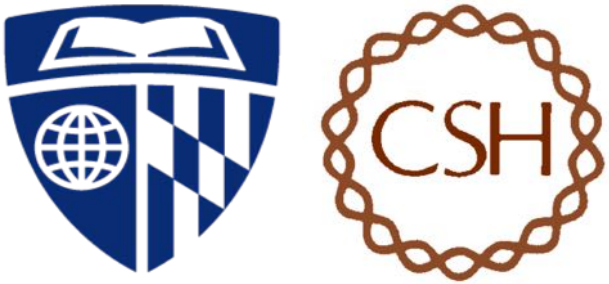
Analyzing -omic Instability in breast cancer with nanopore sequencing of patient-derived organoids

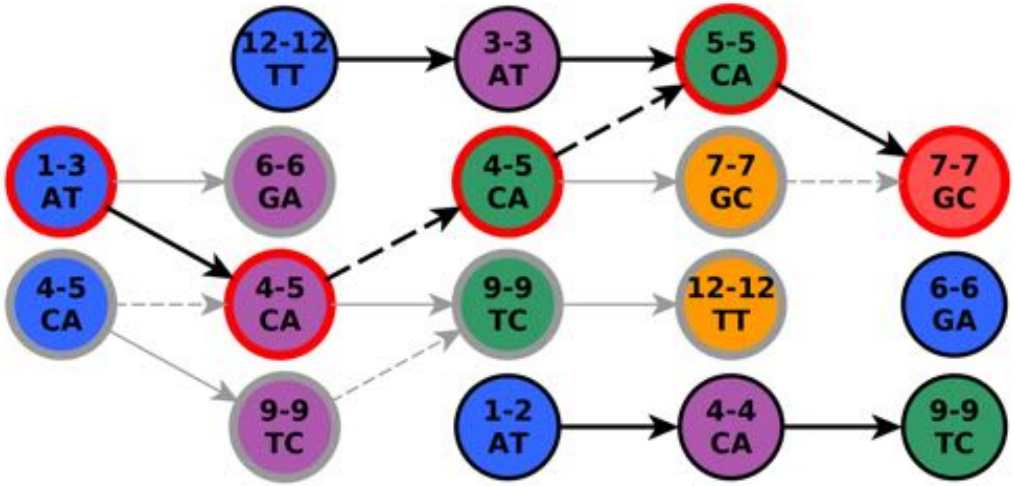
Michael C. Schatz

Nanopore Day DC
March 22, 2018



UNCALLED Signal Alignment

4	3	2	1	0
AT *.083*	**CA** *.063*	AT *.042*	**CA** *.032*	AA *.018*
CG *.041*	GA *.074*	**CA** *.036*	GC *.058*	GA *.045*
GG *.039*	TC *.038*	GT *.027*	GG *.042*	**GC** *.068*
CA *.054*	TT *.043*	TC *.013*	TT *.029*	TC *.024*



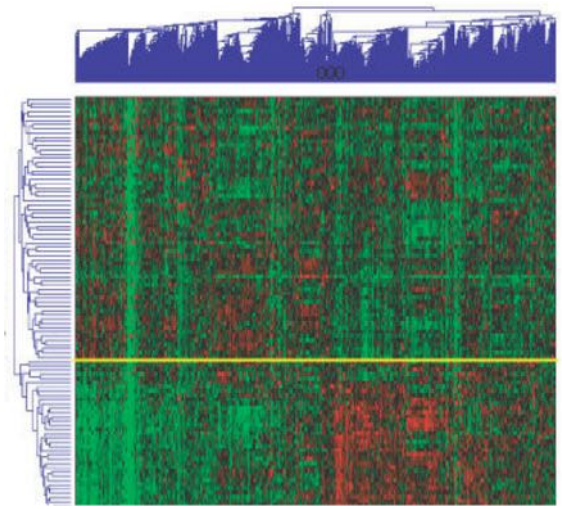
Sam Kovaka

Align raw signal data without base-calling

- Read-until
- Rapid Genotyping
- 95.03% alignment rate
- 99.81% correct
- 675bp / sec / thread

