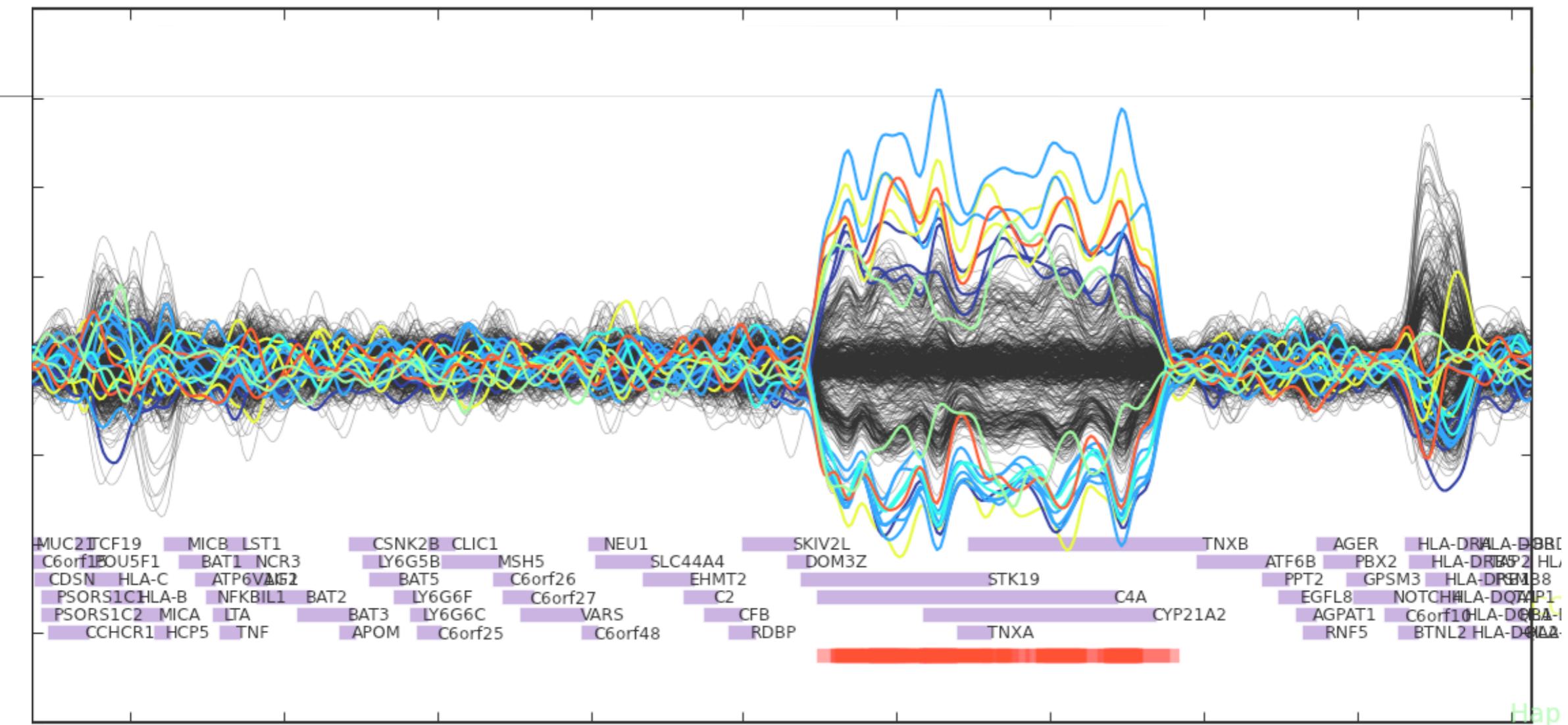
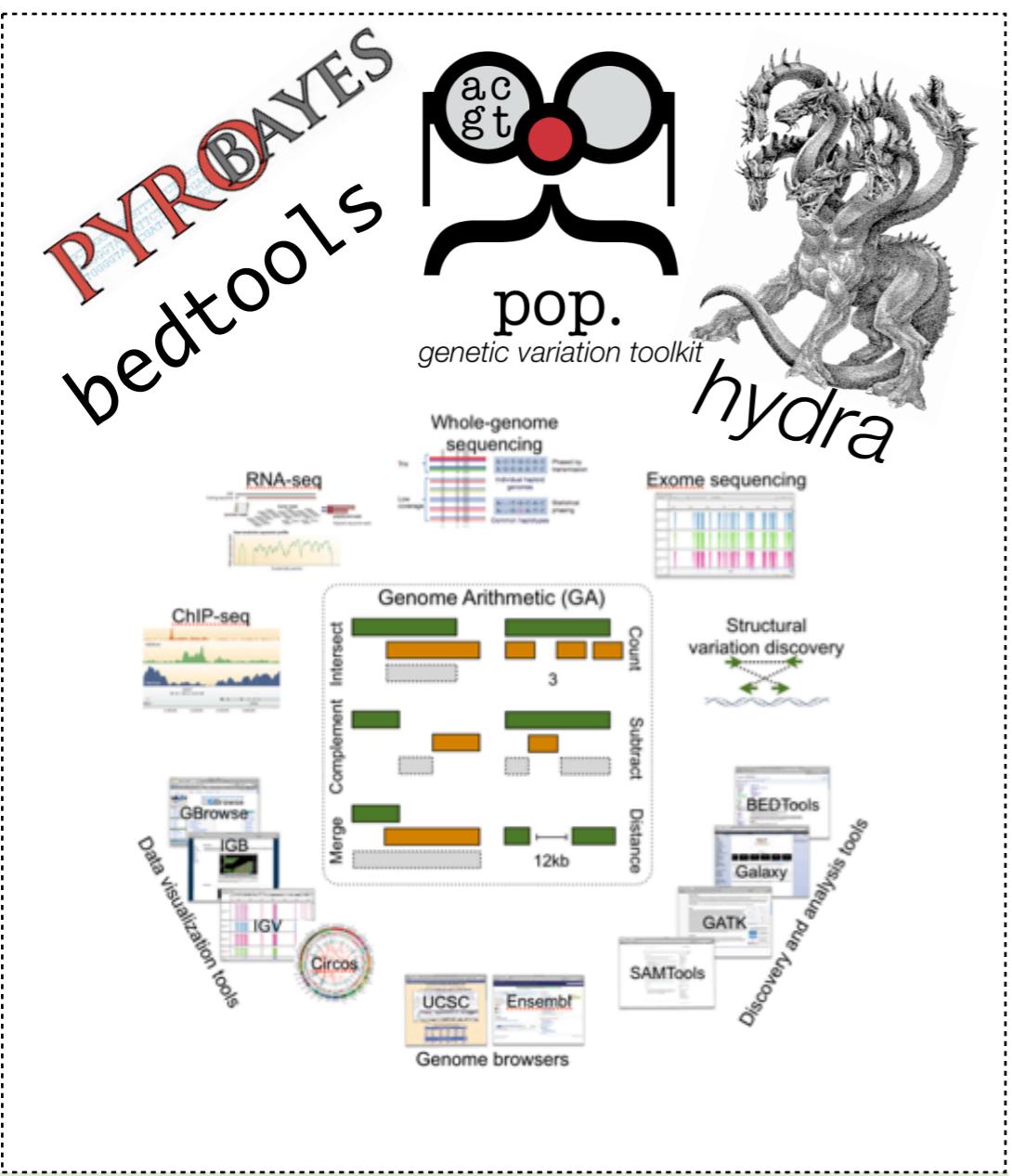


EXPLORING THE Origin and extent of structural variation in human genomes

Aaron Quinlan (arq5x)
Department of Public Health Sciences
Center for Public Health Genomics



Quinlan Lab: computational genomics

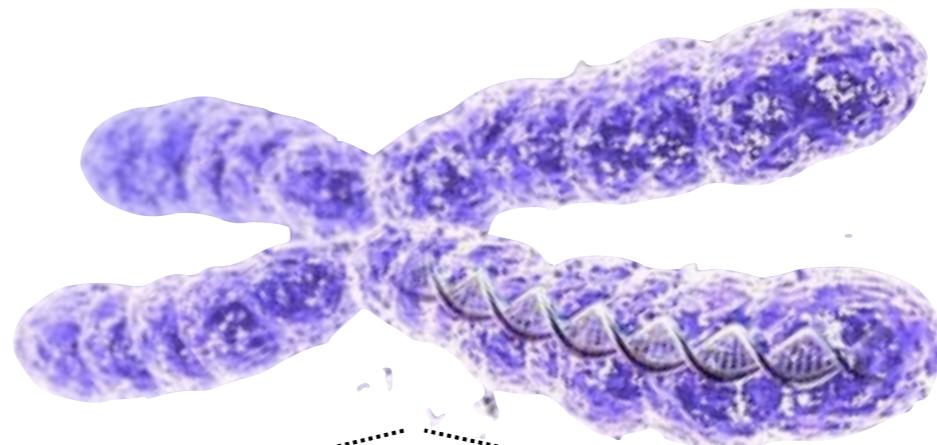


Genomics software and method development

Cancer genomics
Extreme sensitivity to IR
Genome stability in iPS cells
Genetics of lupus
Genetics of T1D
Genome structural variation
Functional structural variation

DNA sequencing to explore genome biology

A brief overview of structural variation



Reference genome

A B C D E

Deletion

A * C D E

Insertion

A B C Q D E

Duplication

A B C D D E

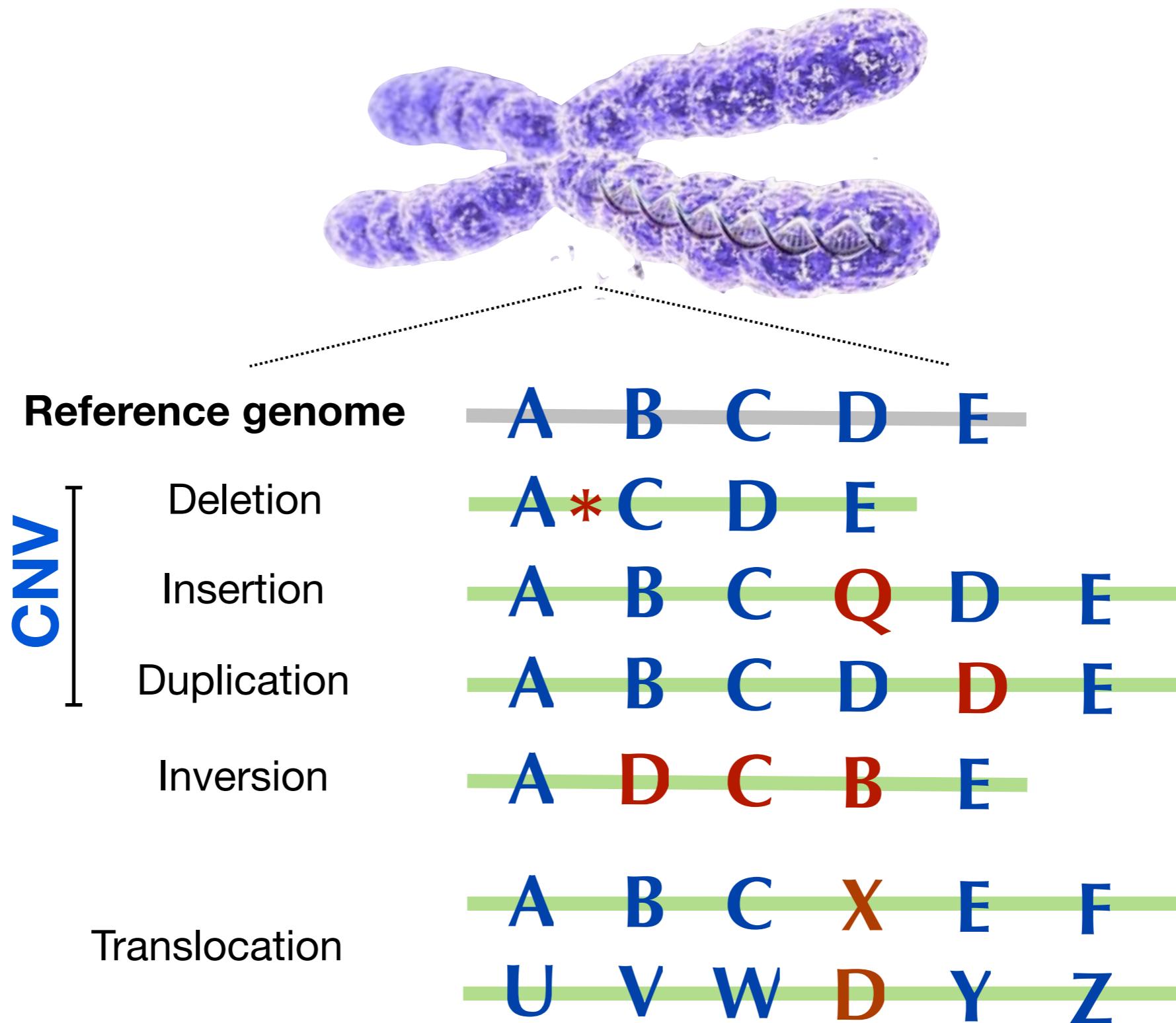
Inversion

A D C B E

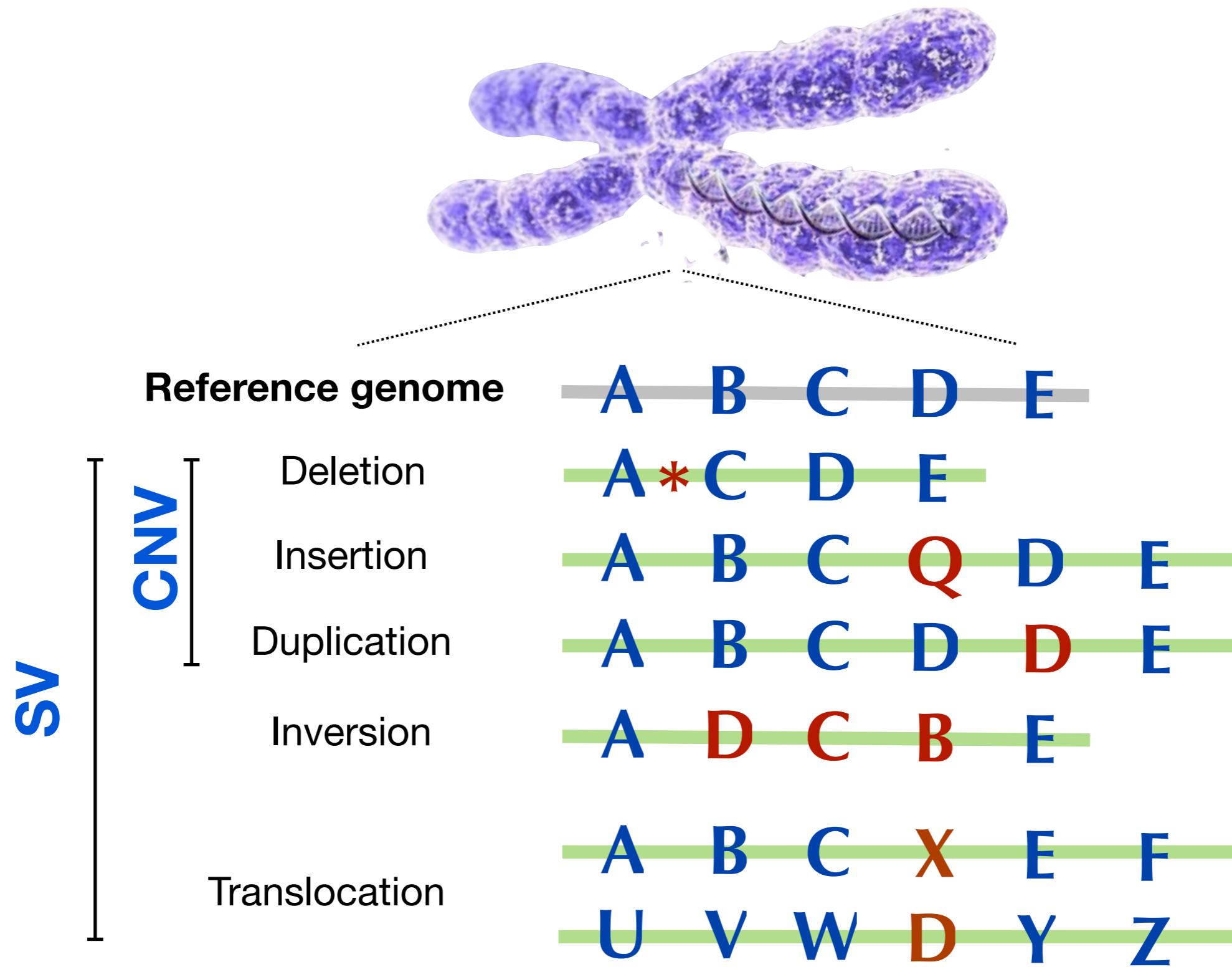
Translocation

A B C X E F
U V W D Y Z

A brief overview of structural variation



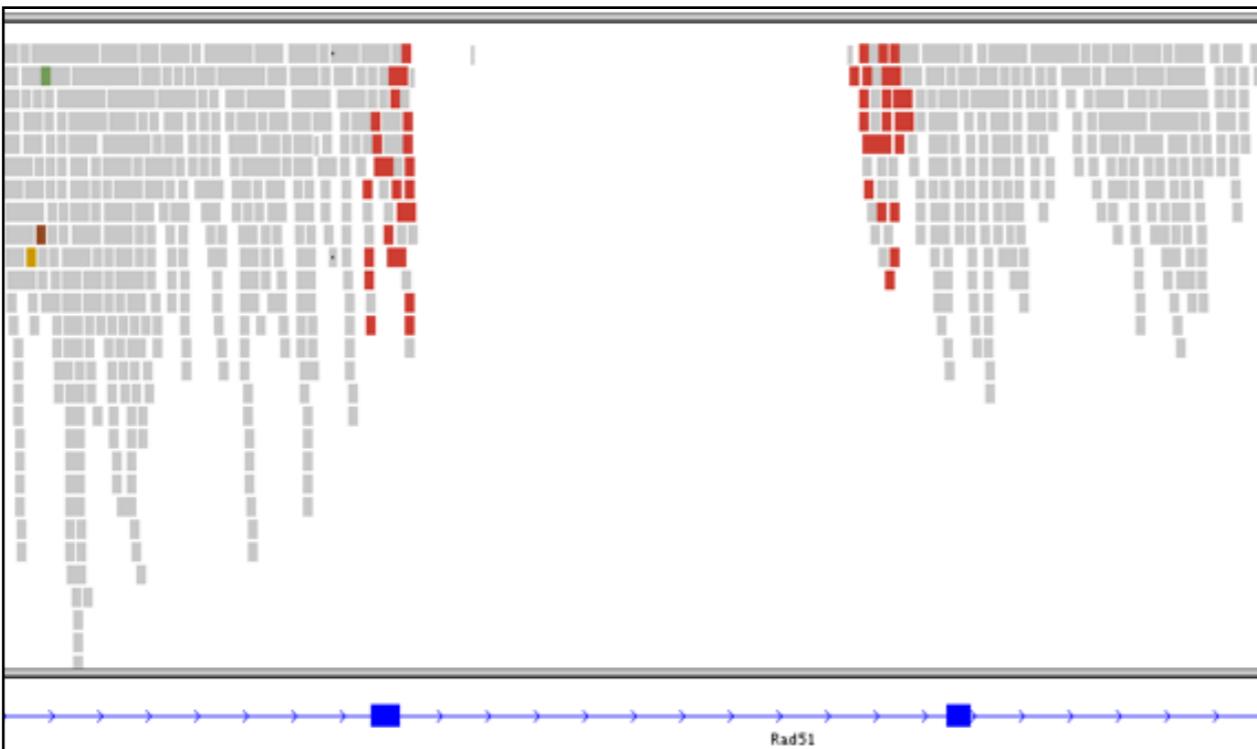
A brief overview of structural variation



A brief overview of structural variation

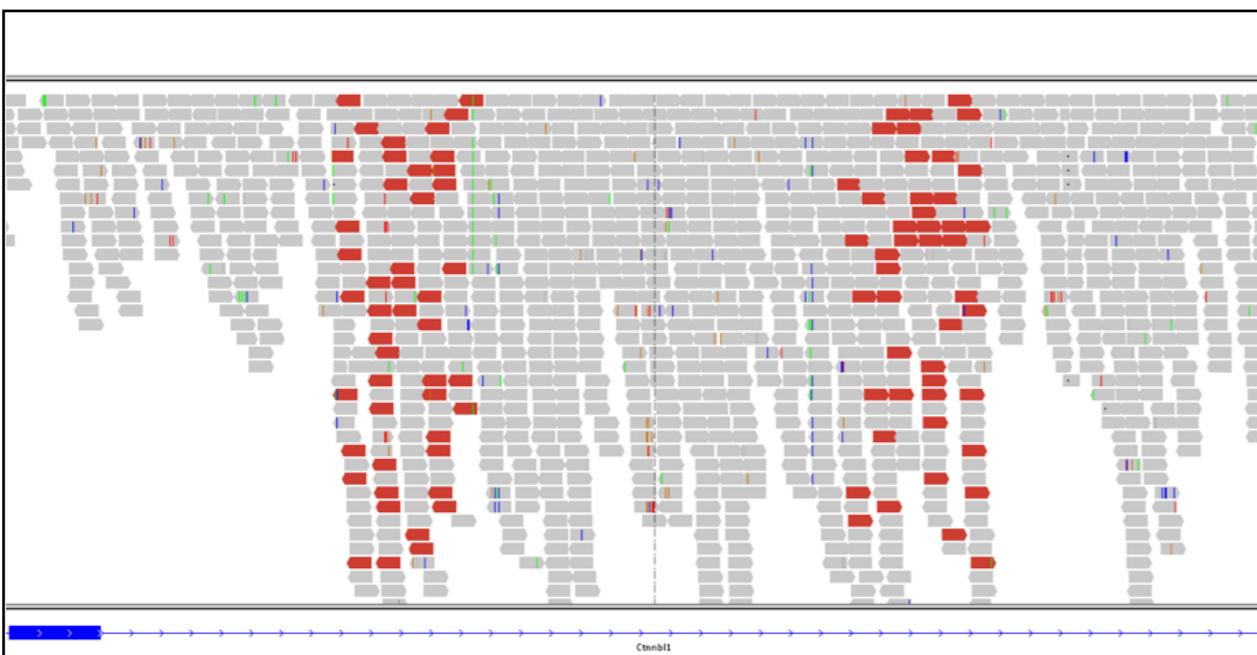
Some examples...

Ref.



Deletion
in experimental
genome

Ref.



Duplication
in experimental
genome

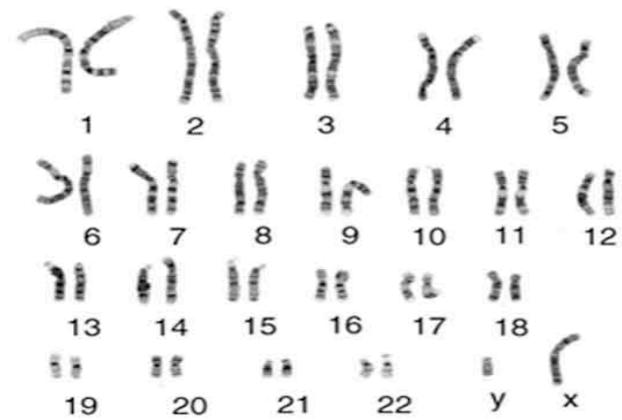
A brief overview of structural variation

Why is SV relevant to genetics?

