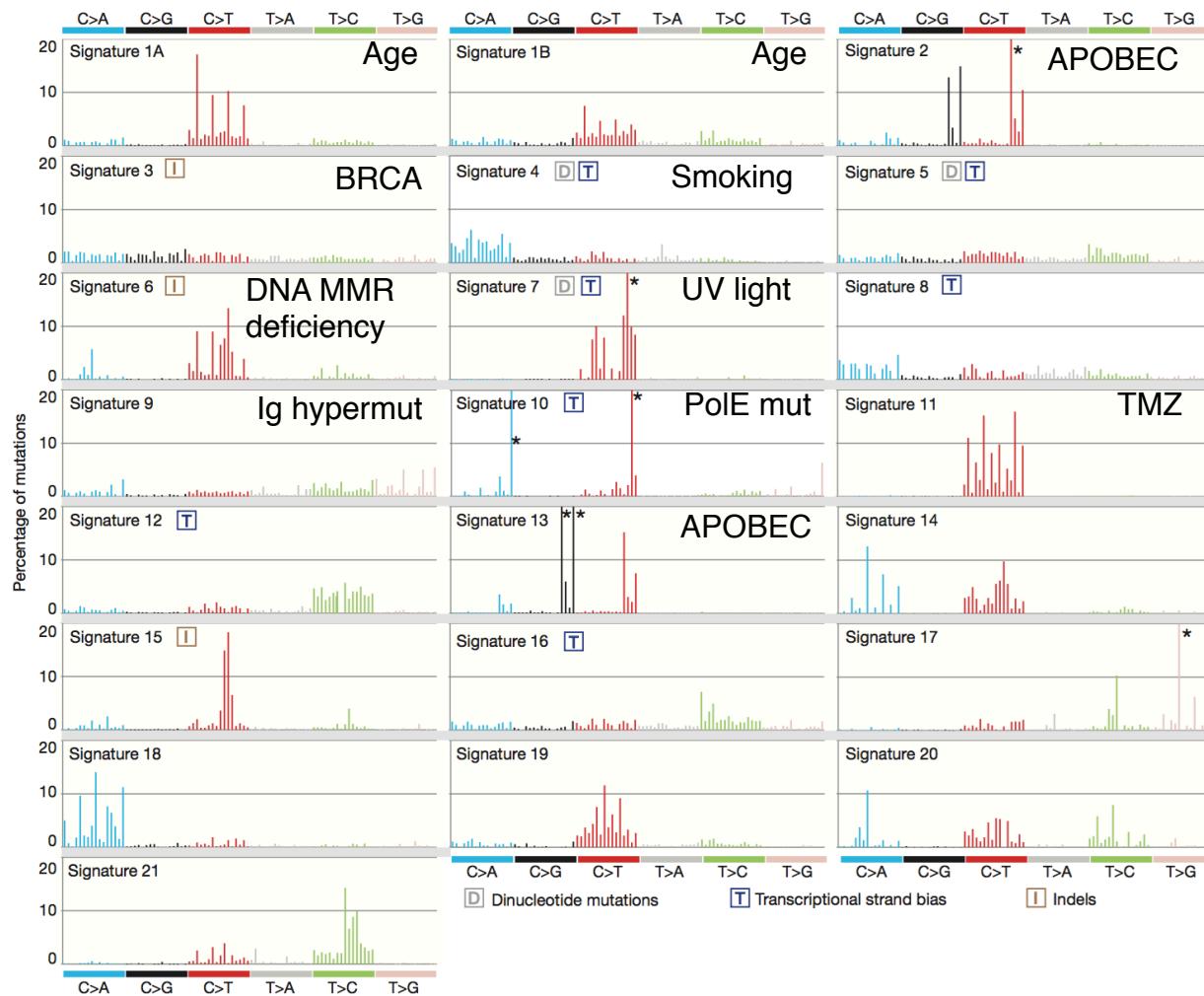


Imielinski et al *Cell* 2012

# Rearrangement burden across ~2700 cancers

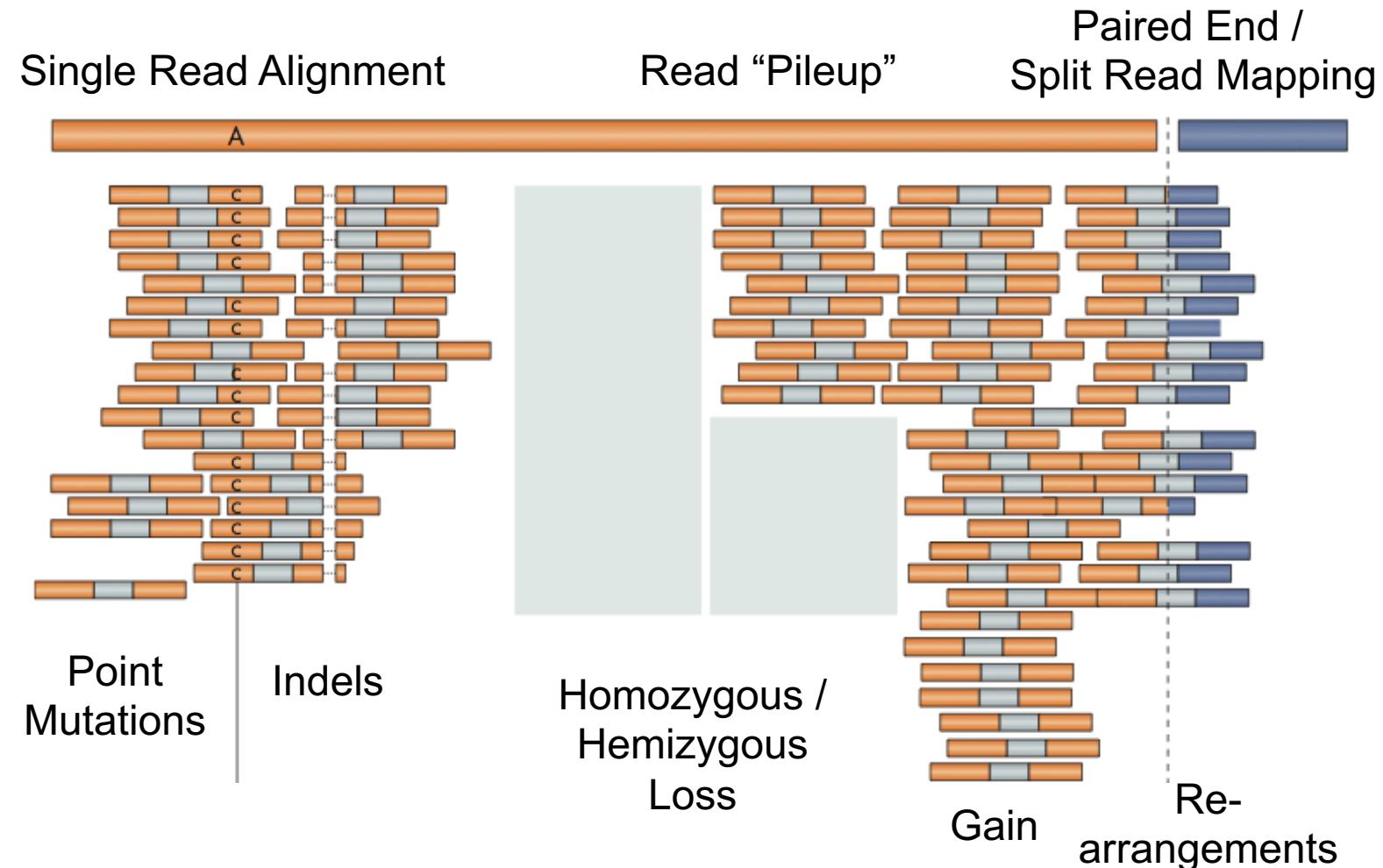


# SNV signatures



Alexandrov et al 2013 *Nature*

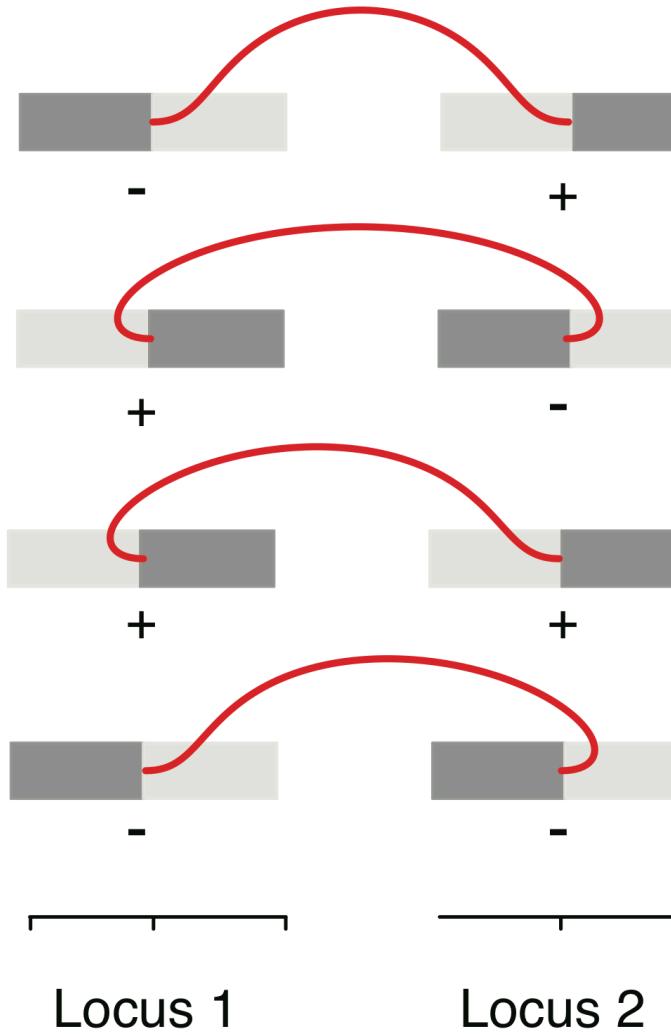
# A primer in somatic variation (from short reads)



Meyerson, Getz, Gabriel (Nature Reviews Genetics 2010)

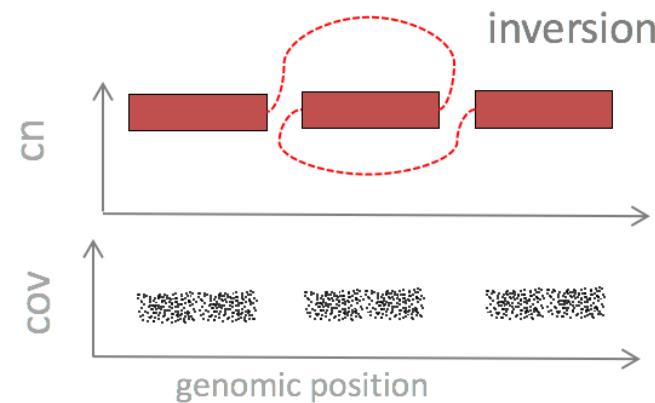
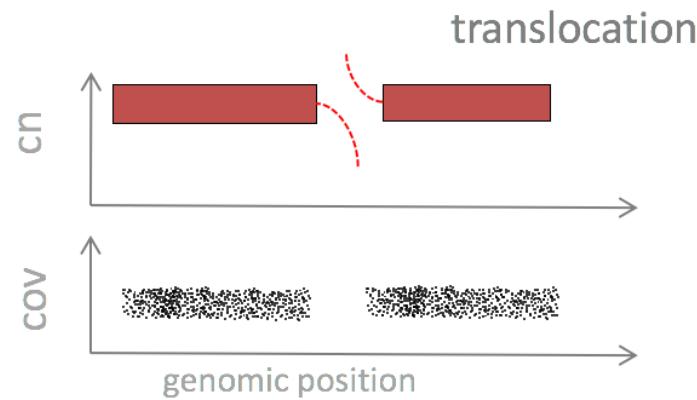
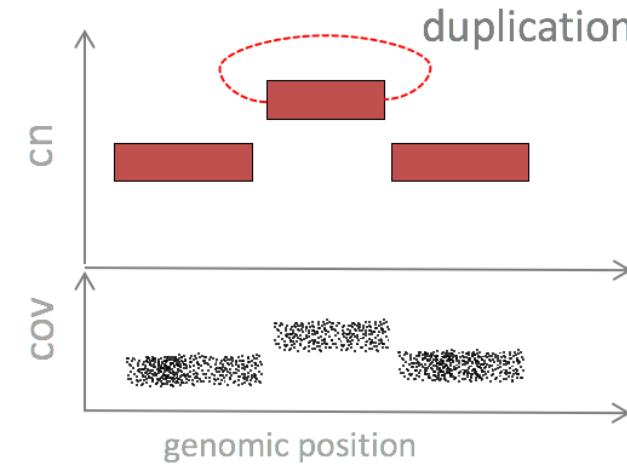
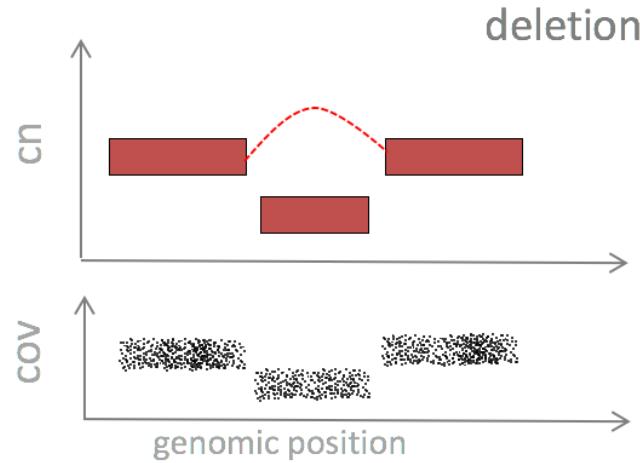
Adapted from Meyerson, Getz, Gabriel (Nature Reviews Genetics 2010)

# How do we “count” rearrangements?

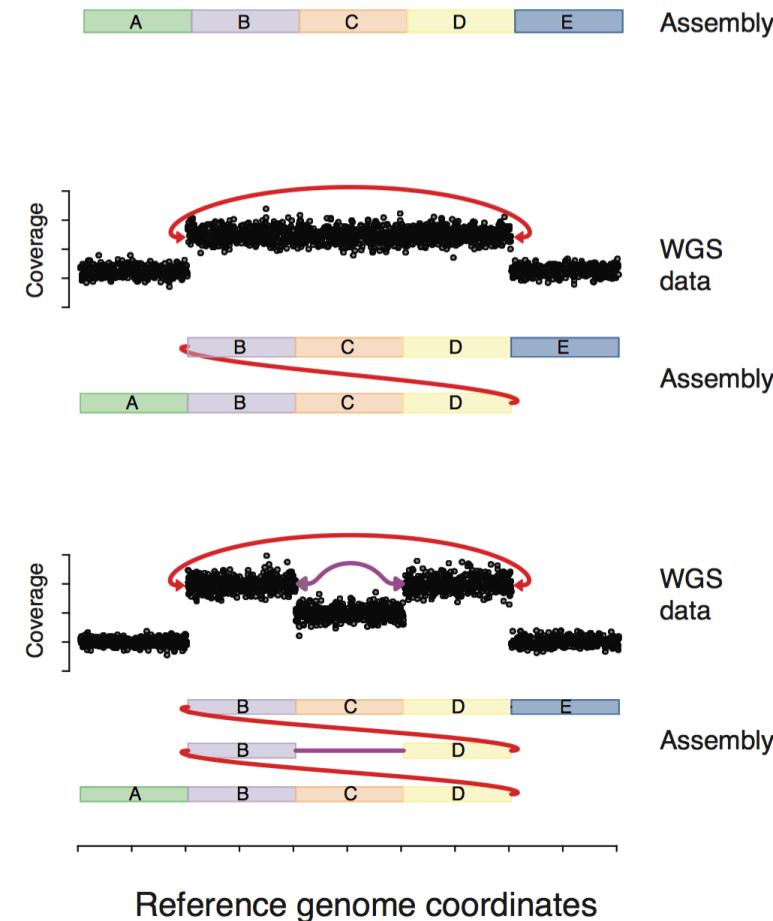
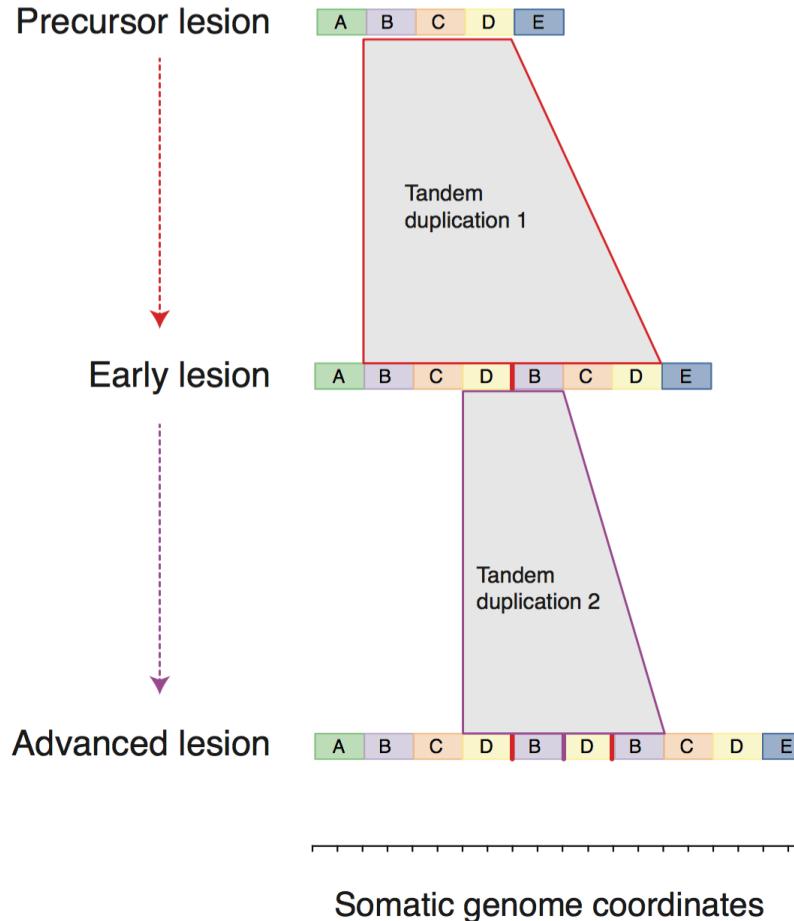


Maciejowski and Imielinski 2017

# Can it be all so simple?



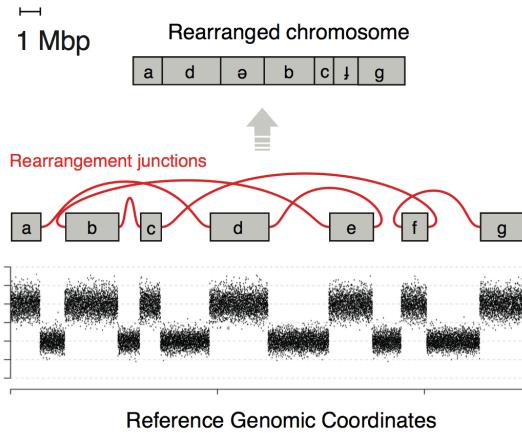
# What is an “event”?



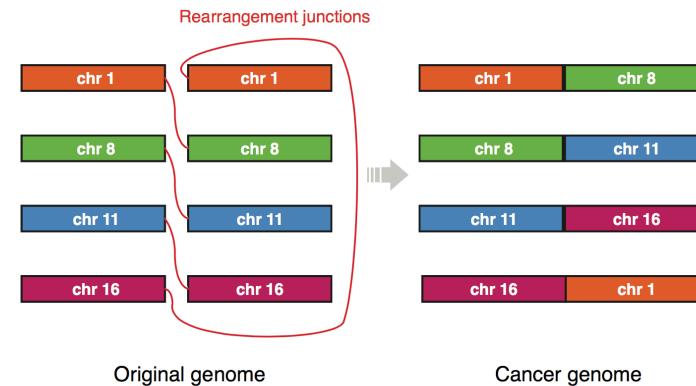
Maciejowski and Imielinski (2017)

# Complex structural variation in cancer

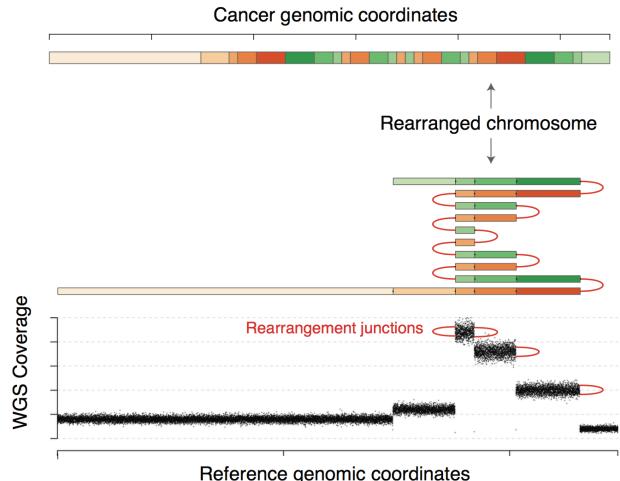
## Chromothripsis



## Chromoplexy

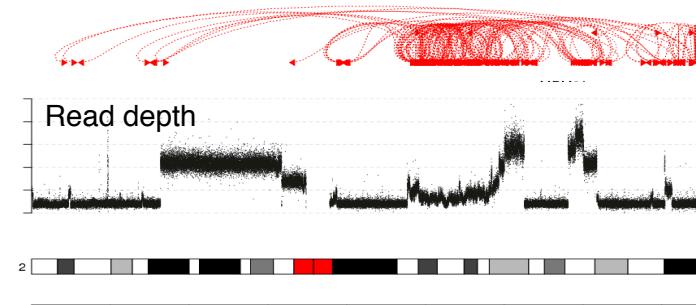


## Breakage-fusion-bridge



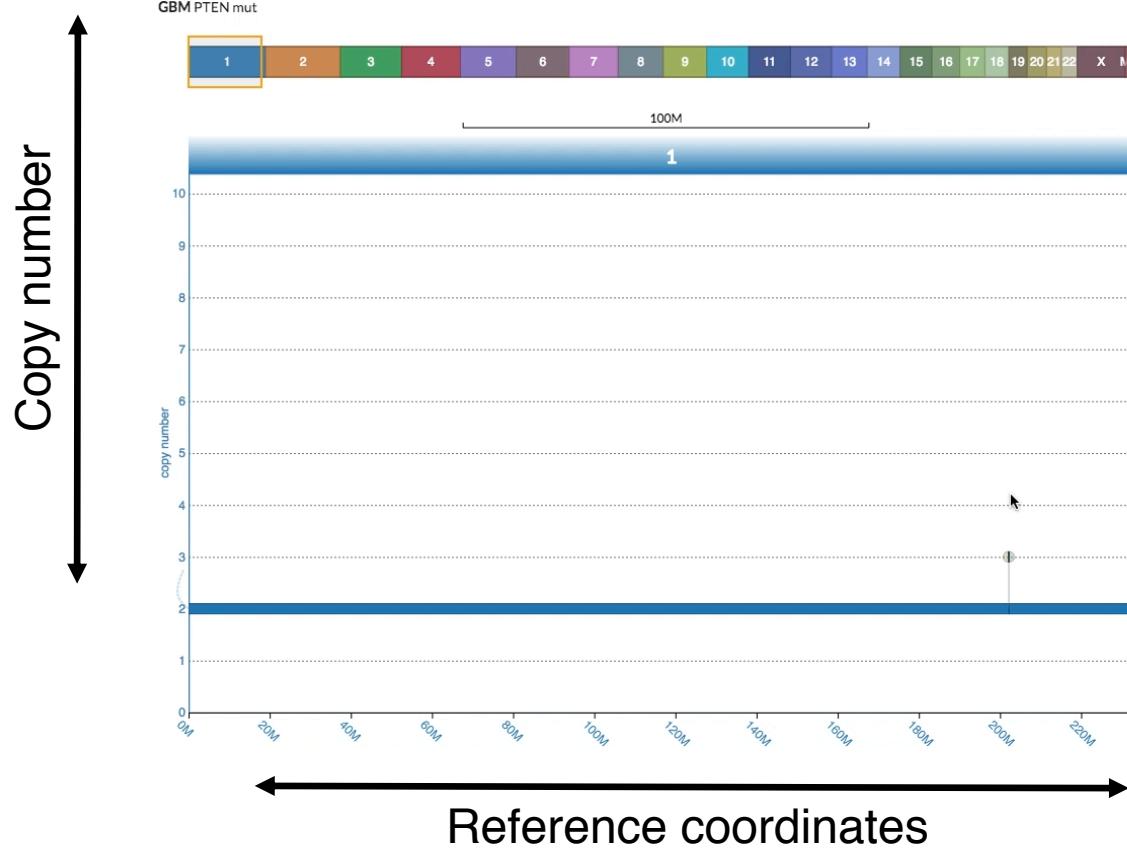
??????