Evolution of Cancer Genomics

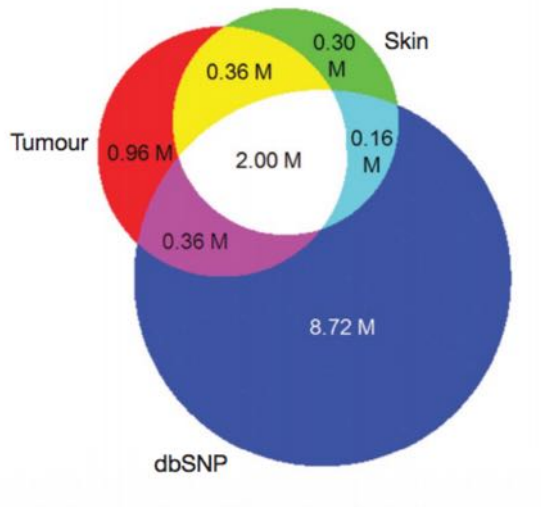
Microarray Profiling



Gene expression profiling predicts clinical outcome of breast cancer

(Van 't Veer *et al*, Nature, 2002)

First Cancer Genome



DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

(Ley *et al*, Nature, 2008)

Pan-Cancer Analysis

Signatures of mutational processes in human cancer

(Alexandrov *et al*, Nature, 2013)

Importance of Structural Variations in Cancer

Copy number changes

Especially amplification & deletions of oncogenes and tumor suppressors

Gene Fusions

Modifies protein sequence & function, potentially alters gene expression by fusing highly expressed transcript with lowly expressed transcript

Prognostic indicator

Greater genome instability generally leads to worse patient outcomes

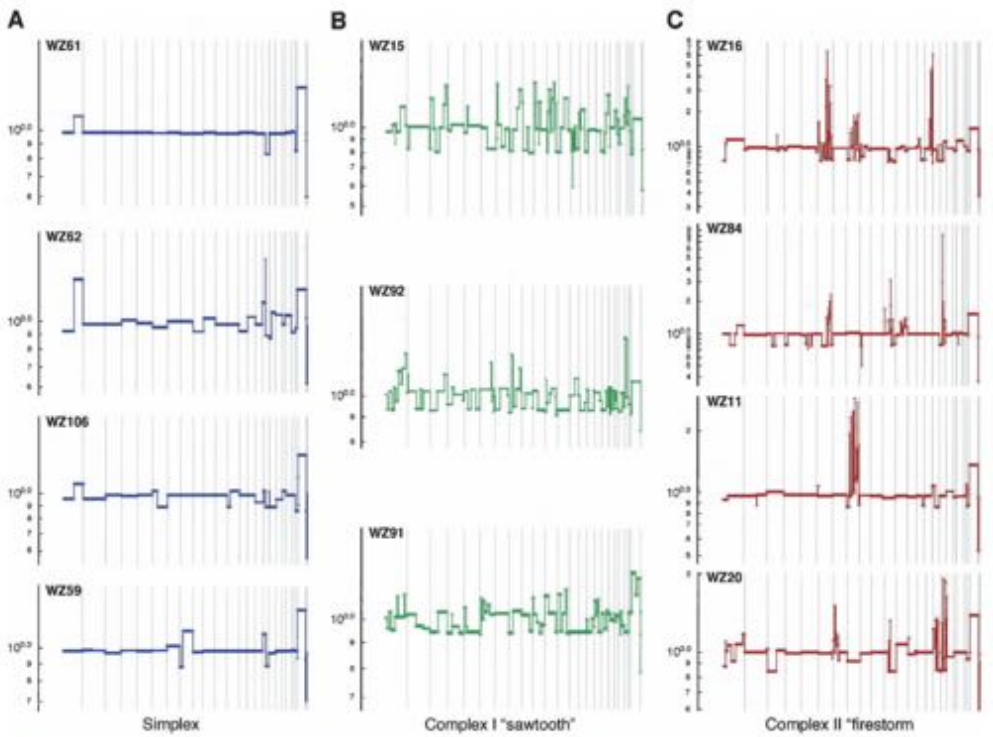


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The *y*-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in *C* is compressed relative to *A* and *B* owing to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from *left* to *right* according to probe position.