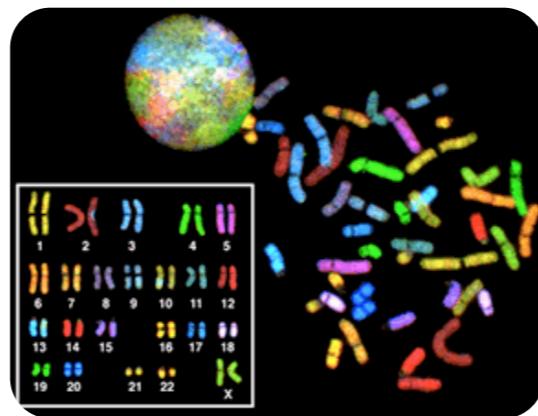
- ▶ **They are common and affect a large fraction of the genome**
 - ▶ Any two humans differ by 3,000 - 10,000 SVs
 - ▶ In total, they impact more base pairs than all single-nucleotide differences.
- ▶ **They are a major driver of genome evolution**
 - ▶ Speciation can be driven by rapid changes in genome architecture
 - ▶ Genome instability and aneuploidy: hallmarks of solid tumor genomes
- ▶ **Genetic basis of traits**
 - ▶ Gene dosage effects.
 - ▶ Neuropsychiatric disease (e.g., autism, schizophrenia)
 - ▶ Spontaneous SVs implicated in so-called “genomic” and developmental disorders
 - ▶ Common SNPs are not proxies for all forms of variation.
 - ▶ Somatic genome instability; age-dependent disease

A brief overview of structural variation

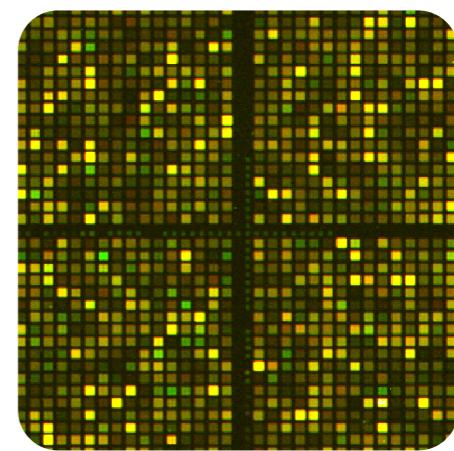
Our understanding is driven by technology



1940s - 1980s
Cytogenetics / Karyotyping



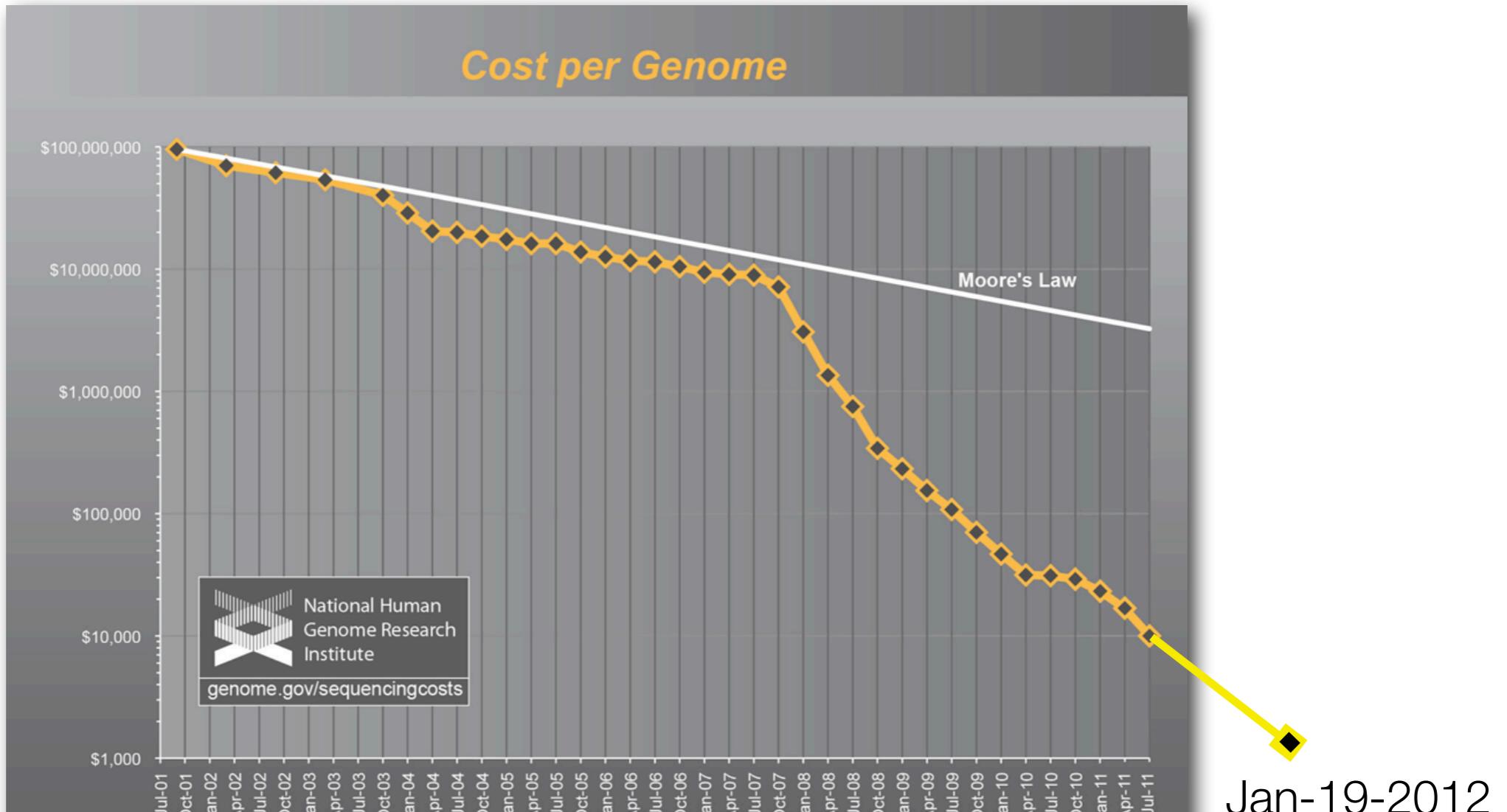
1990s
CGH / FISH /
SKY / COBRA



2000s
Genomic microarrays
BAC-aCGH / oligo-aCGH

A brief overview of structural variation

DNA sequencing is a potent “chromoscope”



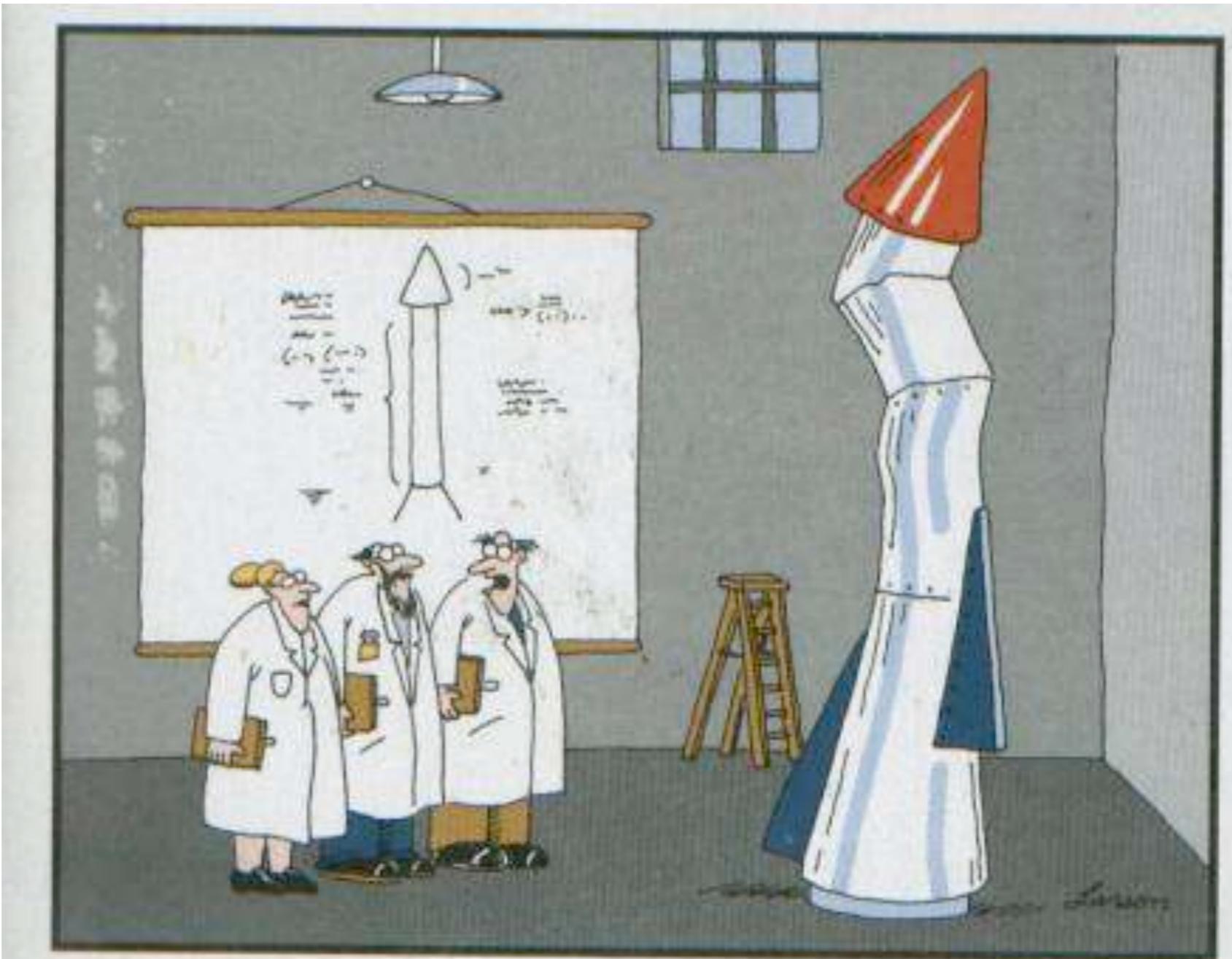
Sequencing a single human genome:

2002: 1 decade, > 50 labs, > \$3 billion (a dollar a base pair).

2012: 1 week, 1 grad student, \$4000. (Cheaper than a colonoscopy!)

A brief overview of structural variation

Then why isn't characterizing SV easy?



**"It's time we face reality, my friends. ...
We're not exactly rocket scientists."**

A brief overview of structural variation

The genome is a complex puzzle and we get small pieces.



"The \$1000 genome, the \$100,000 analysis"

- Elaine Mardis WashU Genome Center

A brief primer of structural variation

SV discovery with modern sequencing

1. Align DNA sequences from sample to human reference genome



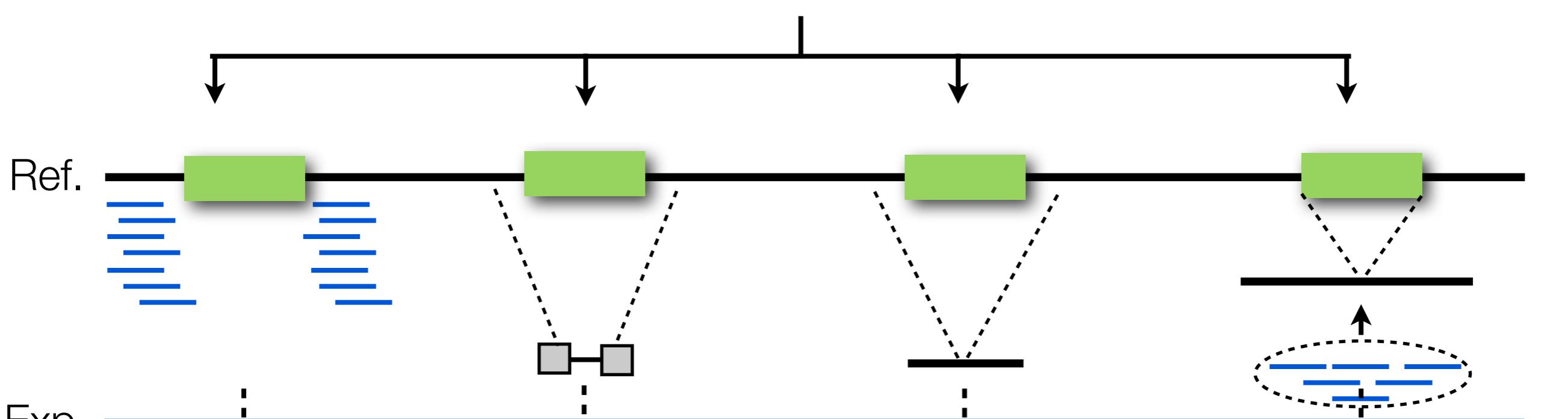
A brief primer of structural variation

SV discovery with modern sequencing

1. Align DNA sequences from sample to human reference genome



2. Look for evidence of structural differences



(a) Depth of
coverage

(b) Paired-end
mapping

(c) Split-read
mapping

(d) *de novo*
assembly

Resolution

High

Low

Difficulty / Cost

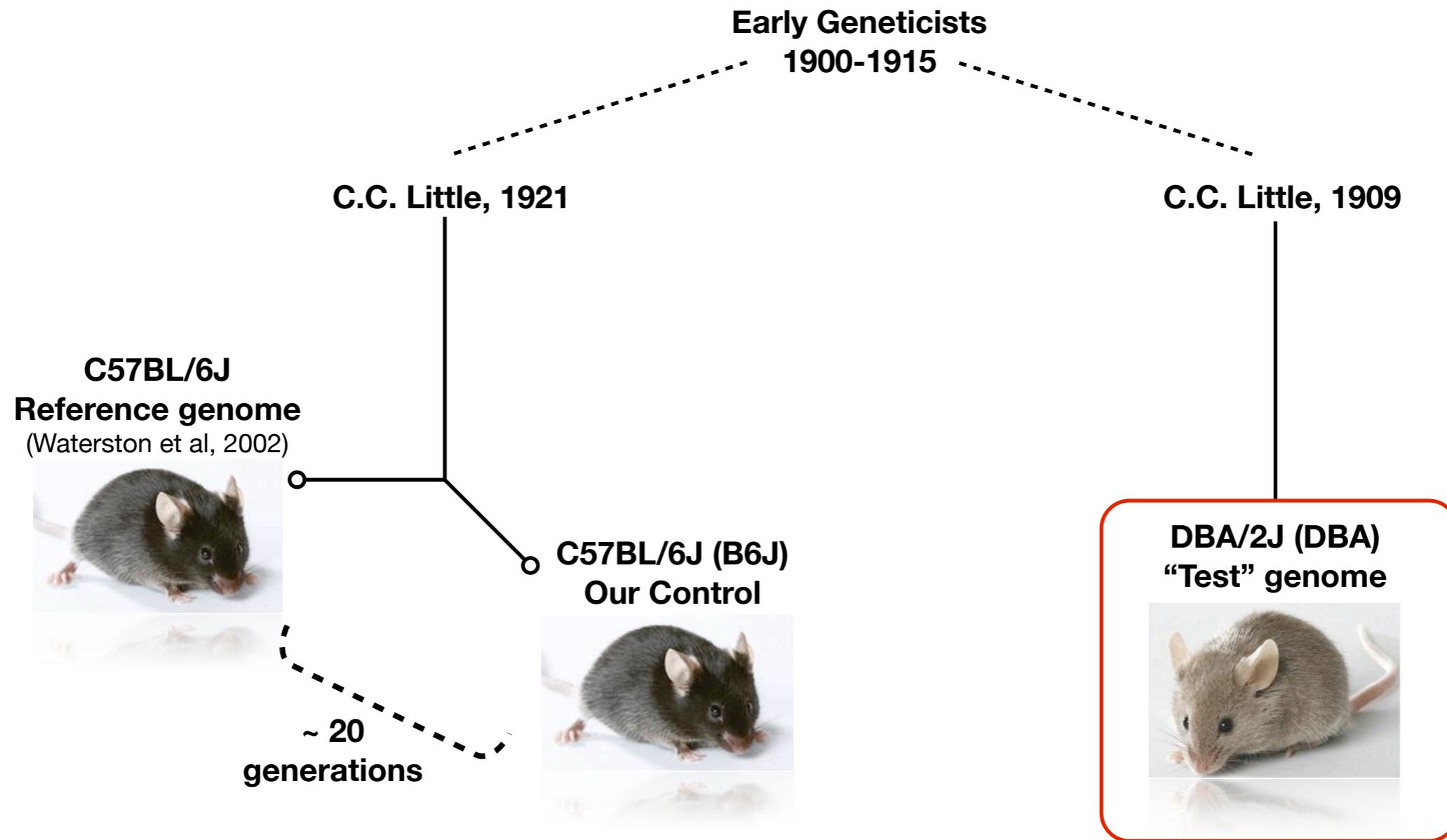
High

Low

Overview

- ▶ 1. SV in the laboratory mouse
- ▶ 2. Somatic SV in tumor genomes
- ▶ 3. The impact of ionizing radiation on single cells.
- ▶ 4. Functional SV from 1000s of “exomes”
- ▶ 5. What next?

1. A pilot study of SV in the laboratory mouse



Goals:

- (1) Unbiased look at all classes of SV.
- (2) How much SV is there? Hotspots?
- (3) Look at precise breakpoints...mechanism

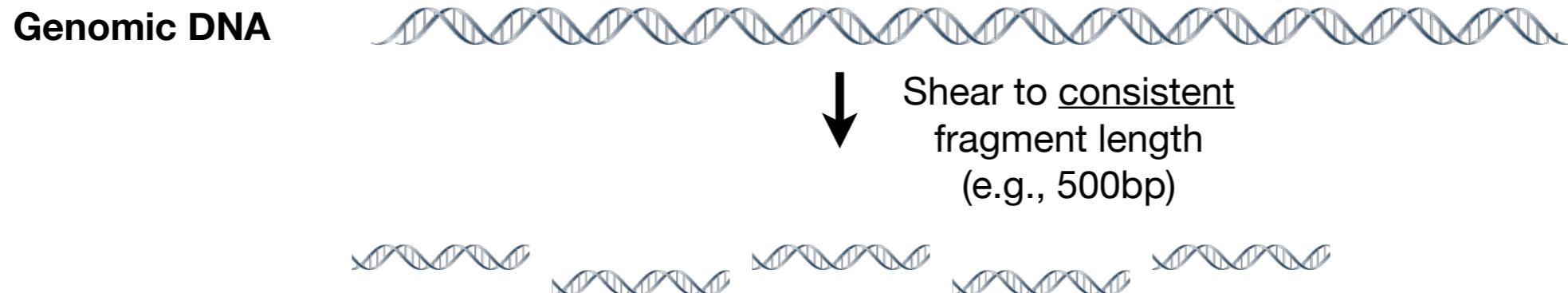
Quinlan et al., 2010. *Genome Research*
Postdoctoral work in Ira Hall's lab.

1. A pilot study of SV in the laboratory mouse.

Paired-end mapping

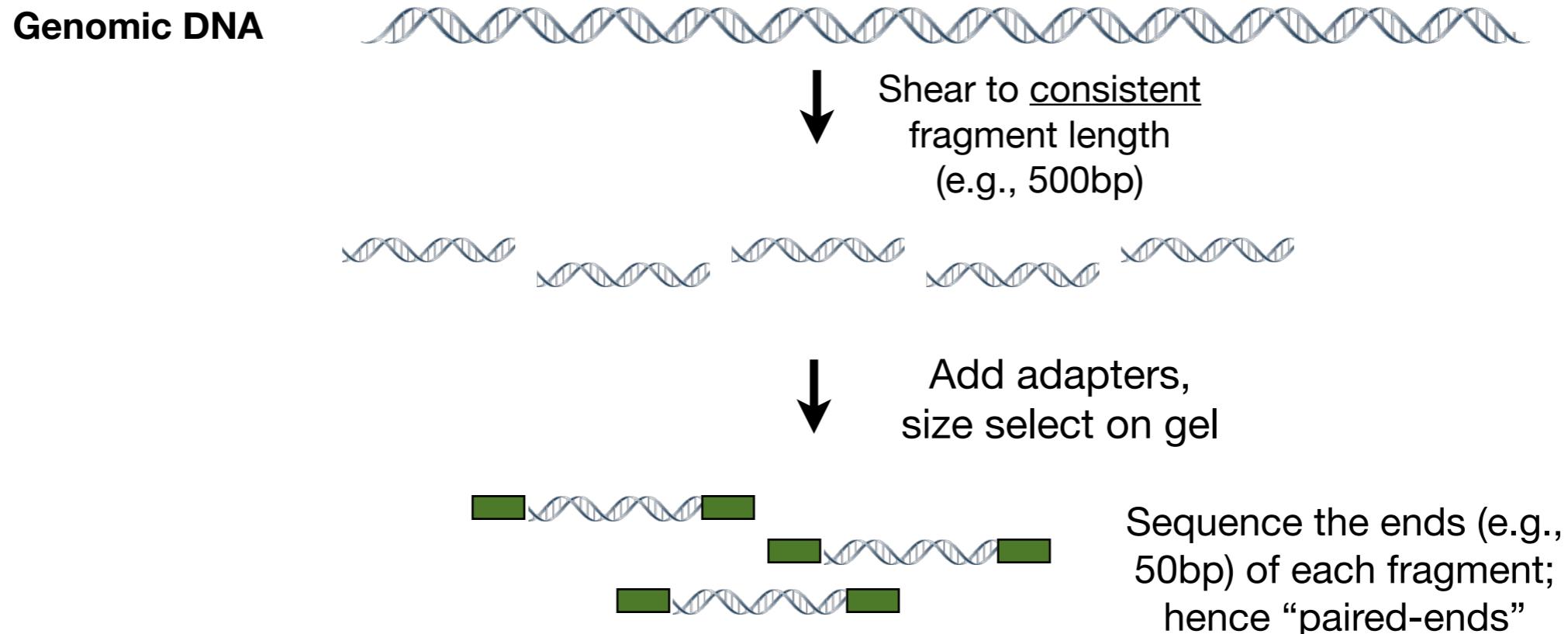
1. A pilot study of SV in the laboratory mouse.

Paired-end mapping



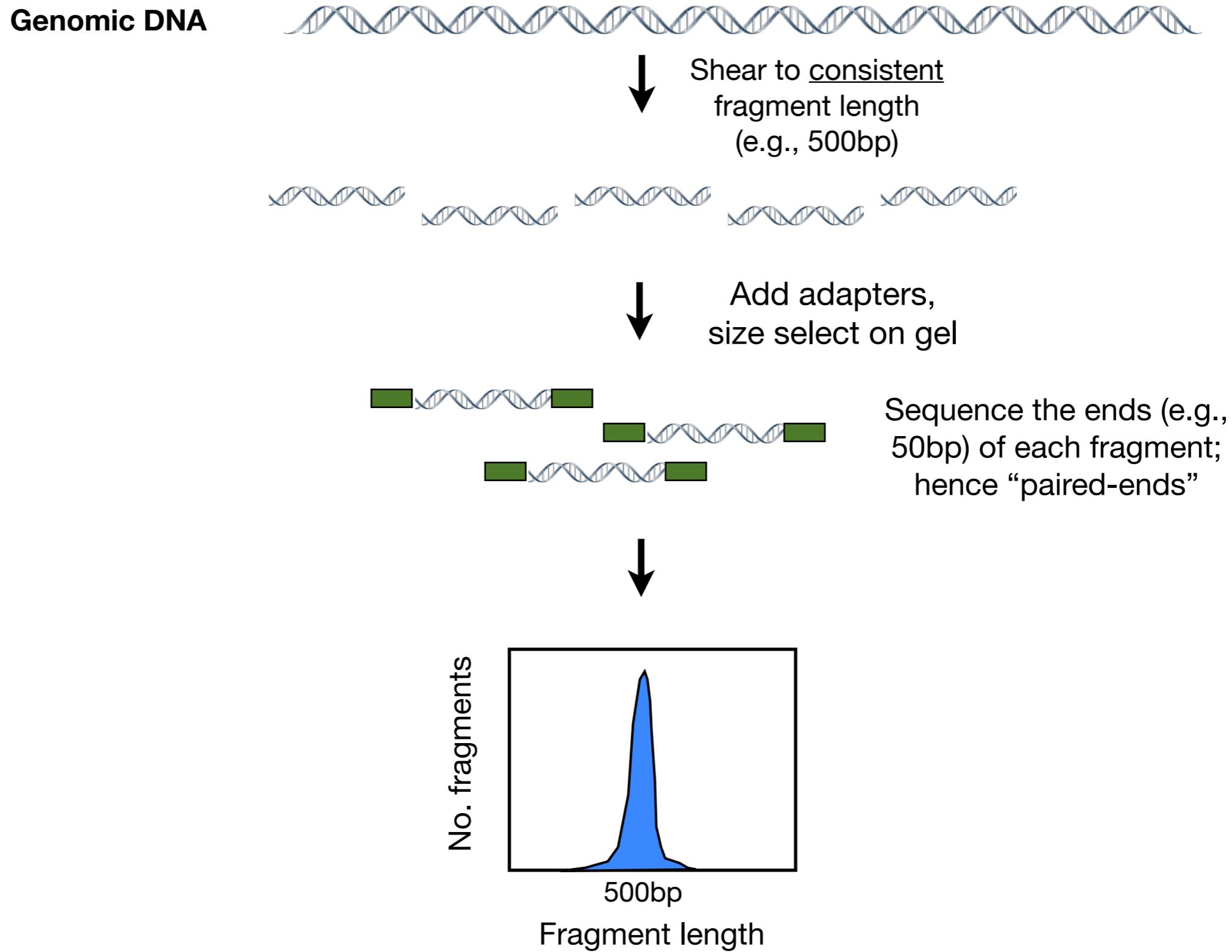
1. A pilot study of SV in the laboratory mouse.