Junctions



Imielinski and Rubin (Nat Clin Oncology 2017)

# Templated insertion chains (TIC)

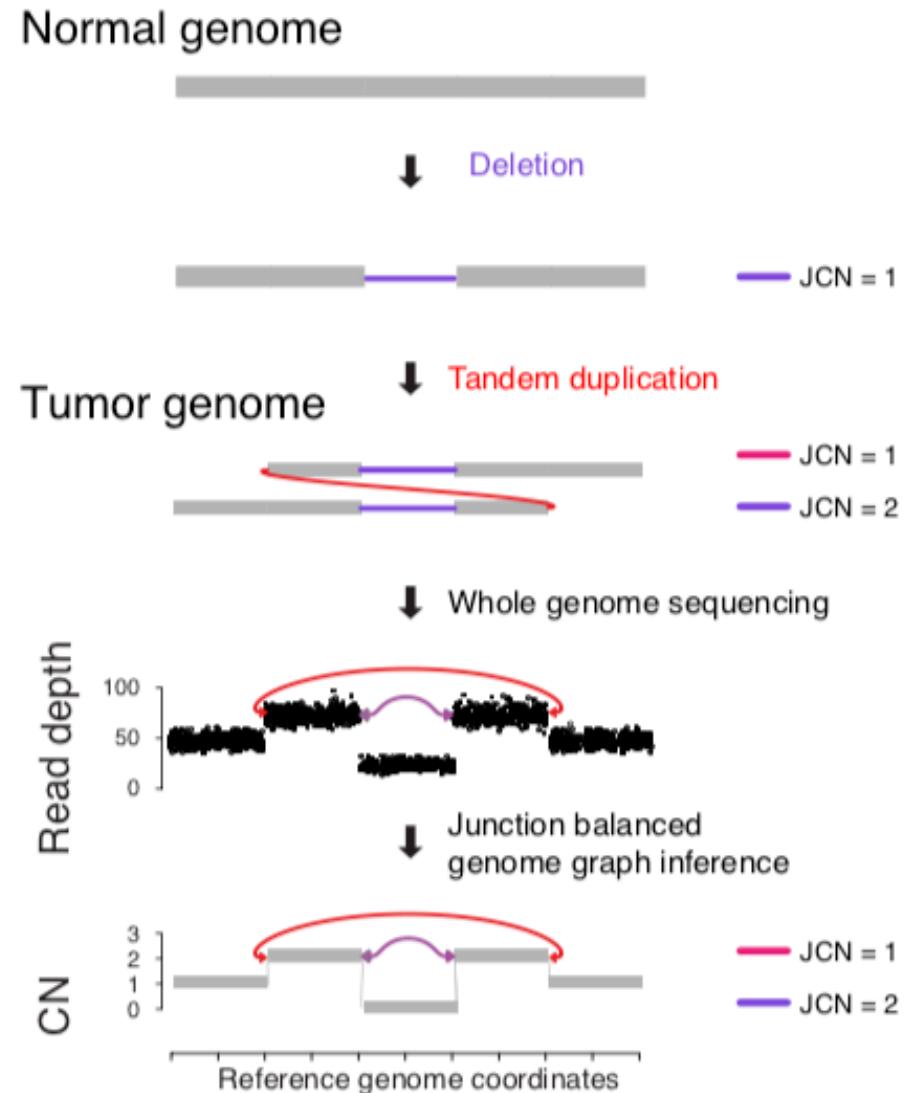


Charalampos  
Xanthopoulakis

<https://github.com/mskilab/gGnome>  
<https://github.com/mskilab/gGnome.js>

[www.mskilab.com/gGraph](http://www.mskilab.com/gGraph)

# Genome graphs and the topology of junction copy number (JCN)

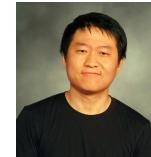


# Genome graph inference (JaBbA)

Kevin  
Hadi



Xiaotong  
Yao

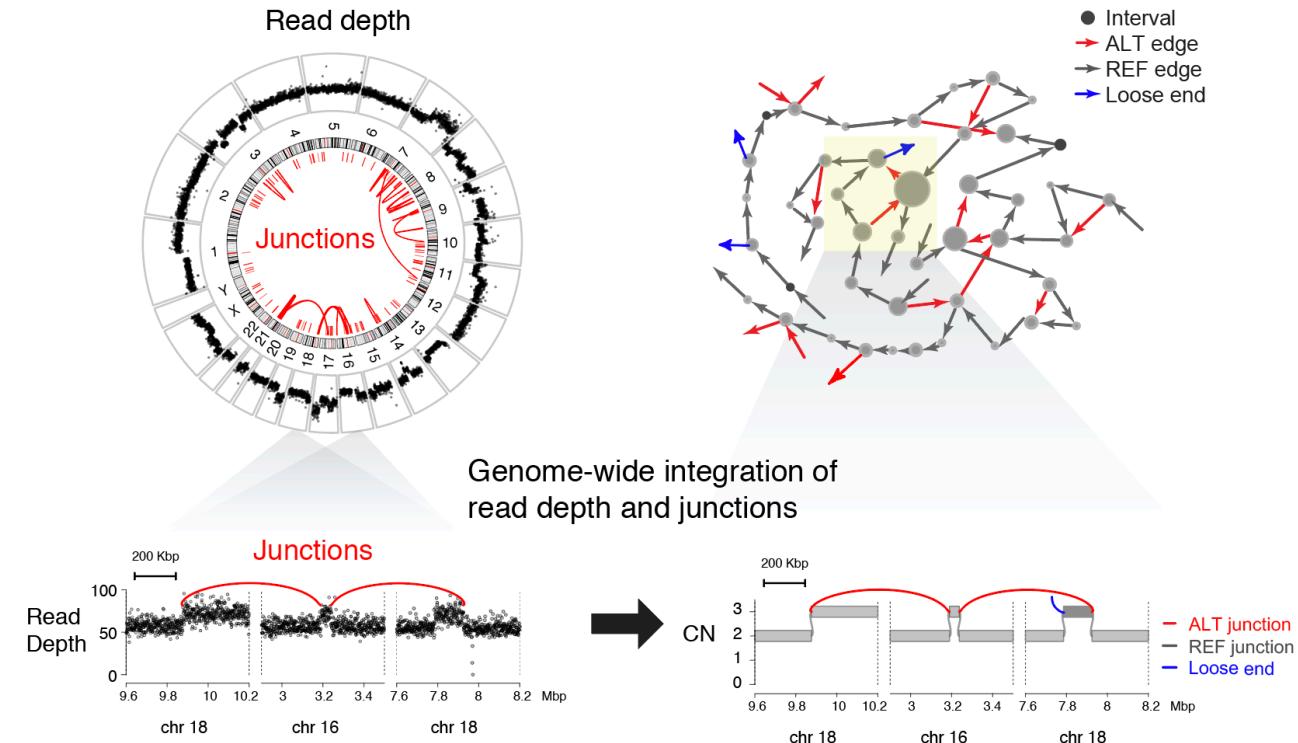


Julie  
Behr



**Input:**  
Whole Genome  
Sequence

**Output:**  
Junction-Balanced  
Genome Graph



<https://github.com/mskilab/JaBbA>

<https://github.com/mskilab/gGnome>

Hadi, Yao, Behr, et al Cell 2020

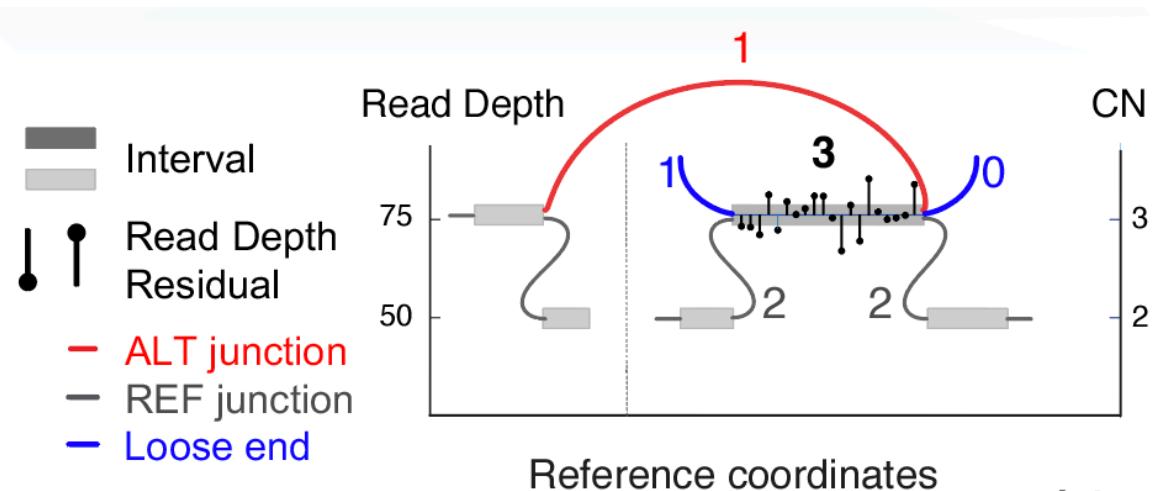
# JaBbA algorithm

Optimization problem:

$$\operatorname{argmin}_{\text{interval CN, junction CN}}$$

$$\sum_{\text{intervals}} \text{Read Depth Residual}^2 + \lambda \# \text{ Incorporated Loose Ends}$$

Subject to:  
Junction Balance Constraints



Junction Balance constraint:  $2 + 0 + 1 = 3 = 2 + 1 + 0$

**Interval CN**

REF Junction CN

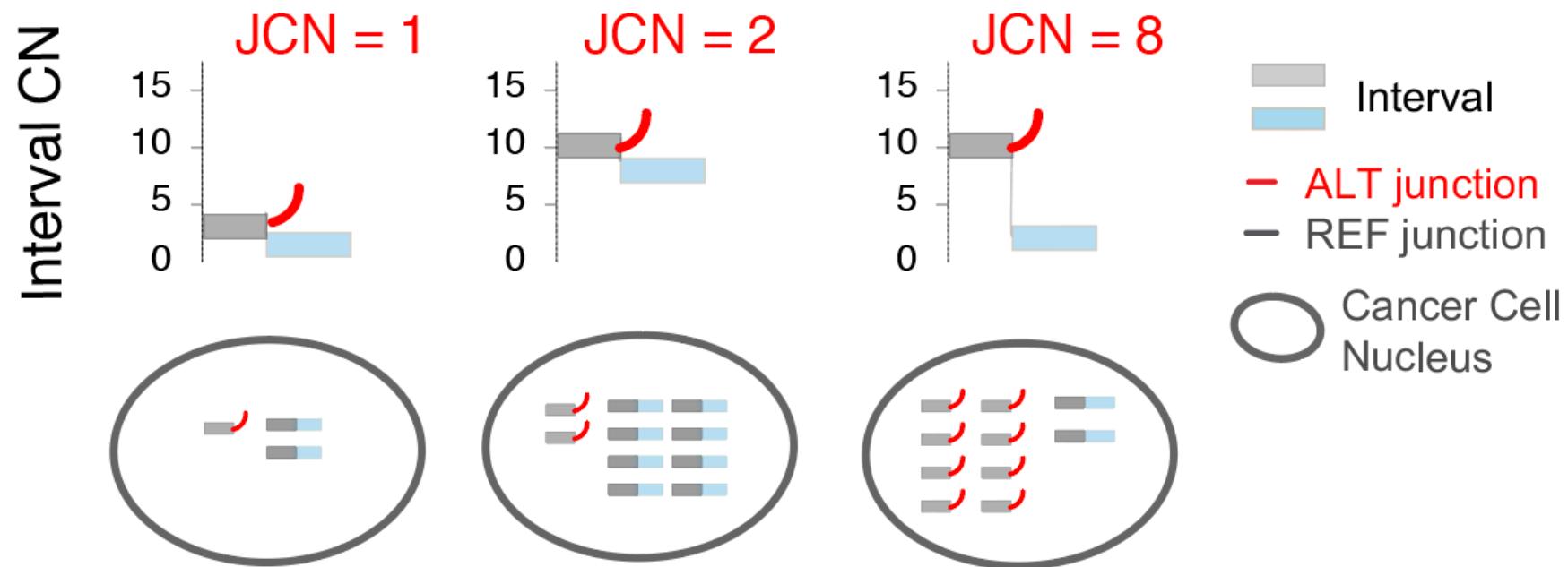
ALT Junction CN

Loose end CN

<https://github.com/mskilab/JaBbA>

<https://github.com/mskilab/gGnome>

## Junction Copy Number

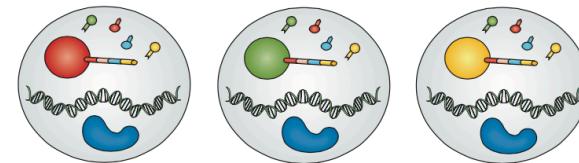


# Linked-read whole genome sequencing (10x genomics)

100 Kbp “synthetic long reads” \$7/Gb

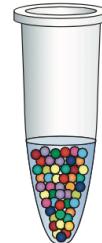
## Emulsion PCR

Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs



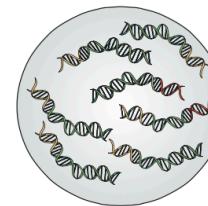
## GEMs

Each micelle has 1 barcode out of 750,000



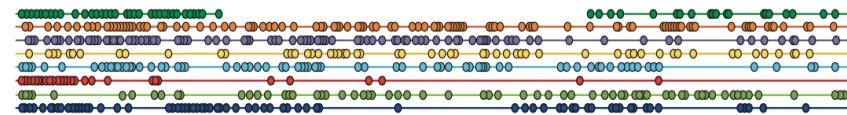
## Amplification

Long fragments are amplified such that the product is a barcoded fragment ~350 bp



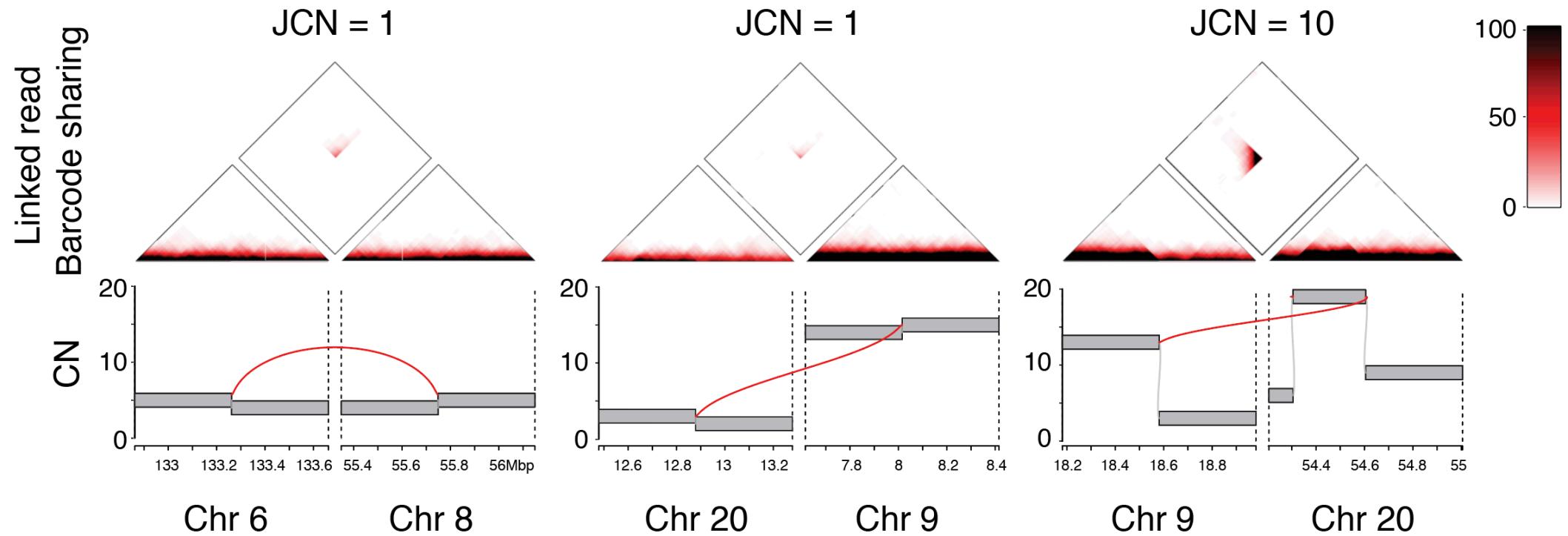
## Pooling

The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation

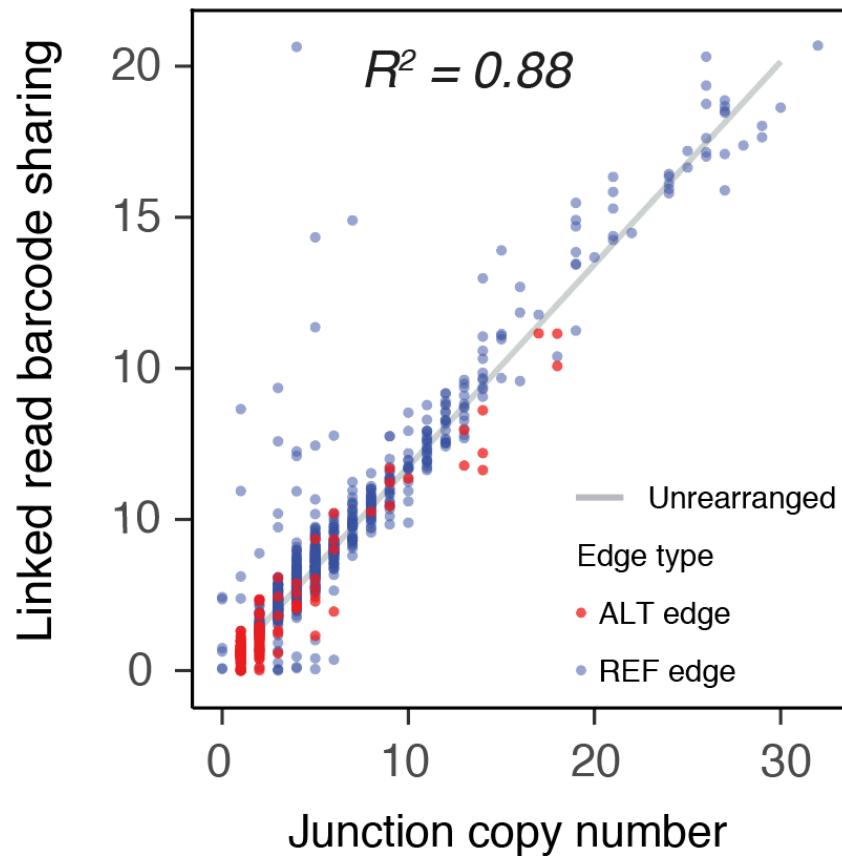


Goodwin, Mcpherson, McCombie *Nature Reviews Genetics* 2016

# WGS (JaBbA) vs linked-reads



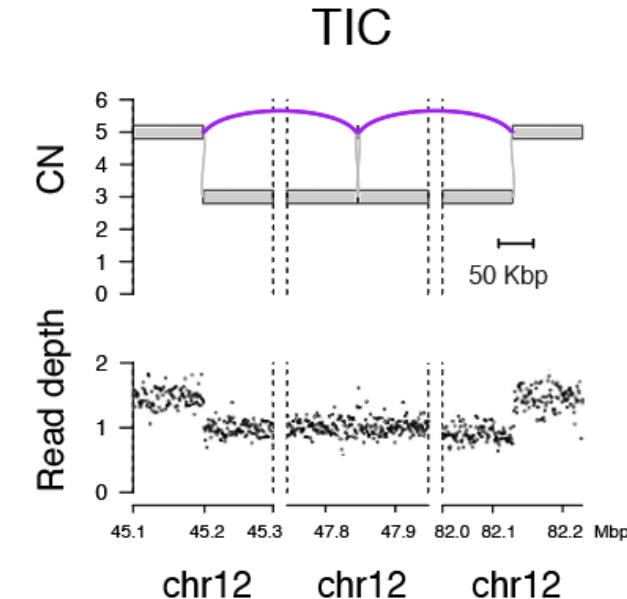
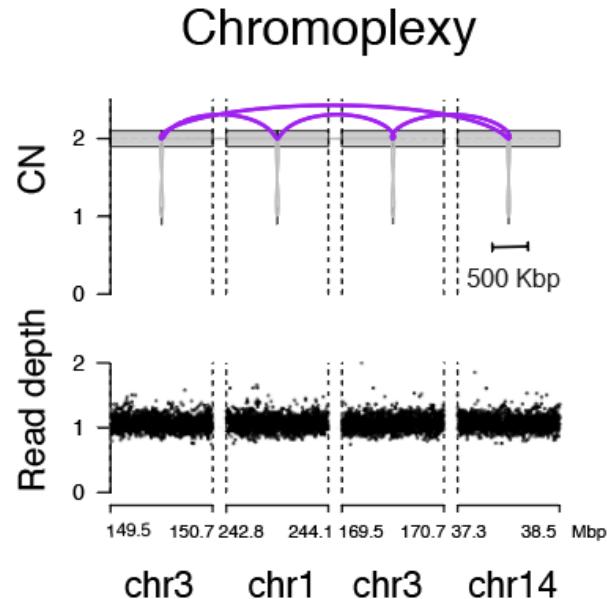
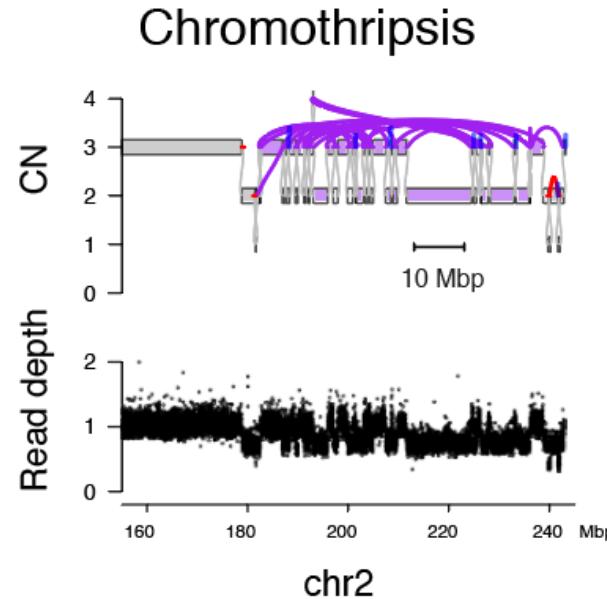
# WGS (JaBbA) vs linked-reads



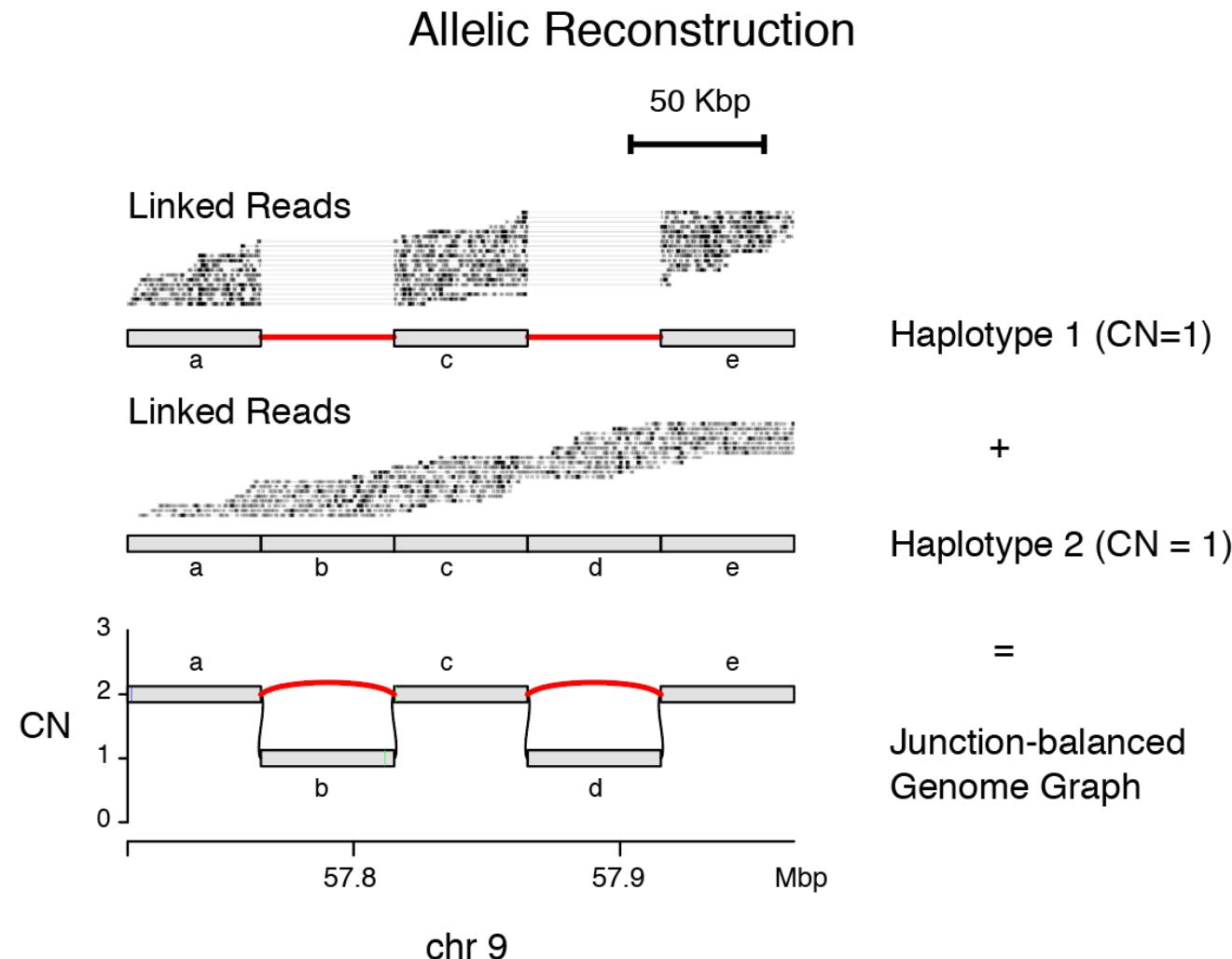
# JaBbA applications

## Complex Structural Variants

Variant subgraph

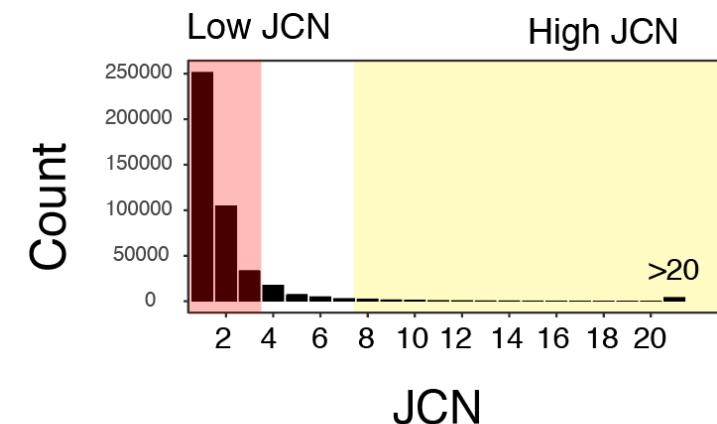
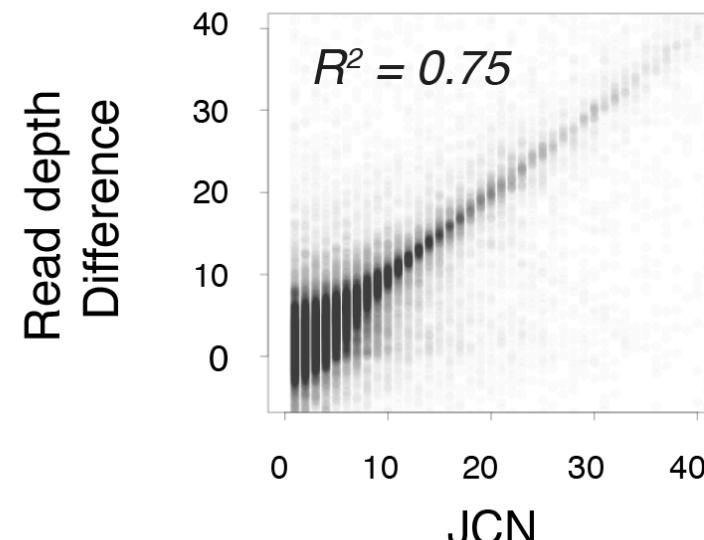
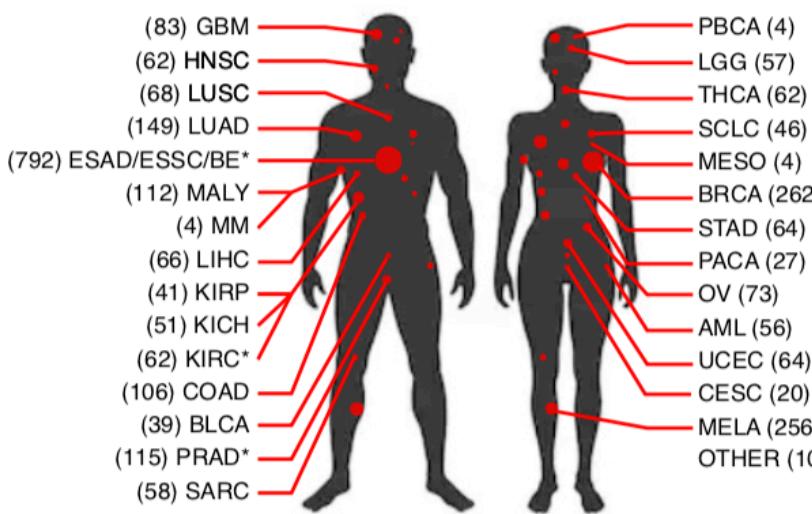