(Hicks *et al,* 2006, Genome Research)

Importance of Structural Variations in Cancer

Copy number changes

Especially amplification & deletions of c

C

M
p
f
l

P

Greater genome instability generally leads to worse patient outcomes

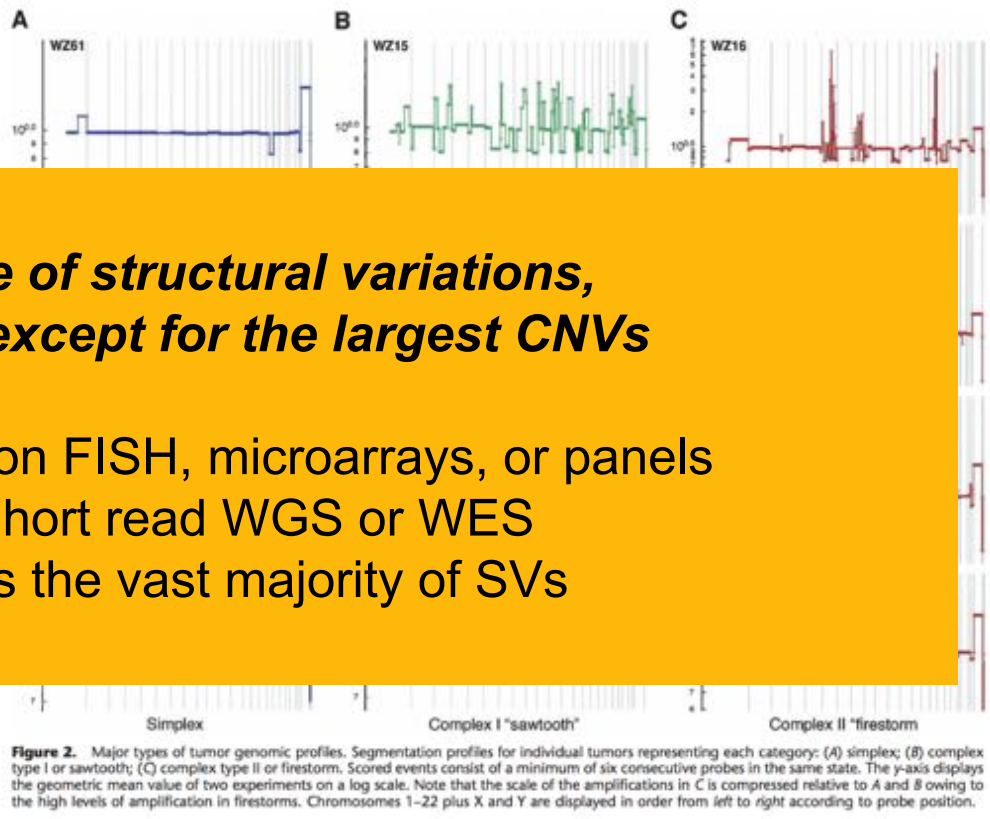


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The *y*-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in *C* is compressed relative to *A* and *B* owing to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from *left* to *right* according to probe position.