Paired-end mapping



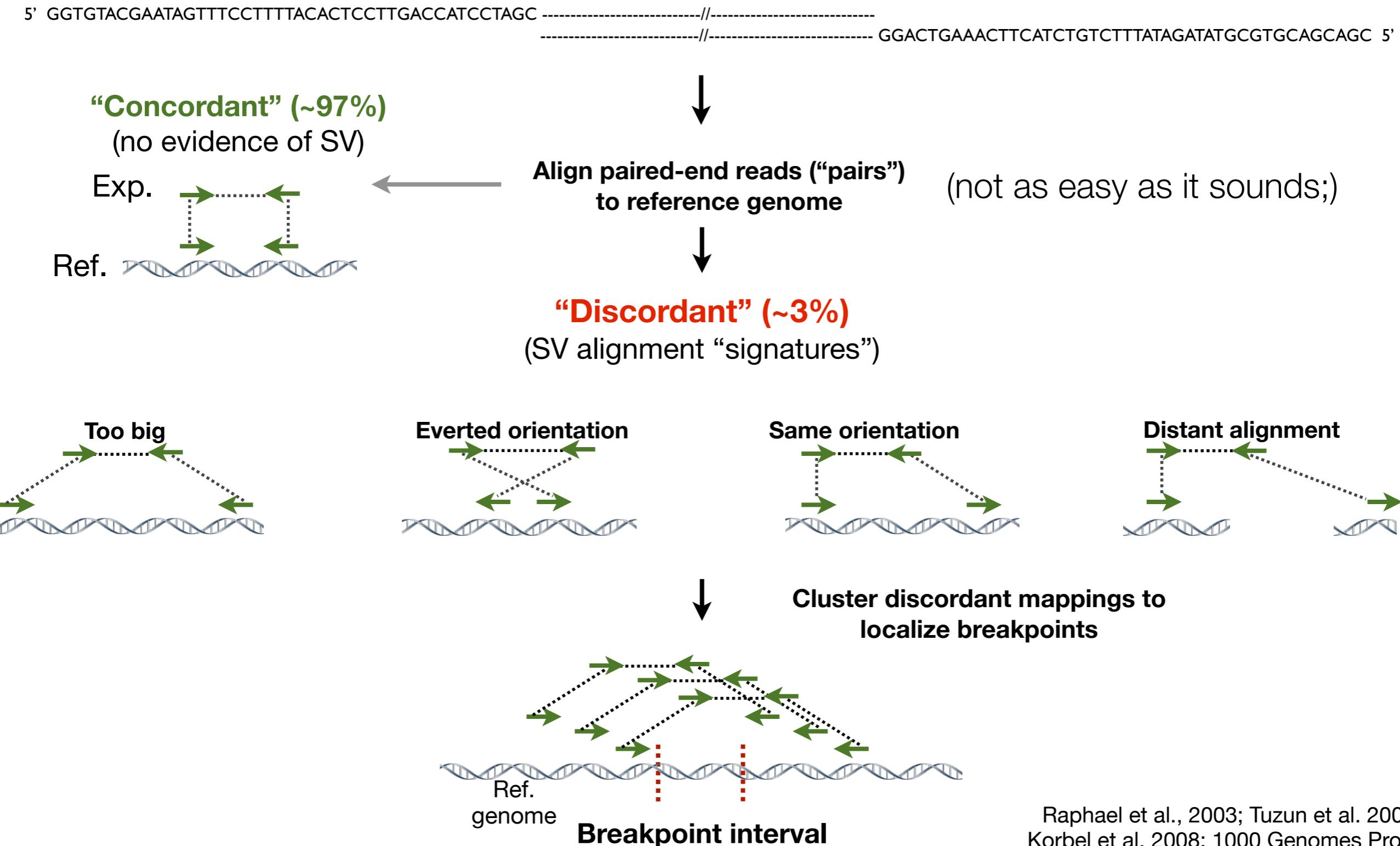
1. A pilot study of SV in the laboratory mouse.

Paired-end mapping



1. A pilot study of SV in the laboratory mouse.

Screen for clusters of “discordant” pairs



1. A pilot study of SV in the laboratory mouse.

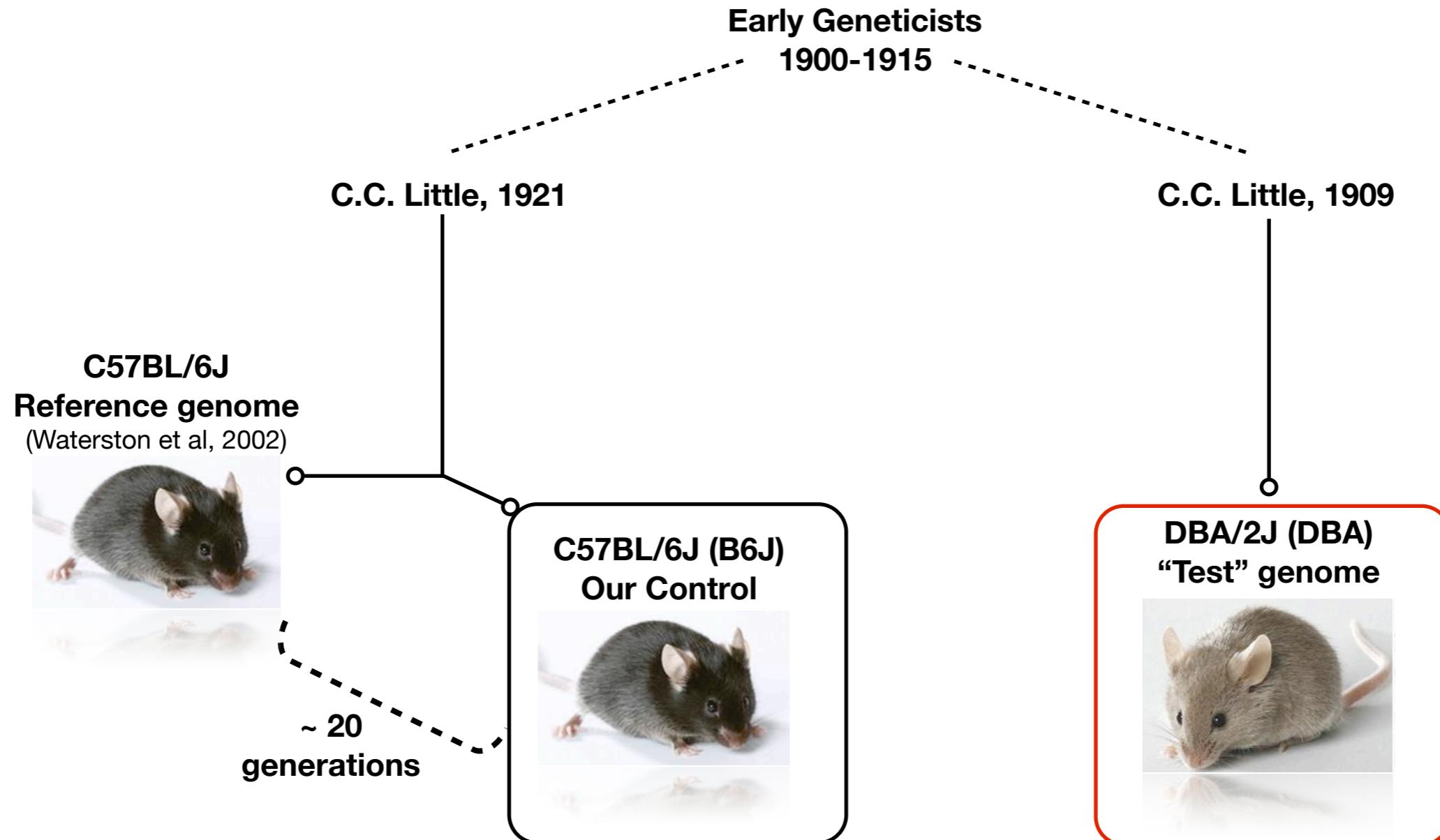
Hydra: software for SV discovery



<http://code.google.com/p/hydra-sv/>

1. A pilot study of SV in the laboratory mouse.

7196 SVs b/w C57BL/6J and DBA/2J



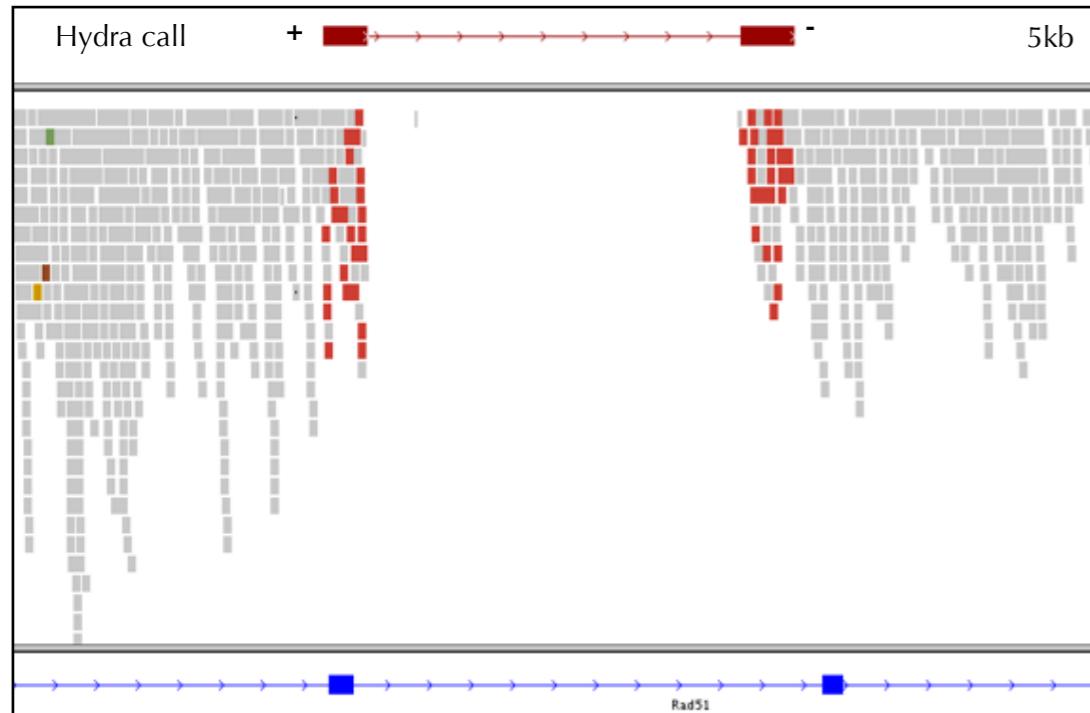
405 calls v. C57BL/6J Ref.
(assembly errors and missing DNA in Ref.)

7196 variants v. C57BL/6J Ref.
(~90% validation rate)

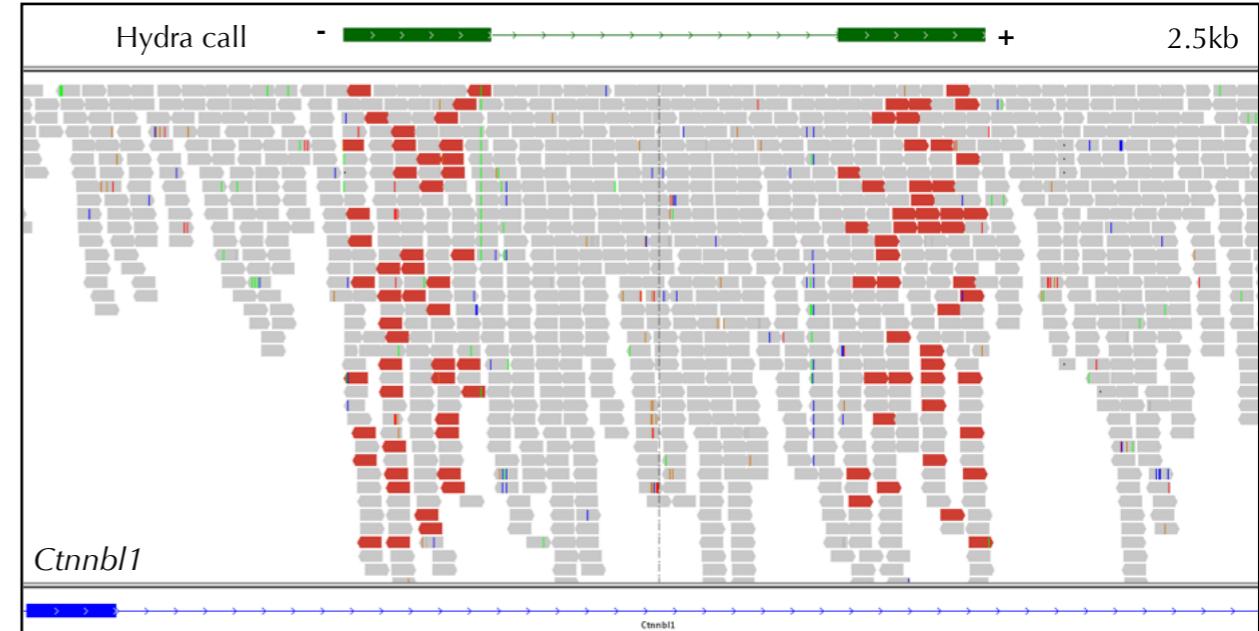
1. A pilot study of SV in the laboratory mouse.

What types of SV did we see?

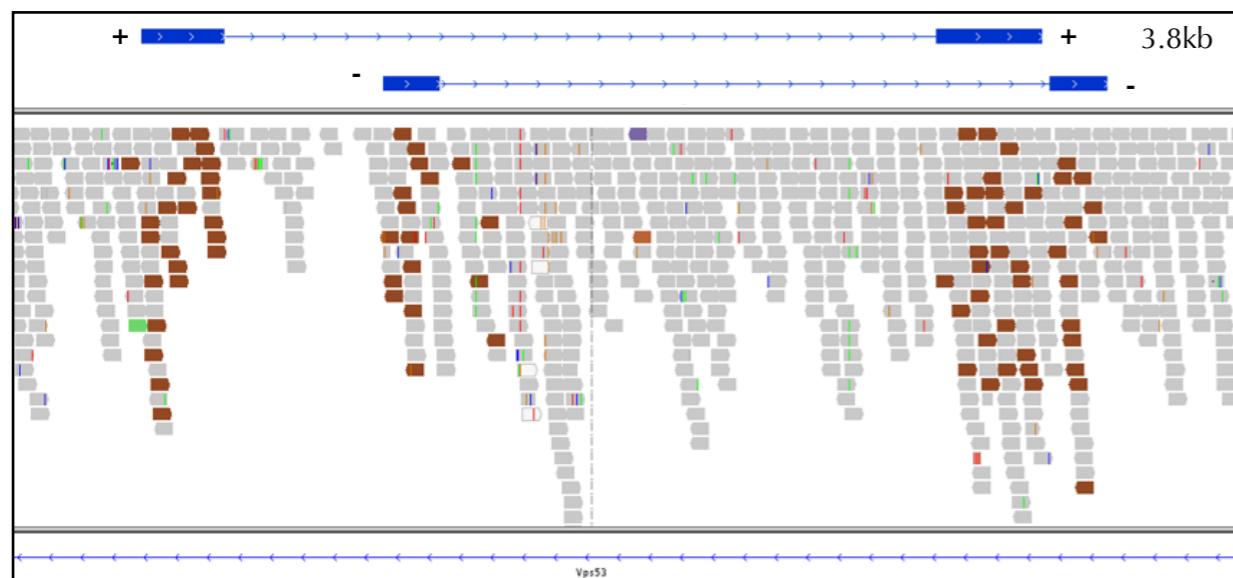
1285 deletions (18%)



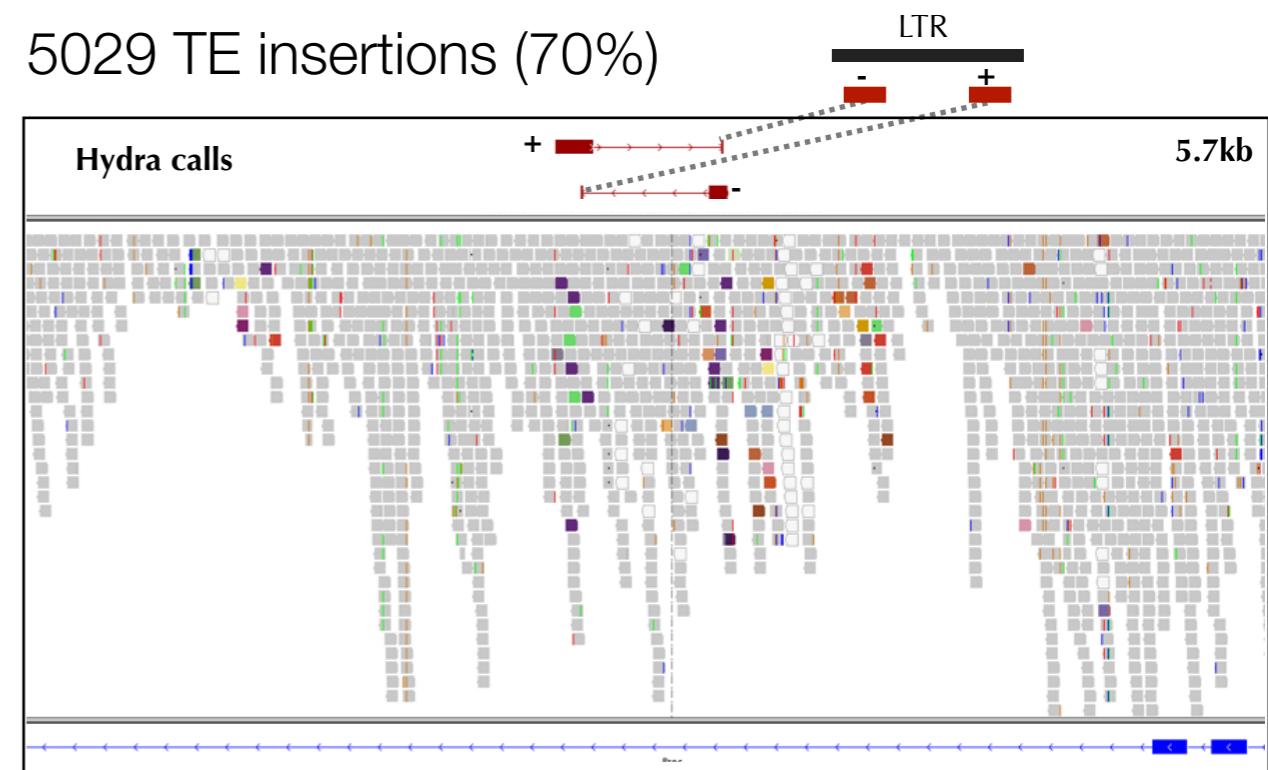
188 duplications (2.6%)



137 inversions (1.9%)