# JaBbA applications

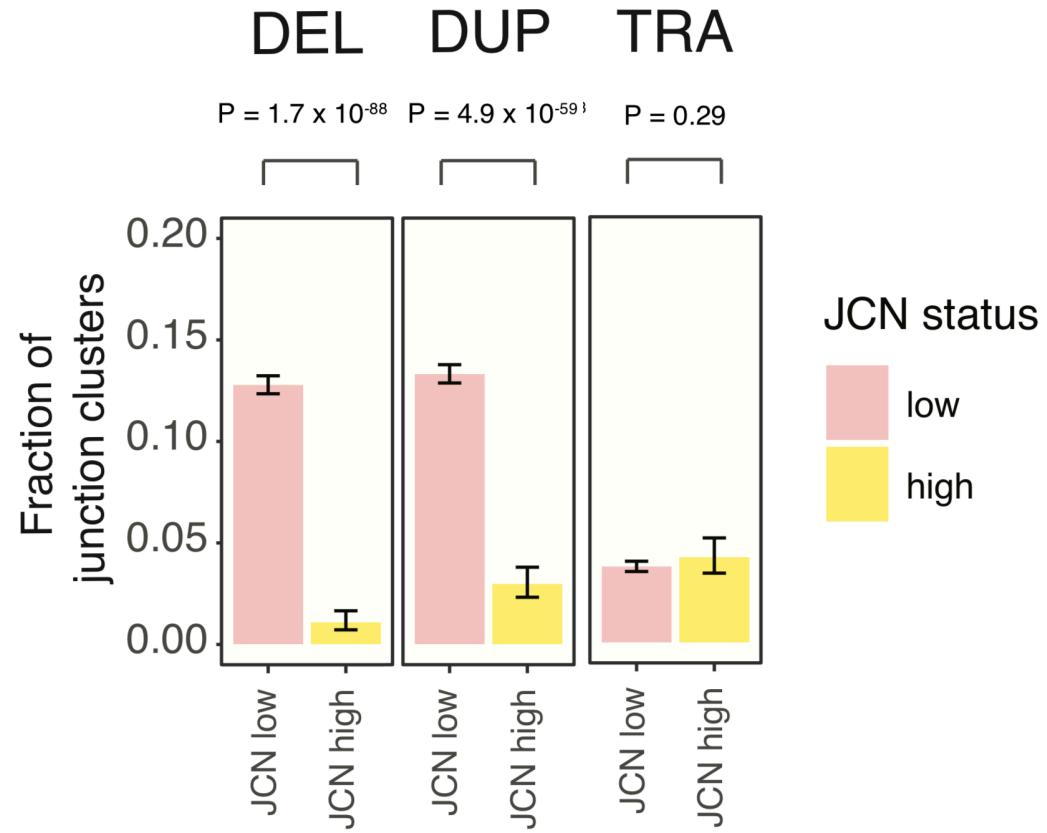


# Landscape of thousands of cancer genome graphs

2,813 WGS samples  
31 tumor types



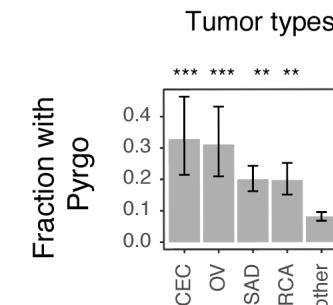
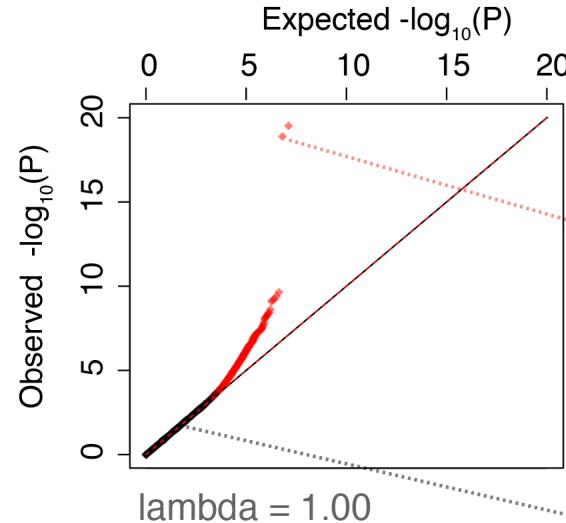
# Clusters of low JCN DUP and DEL



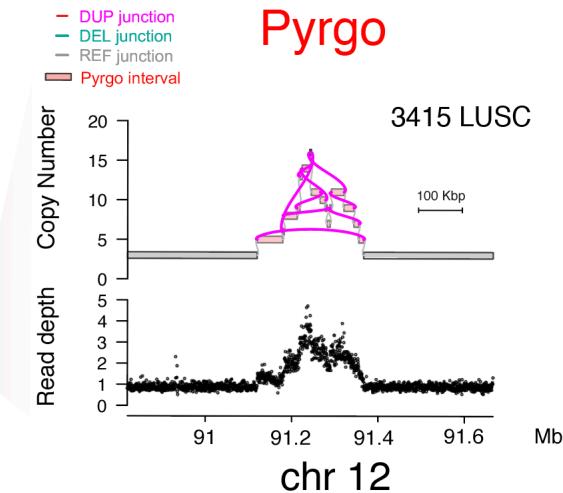
# Duplication “towers”



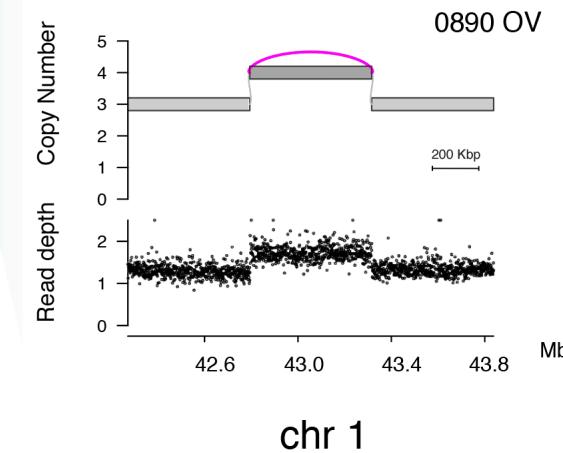
Low JCN  
Duplication Density



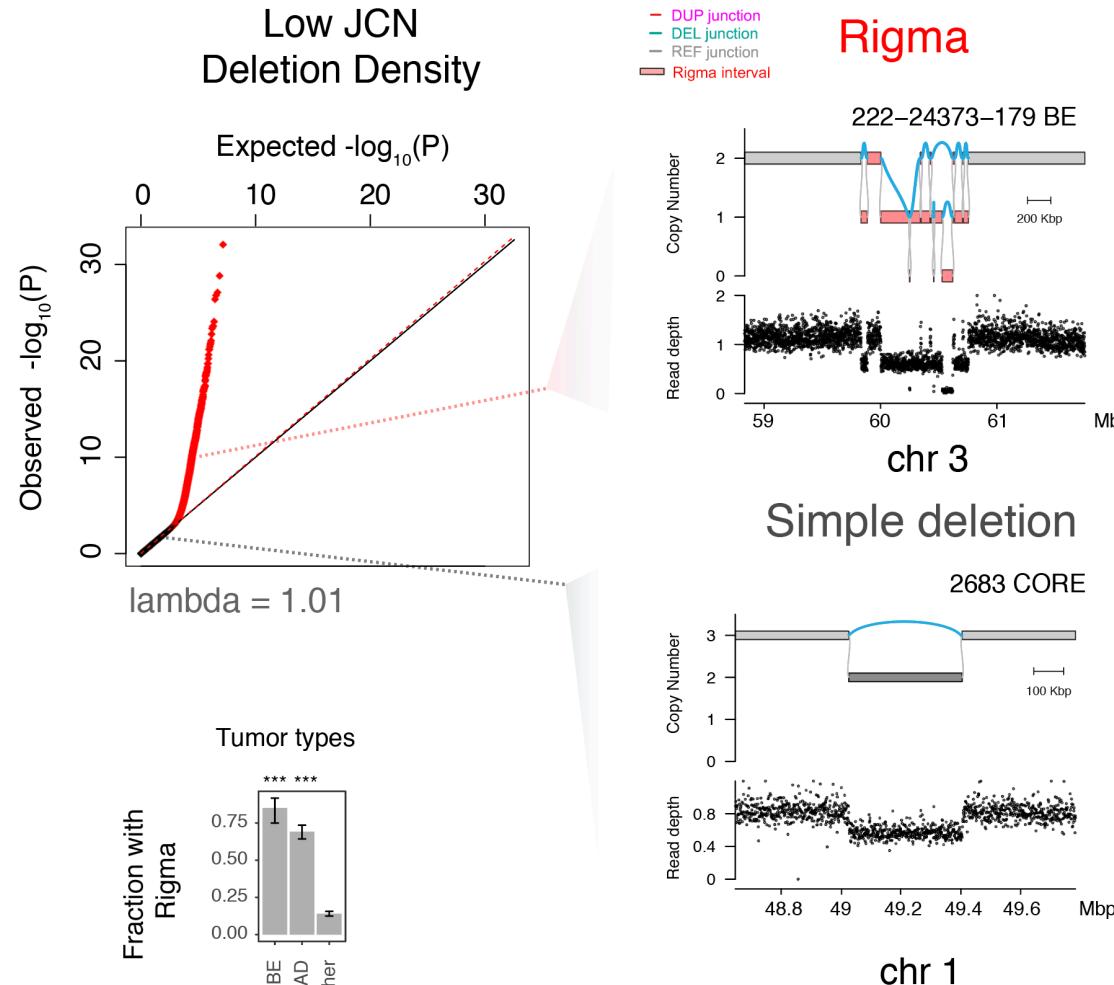
Pyrgo



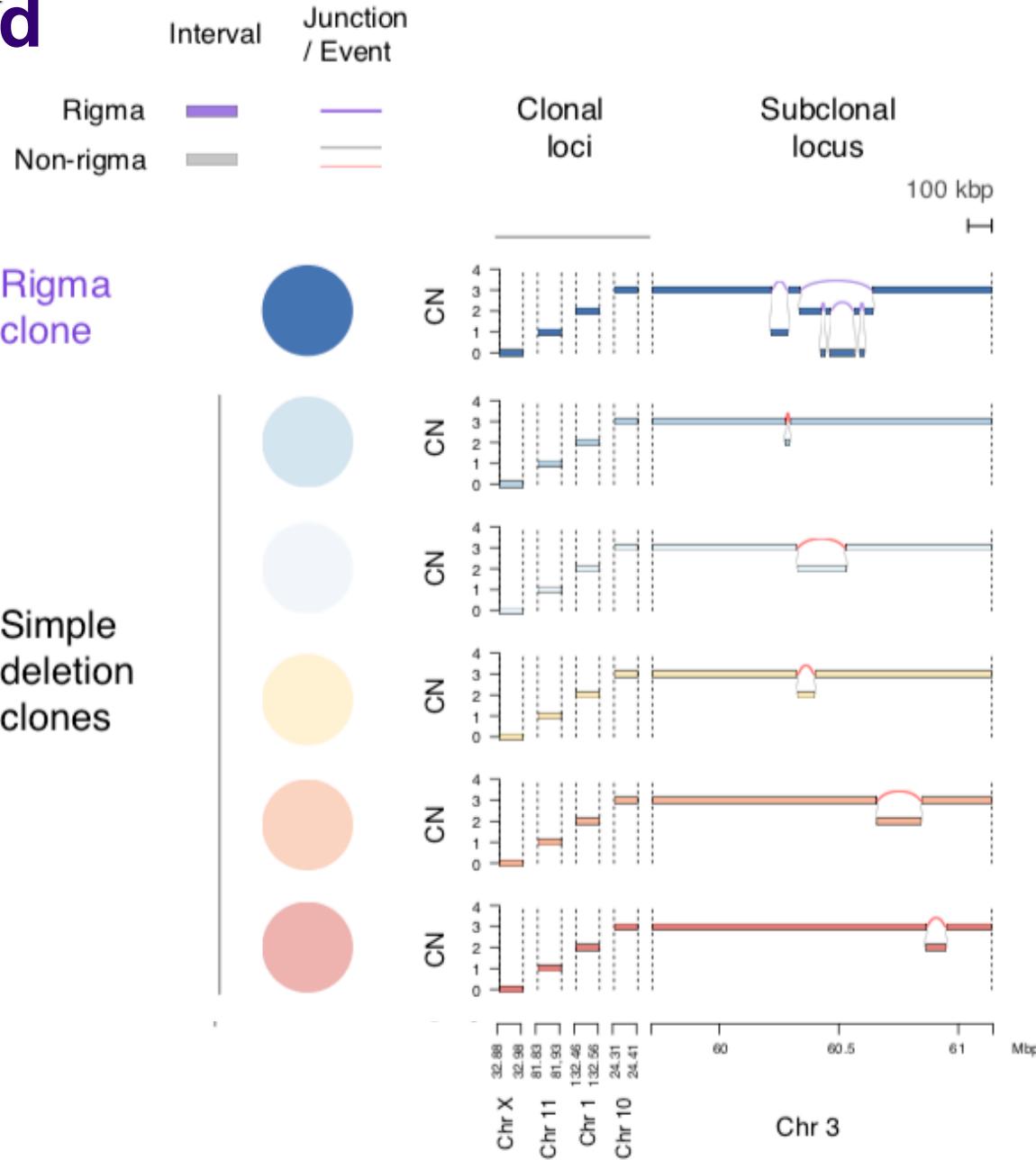
Simple duplication



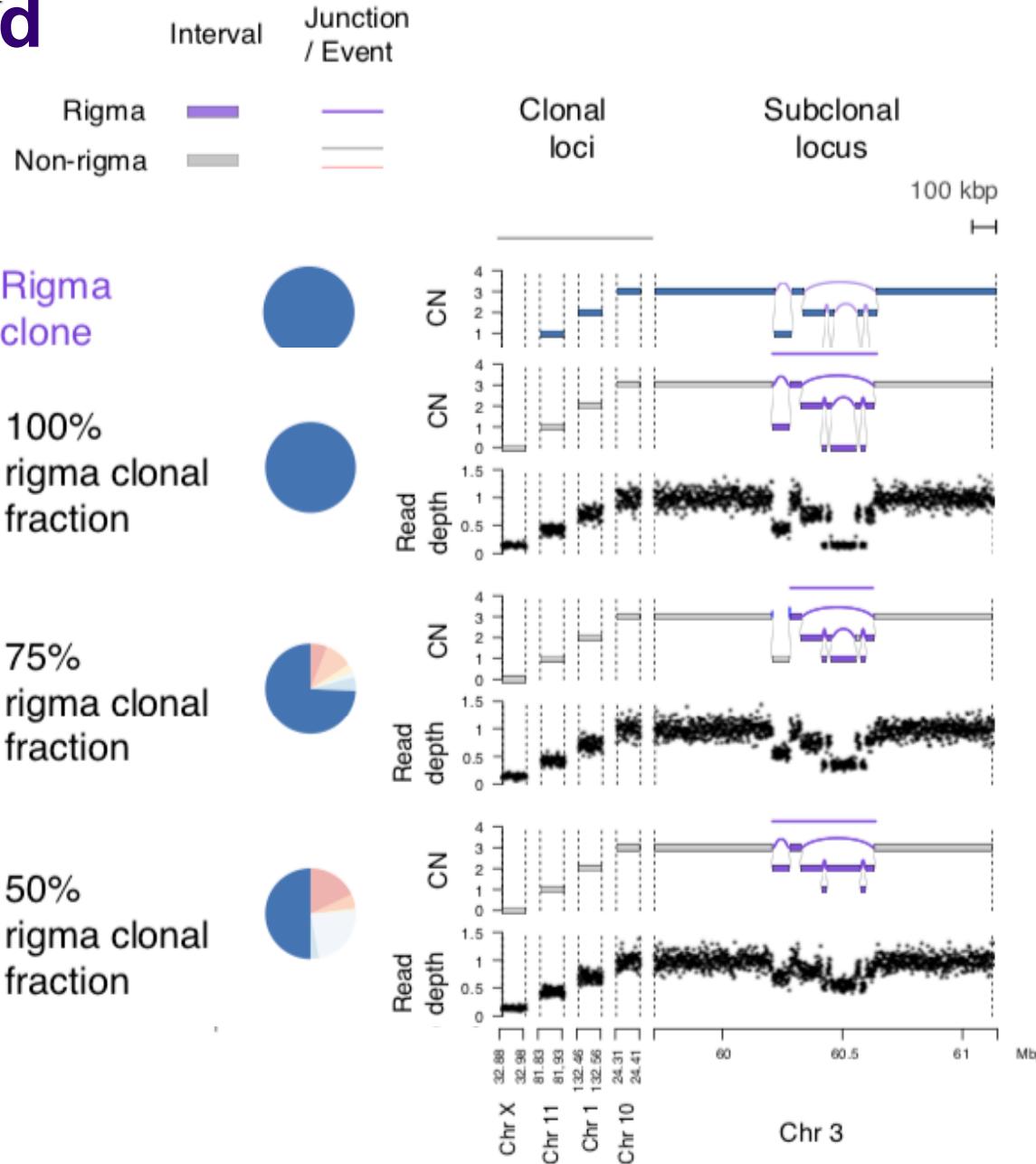
# Deletion “ditches”



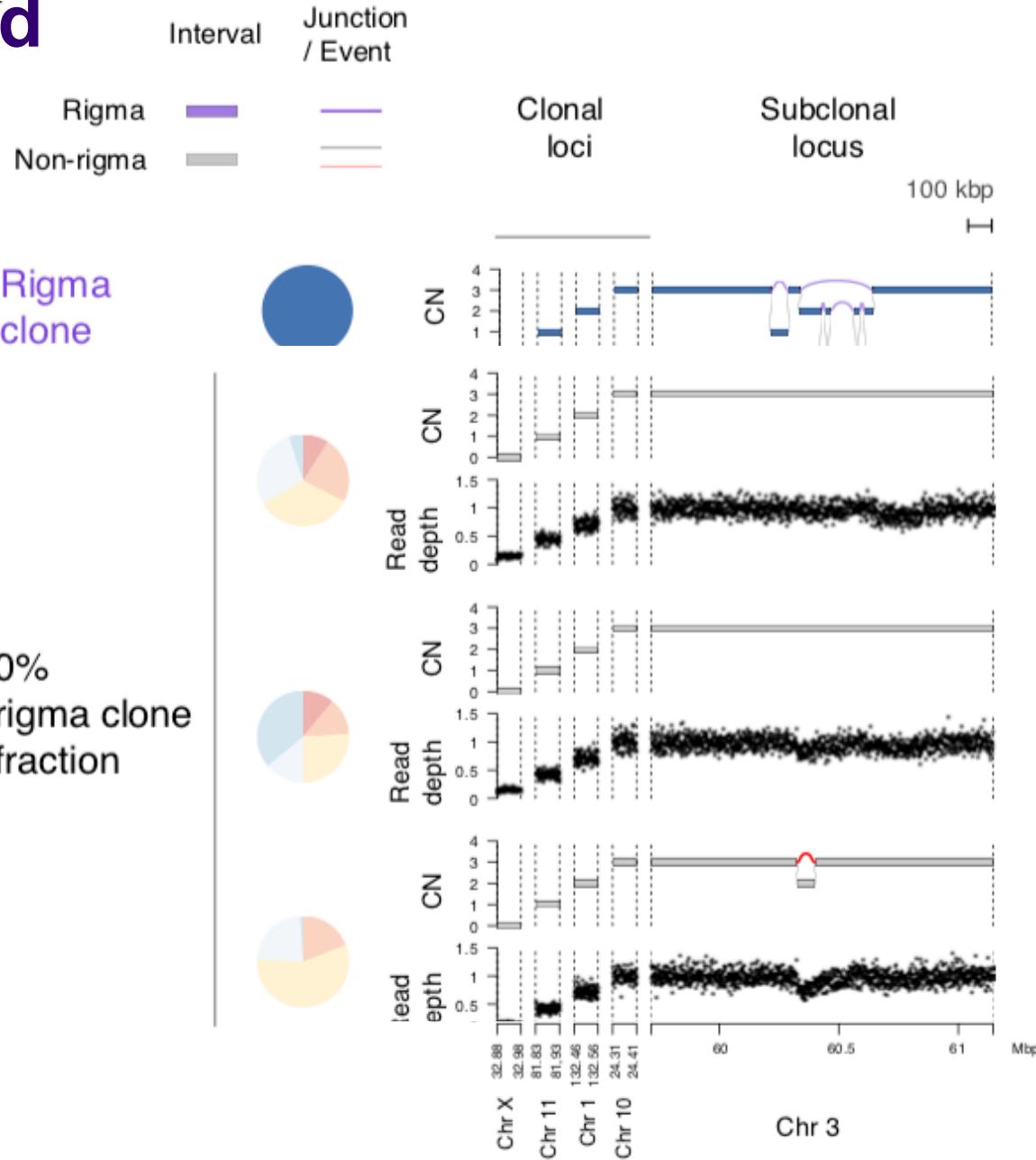
# Clonality and graphs



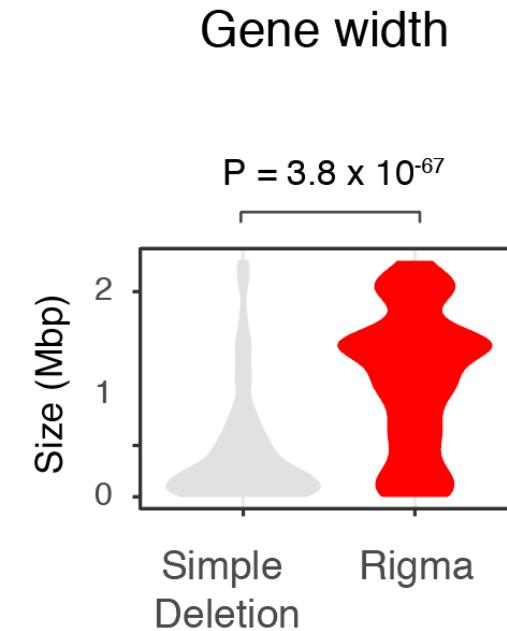
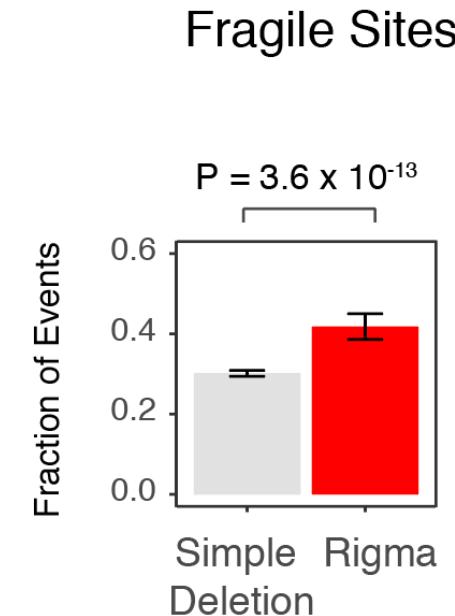
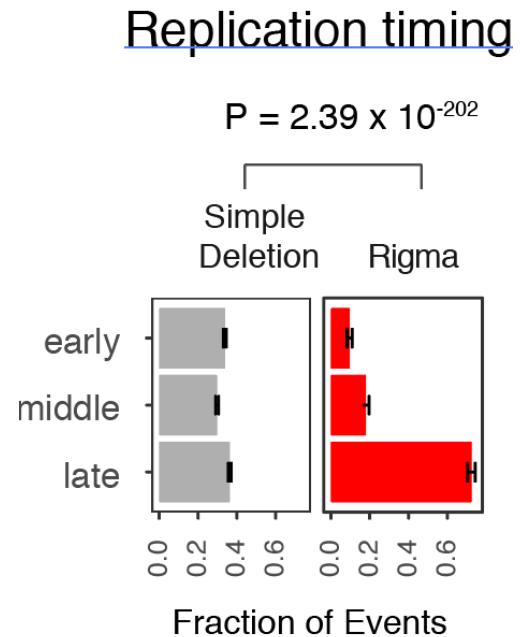
# Clonality and graphs



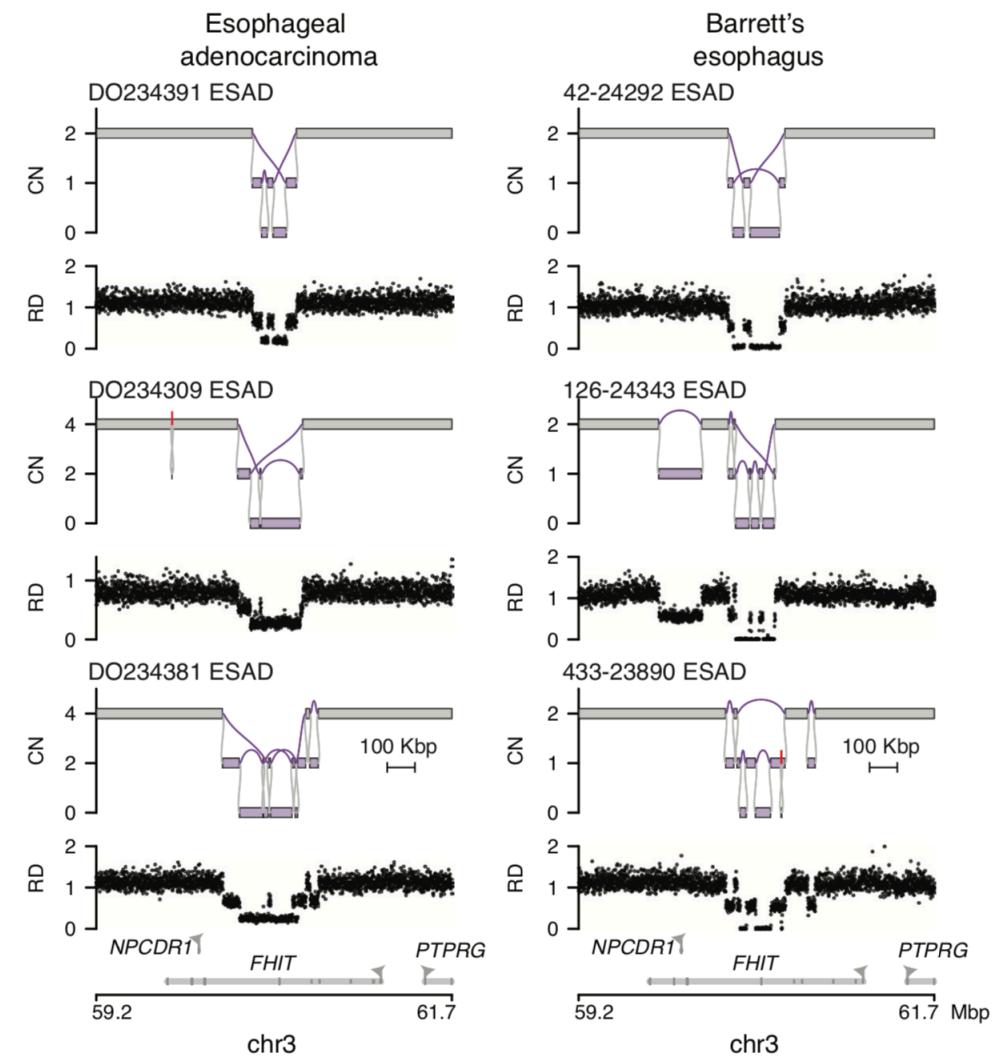
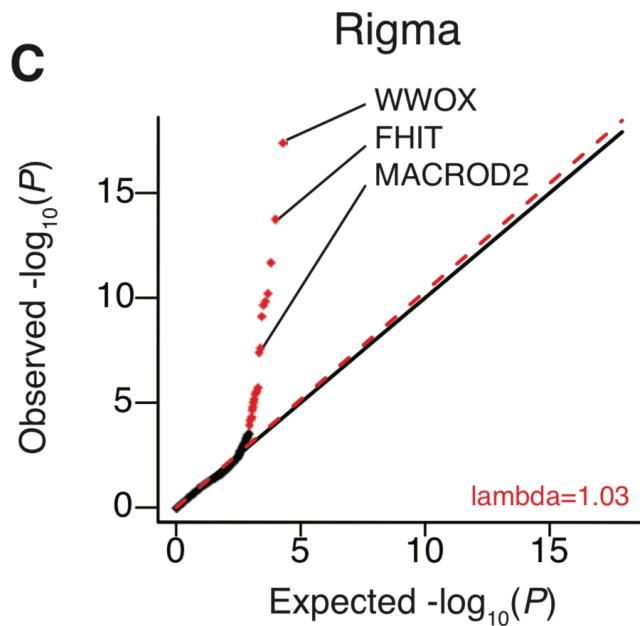
# Clonality and graphs