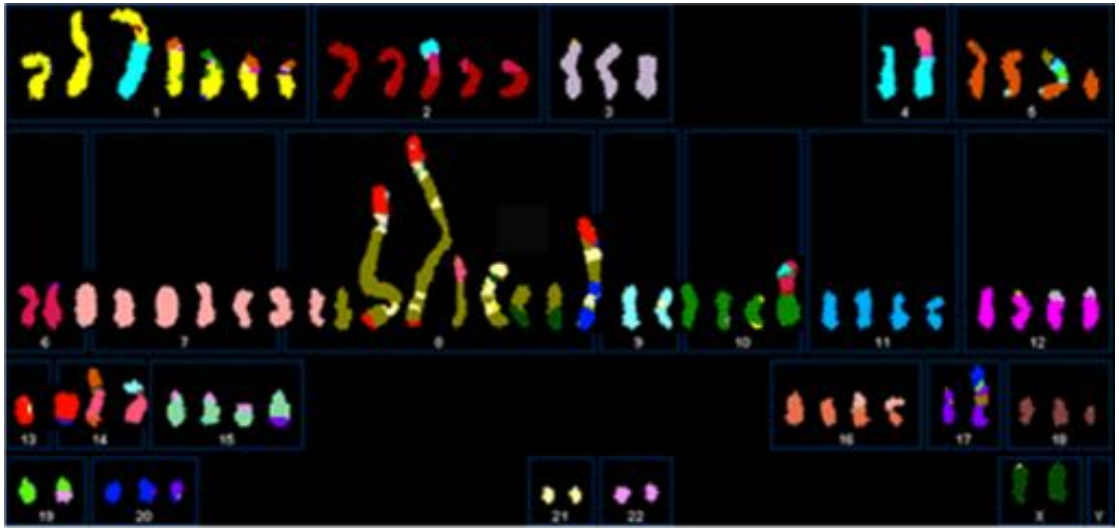
Despite the importance of structural variations, relatively little is known except for the largest CNVs

Clinical standard: low resolution FISH, microarrays, or panels
Research standard: Short read WGS or WES
These approaches miss the vast majority of SVs

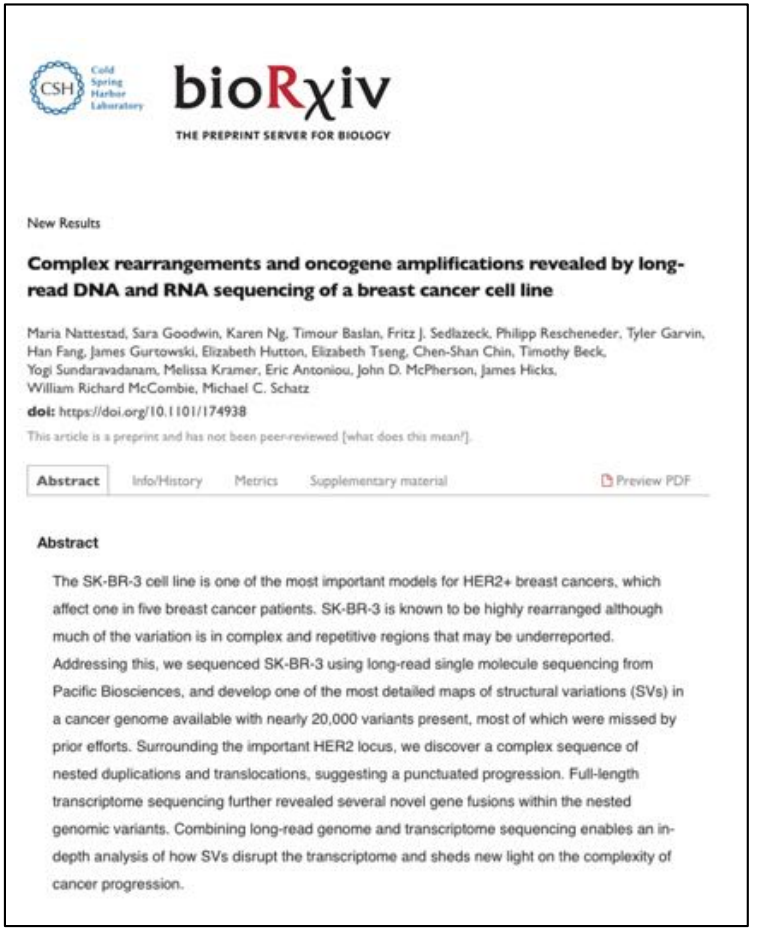
(Hicks *et al,* 2006, Genome Research)