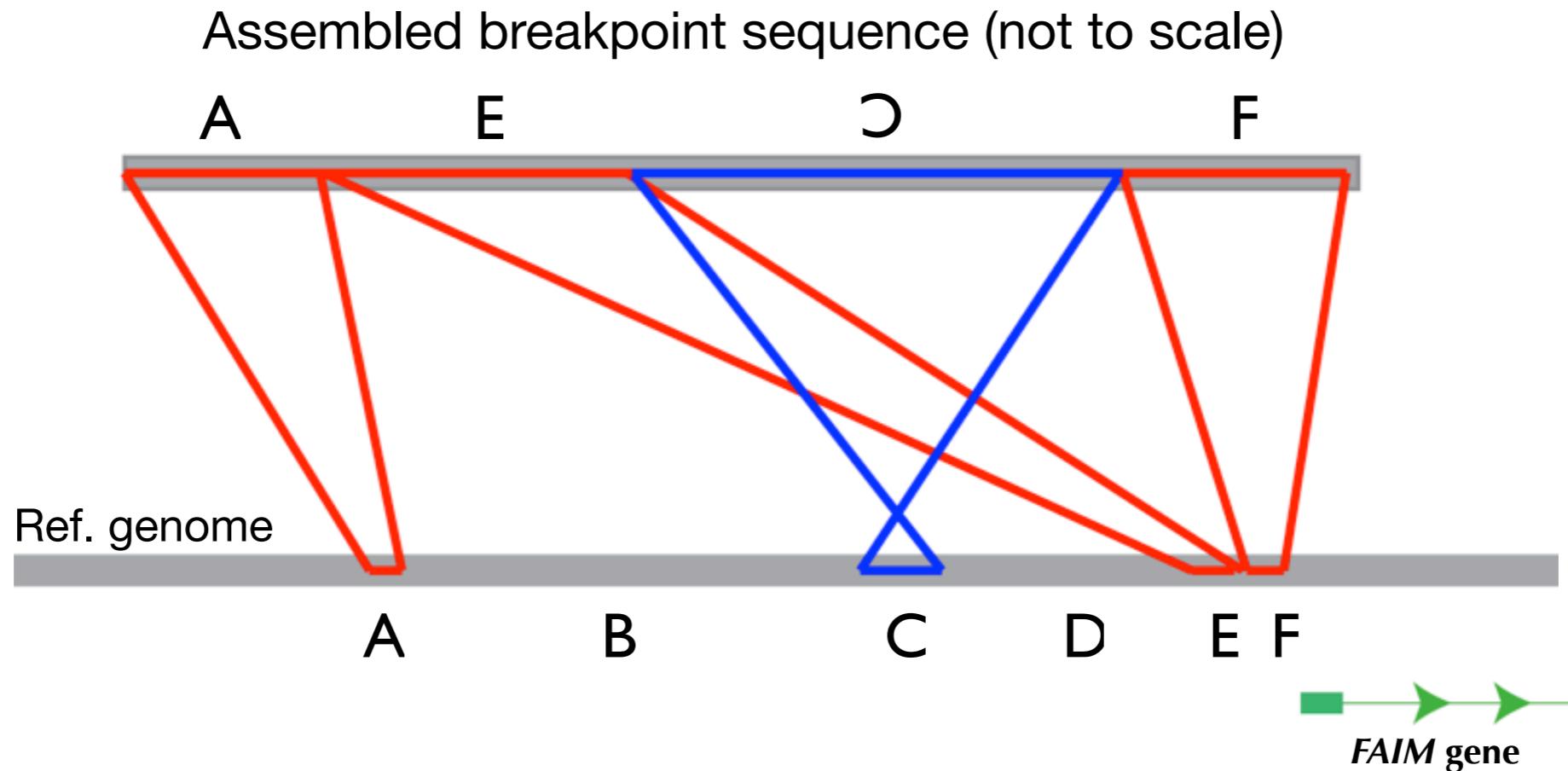


5029 TE insertions (70%)



1. A pilot study of SV in the laboratory mouse.

16% of loci have “complex” structures

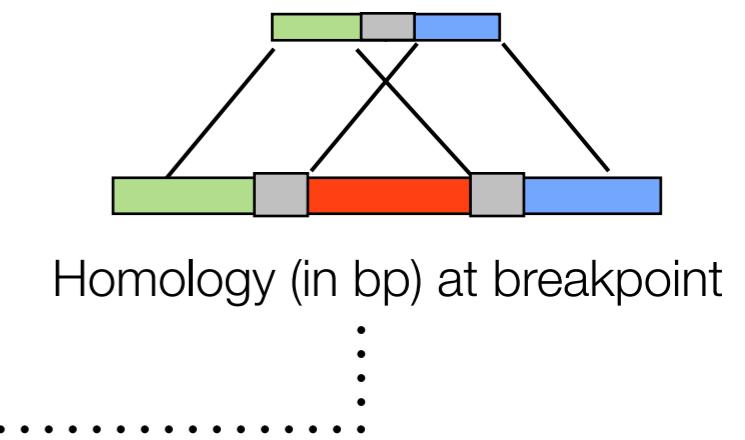
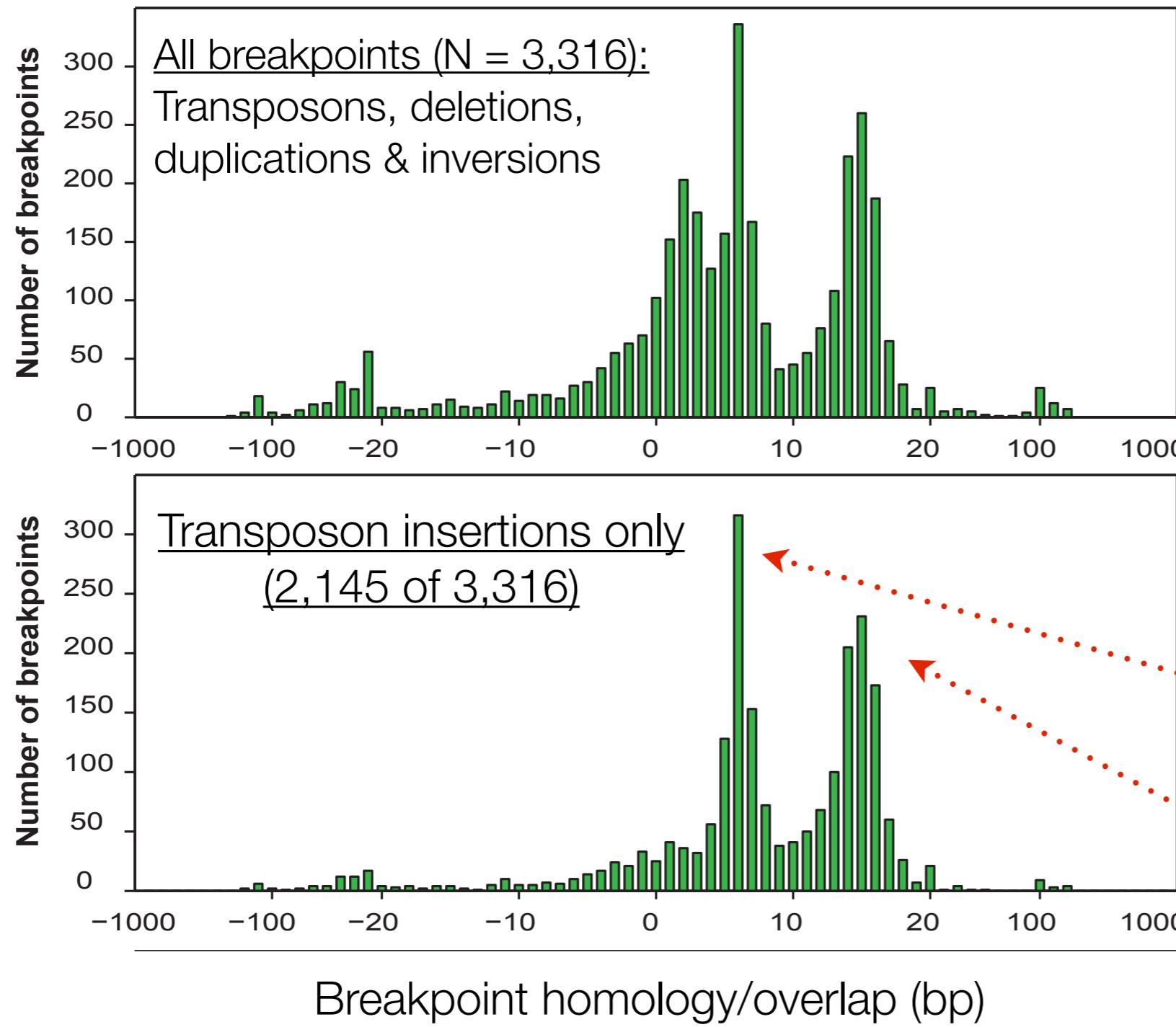


16% of non-mobile element SVs have complex structures:

- multiple intertwined breakpoints
- DNA inserted into the breakpoint from nearby or distant loci
- most likely arise from template switching during DNA replication; e.g., FoSTeS or MMBIR (Lupsik & Hastings, 2009)
- breakpoint patterns are not random; we see consistent patterns

1. A pilot study of SV in the laboratory mouse.

Breakpoint homology suggests mechanism



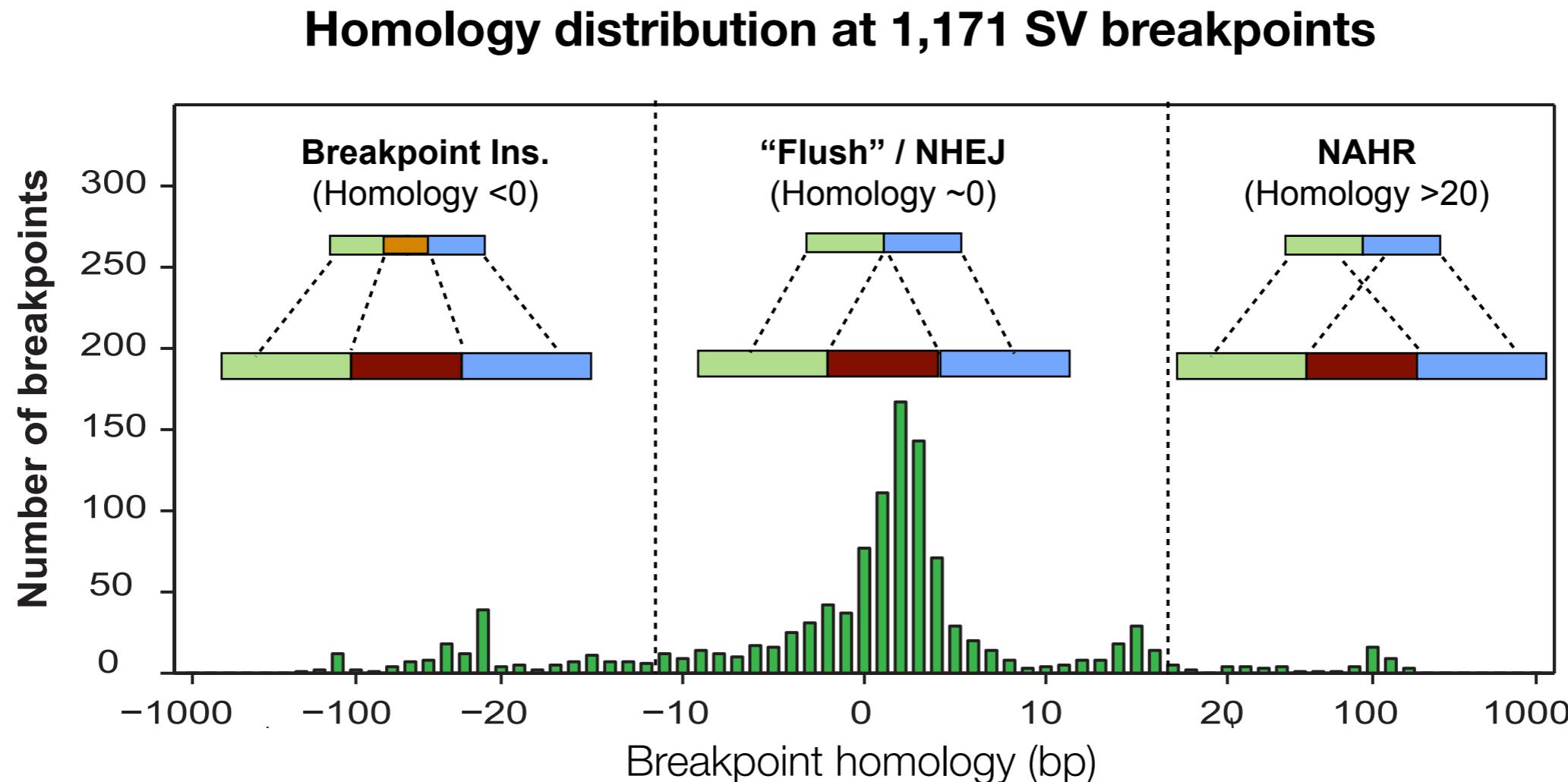
Homology (in bp) at breakpoint

LTR target-site duplications
are ~ 3-7bp

LINE target-site duplications
are ~ 8-20bp

1. A pilot study of SV in the laboratory mouse.

Most non-transposon SVs suggest NHEJ



Results:

- most have “flush” breakpoints = NHEJ or template switching
- NAHR is very rare (4%) - but also hard to see...
- small breakpoint insertions/rearrangements are common (10%)

1. A pilot study of SV in the laboratory mouse.

Summary

- ▶ Found 7200 accurate SVs across the entire mouse genome
- ▶ Able to detect complex genome rearrangements
- ▶ Whole-genome sequencing is a viable approach
- ▶ The Hydra software provides a robust framework for larger human studies
- ▶ The method works...on to human studies.

1a. Genome stability during cellular reprogramming



Cell Stem Cell
Short Article

Genome Sequencing of Mouse Induced Pluripotent Stem Cells Reveals Retroelement Stability and Infrequent DNA Rearrangement during Reprogramming

Aaron R. Quinlan,^{1,2,4} Michael J. Boland,^{3,4} Mitchell L. Leibowitz,¹ Svetlana Shumilina,¹ Sidney M. Pehrson,³ Kristin K. Baldwin,^{3,*} and Ira M. Hall^{1,2,*}

¹Department of Biochemistry and Molecular Genetics

²Center for Public Health Genomics

University of Virginia, Charlottesville, VA 22908, USA

³Department of Cell Biology, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA 92037, USA

⁴These authors contributed equally to this work

*Correspondence: kbaldwin@scripps.edu (K.K.B.), irahall@virginia.edu (I.M.H.)

DOI 10.1016/j.stem.2011.07.018

Next:

- Deriving 30-50 of human iPSC lines with different reprogramming methods
- Whole genome sequencing to measure genome stability

2. Somatic rearrangements in tumor genomes

(or, UVA's 68.6% of 1000 Genomes project)

Goal: accurate detection of somatically acquired rearrangements in cancer genomes.

THE CANCER GENOME ATLAS		
Cancer Type	Tumors	Matched normals
Glioblastoma (GBM)	32	32
Ovarian (OV)	24	24
Acute Myeloid Leukemia (LAML)	52	52
Total	108	108

216 genomes
(all datasets as of 9/2010)

16.89 terabytes
93.8 billion pairs



686 genomes
>20 terabytes
>100 billion read pairs

Population	No. samples used
African-American SW (ASW)	24
Utah with European descent (CEU)	12
Han Chinese (CHB)	22
Southern Han Chinese (CHS)	63
Columbian (CLM)	37
Finnish (FIN)	55
British (GBR)	51
Japanese (JPT)	46
Luhya (LWK)	66
Mexican-American (MXL)	28
Puerto-Rican (PUR)	53
Tuscan (TSI)	1
Yoruba (YRI)	12
Total	470

470 “low-coverage” genomes (min. 1X)
(8/2010 release, >=10Gb BAM)

5.93 terabytes
30.6 billion pairs



Ongoing collaboration with Ira Hall

2. Somatic rearrangements in tumor genomes

Why do we need the 1000 genomes samples if we already have matched normals?

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

		Samples						
		Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Y							
	N							

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

	Samples						
	Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Y	N	N	Y	N	N	N

Inherited, insufficient data in Normal1

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

		Samples						
		Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Y	N	N	Y	N	N	N	
	Y	N						Inherited, insufficient data in Normal1

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

		Samples						
		Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Y	N	N	Y	N	N	N	
	Y	N	N	Y	Y	Y	Y	Inherited, insufficient data in Normal1
	Y	N	N	N	Y	Y	Y	Inherited, found only in 1000G controls

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

		Samples						
		Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Inherited, insufficient data in Normal1	Y	N	N	Y	N	N	N
	Inherited, found only in 1000G controls	Y	N	N	N	Y	Y	Y
	Specific to a <i>single</i> tumor	Y	N	N	N	N	N	N

2. Somatic rearrangement in tumor genomes

Matrix to exclude “cryptic” germline variants.

		Samples						
		Tumor1	Normal1	Tumor2	Normal2	1000G (1)	1000G (2)	1000G (3)
Rearrangements	Y	N	N	Y	N	N	N	Inherited, insufficient data in Normal1
	Y	N	N	N	Y	Y	Y	Inherited, found only in 1000G controls
	Y	N	N	N	N	N	N	Specific to a <i>single</i> tumor
	Y	N	Y	N	N	N	N	Specific to <i>multiple</i> tumors

2. Somatic rearrangement in tumor genomes

Technical challenges

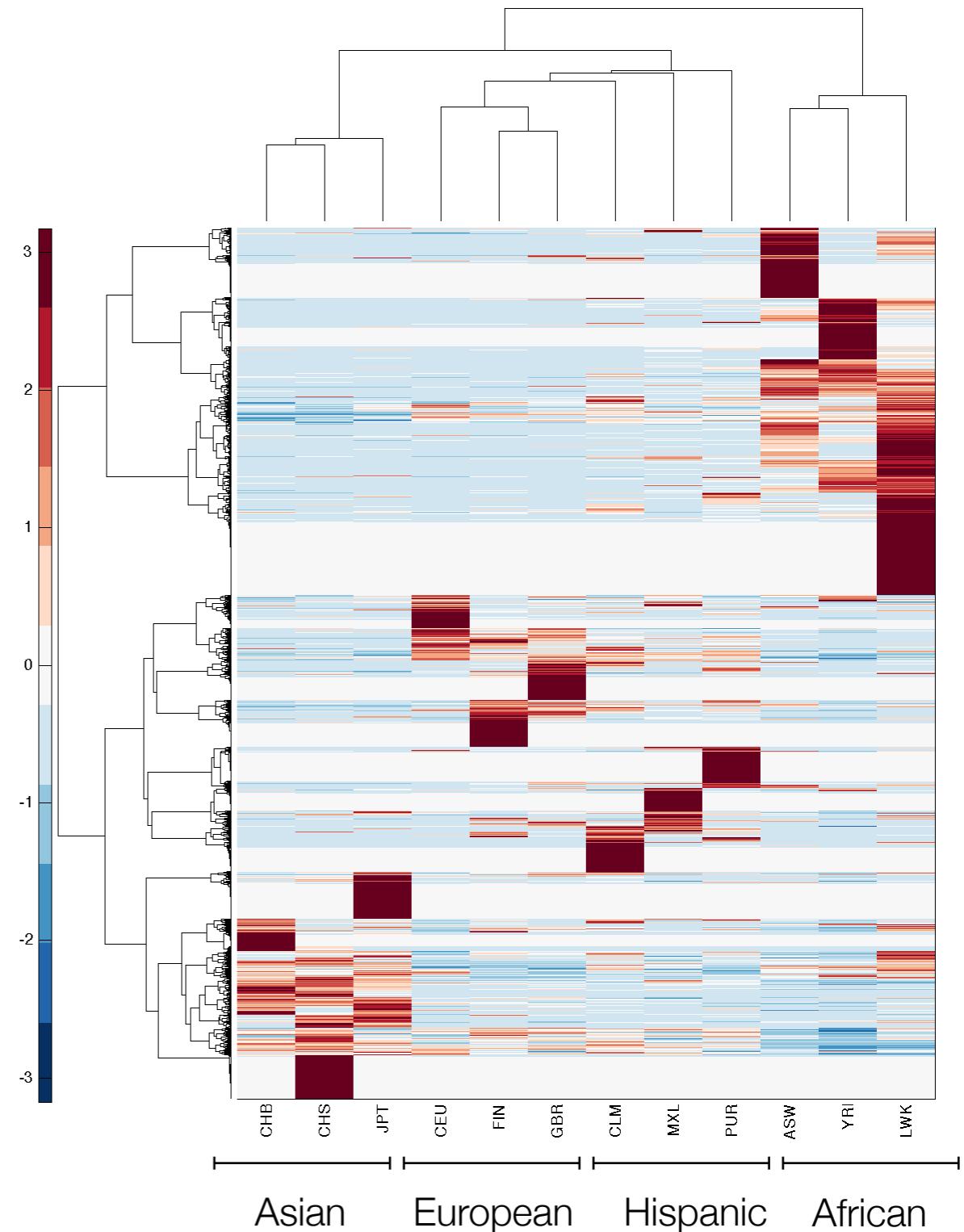
- ▶ Enormous amount of data. A big “compute”
- ▶ Needed to reengineer Hydra (the SV discovery tool)
- ▶ You can’t just download from TCGA...
- ▶ Account for diverse molecular and analytical strategies.

2. Somatic rearrangement in tumor genomes

Breakpoint patterns reflect human demography

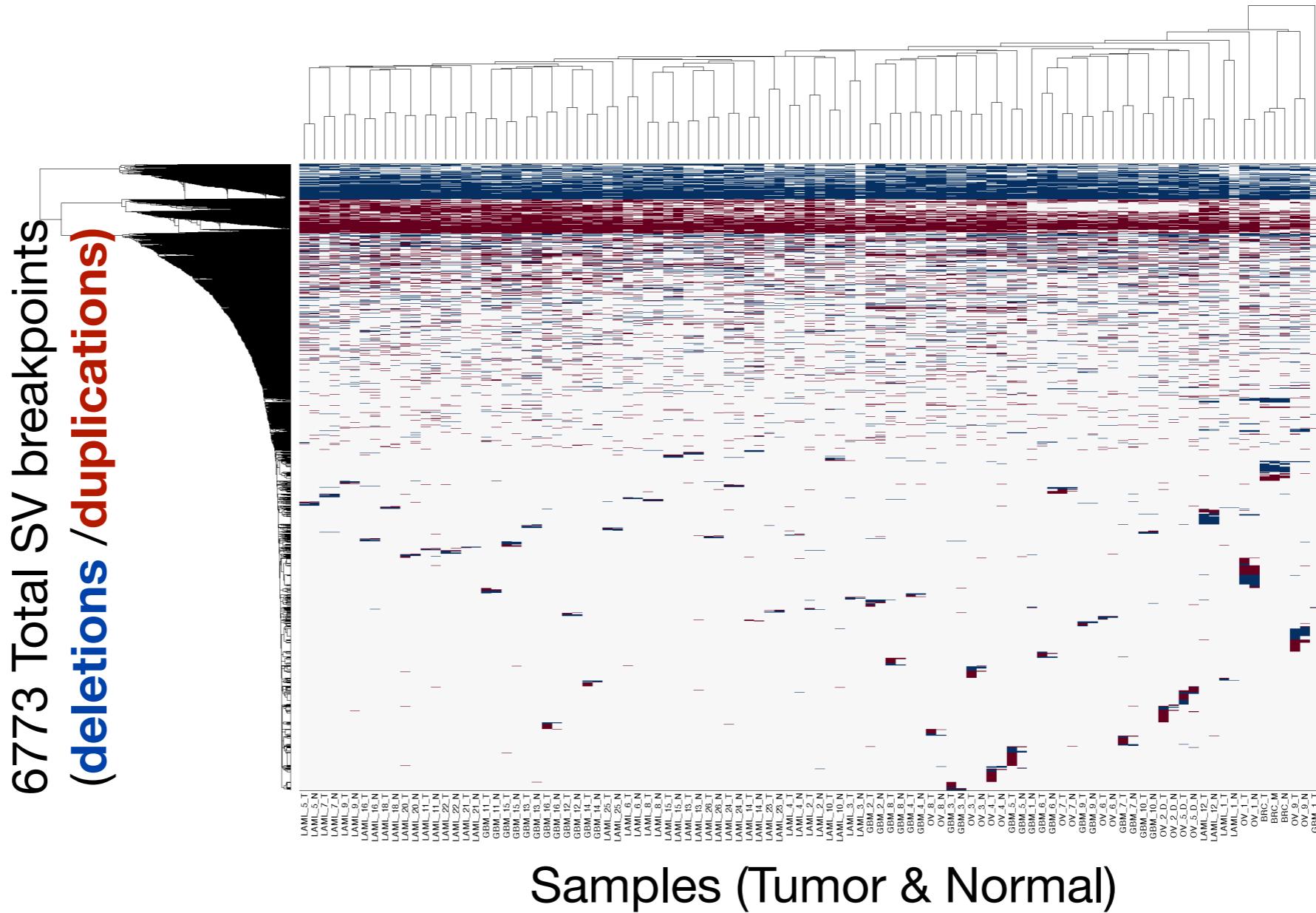
1000 Genomes samples

- mean allele frequency in each of 12 populations
- normalized per row (z-score)
- 7796 breakpoints