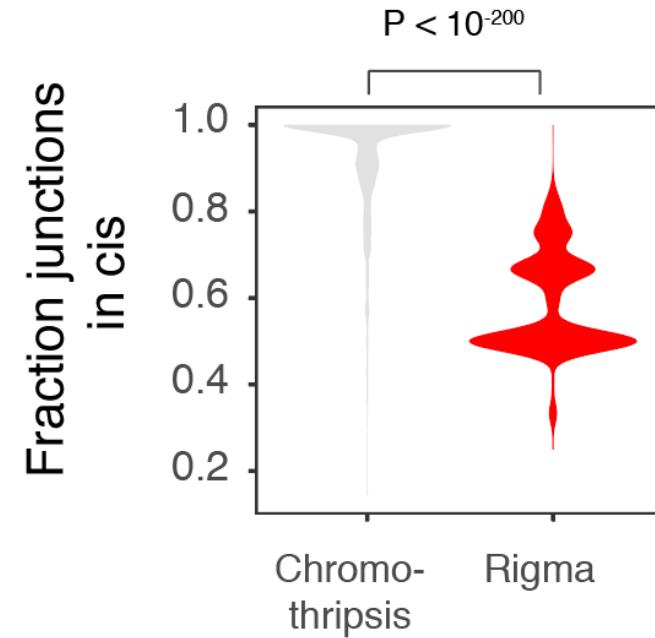
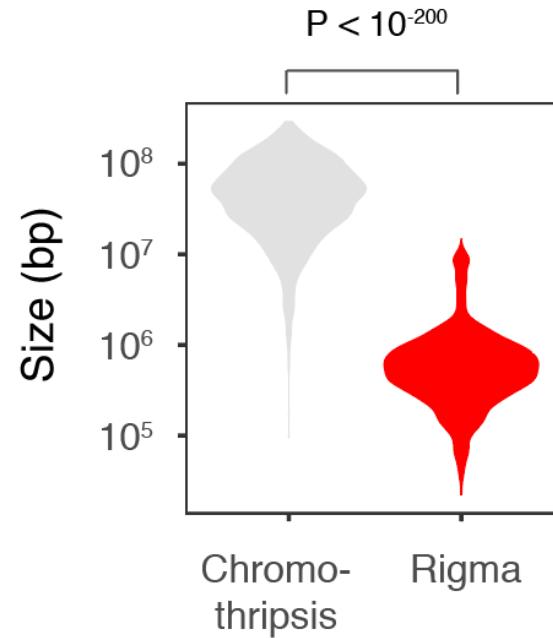
# Rigma = late replicating, fragile sites, long genes



# Recurrent rigma

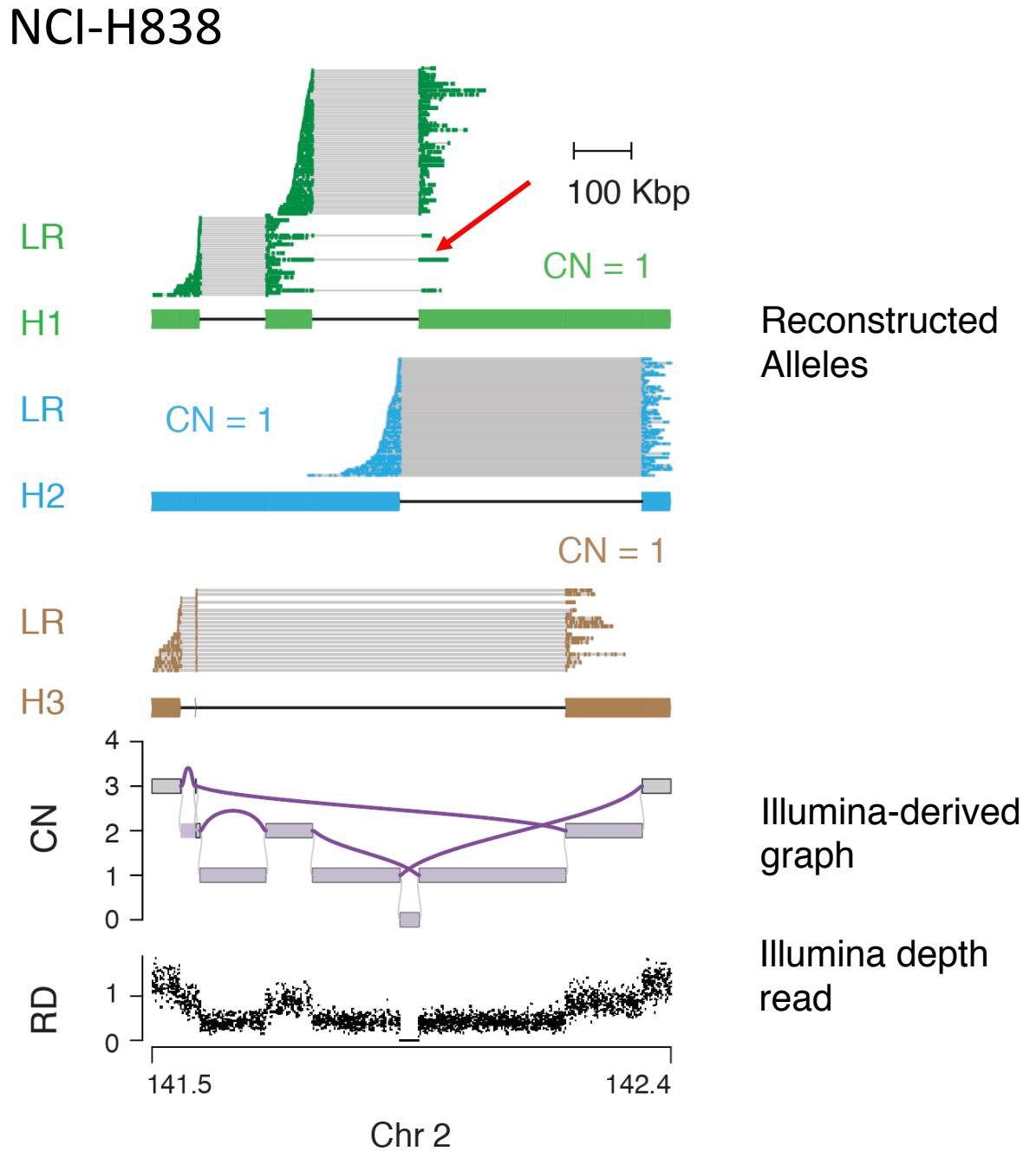


# Rigma != Chromothripsis

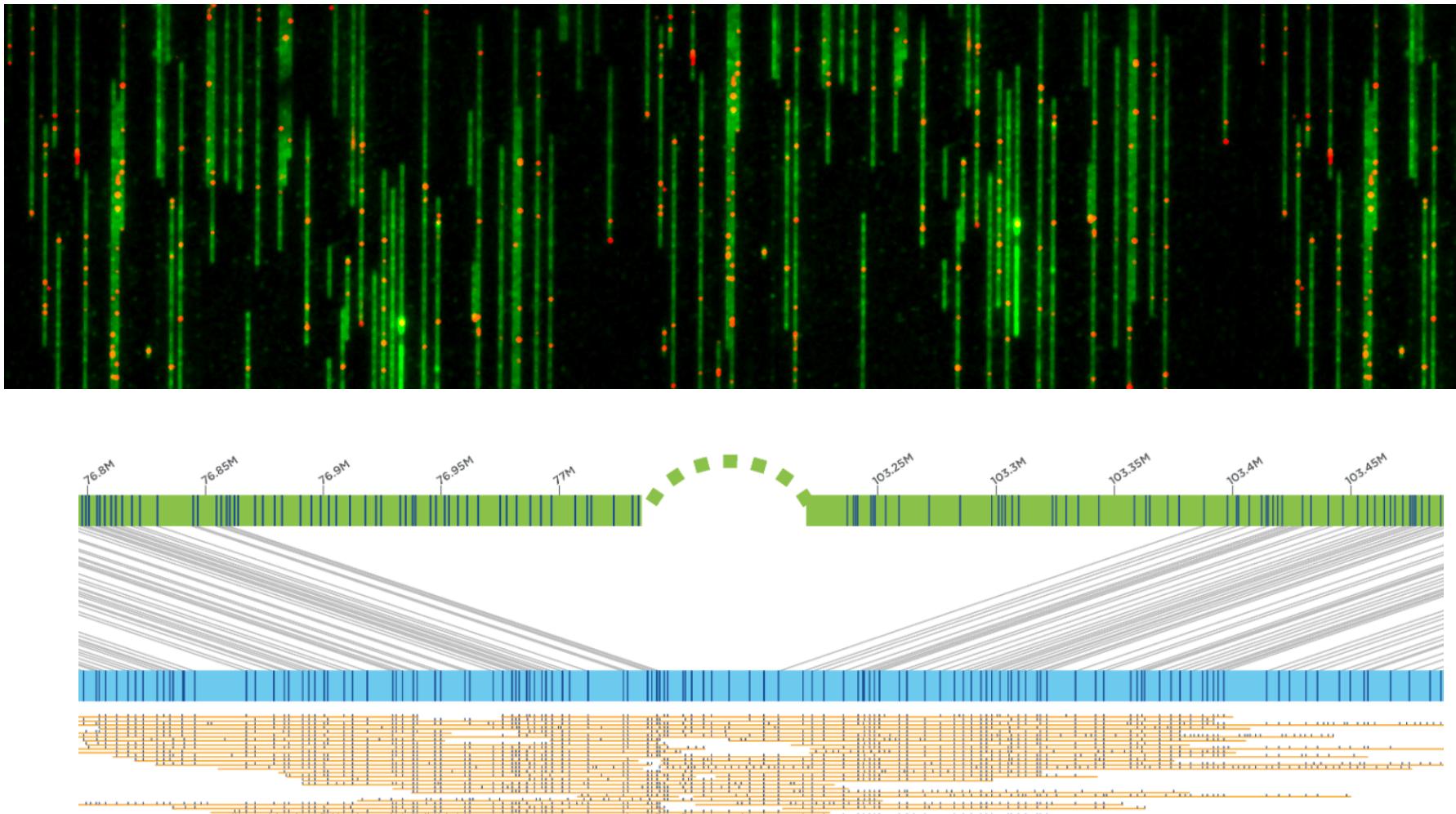


# Rigma junctions occur in *trans*

Rigma



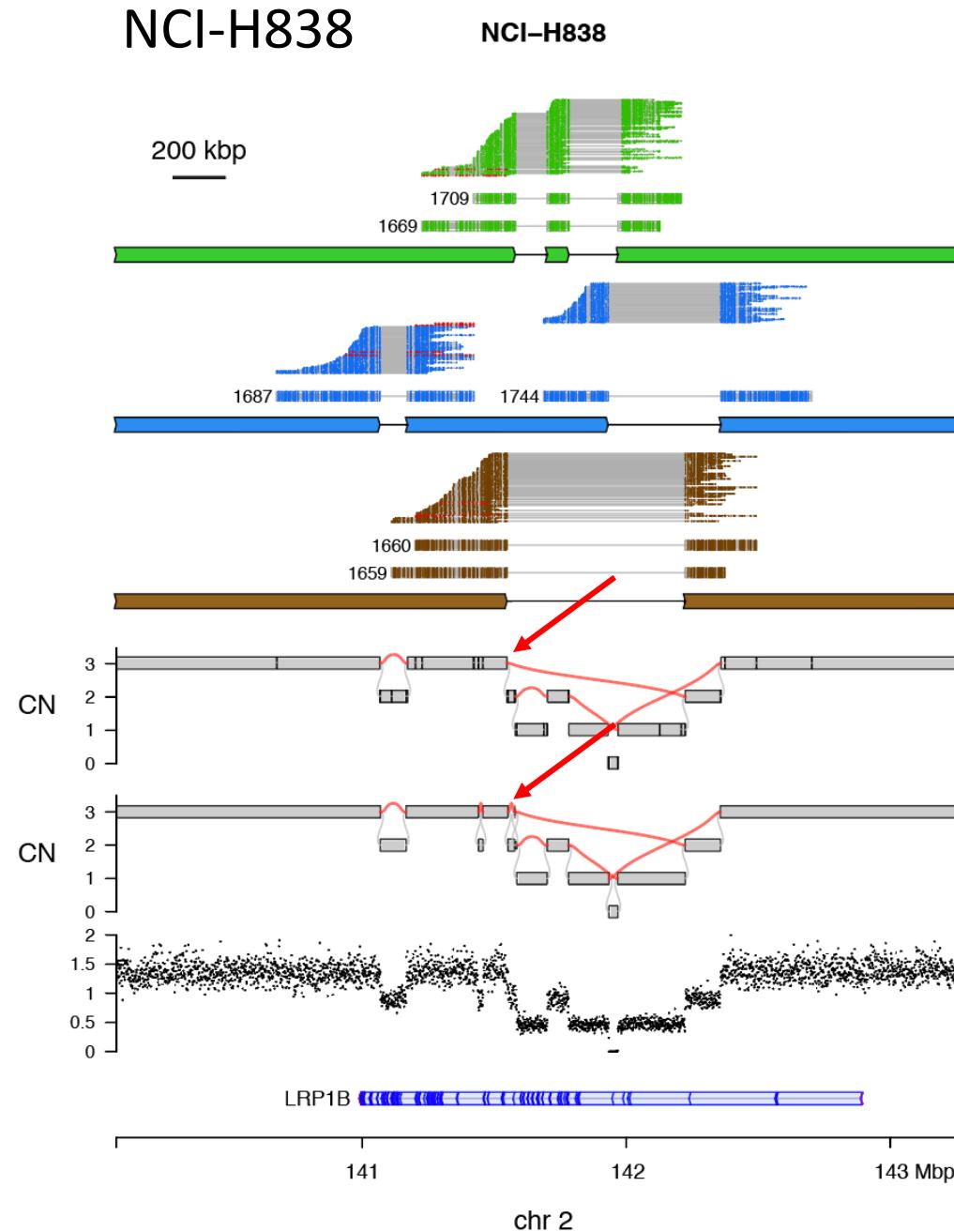
# Optical mapping (BioNano Genomics)



# Rigma junctions occur in *trans*



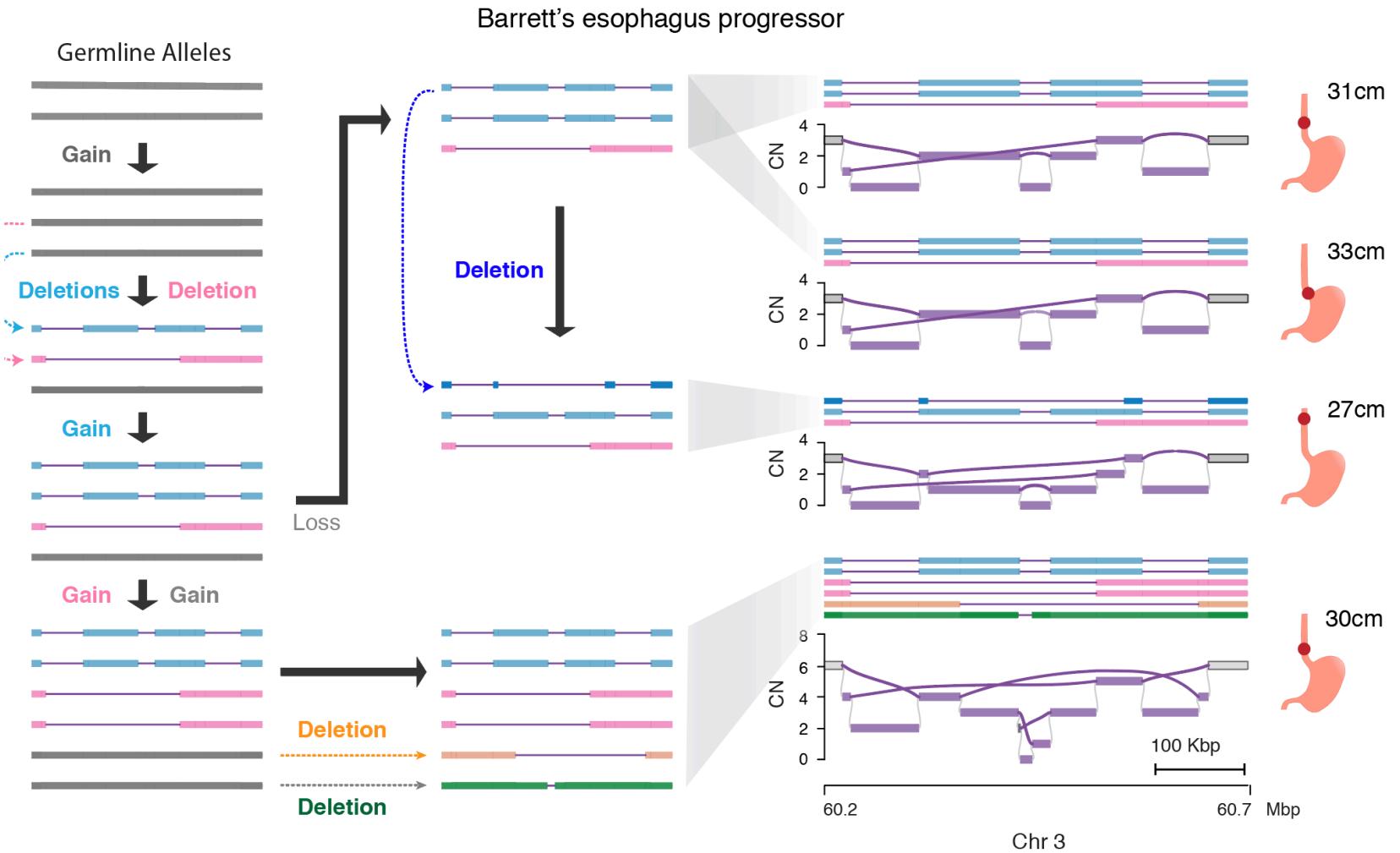
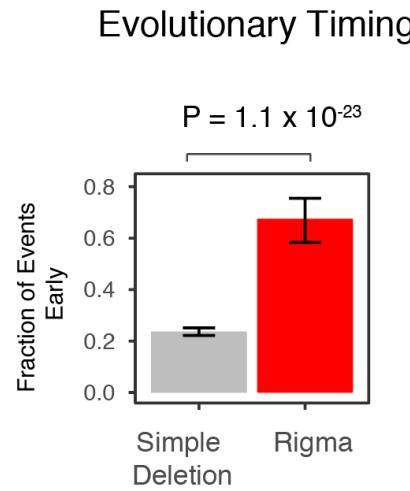
Julie Behr



Rigma

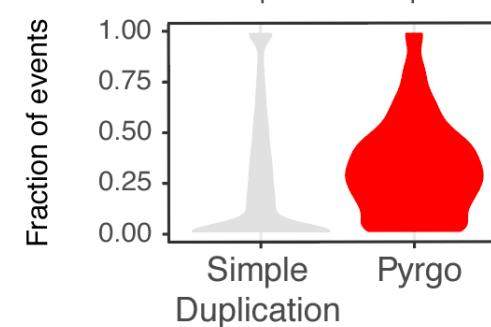


# Rigma events occur early in esophageal adenocarcinoma tumorigenesis



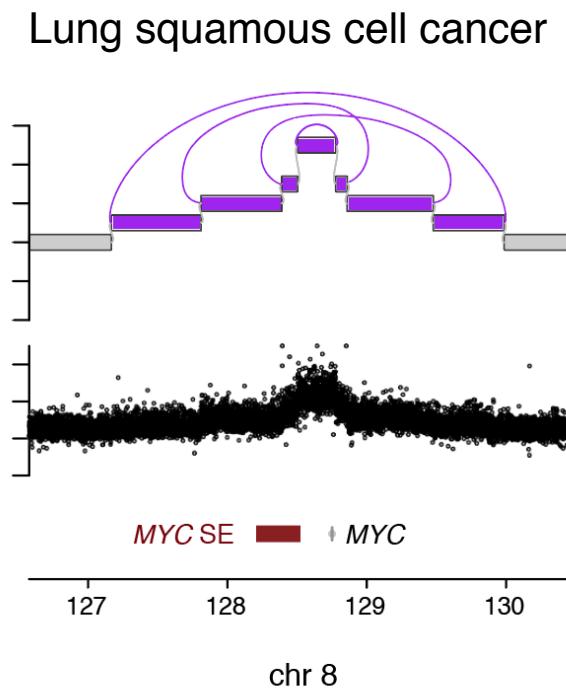
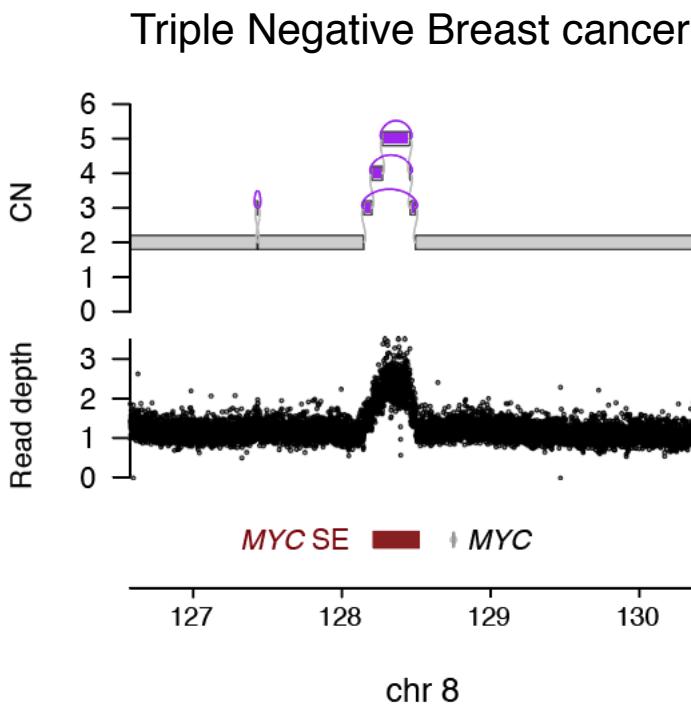
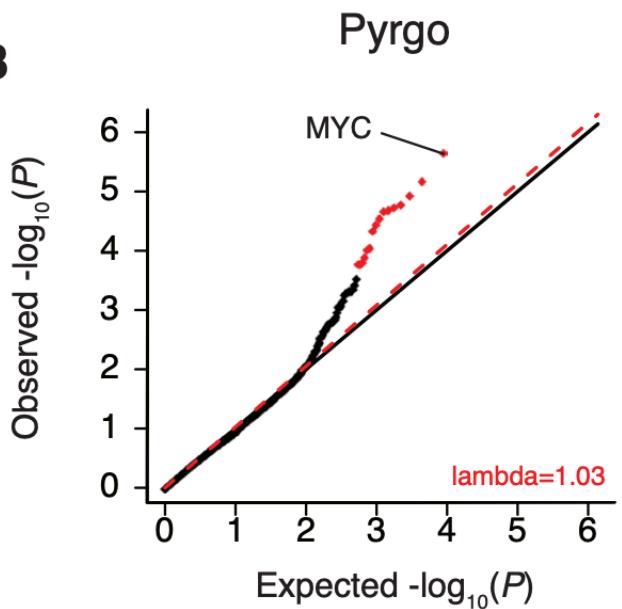
# Pyrgo and Superenhancers

$P = 7.39 \times 10^{-20}$



Super-enhancer  
Overlap

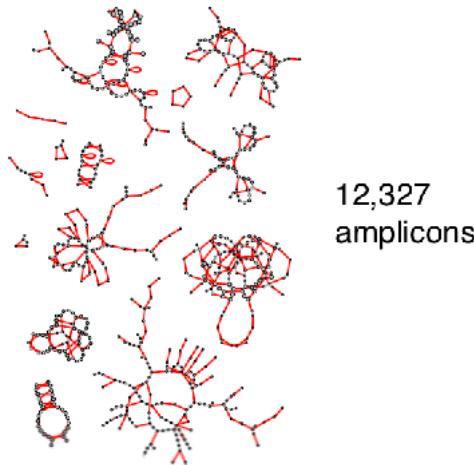
B



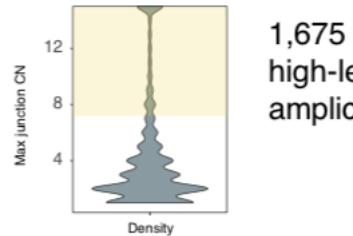
2,779  
junction-balanced  
genome graphs

# High JCN Amplicons

Amplified subgraphs  
(interval CN > 2\*ploidy)

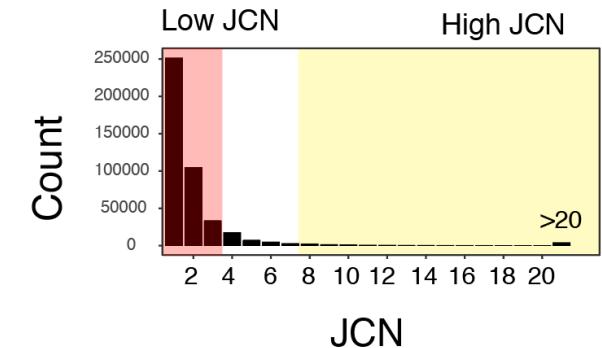
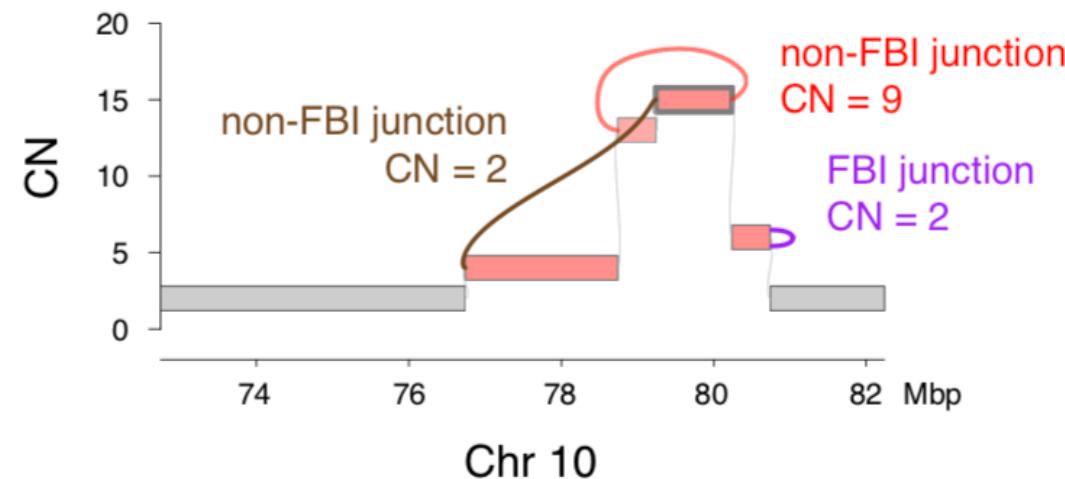