Structural Variations in SKBR3

- SKRB3 cell line was derived by G. Trempe and L. J. Old in 1970 from pleural effusion cells of a patient, a white, Caucasian female
- Most commonly used Her2-amplified breast cancer cell line
- Often used for pre-clinical research on Her2-targeting therapeutics such as Herceptin (Trastuzumab) and resistance to these therapies.



(Davidson et al, 2000)



bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line

Maria Nattestad, Sara Goodwin, Karen Ng, Timour Baslan, Fritz J. Sedlazeck, Philipp Rescheneder, Tyler Garvin, Han Fang, James Gurtowski, Elizabeth Hutton, Elizabeth Tseng, Chen-Shan Chin, Timothy Beck, Yogi Sundaravadanam, Melissa Kramer, Eric Antoniou, John D. McPherson, James Hicks, William Richard McCombie, Michael C. Schatz

doi: https://doi.org/10.1101/174938

Abstract | Info/History | Metrics | Supplementary material | Preview PDF

Abstract

The SK-BR-3 cell line is one of the most important models for HER2+ breast cancers, which affect one in five breast cancer patients. SK-BR-3 is known to be highly rearranged although much of the variation is in complex and repetitive regions that may be underreported. Addressing this, we sequenced SK-BR-3 using long-read single molecule sequencing from Pacific Biosciences, and develop one of the most detailed maps of structural variations (SVs) in a cancer genome available with nearly 20,000 variants present, most of which were missed by prior efforts. Surrounding the important HER2 locus, we discover a complex sequence of nested duplications and translocations, suggesting a punctuated progression. Full-length transcriptome sequencing further revealed several novel gene fusions within the nested genomic variants. Combining long-read genome and transcriptome sequencing enables an in-depth analysis of how SVs disrupt the transcriptome and sheds new light on the complexity of cancer progression.