2. Somatic rearrangement in tumor genomes

Accurate tumor-specific rearrangements

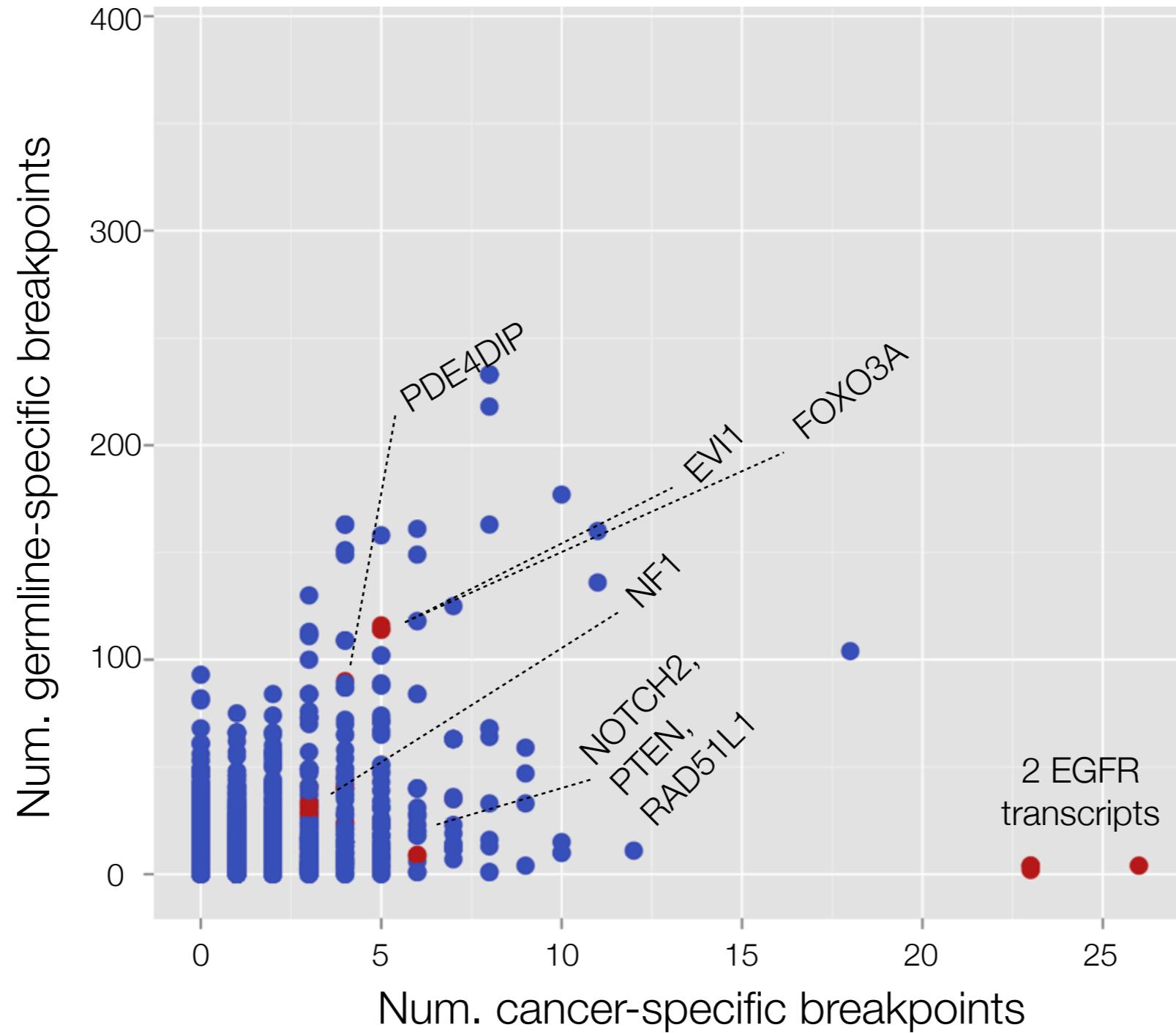


Results:

- 1) All but one tumor/normal pair cluster together
- 2) **1,462** breakpoints private to single tumor sample vs. **only 48** to a single matched normal sample
- 3) False positive somatic calls are **reduced >10 fold** due to the presence of many normal genomes

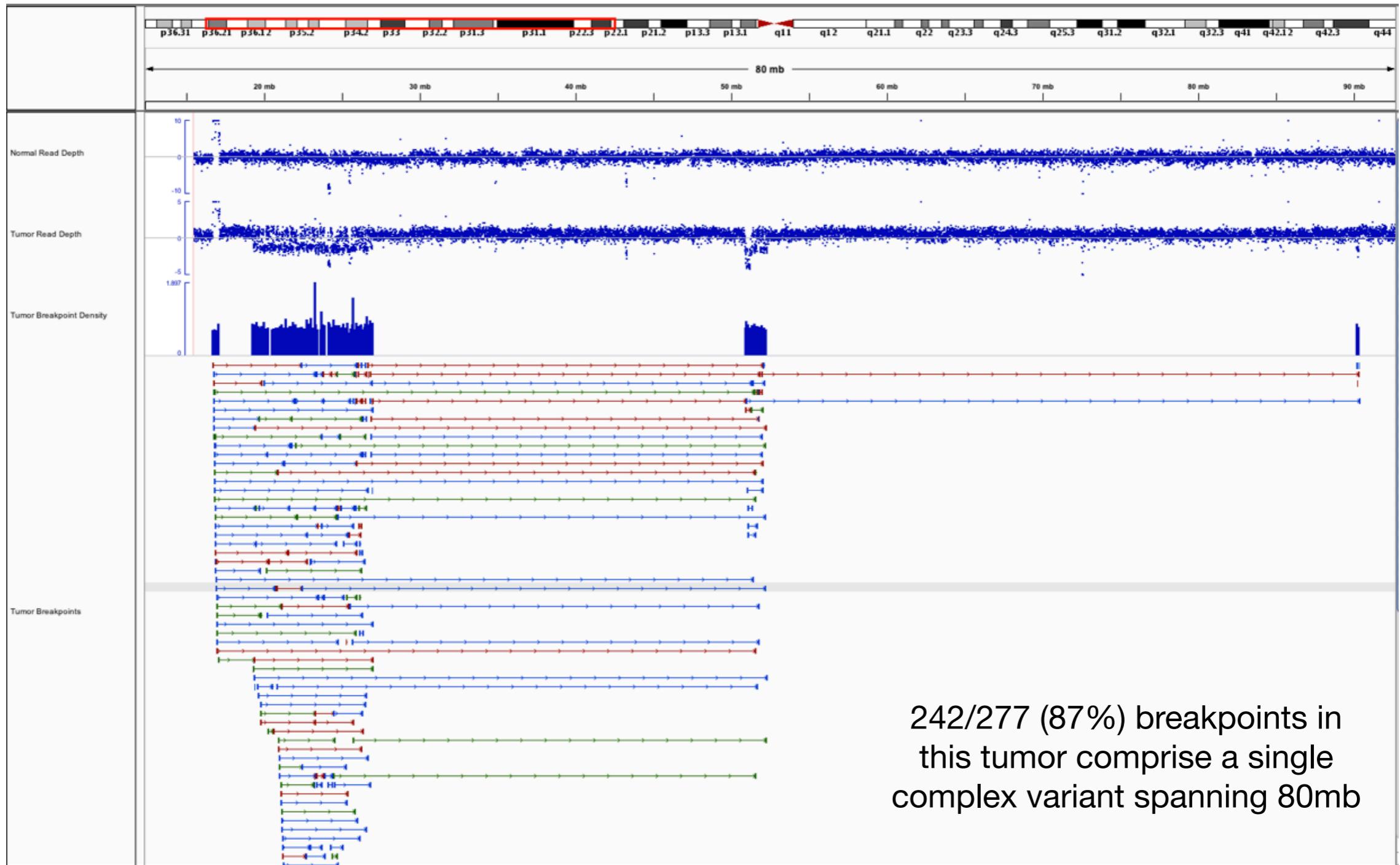
2. Somatic rearrangement in tumor genomes

Usual suspects and new candidates (DRAFT)



2. Somatic rearrangement in tumor genomes

Incredibly complex chromosomal rearrangement



How common is this? What is the cause? chromothripsis, replication, or something else?

2. Somatic rearrangement in tumor genomes

Chromothripsis (chromosome shattering)

Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development

Philip J. Stephens,¹ Chris D. Greenman,¹ Beiyuan Fu,¹ Fengtang Yang,¹ Graham R. Bignell,¹ Laura J. Mudie,¹ Erin D. Pleasance,¹ King Wai Lau,¹ David Beare,¹ Lucy A. Stebbings,¹ Stuart McLaren,¹ Meng-Lay Lin,¹ David J. McBride,¹

Chromothripsis as a mechanism driving complex *de novo* structural rearrangements in the germline[†]

Wigard P. Kloosterman¹, Victor Guryev², Mark van Roosmalen¹, Karen J. Duran¹,
Ewart de Bruijn², Saskia C.M. Bakker³, Tom Letteboer¹, Bernadette van Nesselrooij¹,
Ron Hochstenbach¹, Martin Poot¹ and Edwin Cuppen^{1,2,*}

Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer

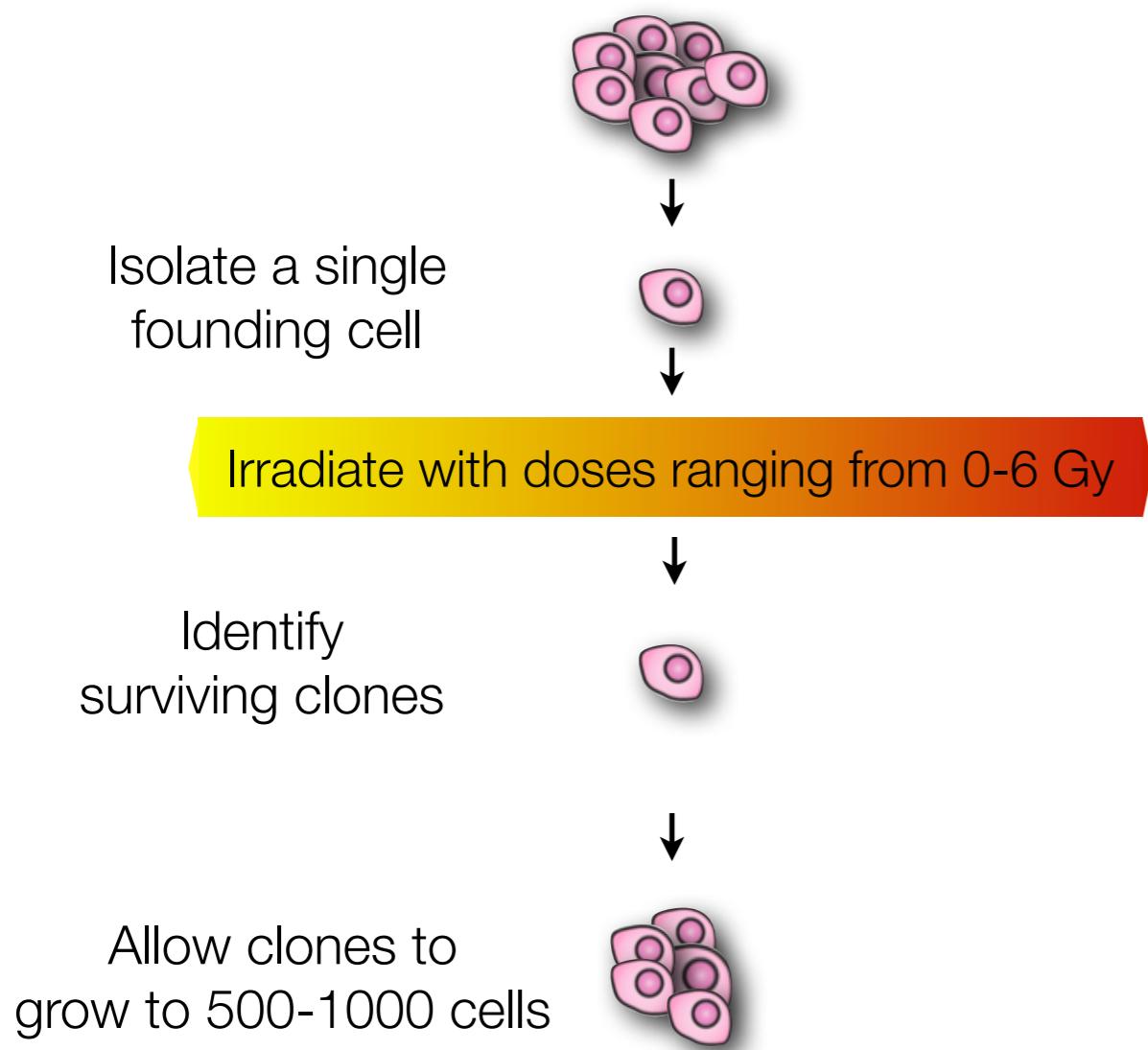
Wigard P Kloosterman¹, Marlous Hoogstraat^{1,2}, Oscar Paling², Masoumeh Tavakoli-Yaraki¹, Ivo Renkens¹,
Joost S Vermaat², Markus J van Roosmalen¹, Stef van Lieshout^{1,2}, Isaac J Nijman³, Wijnand Roessingh²,
Ruben van 't Slot¹, José van de Belt¹, Victor Guryev³, Marco Koudijs², Emile Voest² and Edwin Cuppen^{1,3*}

2. Somatic rearrangement in tumor genomes

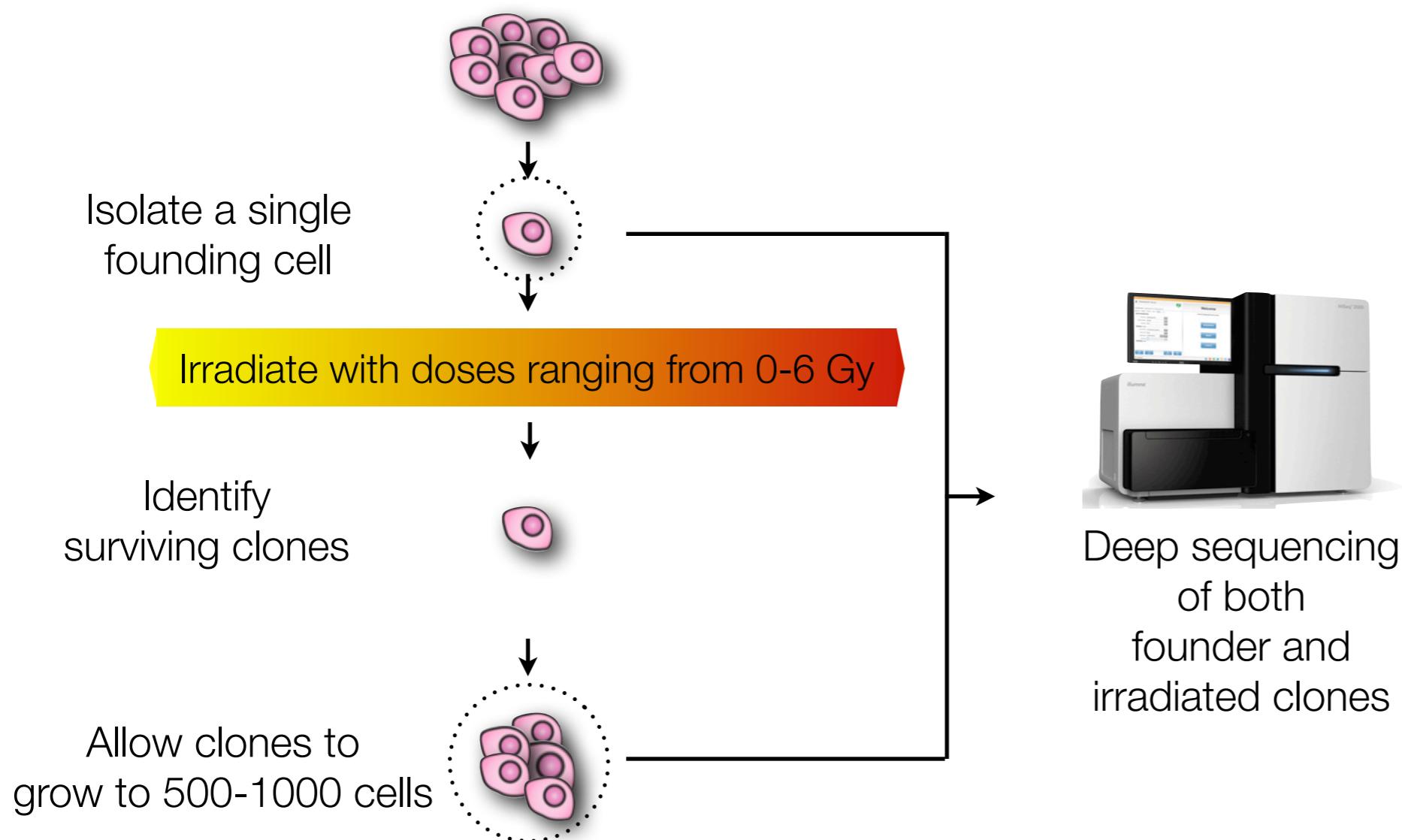
Summary

- ▶ Developed a novel framework for exploring SV in hundreds of genomes at once.
- ▶ Accurate detection of somatic SV in tumors requires more than just a single matched normal.
- ▶ Catching known oncogenes for cancers studied and observing new genes enriched for somatic genome instability
- ▶ Detecting evidence of chromothripsis.

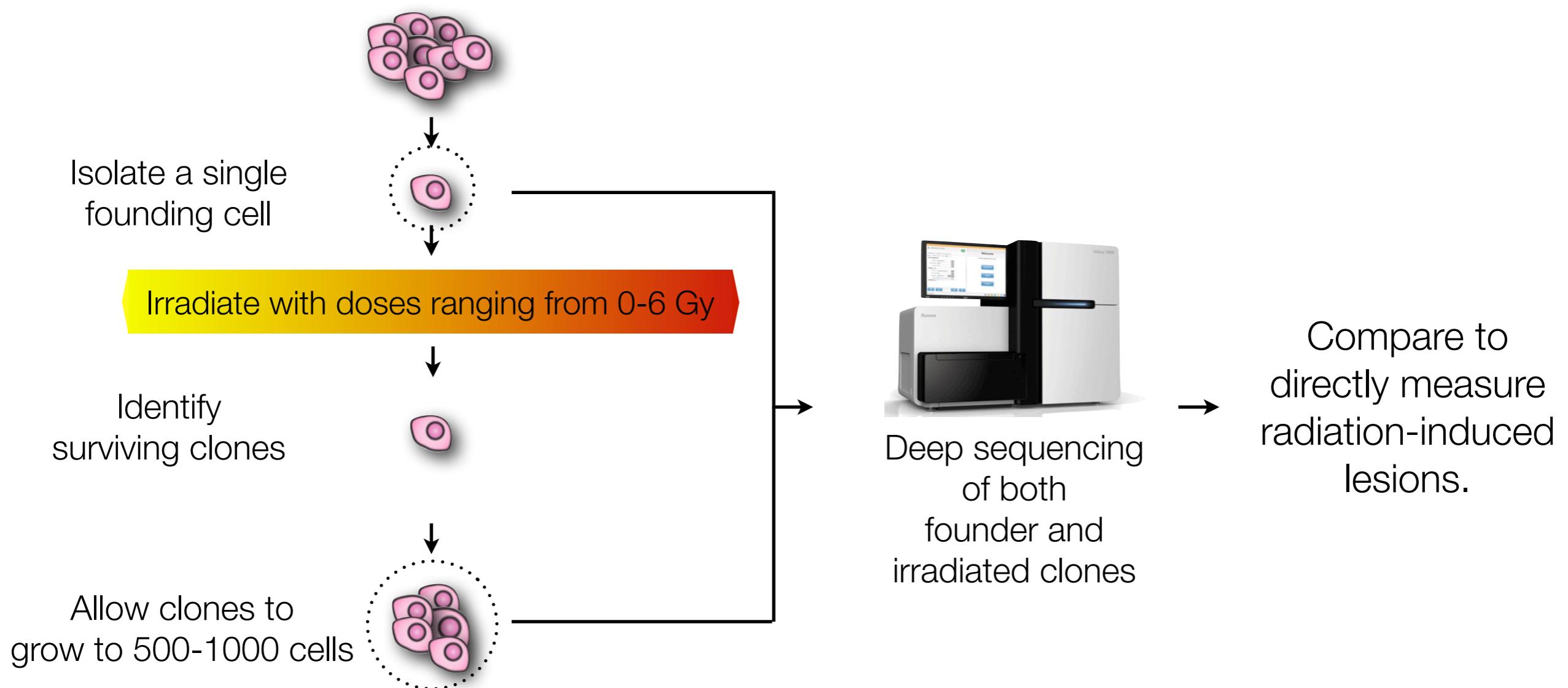
3. Assessing radiation damage in single cells



3. Assessing radiation damage in single cells



3. Assessing radiation damage in single cells



4. Rare SV discovery from 3000+ “exomes”

ex·ome (*noun*) the exonic 1% of genome; that is, the “business end”

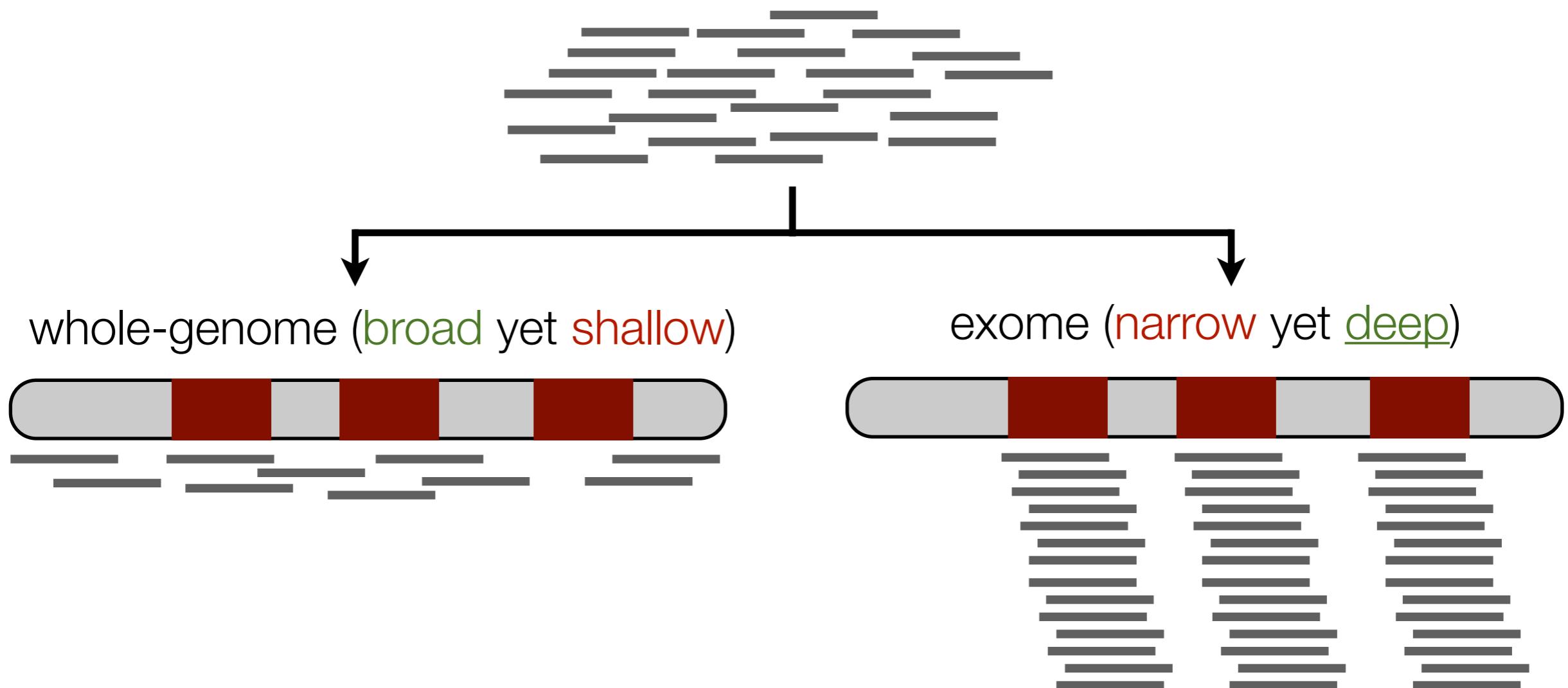
The idea: whole genome sequencing is cheap, but not cheap enough

4. Rare SV discovery from 3000+ “exomes”

ex·ome (*noun*) the exonic 1% of genome; that is, the “business end”

The idea: whole genome sequencing is cheap, but not cheap enough

e.g., fixed budget of 100 million DNA fragments



4. Rare SV discovery from 3000+ “exomes”

The typical human exome has:

~20,000 variants (50/50 syn/non-syn)

~200 per exome have never been seen (rare variants)