Subgraphs with  
max junction CN > 7



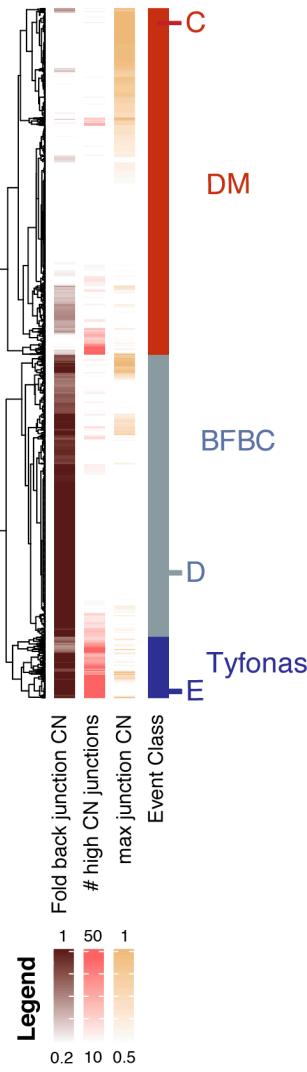
Max CN = 15  
Max Junction CN = 9/15  
FBI CN = 2/15  
# high CN junctions = 1

FBI junction  
High CN, non FBI junction  
Low CN, non FBI junction  
Amplicon interval



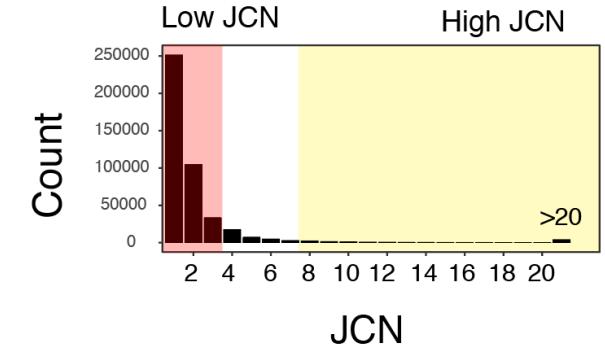
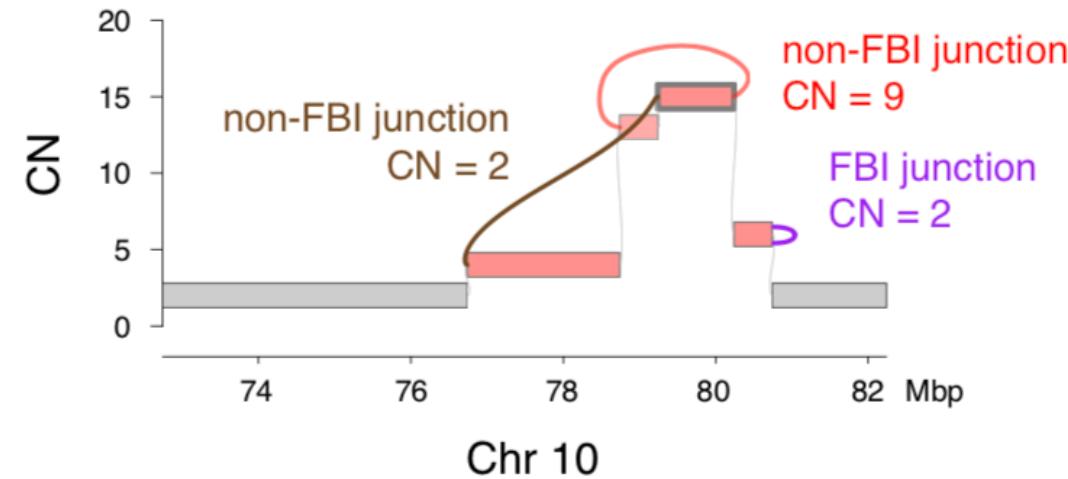
# High JCN Amplicons

Amplicons

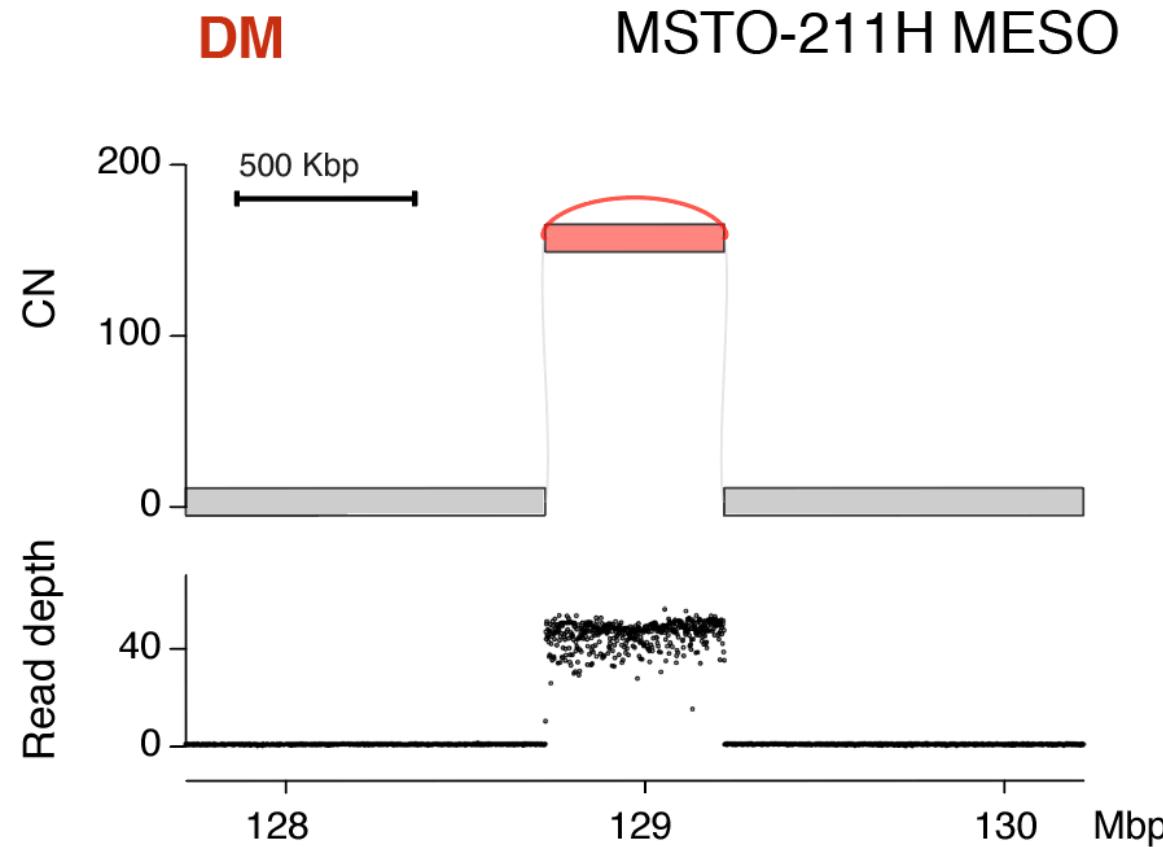
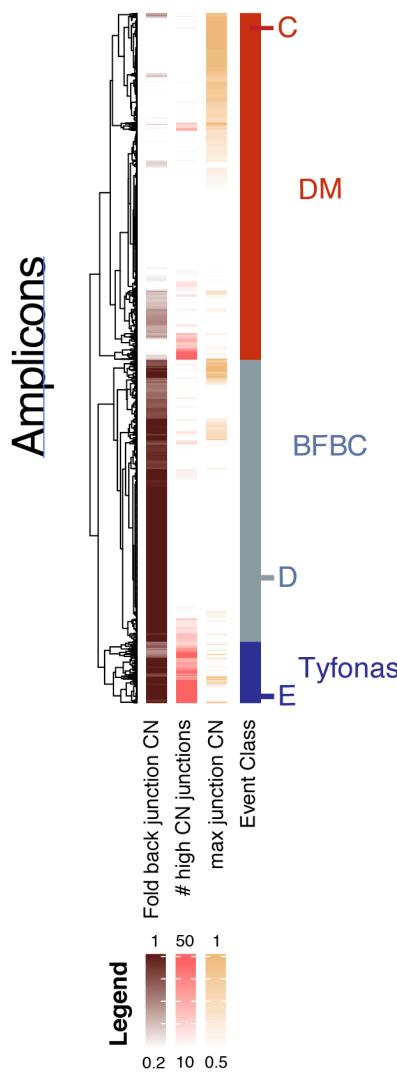


Max CN = 15  
 Max Junction CN = 9/15  
 FBI CN = 2/15  
 # high CN junctions = 1

- FBI junction
- High CN, non FBI junction
- Low CN, non FBI junction
- Amplicon interval

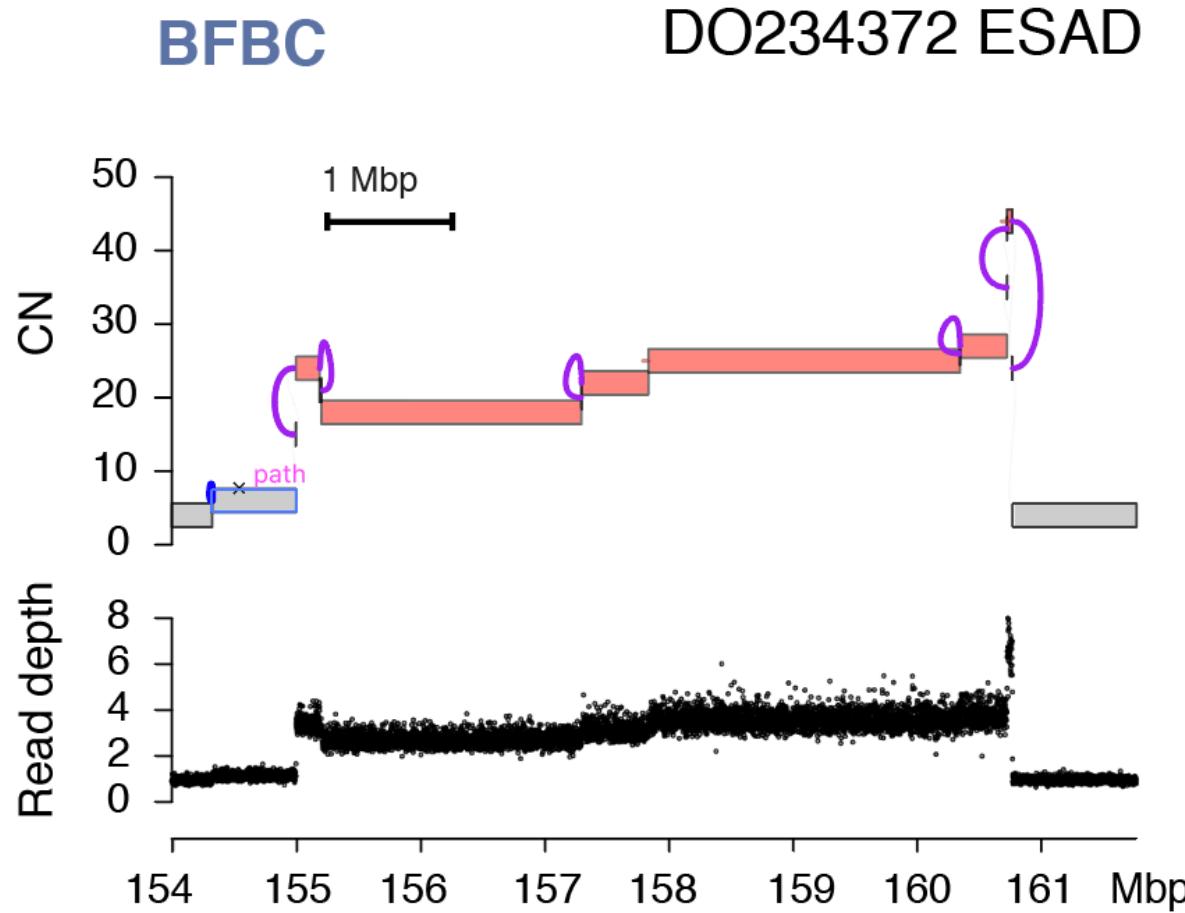
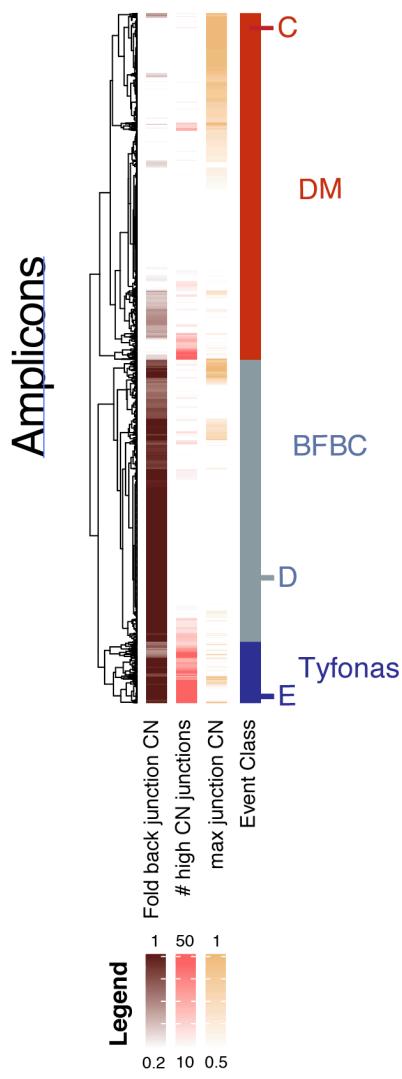


# High JCN Amplicons



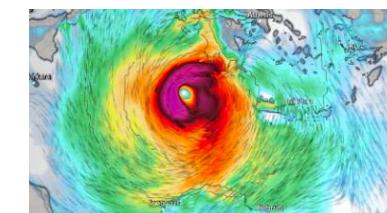
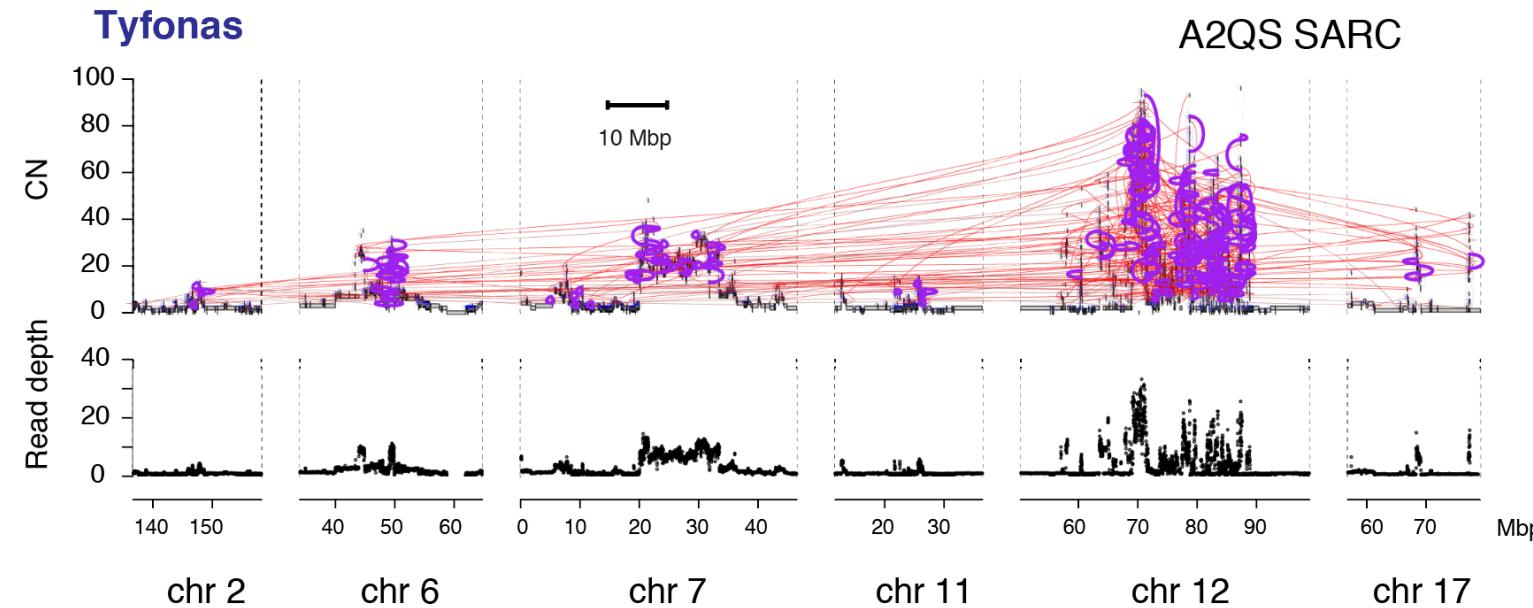
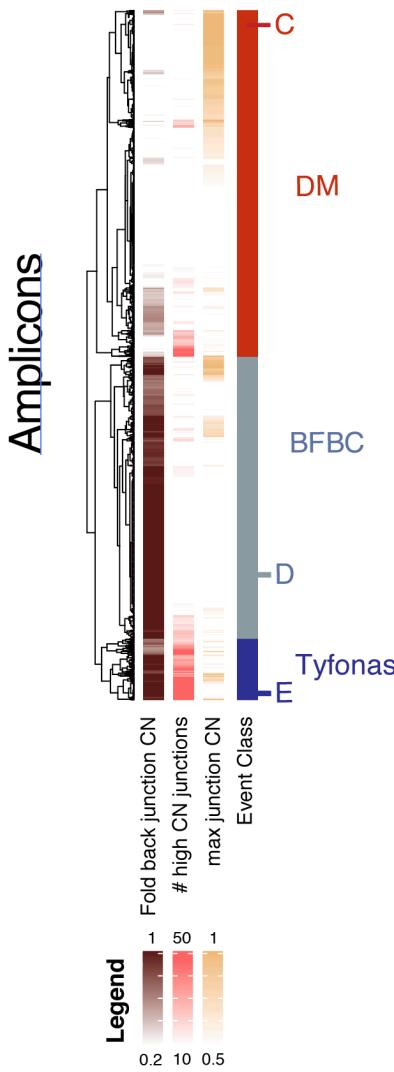
- FBI junction
- High CN, non FBI junction
- Low CN, non FBI junction
- Loose end
- Amplicon interval

# High JCN Amplicons



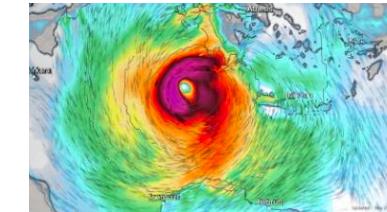
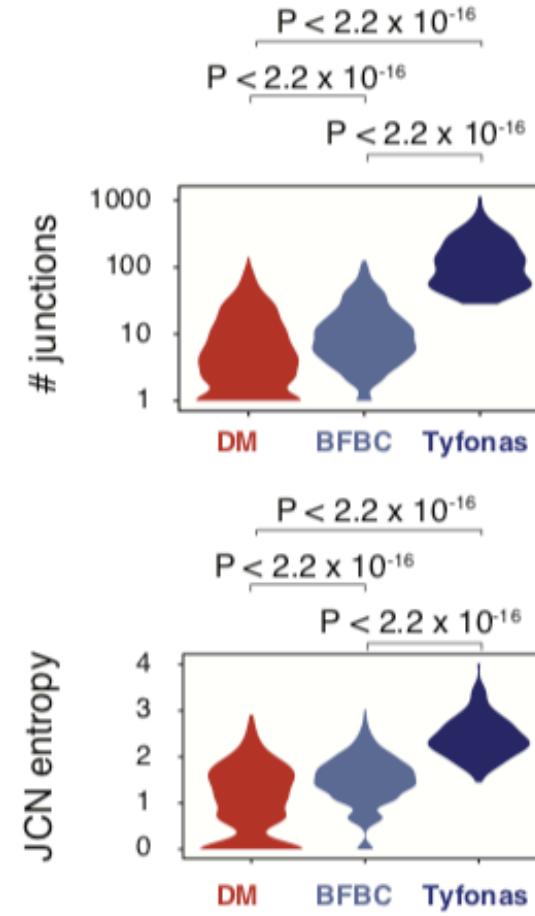
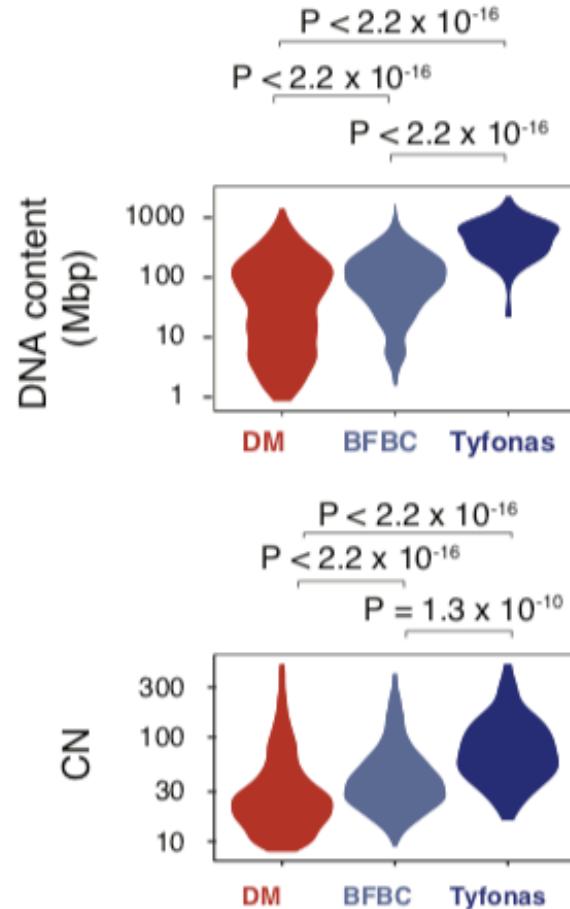
- FBI junction
- High CN, non FBI junction
- Low CN, non FBI junction
- Loose end
- Amplicon interval

# Genomic “typhoons”



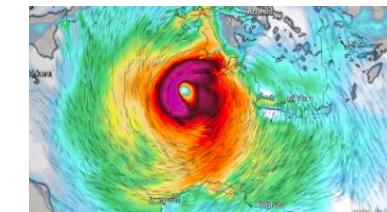
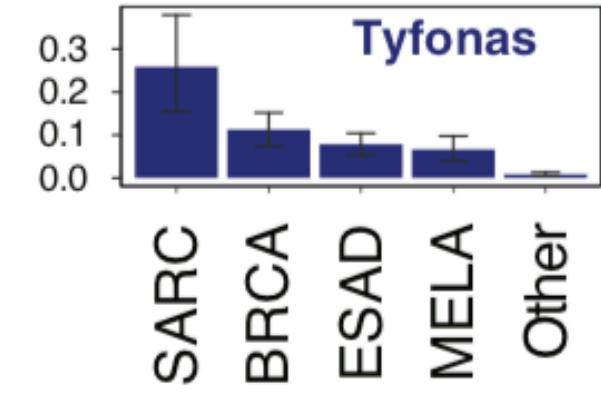
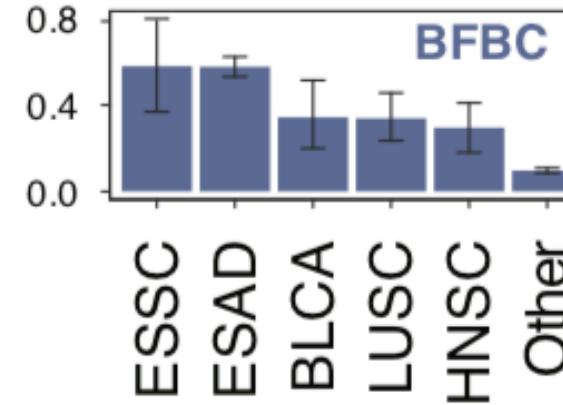
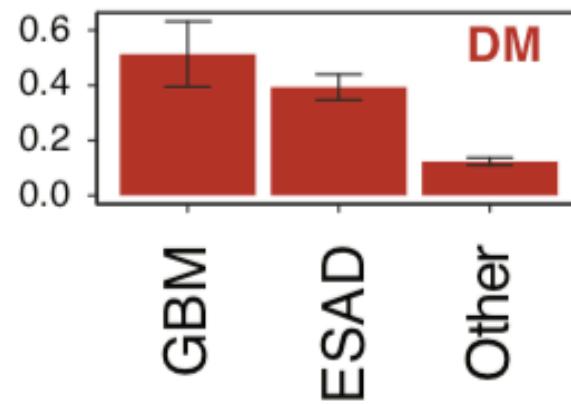
- FBI junction
- High CN, non FBI junction
- Low CN, non FBI junction
- Loose end
- Amplicon interval

# Tyfonas are larger and have a more diverse distribution of JCN than DM, BFBC

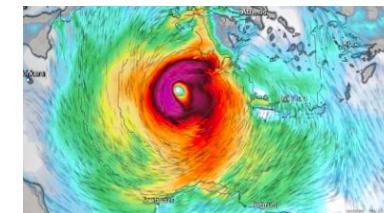
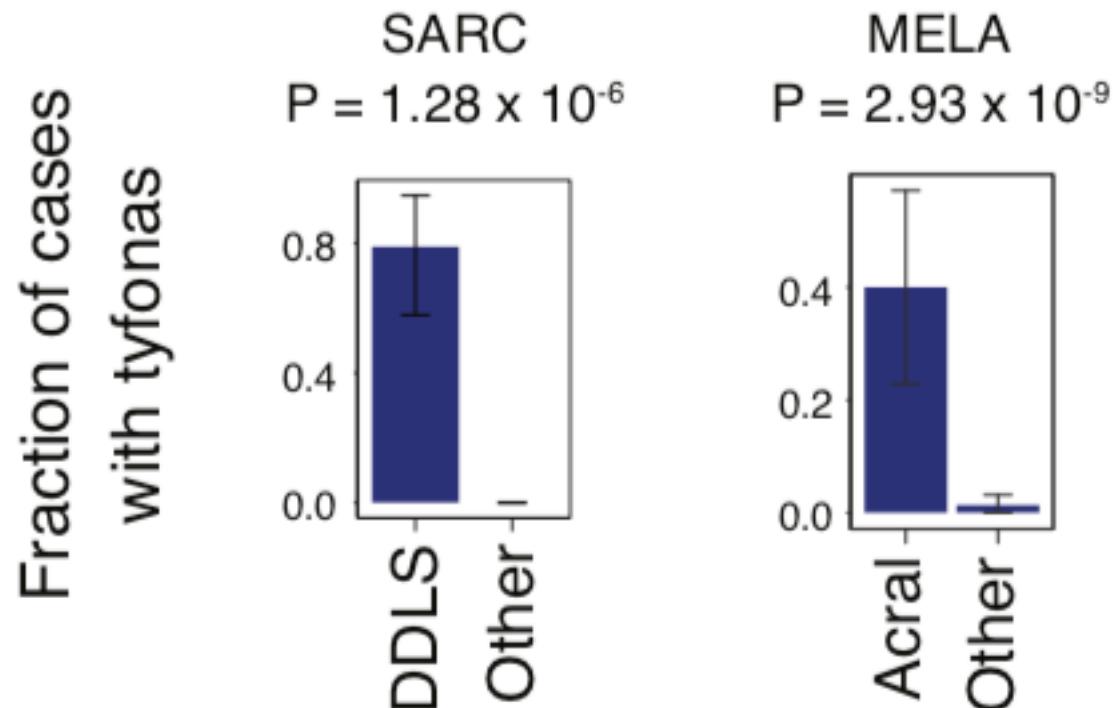