Structural Variations in SKBR3

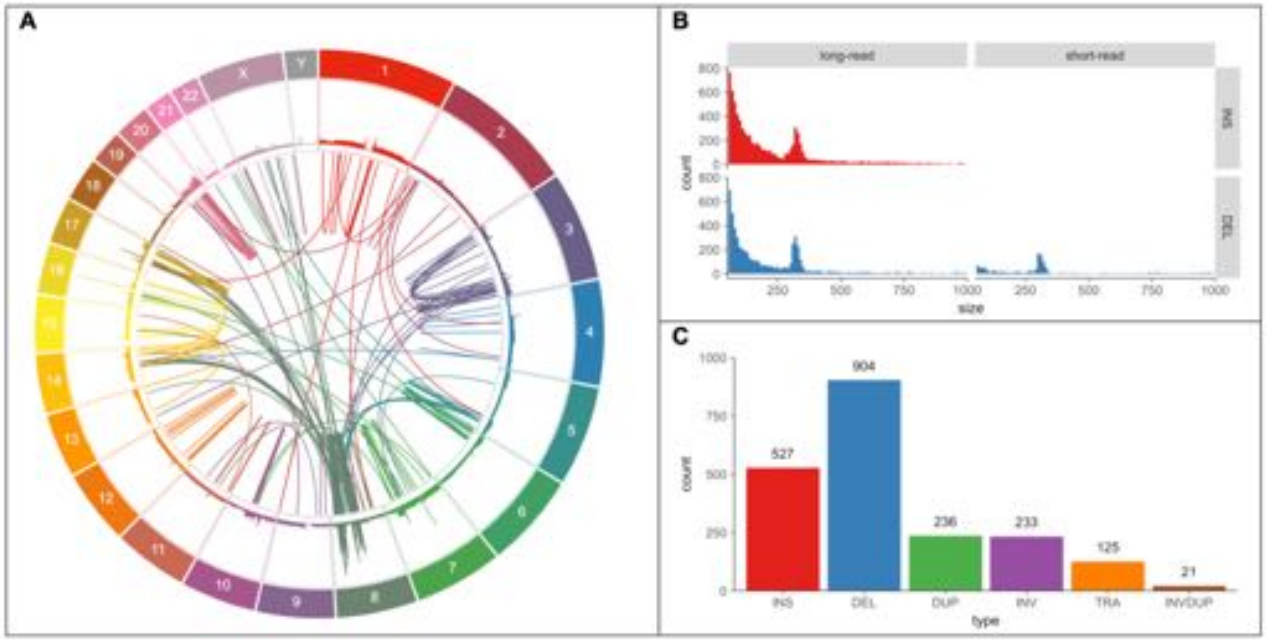


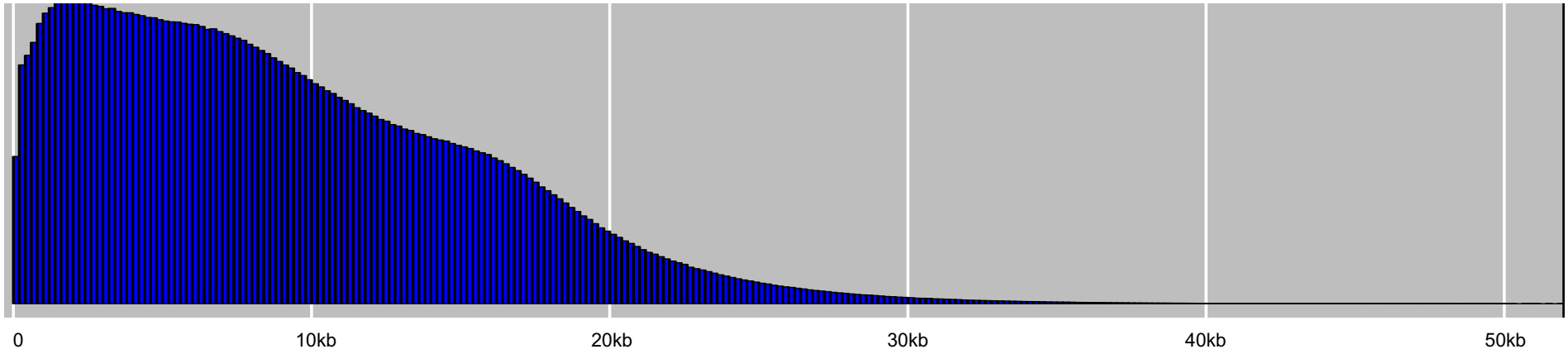
Figure 1 | Variants found in SK-BR-3 with PacBio long-read sequencing. (A) Circos plot showing long-range (larger than 10 kbp or interchromosomal) variants found by Sniffles from split-read alignments, with read coverage shown in the outer track. (B) Variant size histogram of deletions and insertions from size 50 bp up to 1 kbp found by log-read (Sniffles) and short-read (Survivor 2-caller consensus) variant-calling, showing similar size distributions for insertions and deletions from long reads but not for short reads where insertions are entirely missing. (C) Sniffles variant counts by type for variants above 1 kbp in size, including translocations and inverted duplications.