4. Rare SV discovery from 3000+ “exomes”

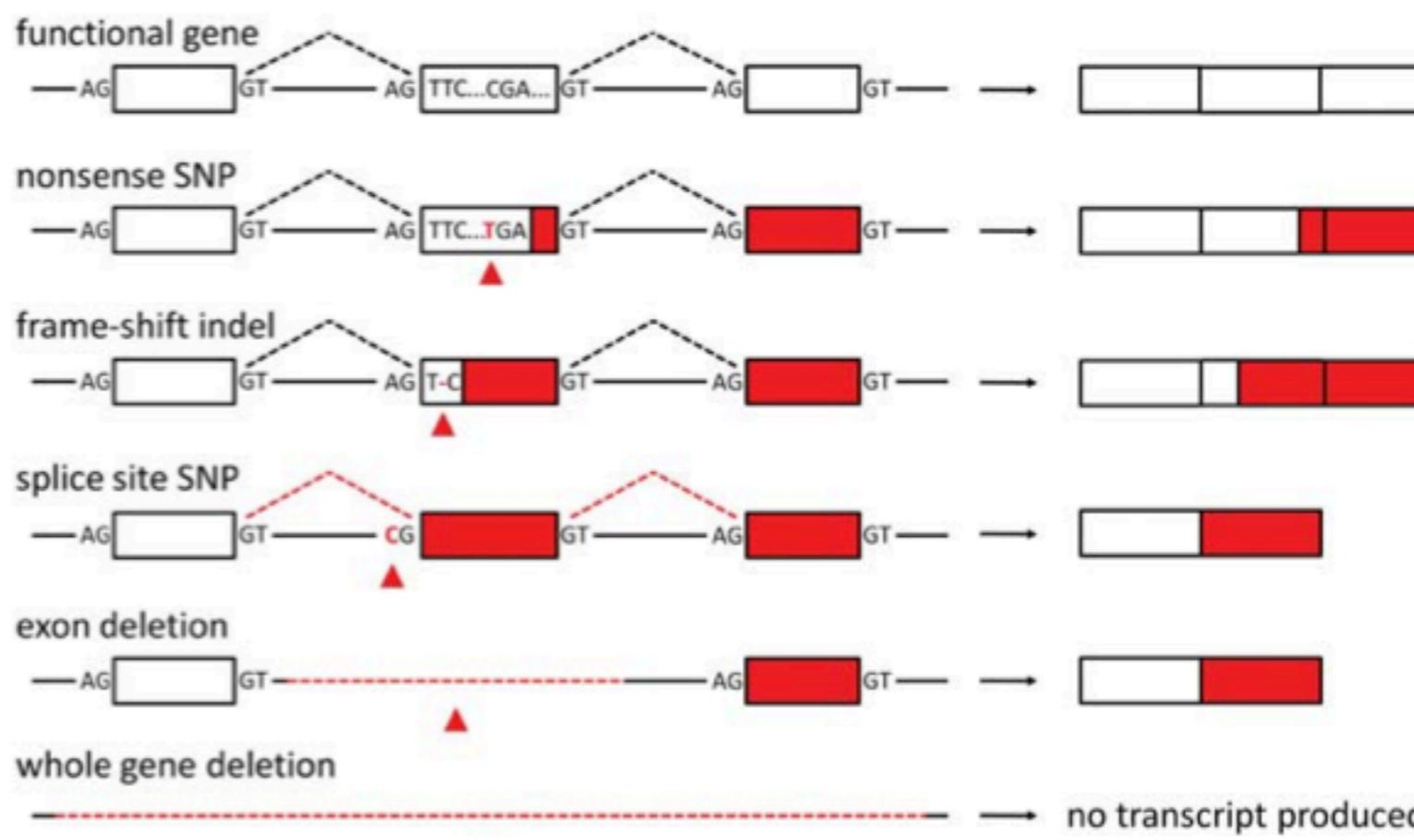
The typical human exome has:

~20,000 variants (50/50 syn/non-syn)

~200 per exome have never been seen (rare variants)

>=120 loss-of-function variants

SNP/INDEL
SV



~60 / exome

~40 / exome

~20 / exome

4. Rare SV discovery from 3000+ “exomes”

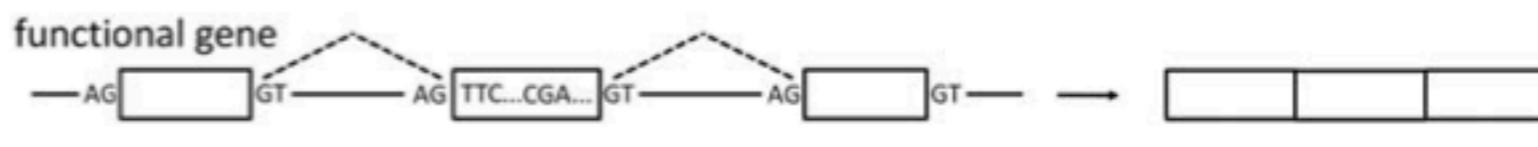
The typical human exome has:

~20,000 variants (50/50 syn/non-syn)

~200 per exome have never been seen (rare variants)

>=120 loss-of-function variants

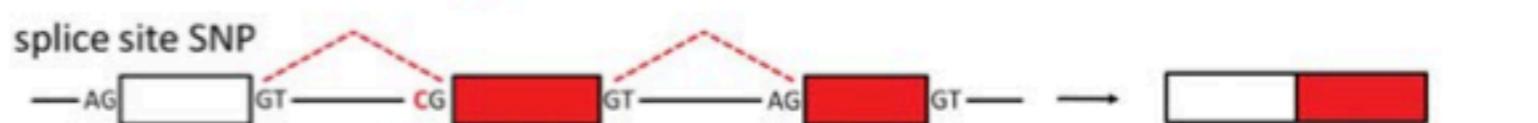
SNP/INDEL



~60 / exome



~40 / exome



~20 / exome

SV

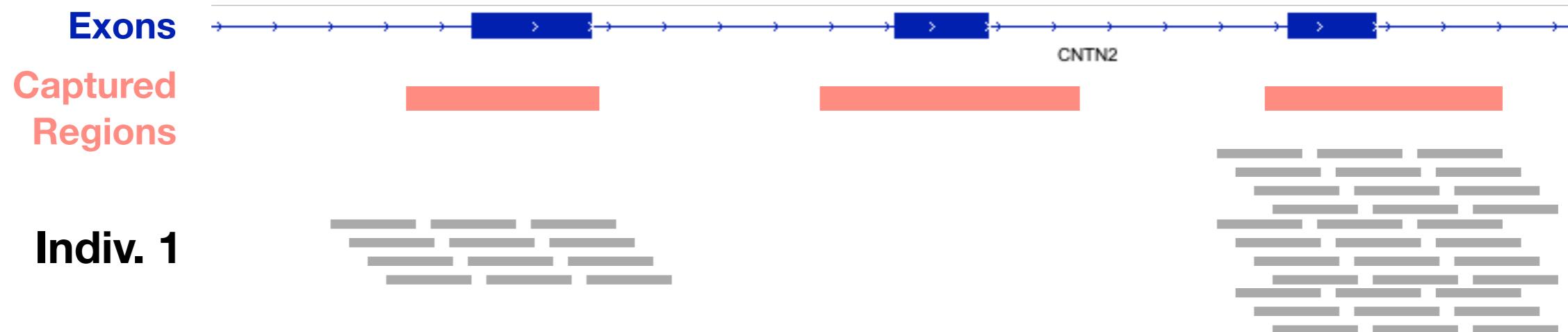


mostly ignored



4. Rare SV discovery from 3000+ “exomes”

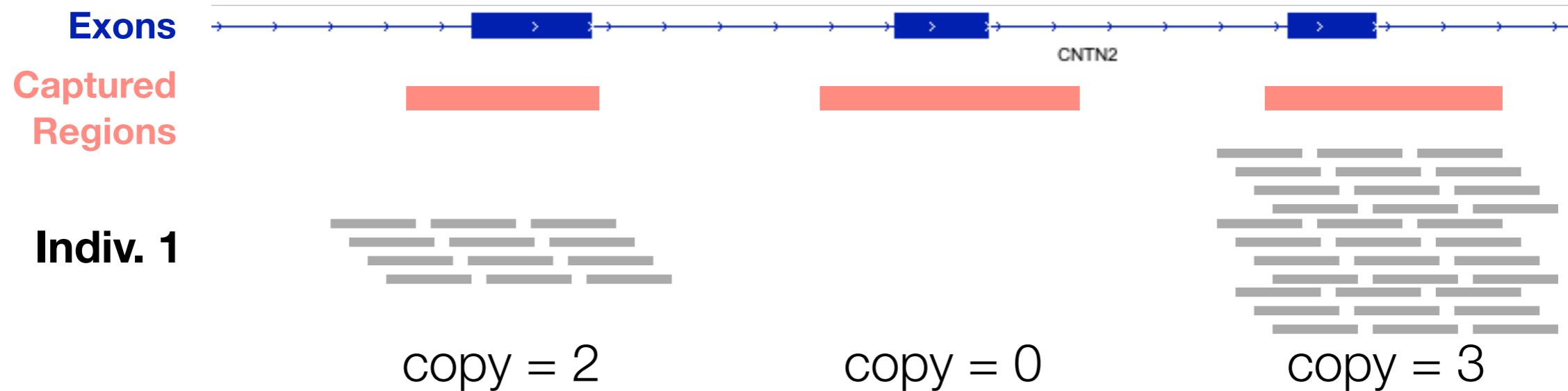
How do we detect SV among exomes?



Basic principle: Sequence depth indicates copy-number

4. Rare SV discovery from 3000+ “exomes”

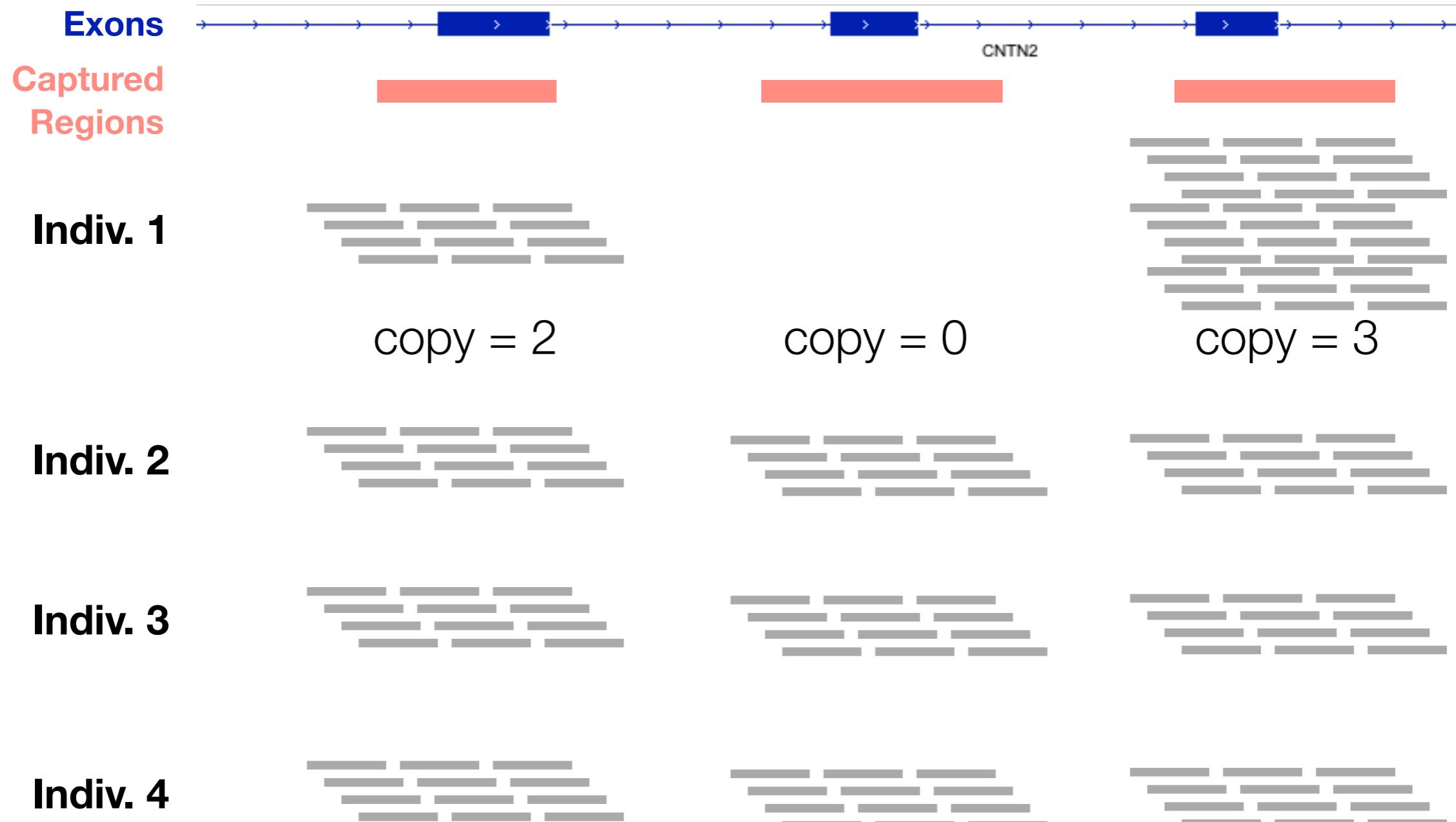
How do we detect SV among exomes?



Basic principle: Sequence depth indicates copy-number

4. Rare SV discovery from 3000+ “exomes”

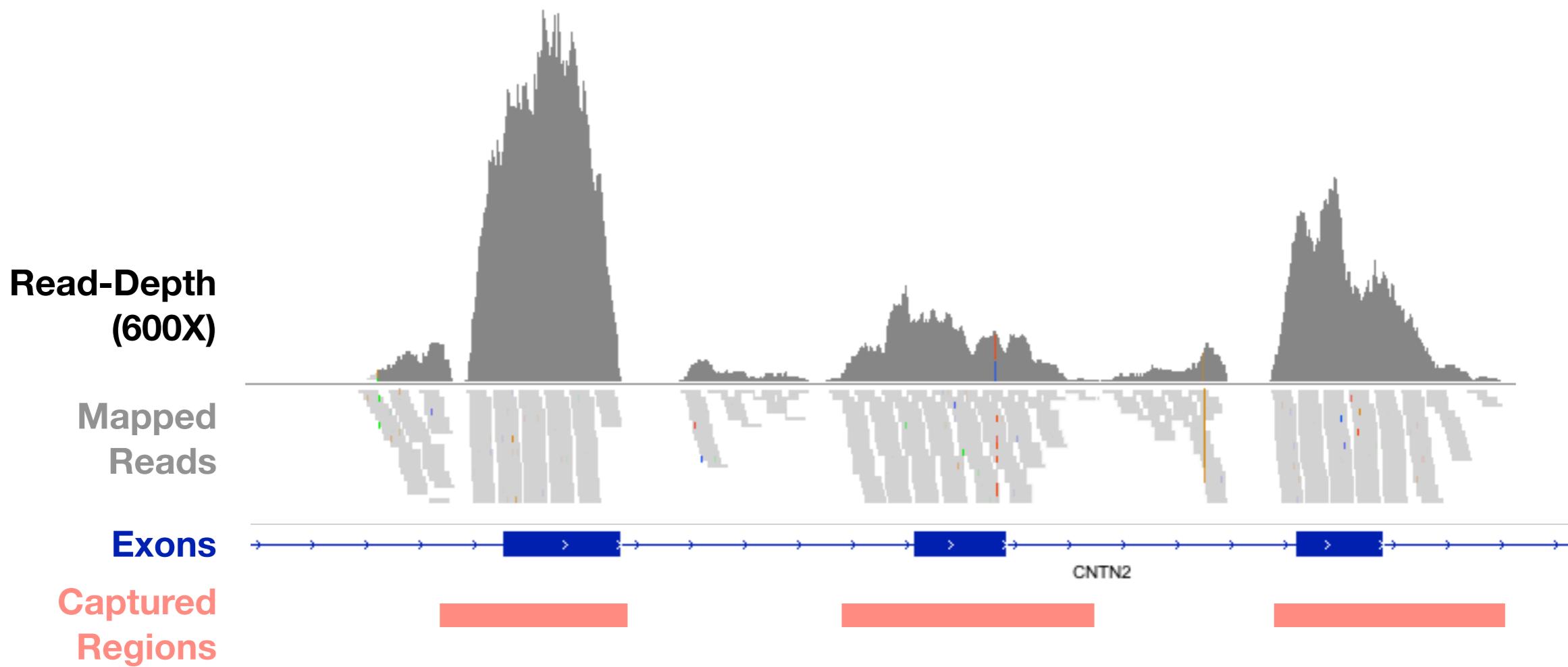
How do we detect SV among exomes?



Basic principle: Sequence depth indicates copy-number

4. Rare SV discovery from 3000+ “exomes”

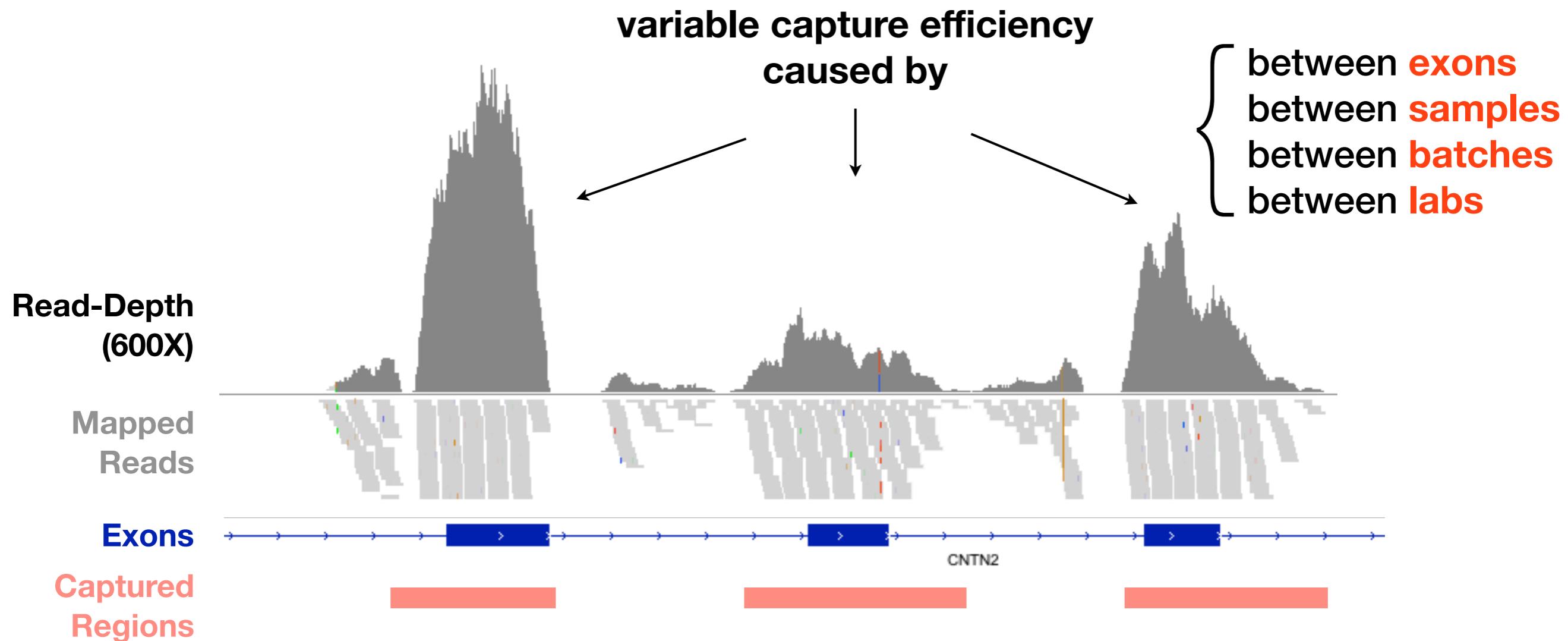
Sounds great, but here's the rub...



Multiple sources of **artificial** variance in sequence coverage

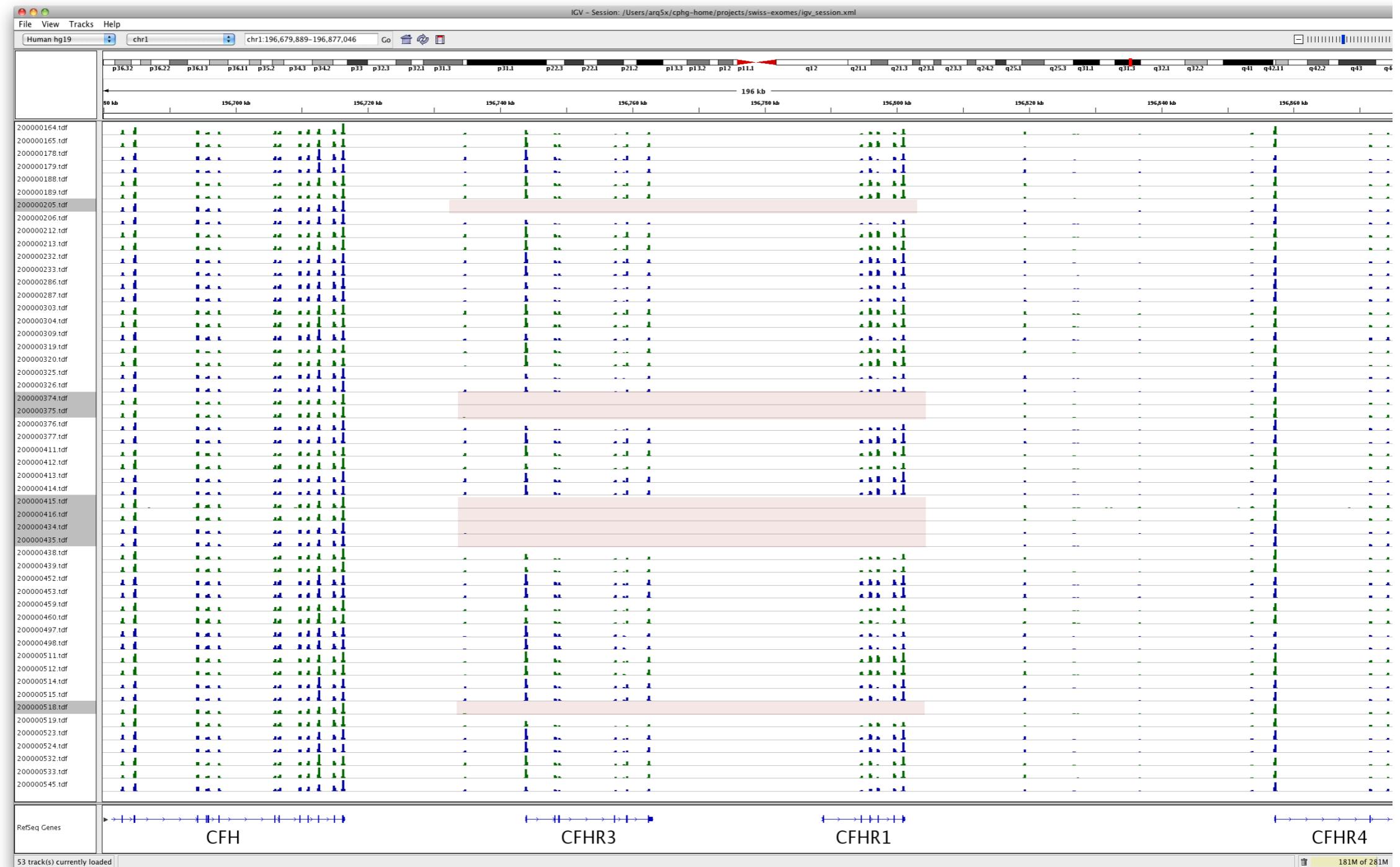
4. Rare SV discovery from 3000+ “exomes”

Sounds great, but here's the rub...