# Tyfonas enriched in distinct tumor types

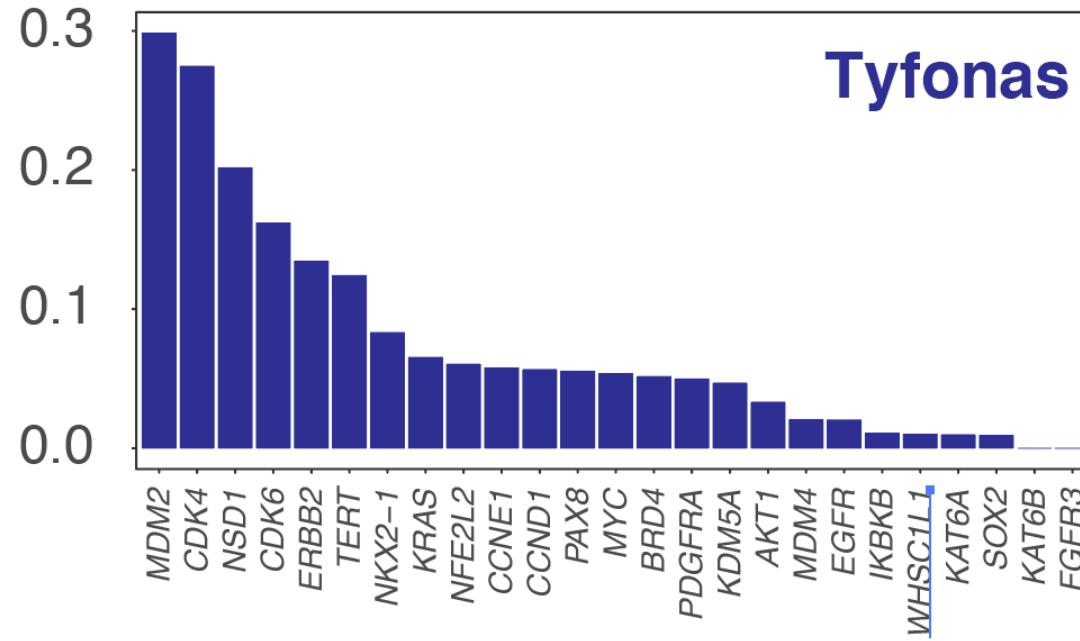
Fraction of  
cases with event



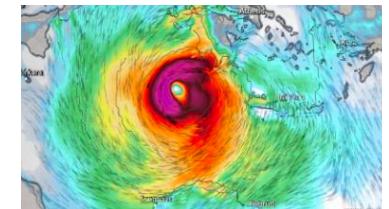
# Tyfonas enriched in distinct tumor (sub)types



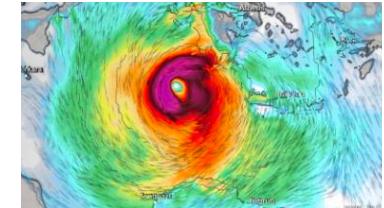
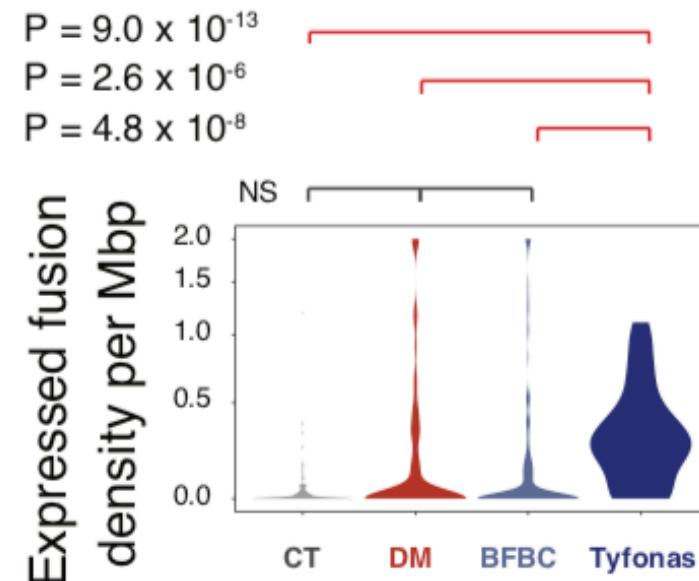
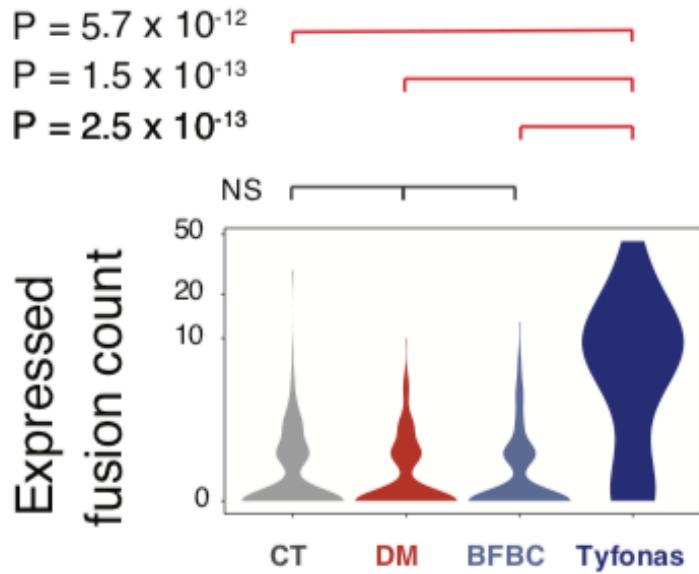
# Tyfonas amplify known oncogenes



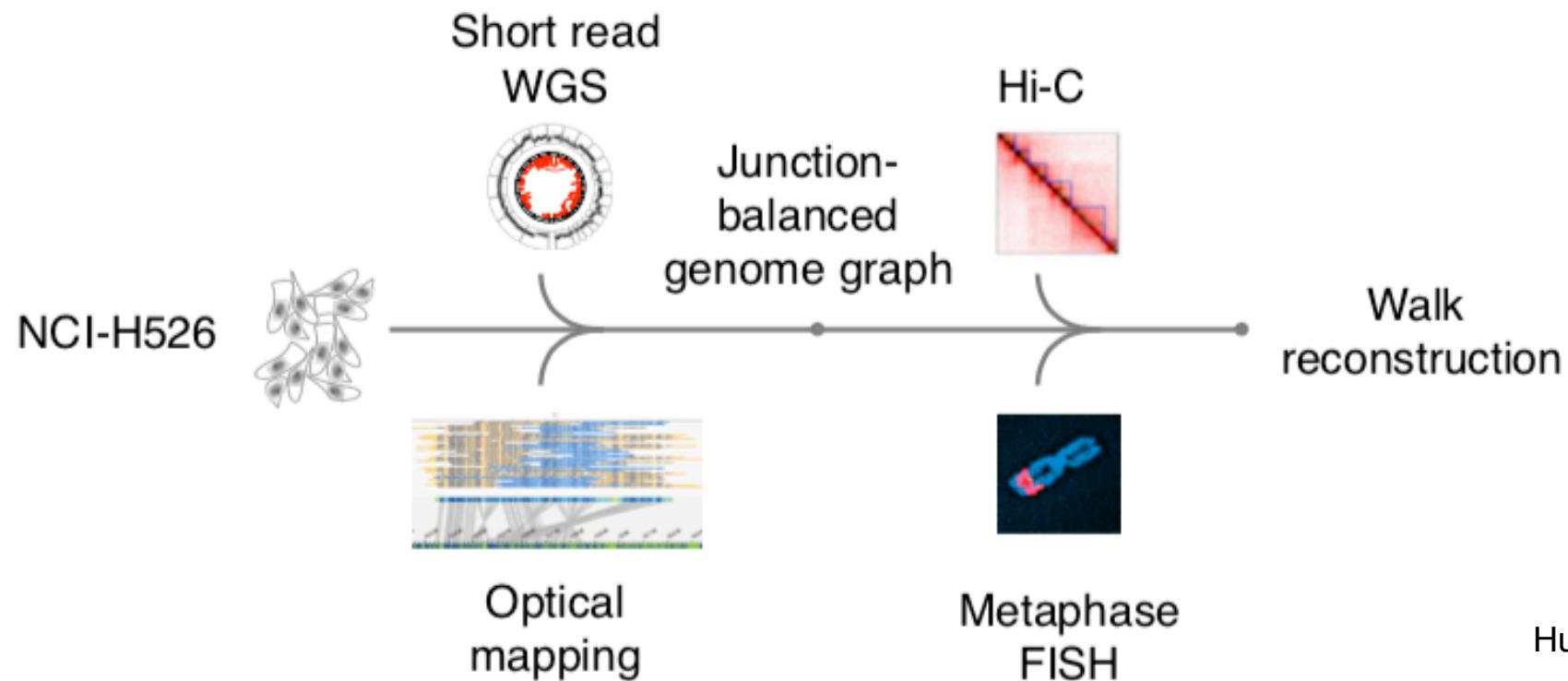
GISTIC amplification peaks



# Tyfonas are enriched in expressed protein-coding fusions



# Tyfonas reconstruction



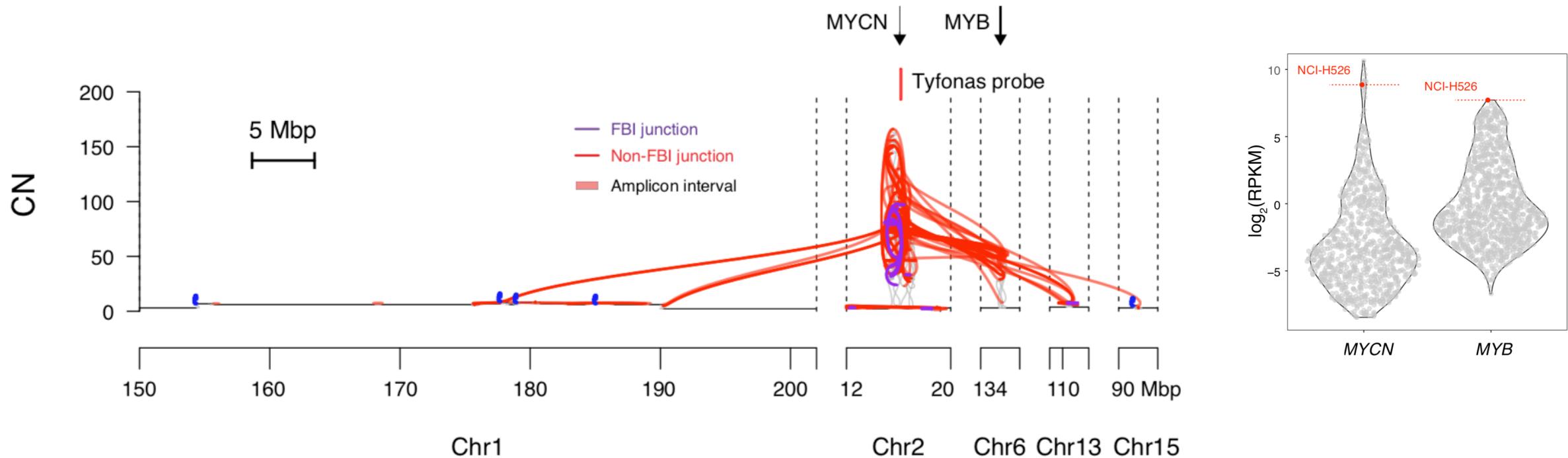
Huasong  
Tian



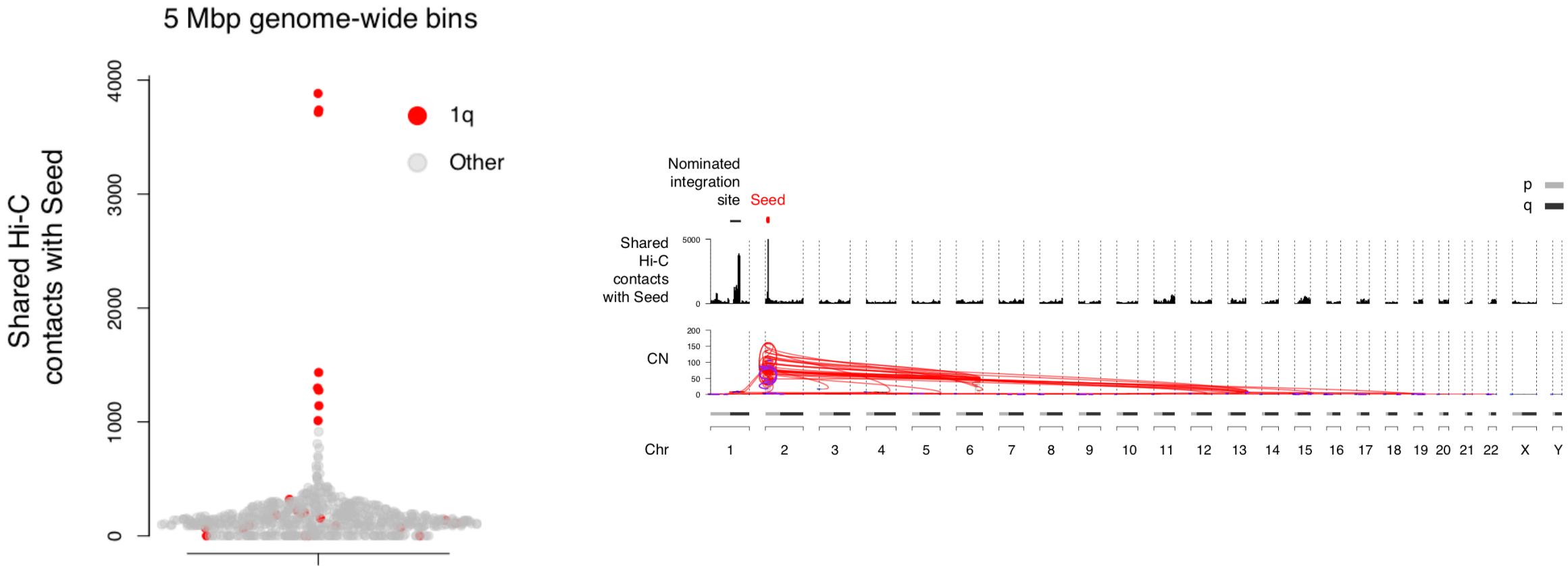
Julie  
Behr



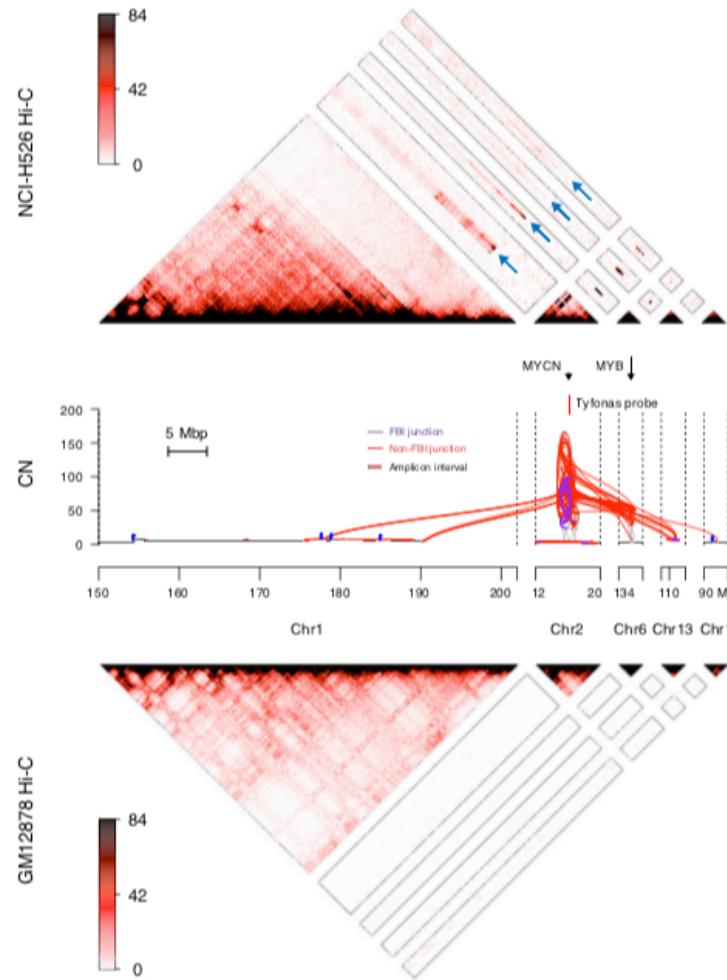
# SCLC tyfonas (NCI-H526)



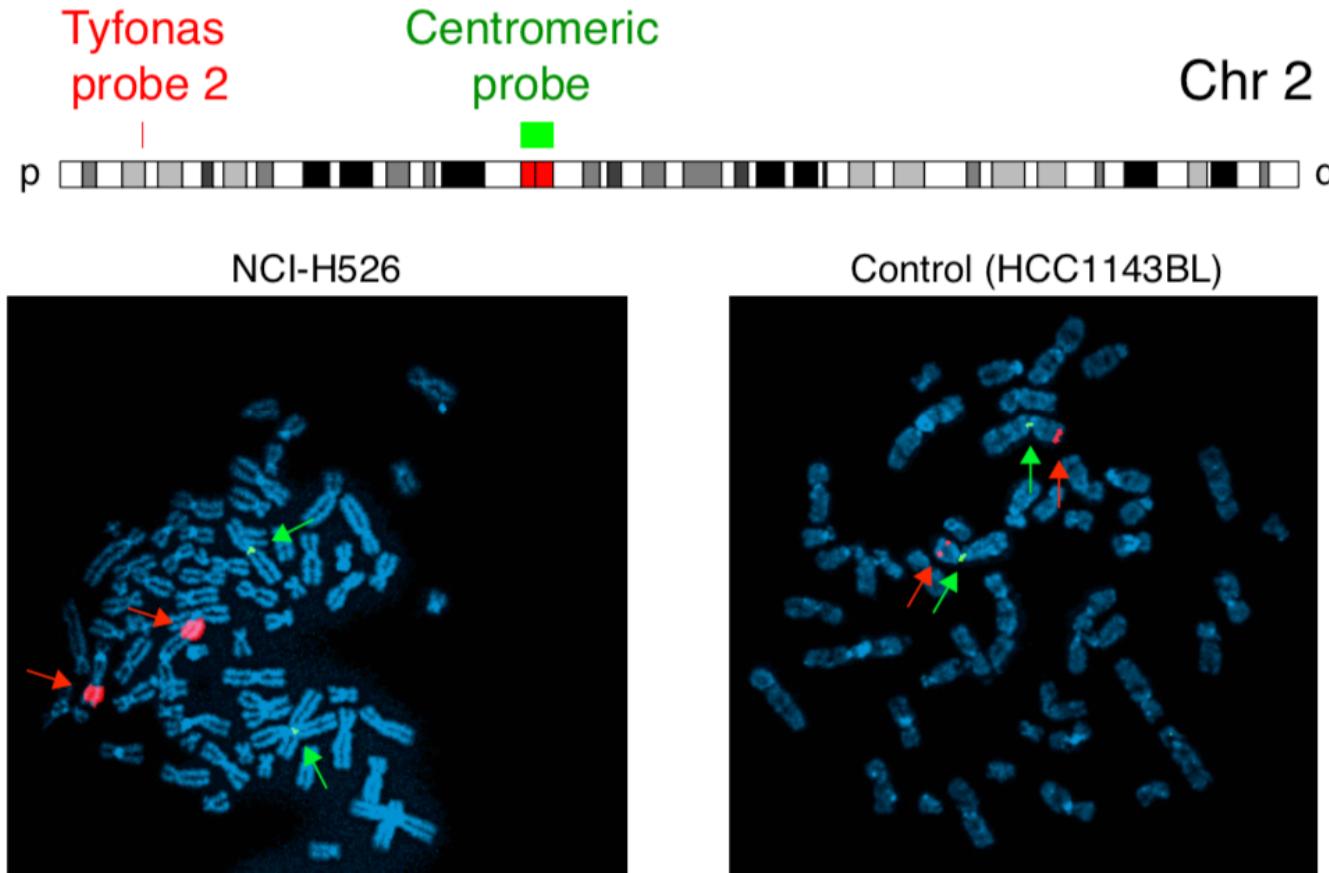
# Hi-C suggests chromosomal integration of tyfonas



# Hi-C suggests chromosomal integration of tyfonas



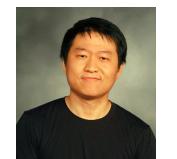
# FISH confirms chromosomal integration of tyfonas



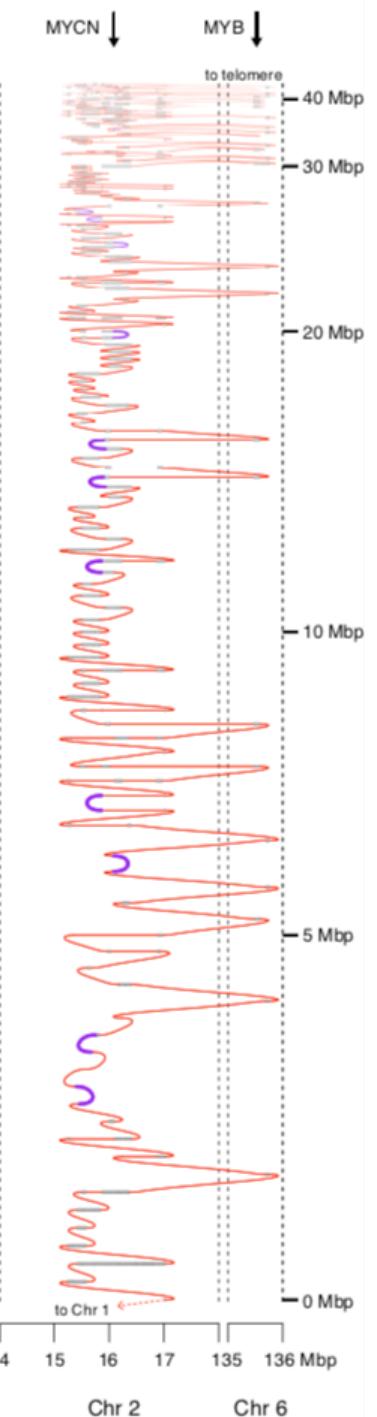
Sarah  
Kudman



Xiaotong  
Yao

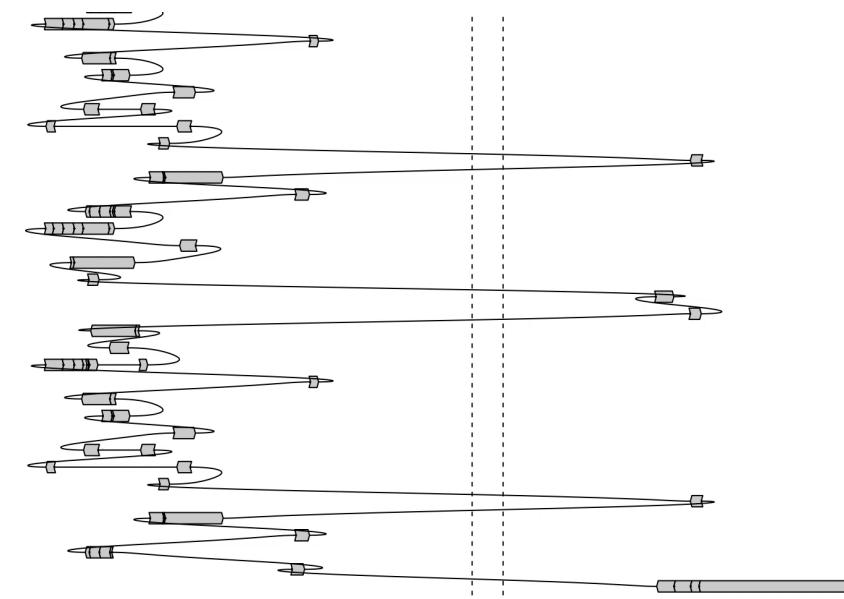


# Tyfonas reconstruction



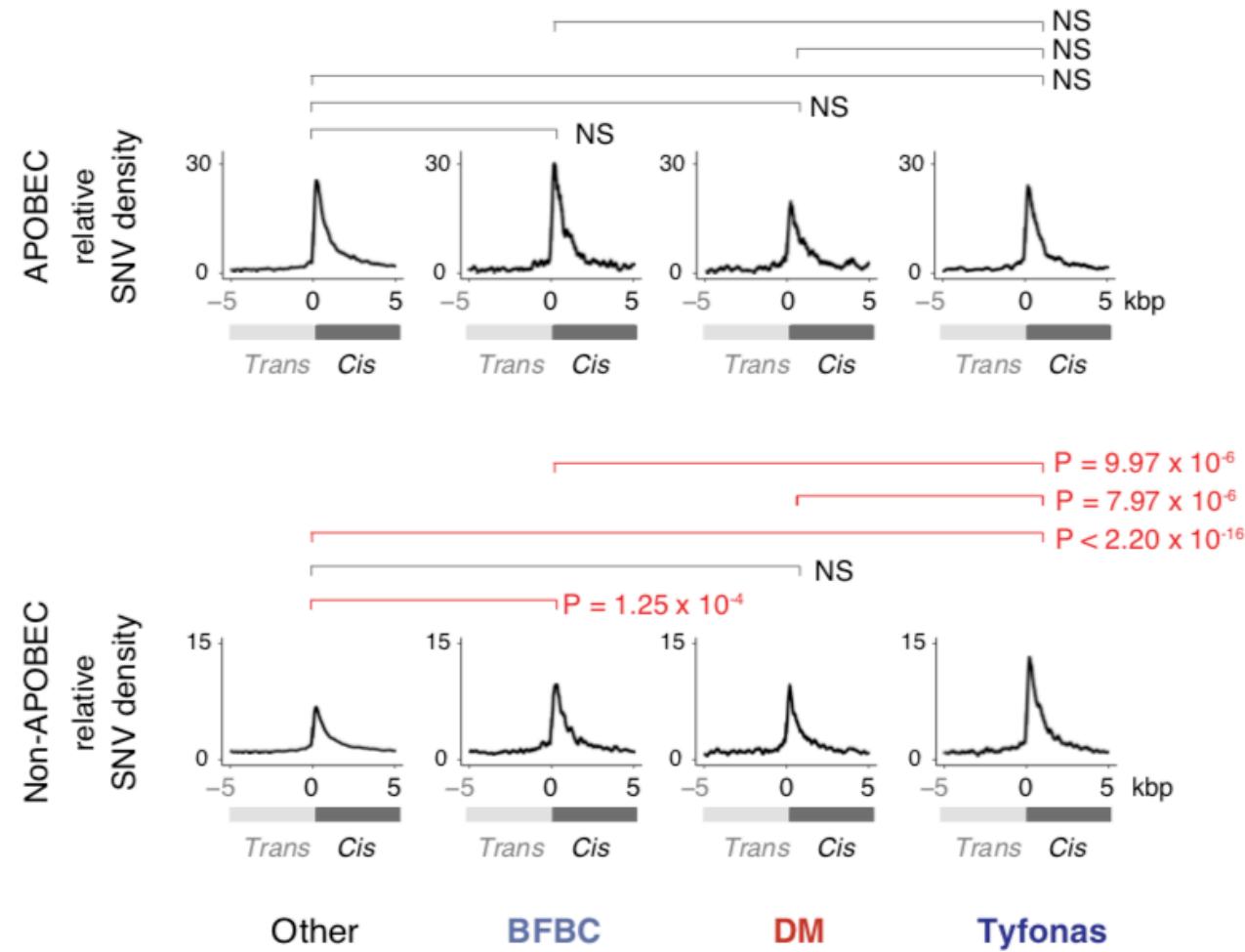
Chr 2

Chr 6



Julie Behr

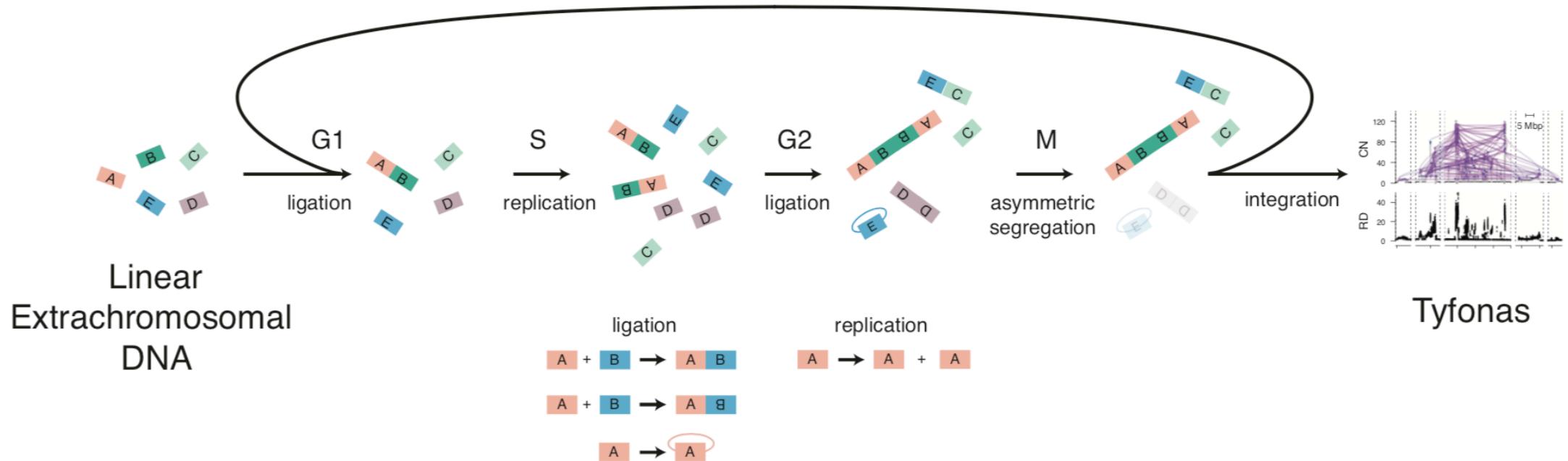
# Tyfonas are enriched in a novel breakend hypermutation process