Multiple sources of **artificial** variance in sequence coverage

4. Rare SV discovery from 3000+ “exomes” ...but homozygous deletions should be easy, right?



CFHR3 and CFHR1 del. enriched in siblings with ischemic stroke)

4. Rare SV discovery from 3000+ “exomes”

The Exome Sequencing Project (ESP)

- ▶ Cardiovascular disease, stroke, lung disorders.
- ▶ Co-chair (w/ Evan Eichler, UW) of Structural Variation Group
 - ▶ Catalog copy-number polymorphic genes/exons.
 - ▶ Identify candidates that may underlie studied phenotypes.
- ▶ Capture and sequencing done @ UW and Broad Institute.
- ▶ **Pilot:** CNVs in 575 high quality samples (HiSeq / Nimblegen)
- ▶ **Current:** Expanding study to ~3300 samples. Increasing method resolution

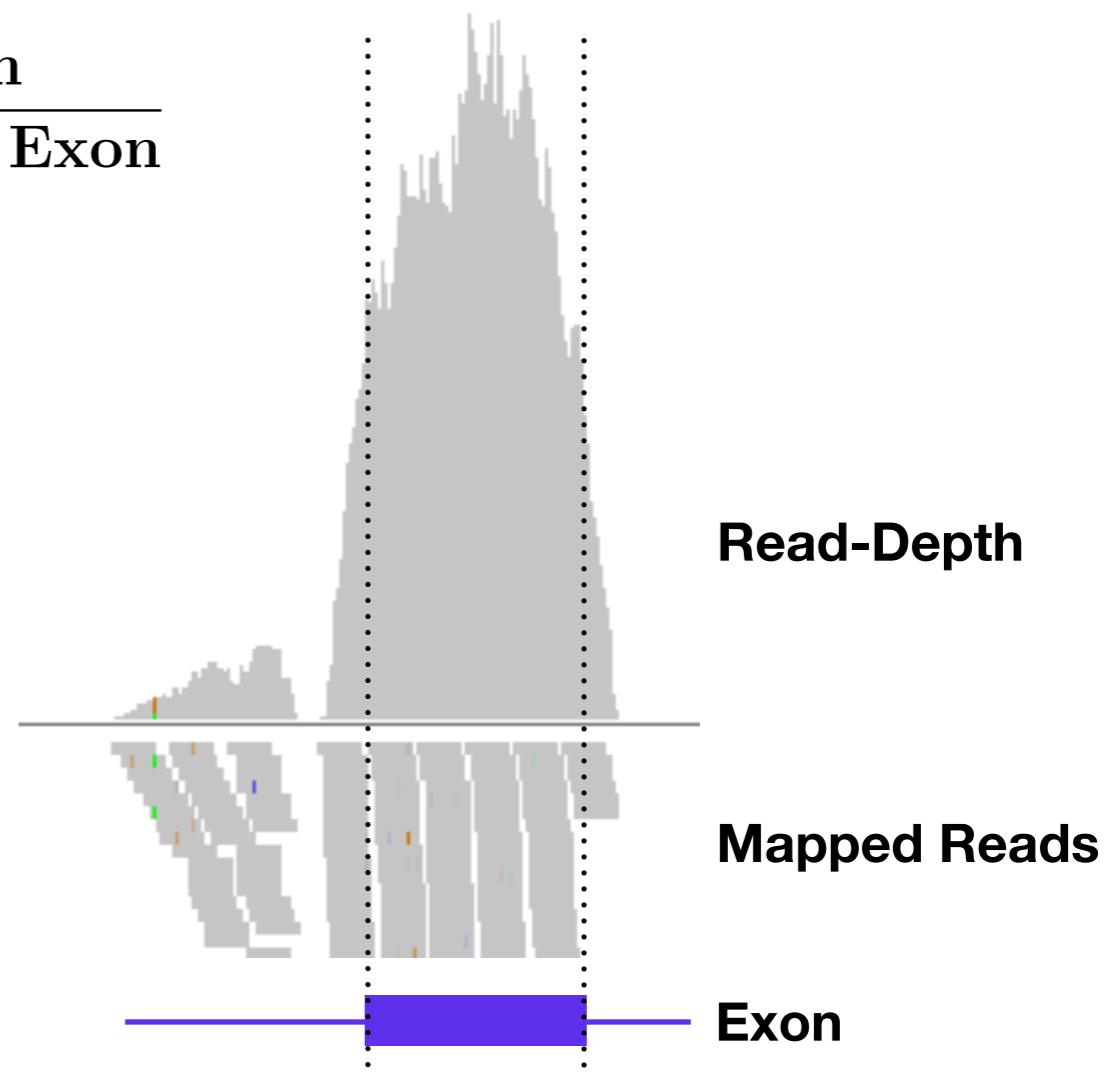


4. Rare SV discovery from 3000+ “exomes”

RPKM & Singular Value Decomposition (SVD)

$$RPKM = \frac{10^9 \cdot \text{Reads Mapping to Exon}}{\text{Total Mapped Reads} \cdot \text{Length of Exon}}$$

$$zRPKM = \frac{RPKM_{exon,sample} - \mu_{exon}}{\sigma_{exon}}$$

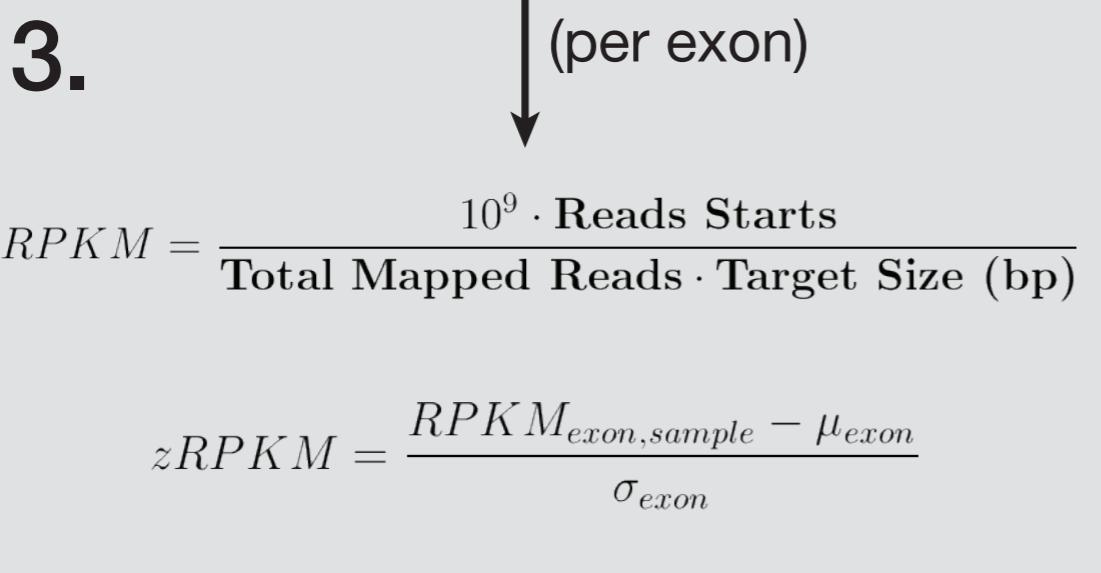
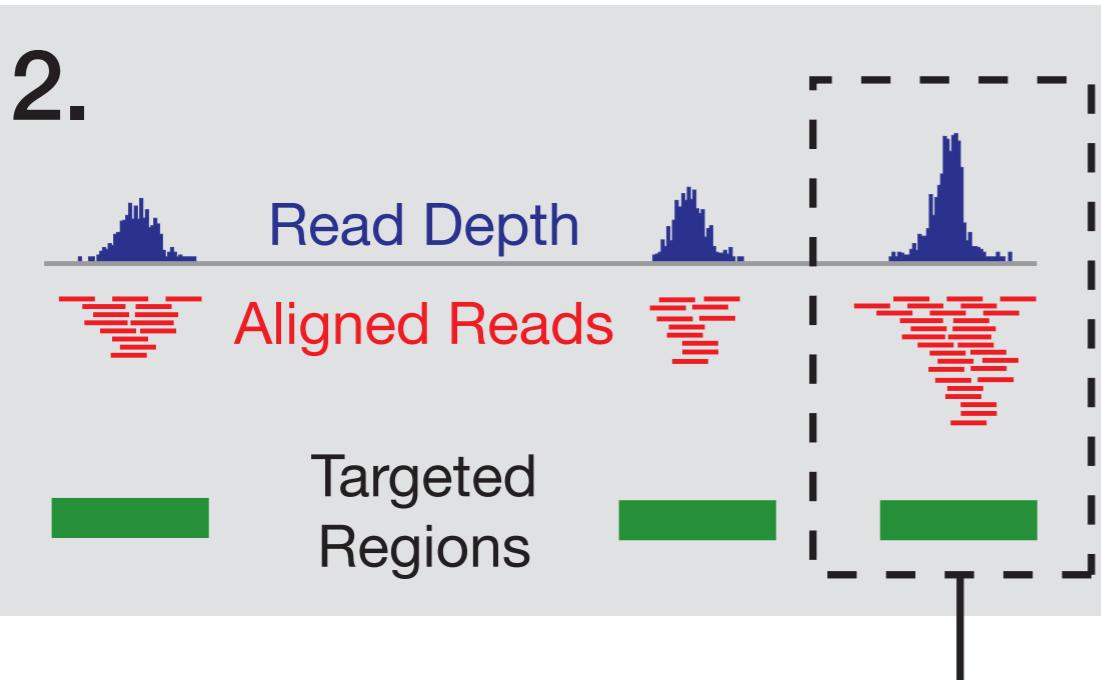


- ▶ RPKM: **R**eads **P**er **K**ilobase of exon per **M**illion mapped reads
- ▶ RPKM is a normalized measure of read depth over an exon

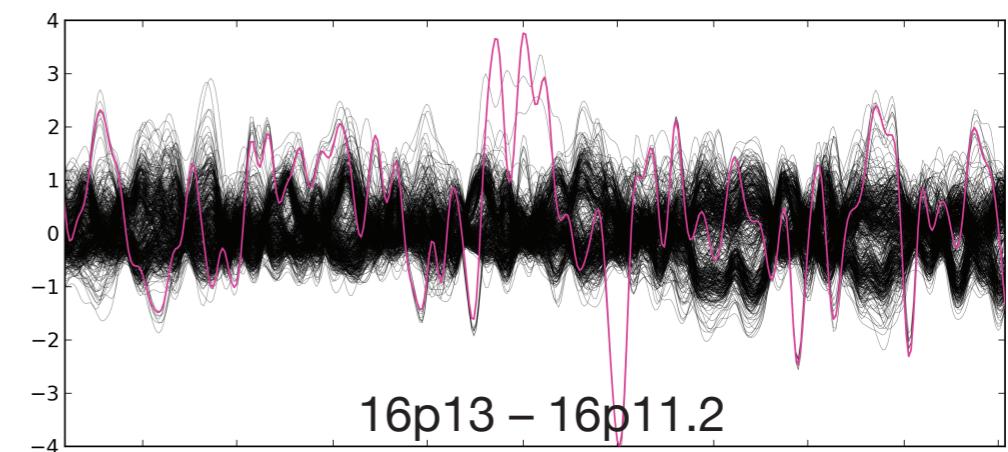


1. FASTQ file contains sequencing data

mrsFAST aligner ↓



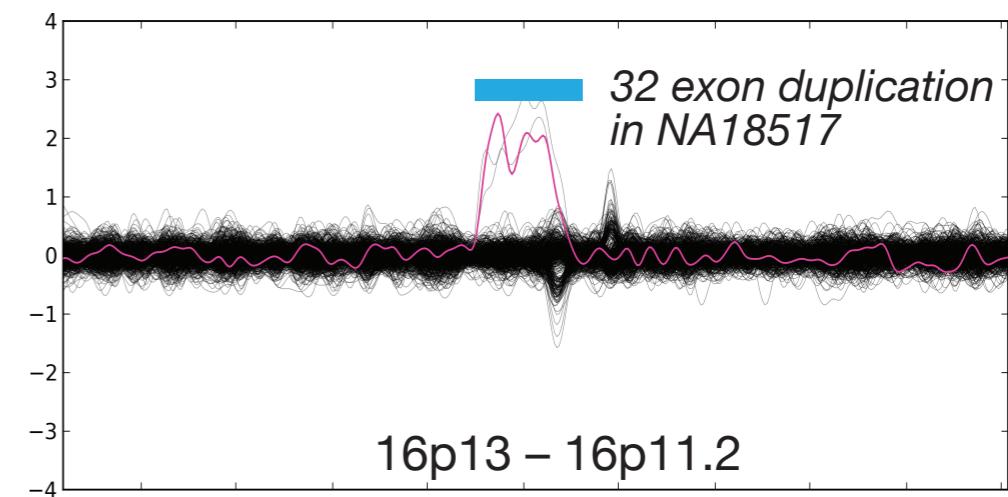
4.



U S V^T Singular Value Decomposition

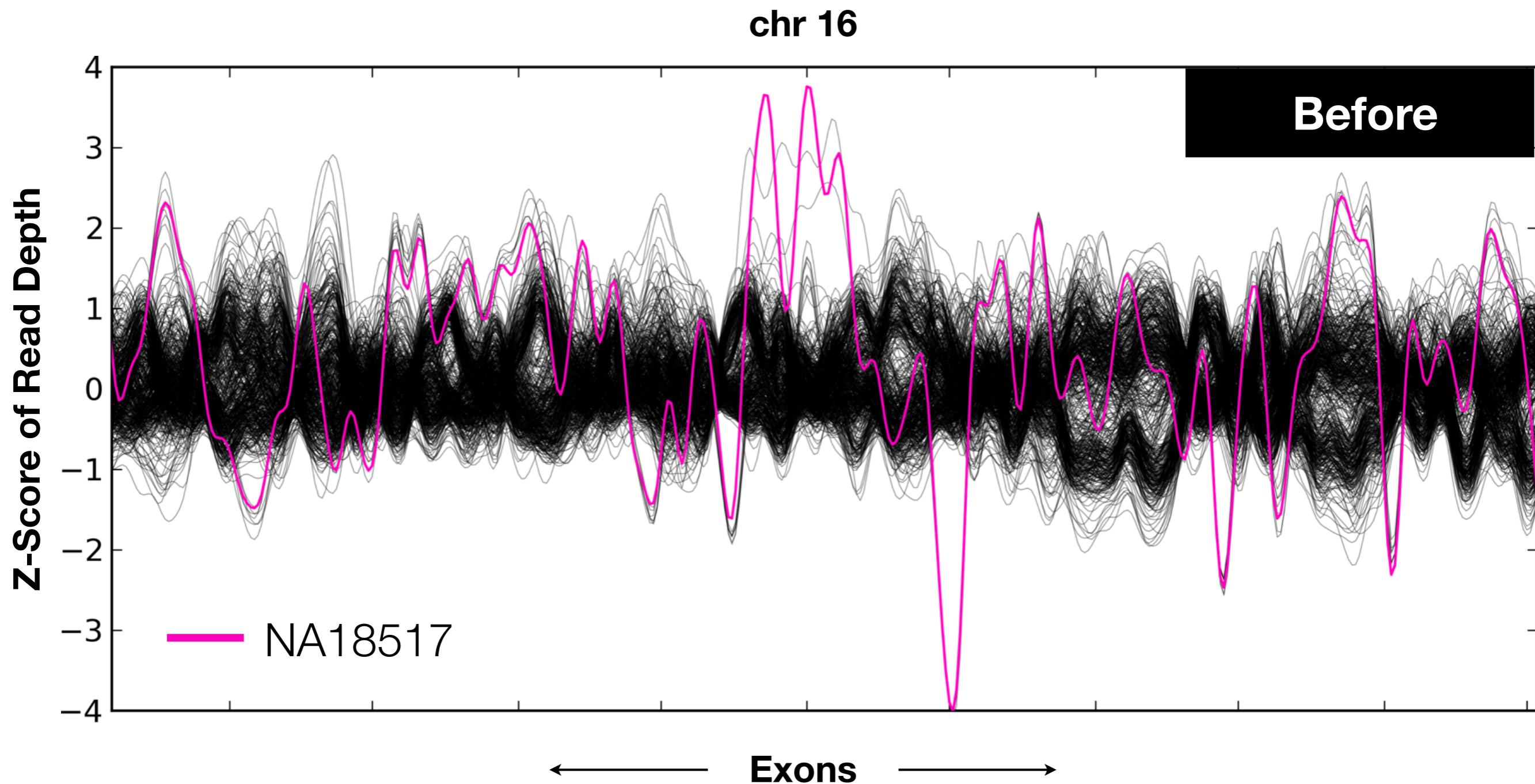
Zero strongest k singular values

U • $\begin{matrix} S' \\ 0_0 \\ \vdots \\ 0_n \end{matrix}$ • V^T Reconstruct data matrix



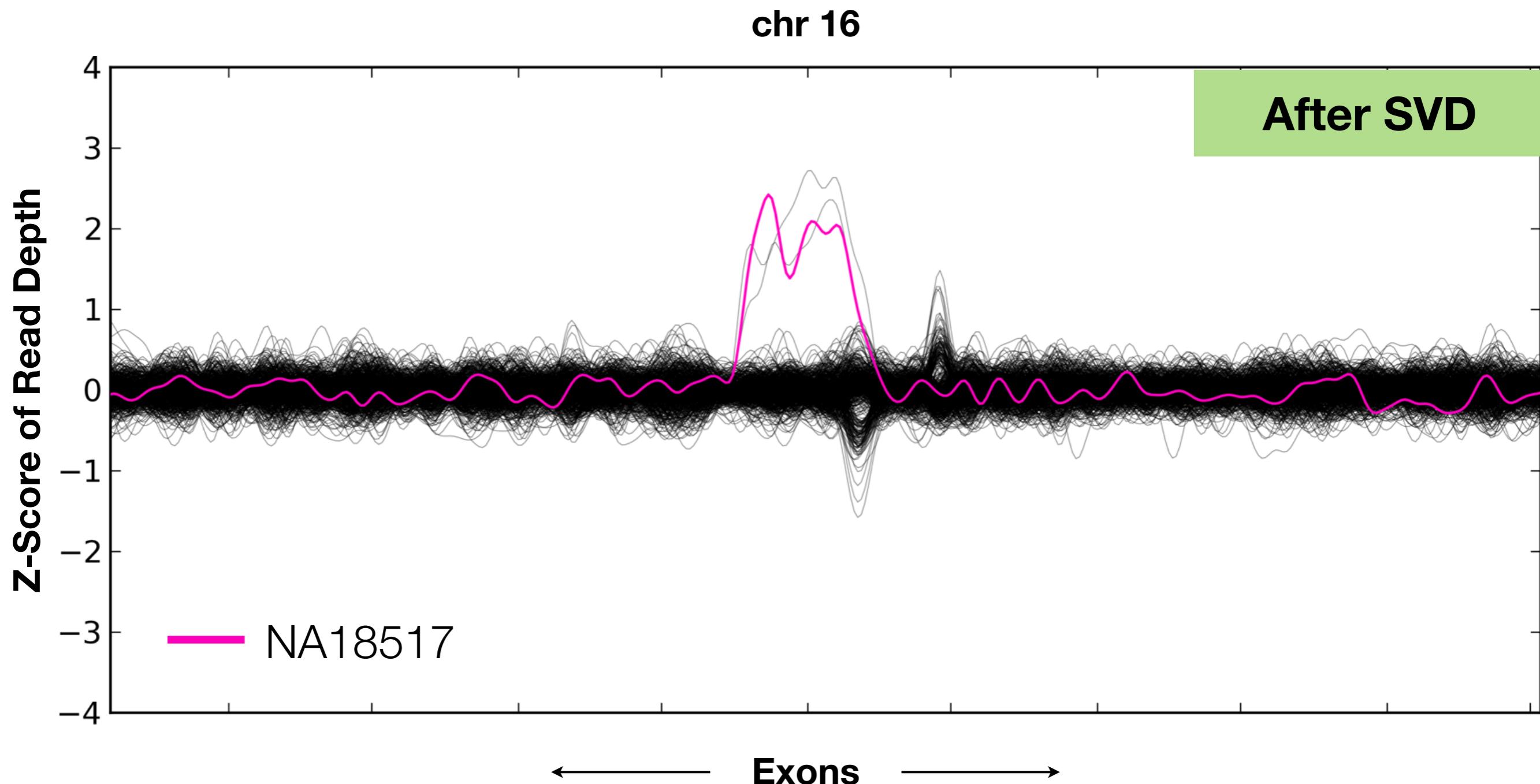
4. Rare SV discovery from 3000+ “exomes”

RPKM normalization is not enough!