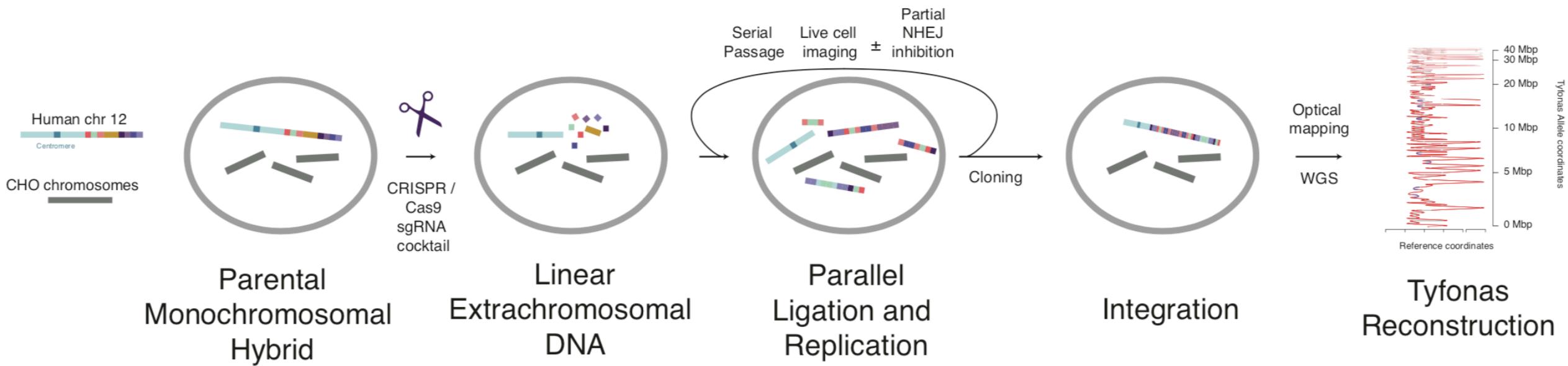
Kevin  
Hadi



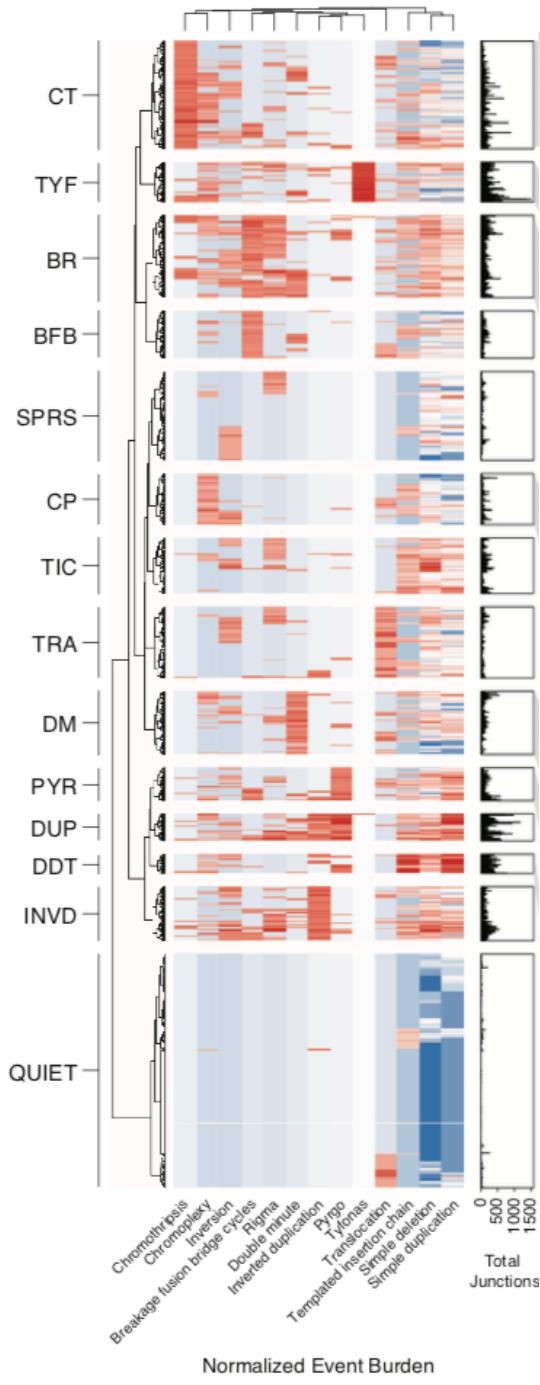
# Tyfonas model



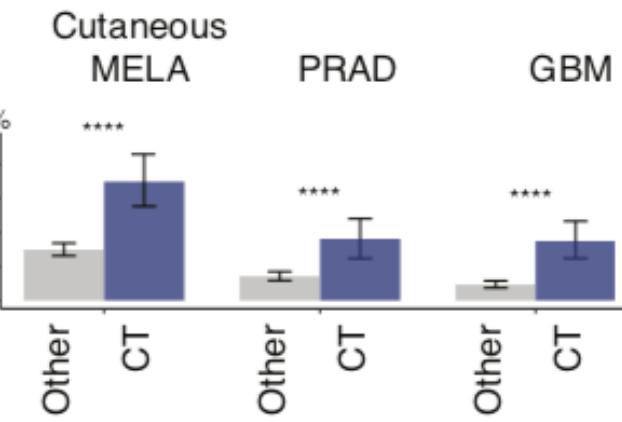
# Monochromosomal hybrid model



# Genome graph features define 13 clusters across ~2813 cancers



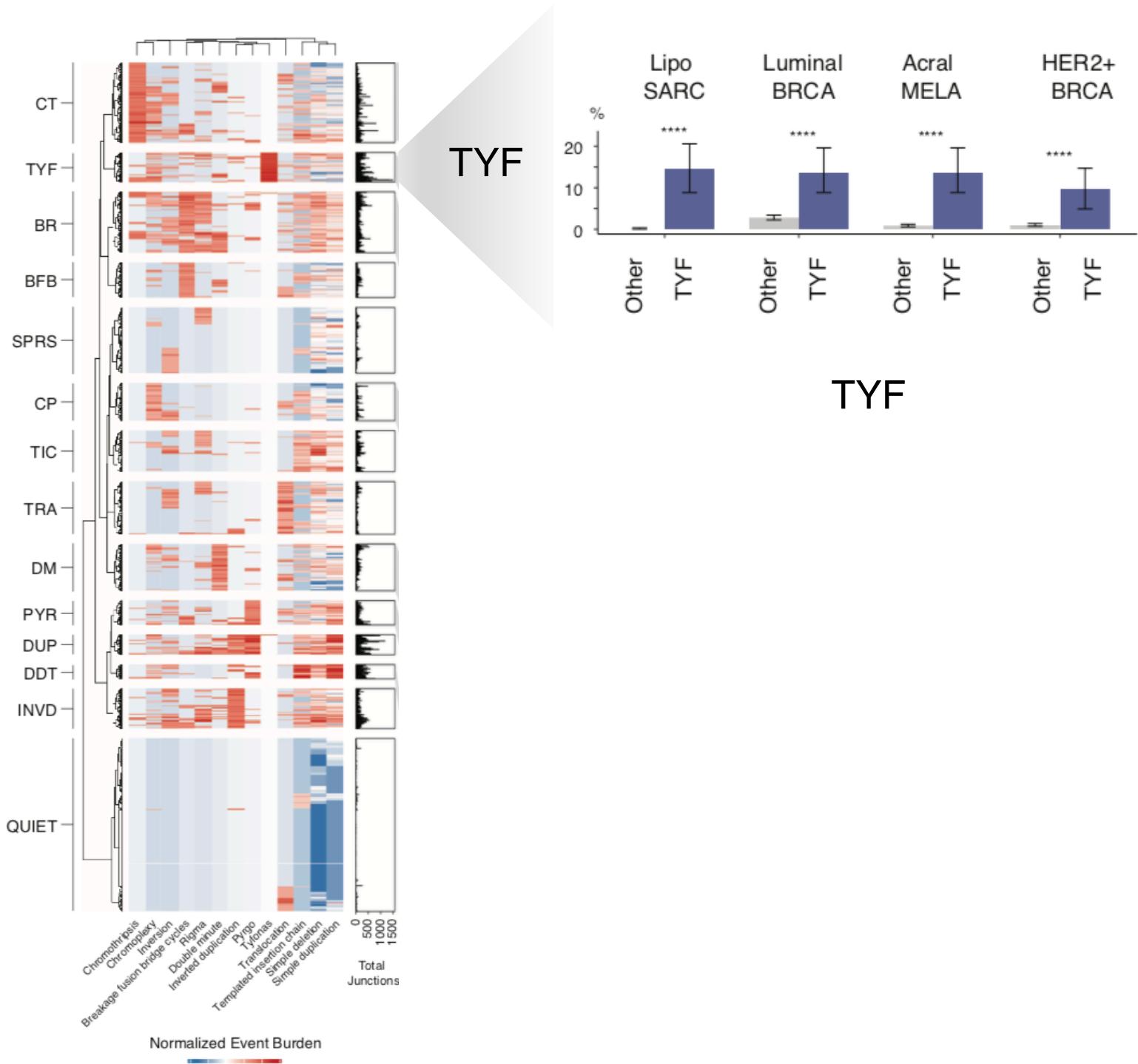
CT



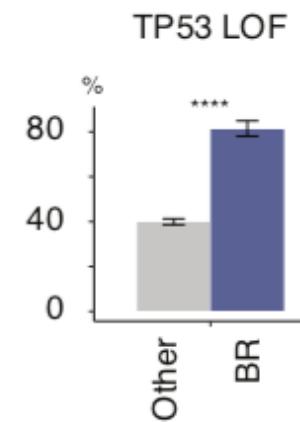
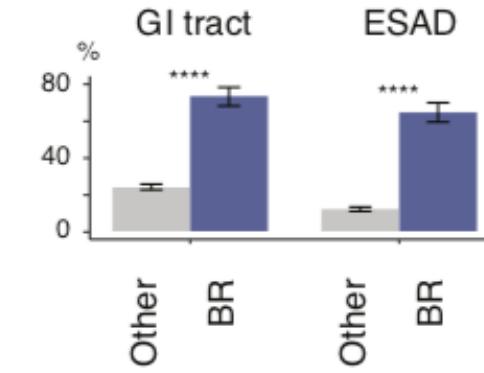
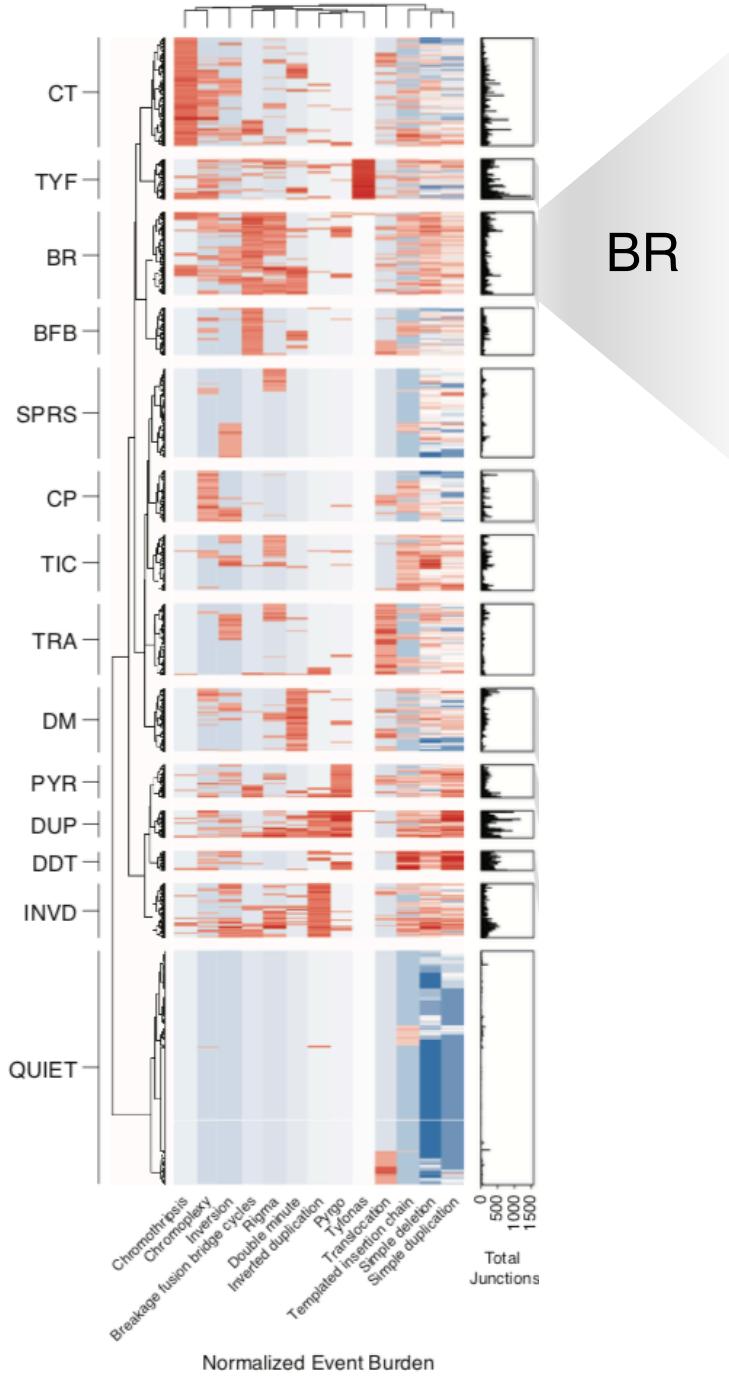
Xiaotong  
Yao



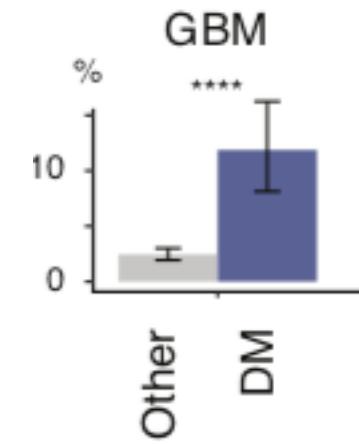
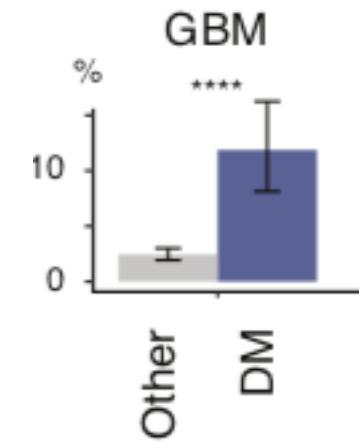
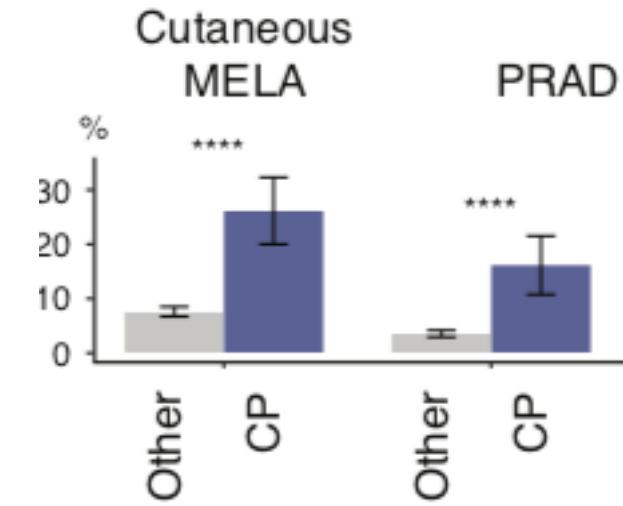
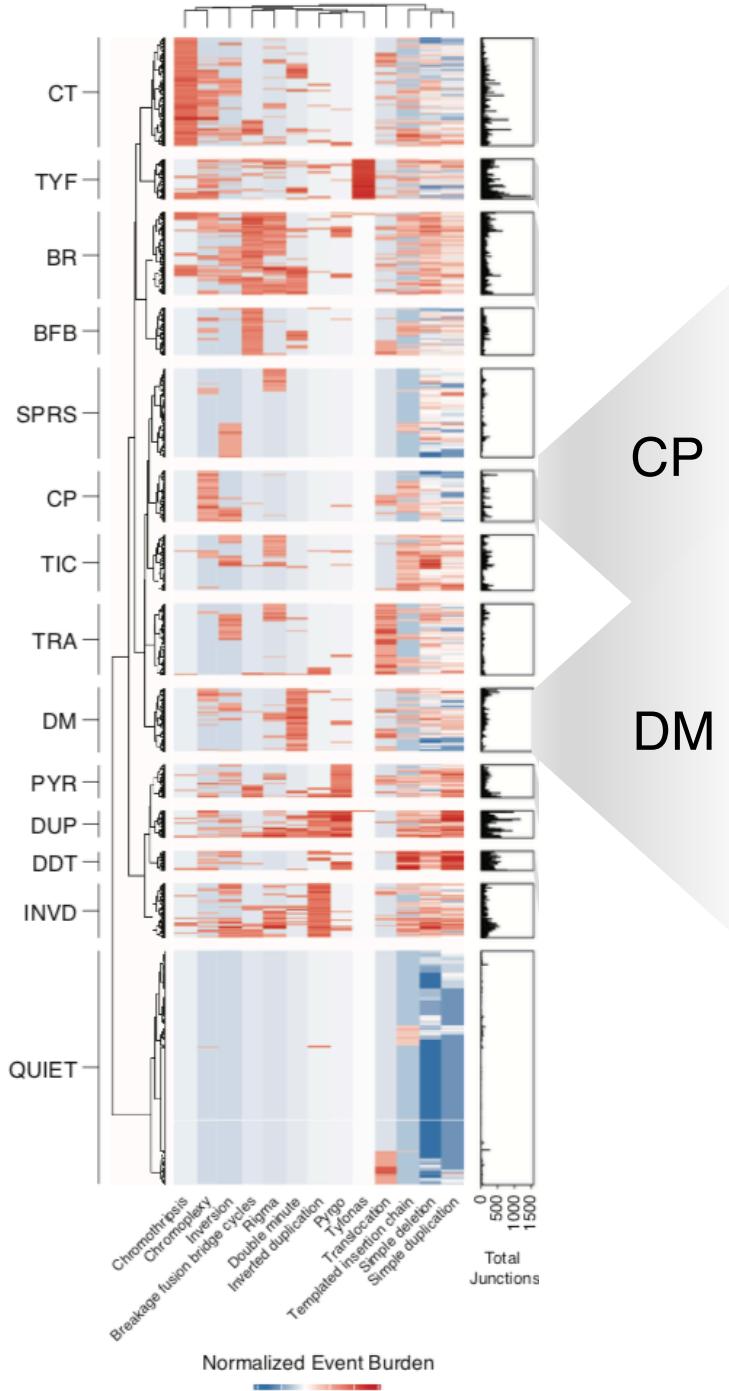
# Genome graph features define 13 clusters across ~2813 cancers



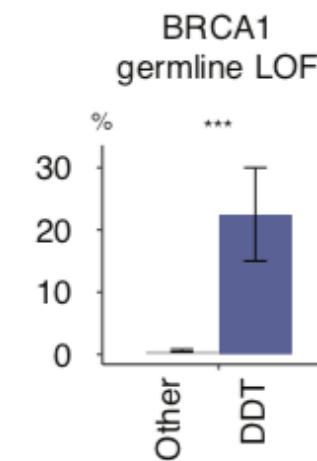
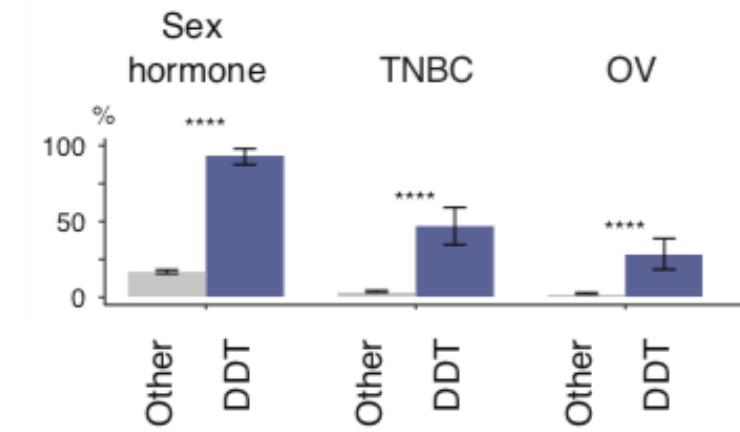
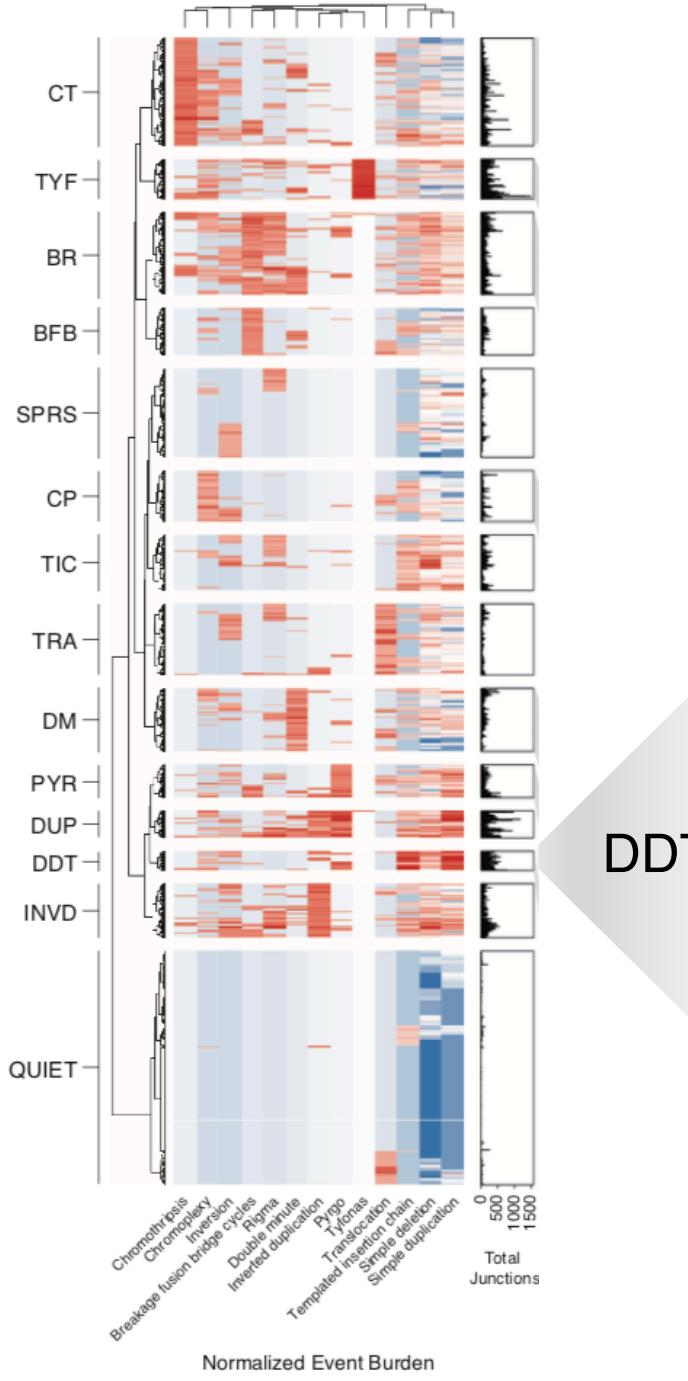
# Genome graph features define 13 clusters across ~2813 cancers