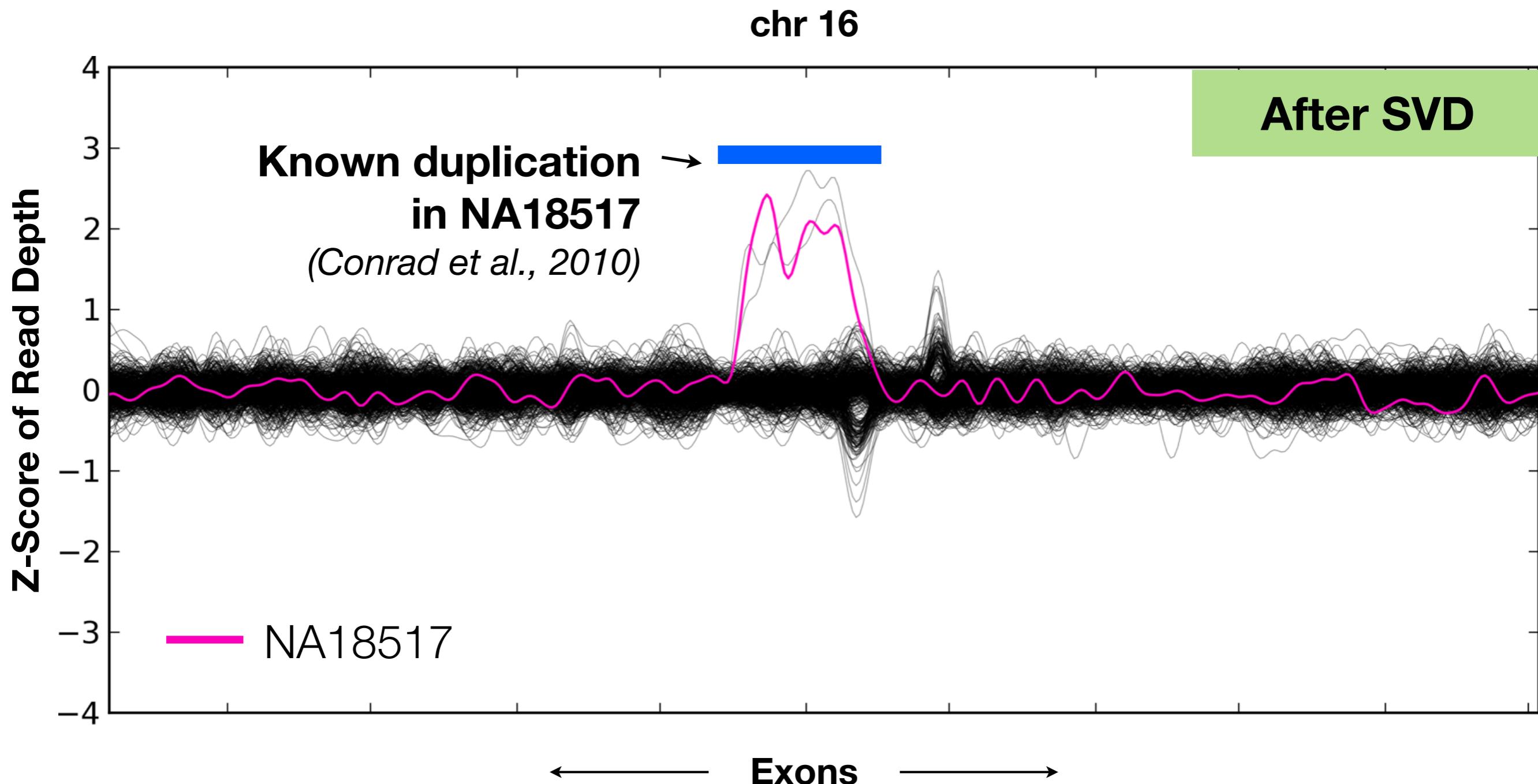
4. Rare SV discovery from 3000+ “exomes”

Ah, that's better: after SVD transformation



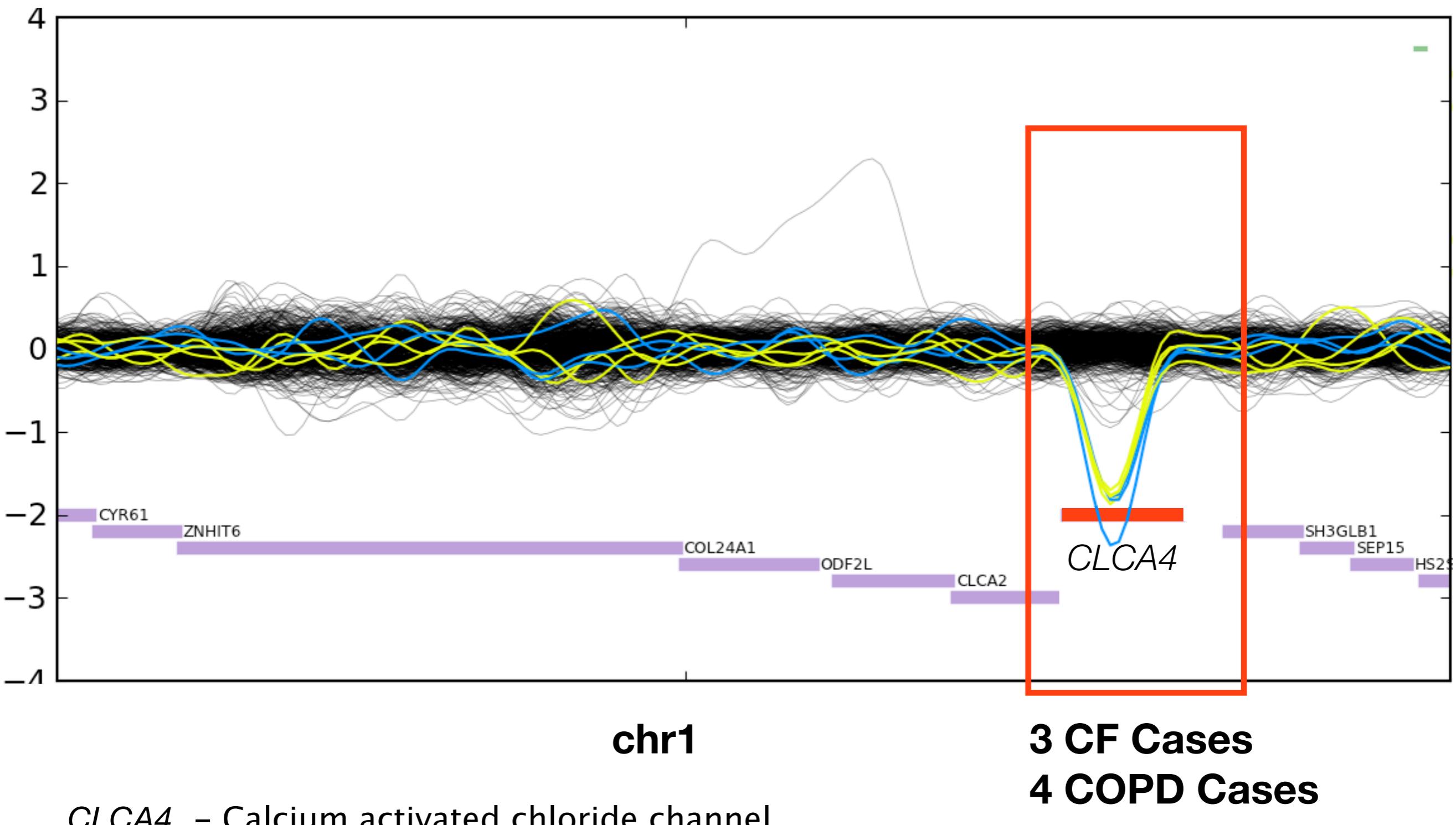
4. Rare SV discovery from 3000+ “exomes”

Ah, that's better: after SVD transformation



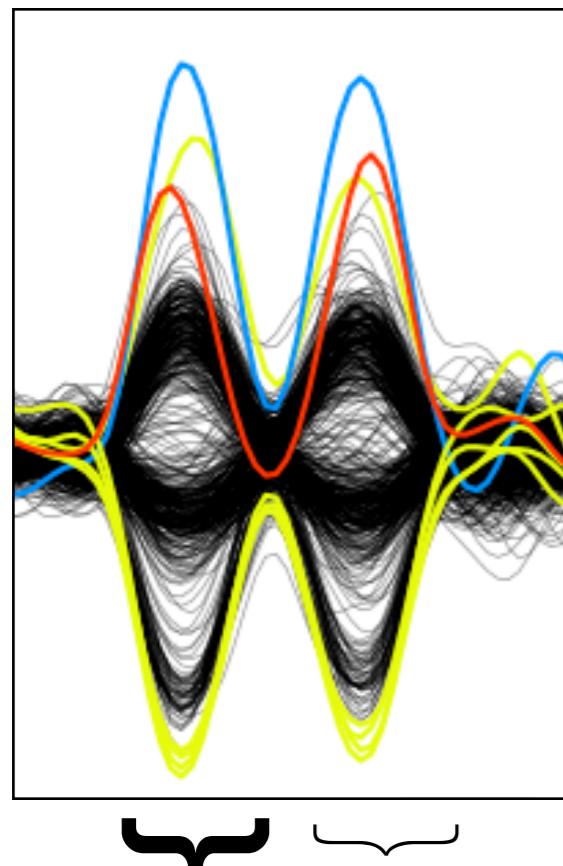
4. Rare SV discovery from 3000+ “exomes”

Deletion of exons in **CLCA4**



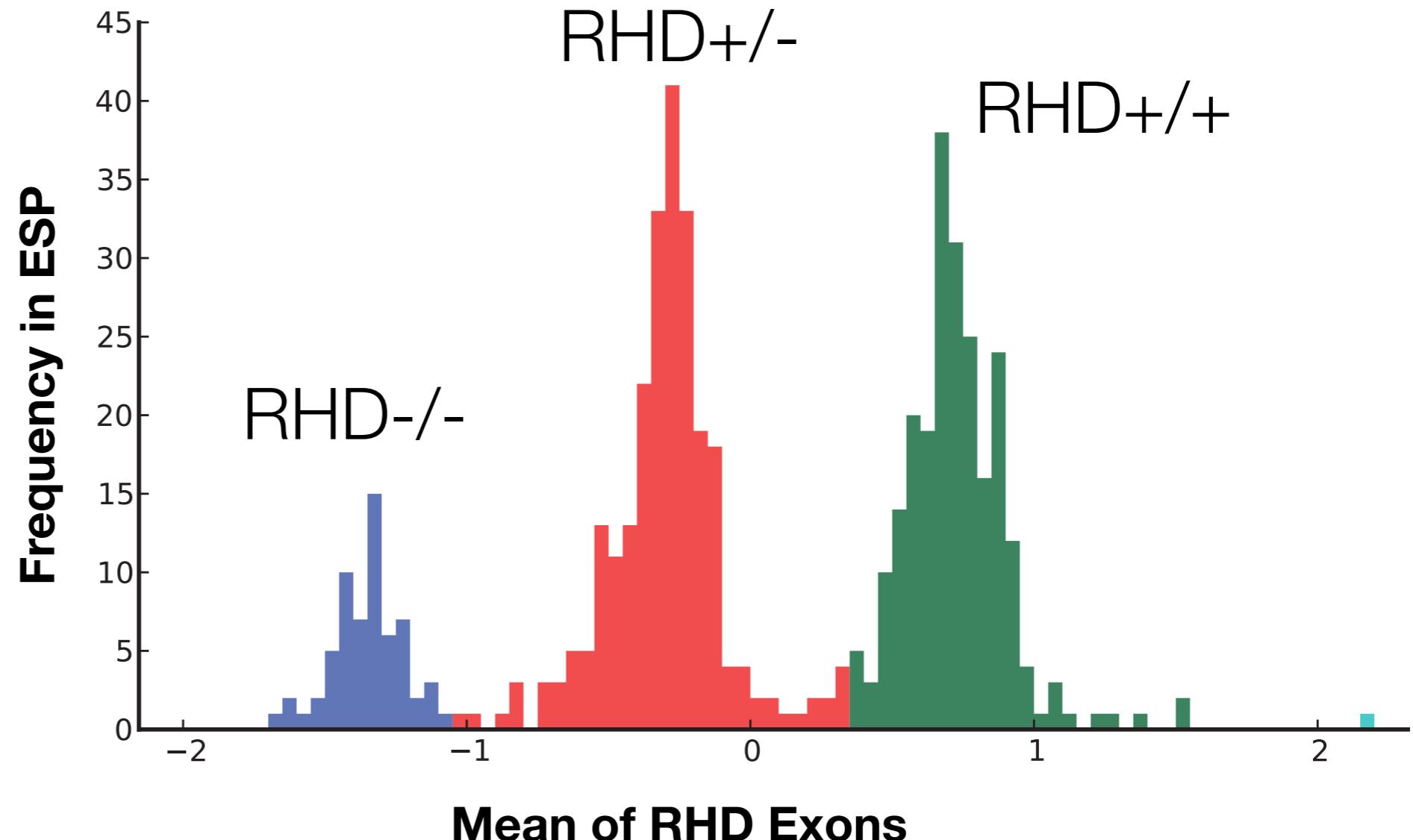
4. Rare SV discovery from 3000+ “exomes”

Genotyping the RHD (Rhesus factor D) locus



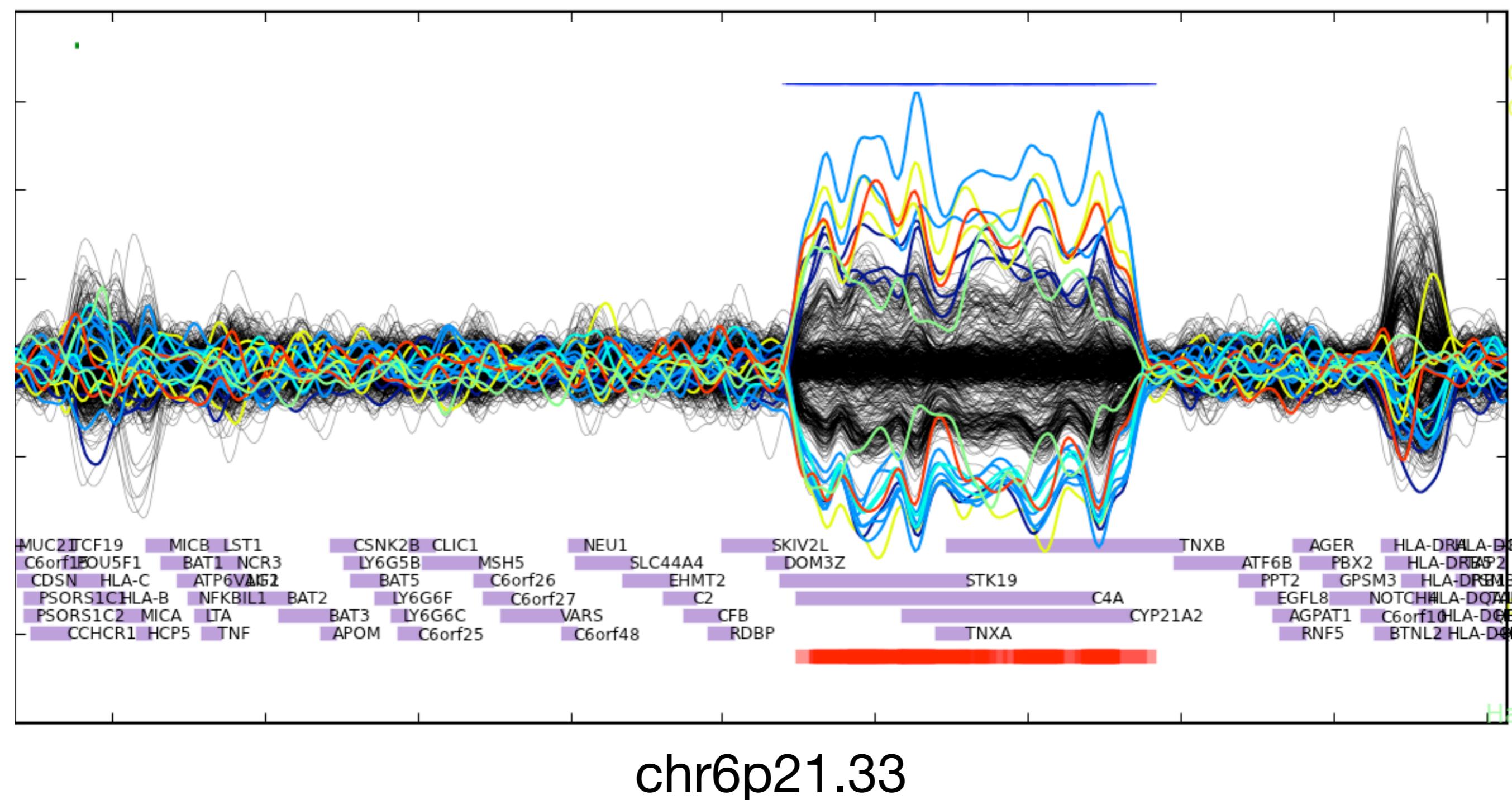
RHD **RHCE**

paralog



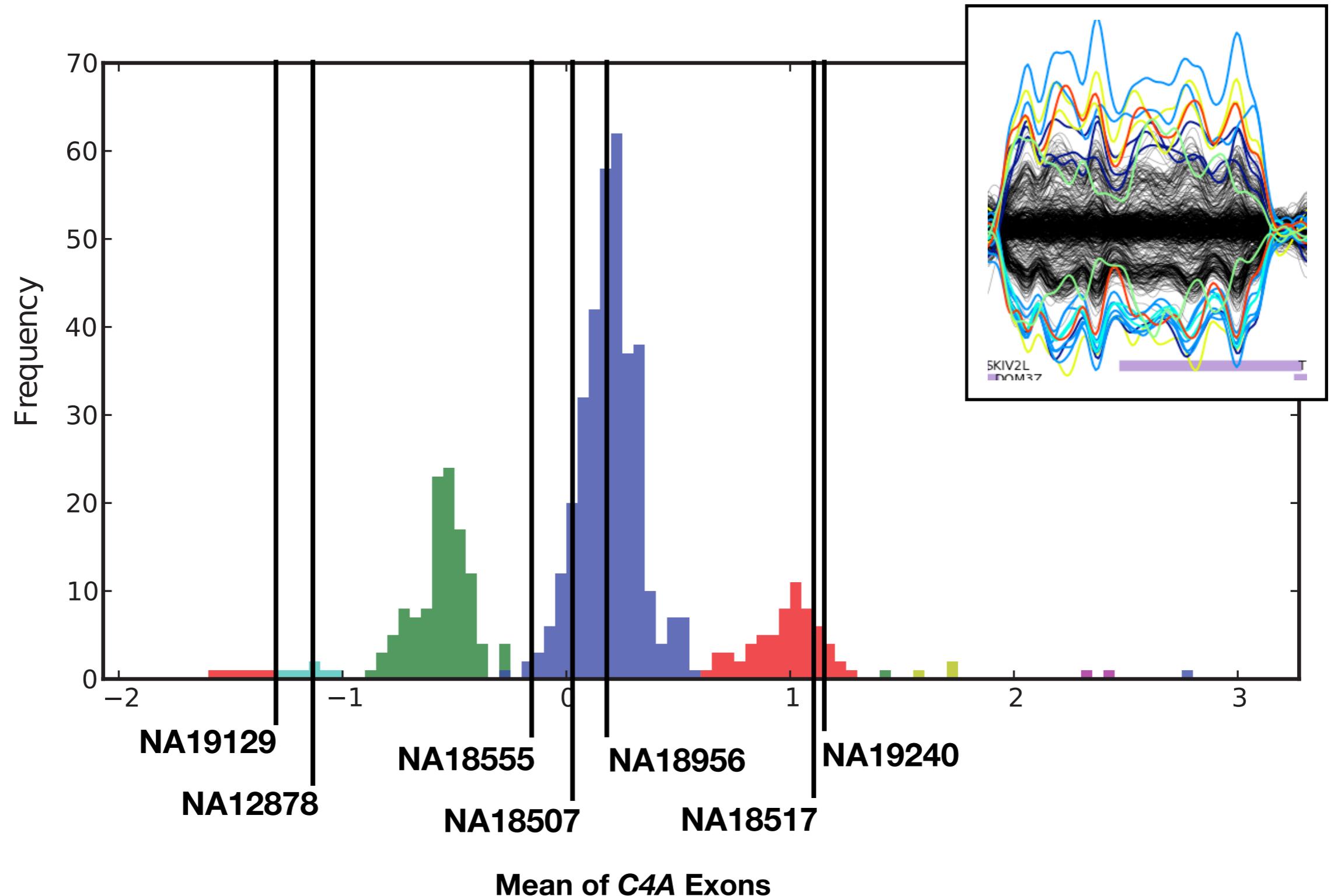
4. Rare SV discovery from 3000+ “exomes”

C4A locus



4. Rare SV discovery from 3000+ “exomes”

Genotyping the C4A locus



4. Rare SV discovery from 3000+ “exomes”

So far, we've identified over affected genes

ACOT1	CLEC18A	FRMPD2L1	KIR2DL4	LRRC37A2	OR4P4	PSGII	TOR1A
ACTR3B	CLEC18C	GBP3	KIR2DS4	MCAM	OR4S2	RELL1	TRGV3
AG10	CLYBL	GHR	KIR3DL1	METTL10	OR52N5	RGPD3	TTC35
ALPPL2	CNTNAP3	GOLGA6D	KIR3DL1/2V	MOXD1	OR5P2	RHCE	TTC6
AMY1A	CT45A1	GOLGA8A	KIR3DL3	MRC1	P450-CYP21B	RHD	TUBA3E
AMY1B	CYP21A2	GOLGA8E	KIR3DP1	NBPF1	PAGE2	RSPH10B	UBTFL1
ANKRD30B	CYP2A6	GPAT2	KIR3DS1	NBPF10	PCDHA10	RSPH10B2	UGT2B11
ANKRD36	DKFZp564C24	GSTM1	KLRC2	NBPF14	PCDHA9	SIGLEC11	UGT2B28
ANKRD36B	DKFZp586I102	GSTT1	KRT1B	NBPF3	PDSS1	SIGLEC14	VCX
ARHGEF5L	DMBT1	GTF2IRD2	KRT77	NEB	PI4KA	SIGLEC5	VCX3A
BTNL3	FAM153A	HERC2P3	KRTAP9-8	NOMO2	POLR2J	SIRPB1	VCX3B
BTNL8	FAM153B	HLA-DQ	KRTAP9-9	NOMO3	POM121C	SLC25A24	WBSCR16
C4A	FAM21A	HLA-DQB1	LCE1D	NXF5	POMZP3	SPAG11A	ZAN
C4B	FAM21B	HLA-DRB1	LCE3B	OBP2A	POTED	SPAG11B	ZDHHC20
C7orf28B	FAM75A6	HP	LCE3C	OBP2B	POTEE	SPANXC	ZNF717
CCDC74A	FAM86C	HUAT	LGALS9	OBP2C	POTEH	SPINK5L2	pp14737
CCNC	FCGBP	IL9R	LGALS9B	OPN1LW	PRAMEF10	SYCP1	
CDC2L1	FCGR2C	KCP	LGALS9C	OR11H12	PRAMEF4	TARP	
CDC2L2	FCGR3A	KIR-K15	LILRA3	OR2T10	PRB1	TAS2R43	
CFHR1	FCGR3B	KIR-K36	LILRA6	OR2T11	PSG1	TBC1D26	
CFHR3	FFAR3*	KIR2DL1	LILRB3	OR2T35	PSG11	TCP10	
CKMT1A	FOXD4L1	KIR2DL2	LOC391322	OR4C11	PSG4	TCRg	
CKMT1B	FRMPD2	KIR2DL3	LRRC37A	OR4F5	PSG9	TNXB	

3. Rare SV discovery from 3000+ “exomes”

Summary

- ▶ **Exome-based read-depth methods:**

- ▶ Can **discover rare CNV variants** affecting exons
- ▶ **Genotype** common copy number polymorphic sites, **up to ~8 copies**
- ▶ **Find single-exon homozygous deletions**

- ▶ **Future directions:**

- ▶ Paralog-specific identification of highly-duplicated genes
- ▶ Refined detection of CNV breakpoints and single-exon deletions
- ▶ Integrate two approaches for improved detection/genotyping power.
- ▶ Scale to 7000 samples.

Recap

- ▶ Our genome is far more plastic than previously appreciated.
- ▶ New insights into genome biology are possible by integrating large and complex datasets.
- ▶ Modern DNA sequencing is a cheap and potent microscope for understanding genome biology and the genetic basis of traits.
- ▶ Genetics research demands quantitative skills and infrastructure.
 - ▶ Analysis of the data is research in and of itself.
 - ▶ Few people have the necessary training.
 - ▶ Computers; storage; trainees with relevant background.

It's an exciting time for genomics. What next?

- ▶ Bring the genetics/genomics scientists together with clinicians to drive translational research.
 - ▶ Many opportunities for med. school collaborations
- ▶ Applying DNA sequencing to the development of fast, cheap molecular diagnostics.
 - ▶ e.g. HLA-typing, BRCA screens,
- ▶ 30,000 genomes in 2012. We're not ready.
 - ▶ Need better methods for comparing many whole genomes.
- ▶ Single cell DNA and RNA characterization.

Acknowledgements

(arg5x@virginia.edu)

Cancer / iPS / Mouse SV

Ira Hall *Univ. of Virginia*

Ankit Malhotra *Univ. of Virginia*

Michael Lindberg *Univ. of Virginia*

Royden Clark *Univ. of Virginia*

Svetlana Sokolova *Univ. of Virginia*

Mitchell Leibowitz *Univ. of Virginia*

Exome CNV

Nik Krumm *Univ. of Washington*

Evan Eichler *Univ. of Washington*

Debbie Nickerson *Univ. of Washington*

Chris Carlson *Fred Hutchinson CRC*

Mark Rieder *Univ. of Washington*

Josh Smith *Univ. of Washington*

Peter Sudmant *Univ. of Washington*

Other Collaborators

Pat Concannon *Univ. of Virginia*

Steve Rich *Univ. of Virginia*

Suna Onengut-Gumuscu *Univ. of Virginia*

Chris Moskaluk *Univ. of Virginia*

Shu-Man Fu *Univ. of Virginia*

Gabor Marth *Boston College*

James Robinson *Broad Institute*

James Taylor *Emory*

Nick Navin *MD Anderson*

Kristin Baldwin *Scripps*

Quinlan Lab

Uma Paila *Postdoctoral Res. Associate*

Data collection is easy. Interpretation is formidable.
“Routine” bioinformatics is a myth.



Ion Torrent “Proton”

\$1000 genome on a bench in 3 hours.



Illumina Hi/Mi Seq

Proven technology; \$1000 genome in 24hr.
(MiSeq at UVa core facility)

Data collection is easy. Interpretation is formidable.

“Routine” bioinformatics is a myth.



[Ion Torrent “Proton”](#)

\$1000 genome on a bench in 3 hours.



[Illumina Hi/Mi Seq](#)

Proven technology; \$1000 genome in 24hr.
(MiSeq at UVa core facility)



“The \$1000 genome, the \$100,000 analysis”

- *Elaine Mardis WashU Genome Center*

Computational demands are staggering and the analyses are extremely intricate