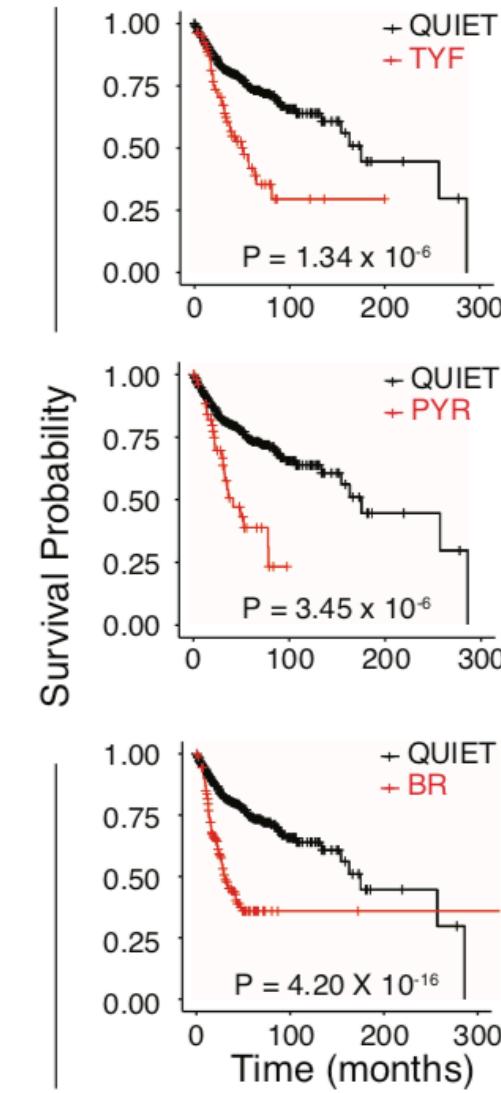
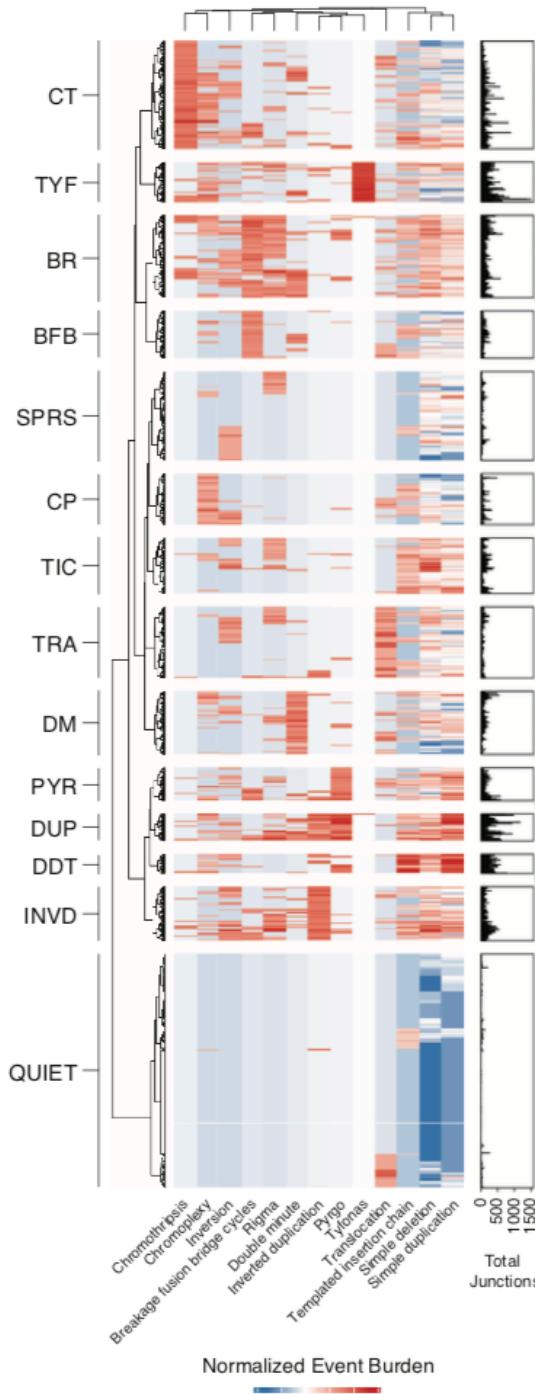
# Genome graph features define 13 clusters across ~2813 cancers



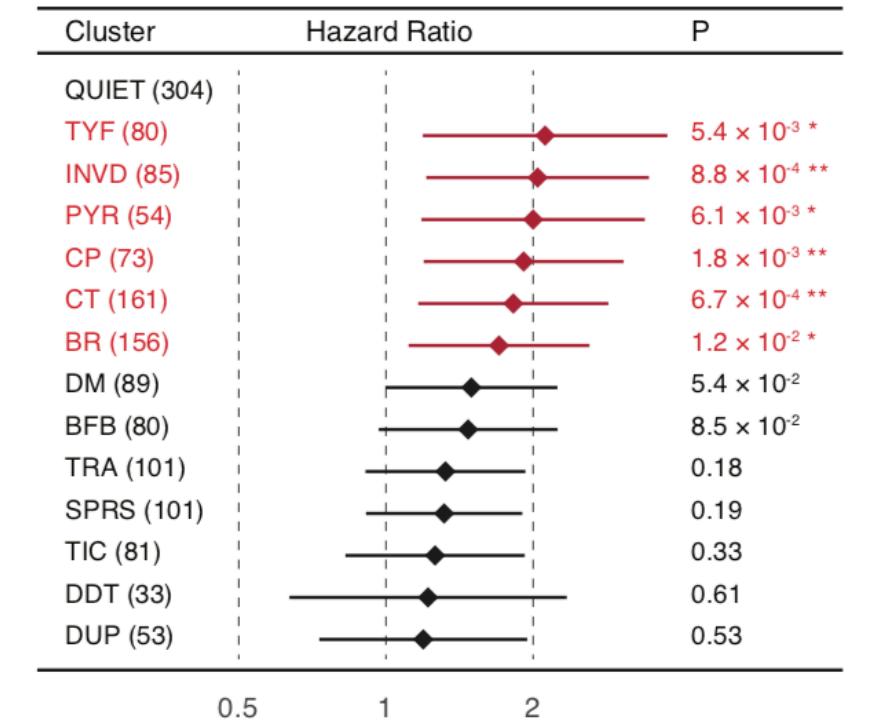
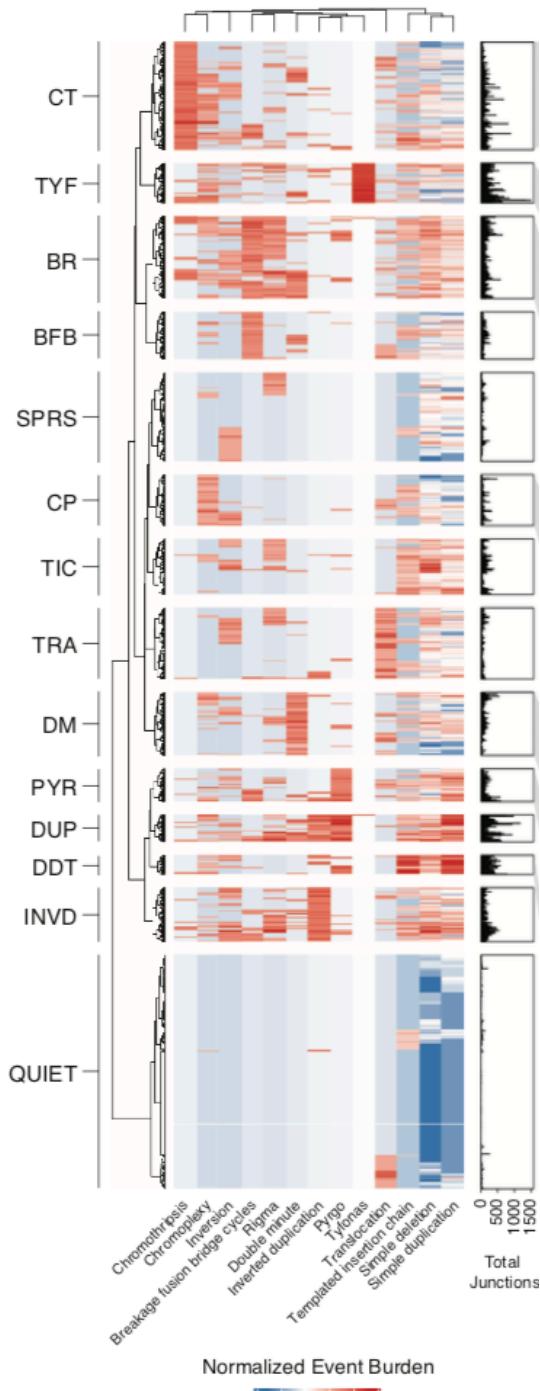
# Genome graph features define 13 clusters across ~2813 cancers



# Genome graph features define 13 clusters across ~2813 cancers



# Genome graph features define 13 clusters across ~2813 cancers



# Conclusions

- Analysis of ~2813 junction balanced genome graphs reveals novel patterns of complex genomic rearrangements (*rigma*, *pyrgo*, *tyfonas*)
- Rigma are late-replicating ditches of clustered *trans* deletion junctions arising early in esophageal adenocarcinoma
- Pyrgo are duplication towers enriched in superenhancers including *MYC*
- Tyfonas are likely extrachromosomal clusters of high copy fold-back inversions that are enriched in acral melanoma and expressed protein-coding fusions
- Tyfonas can be chromosomally integrated and are enriched in a breakend hypermutation process distinct from BFBC and DM
- Complex rearrangement patterns reveal prognostically significant categories, including BFBC / Rigma and DDT clusters

# Linked-read whole genome sequencing (10x genomics)

100 Kbp “synthetic long reads” \$7/Gb

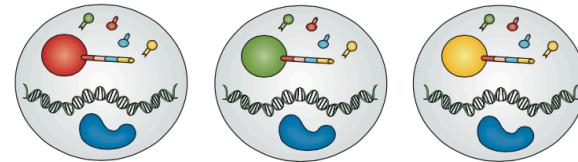
Stay tuned:

Kevin  
Hadi



Genome graph  
signatures  
of homologous  
repair deficiency

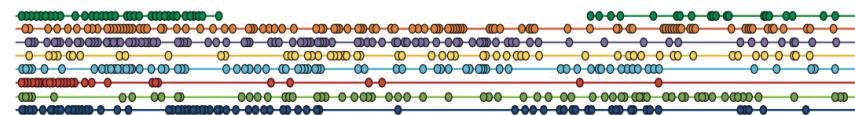
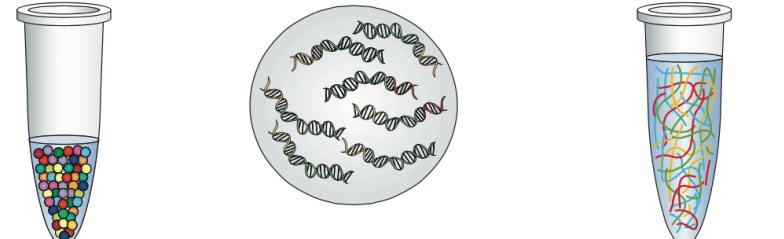
**Emulsion PCR**  
Arbitrarily long DNA  
is mixed with beads  
loaded with  
barcoded primers,  
enzyme and dNTPs



**GEMs**  
Each micelle  
has 1 barcode  
out of 750,000

**Amplification**  
Long fragments are  
amplified such that the  
product is a barcoded  
fragment ~350 bp

**Pooling**  
The emulsion is  
broken and DNA is  
pooled, then it  
undergoes a standard  
library preparation



Goodwin, Mcpherson, McCombie *Nature Reviews Genetics* 2016

# Long-read sequencing (Pac Bio, ONT)

Stay tuned:

Aditya  
Deshpande



High  
order  
chromatin  
structure

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

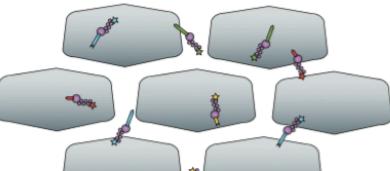
12-20 Kbp reads, \$85-400/Gb\*\*

Aa Pacific Biosciences

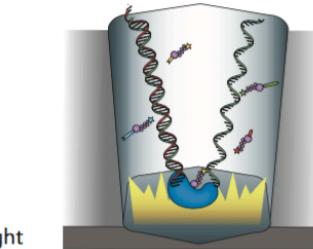
**SMRTbell template**  
Two hairpin adapters allow continuous circular sequencing



**ZMW wells**  
Sites where sequencing takes place



**Labelled nucleotides**  
All four dNTPs are labelled and available for incorporation



**Modified polymerase**  
As a nucleotide is incorporated by the polymerase, a camera records the emitted light

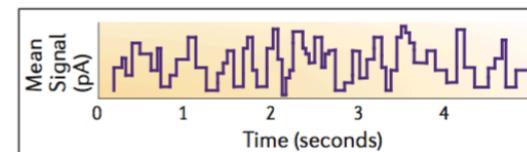
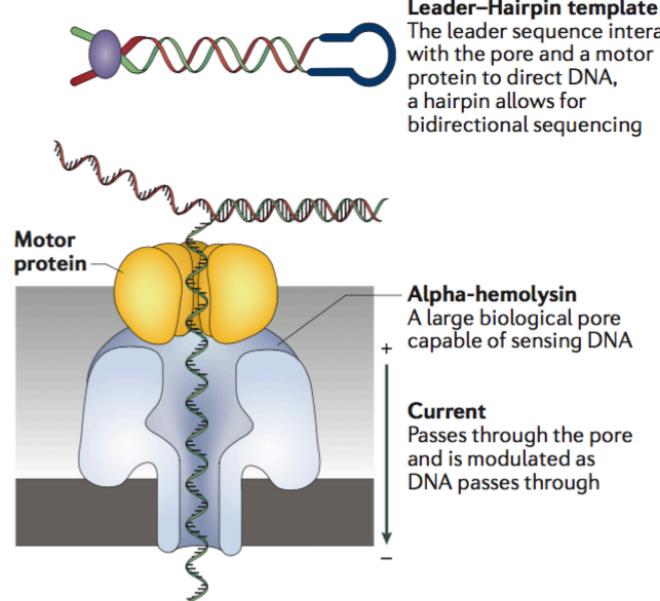


**PacBio output**  
A camera records the changing colours from all ZMWs; each colour change corresponds to one base

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

Ab Oxford Nanopore Technologies

**Leader-Hairpin template**  
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing

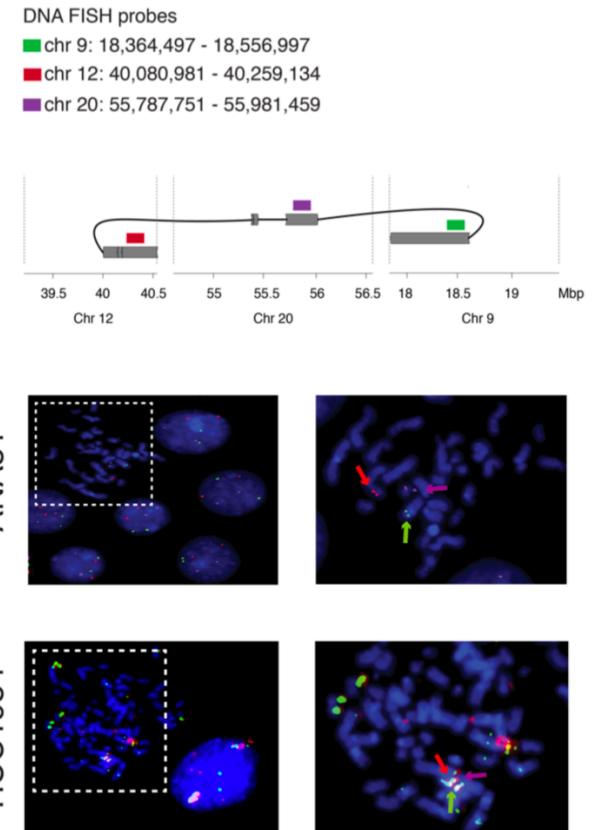
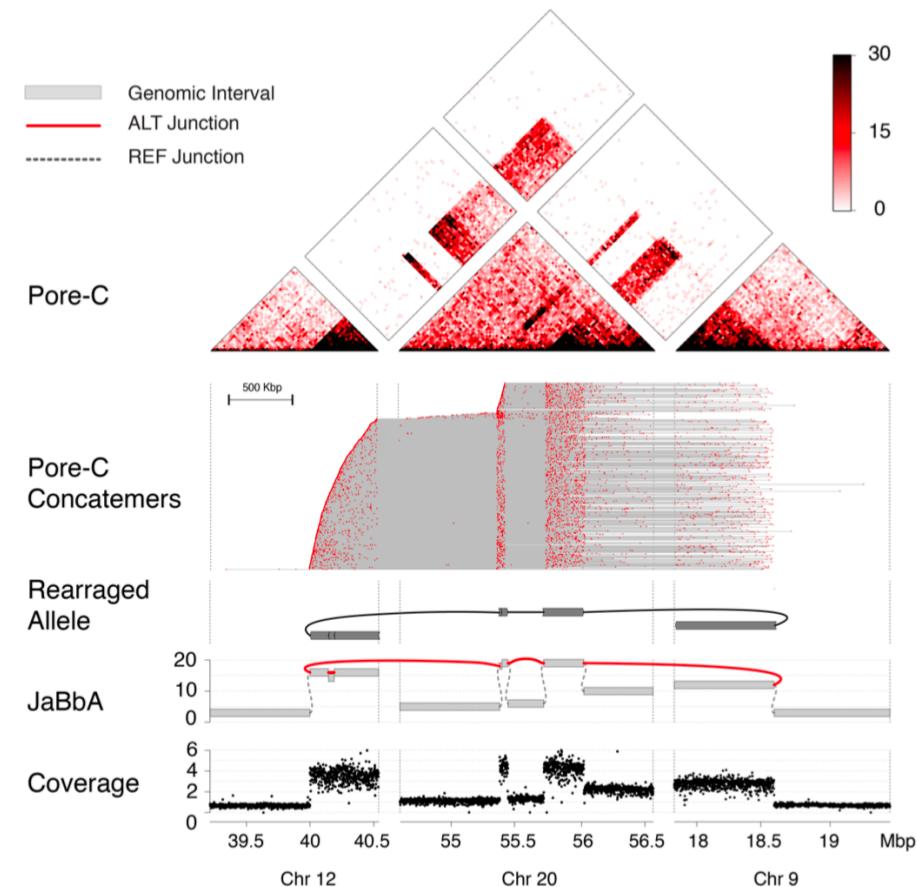
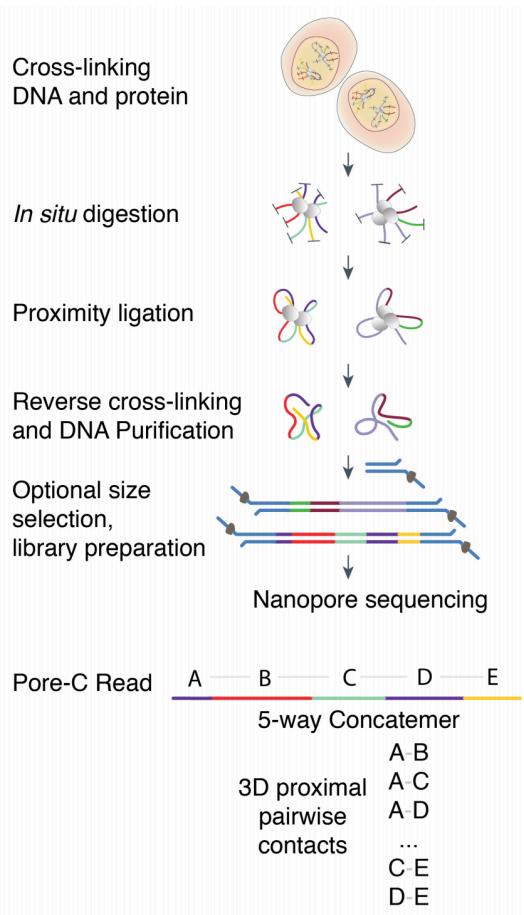


**ONT output (squiggles)**  
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

Goodwin, Mcpherson, McCombie *Nature Reviews Genetics* 2016

# Hi-C → Pore-C

Deshpande, Ulahannan, Pendleton, *et al* (in review)



Aditya  
Deshpande



Netha  
Ulahannan

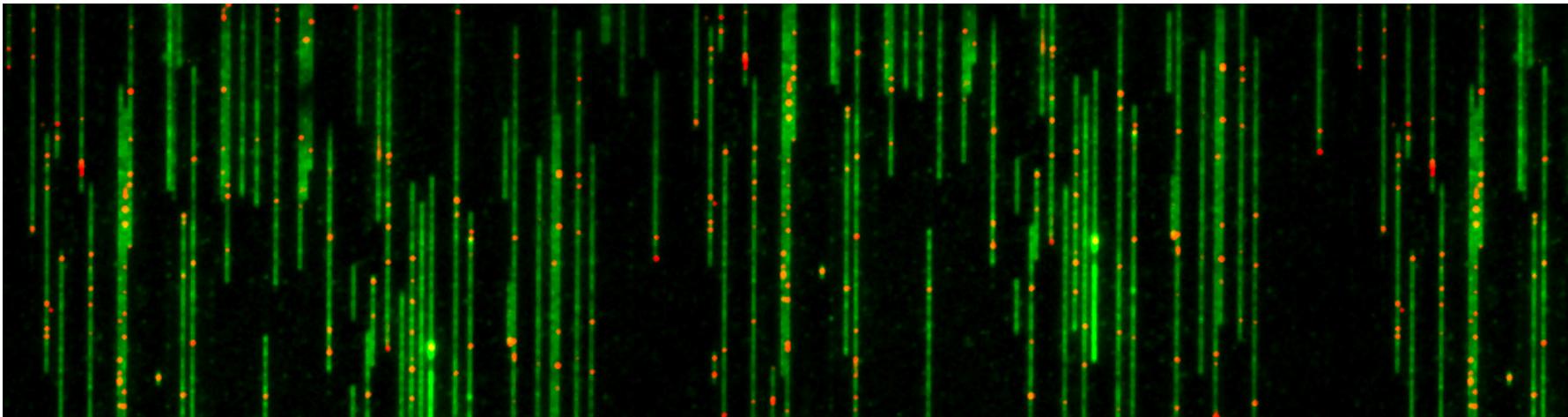


Matt  
Pendleton



Eoghan  
Harrington

# Optical mapping (BioNano Genomics)

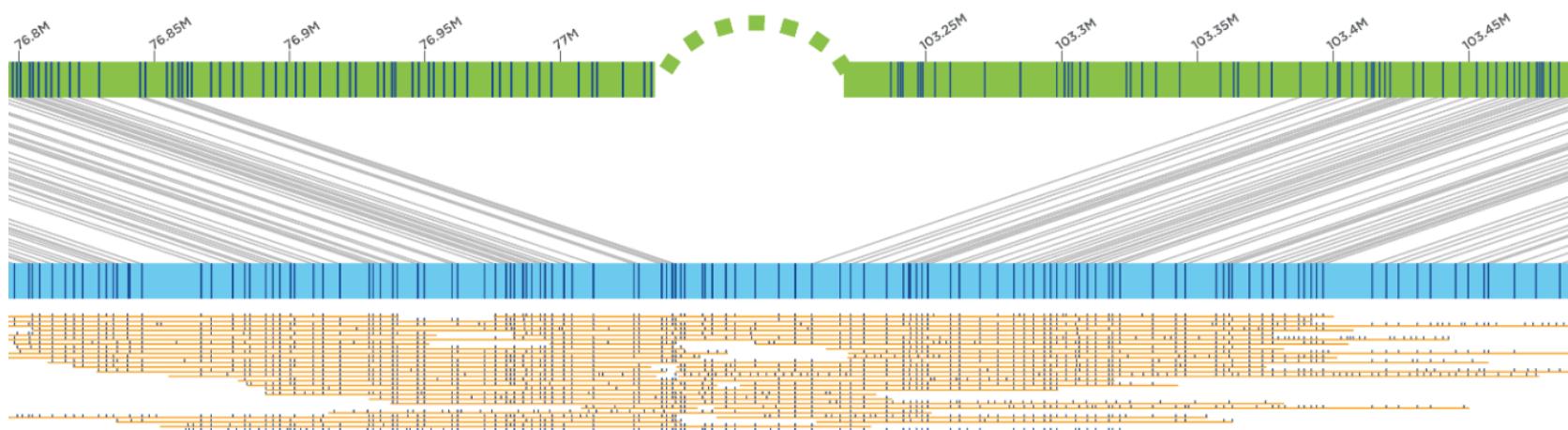


Stay tuned:

Julie  
Behr

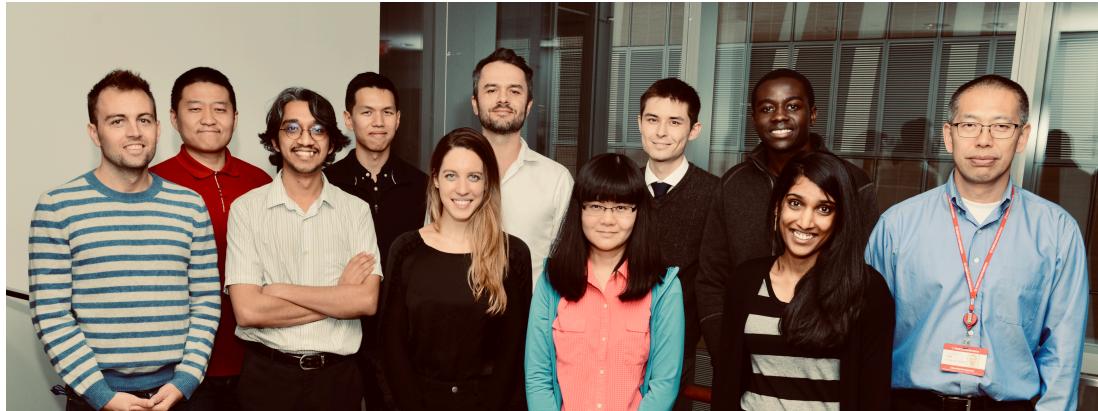


“Loose  
ends” in  
Cancer  
Genomes



# Marcin Imielinski Lab

<http://github.com/mskilab> [www.mskilab.org](http://www.mskilab.org)



Julie Behr (Tri-I CBM PhD student)

Aditya Deshpande (Tri-I CBM PhD student)

Kofi Gyan (Tri-I CBM PhD student)

Kevin Hadi (WCM PBSB PhD student)

Joel Rosiene (medical student)

Alon Shaiber (genomics data scientist)

Huasong Tian (research scientist)

Charalampos Xanthopoulakis (soft eng)

Xiaotong Yao (Tri-I CBM PhD student)



# Marcin Imielinski Lab

<http://github.com/mskilab> [www.mskilab.org](http://www.mskilab.org)



## Key Collaborators:

**Weill Cornell EIPM:** Olivier Elemento, Juan Miguel Mosquera, David Wilkes, Andrea Sboner

**Fred Hutch:** Patty Galipeau, Brian Reid, Tom Paulson, Xiaohong Li

**CCLE / Broad:** Mahmoud Ghandi, Franklin Huang

**NYGC:** Bob Darnell, Kazimierz Wrzeszczynski

**Other:** Ed Reznik, Bud Mishra, Rameen Beroukhim

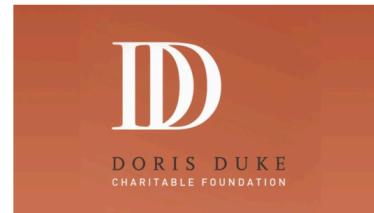


**Marcin Imielinski Lab**  
<http://github.com/mskilab>   [www.mskilab.org](http://www.mskilab.org)



BURROUGHS  
WELLCOME  
FUND

BURROUGHS WELLCOME  
FUND



STARR CANCER CONSORTIUM



Memorial Sloan Kettering  
Cancer Center



Melanoma  
Research Alliance