- Finding 10s of thousands of additional variants in the cancer
- PCR validation confirms high accuracy of long read calls
- Detect many novel gene fusions
- With improved SV analysis, can infer the progression of the cancer

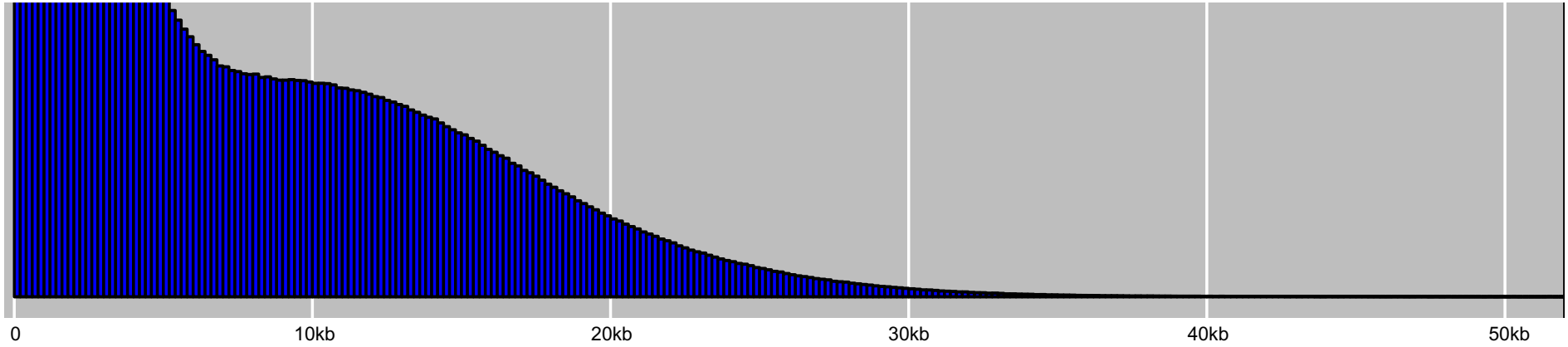
Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line
Nattestad, M et al (2017) bioRxiv https://doi.org/10.1101/174938

Long Read Sequencing of SKBR3

PacBio RSII: 26.3M reads, 72.6X coverage, n50=13,336 bp

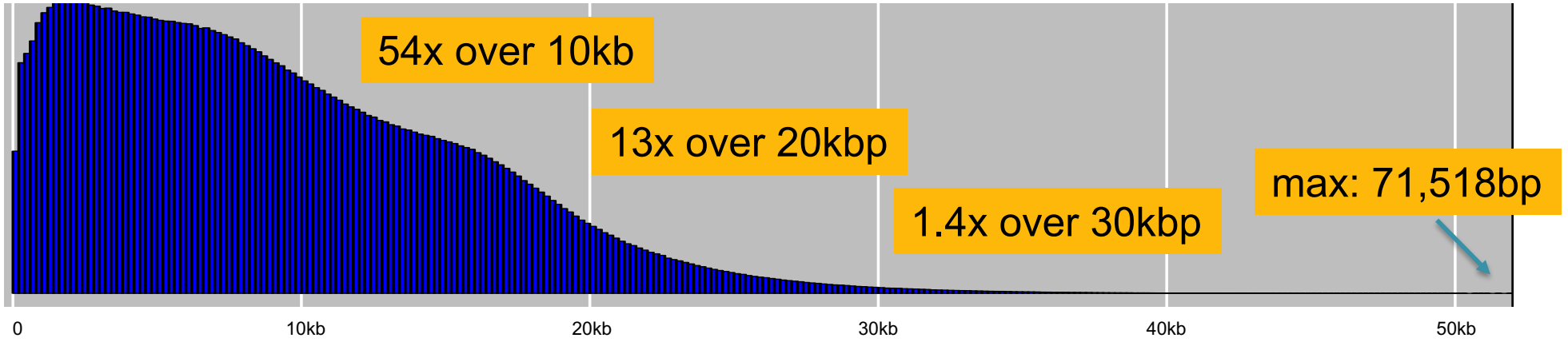


Oxford Nanopore GridION: 13.6M reads, 31.8X coverage, n50=13,350 bp

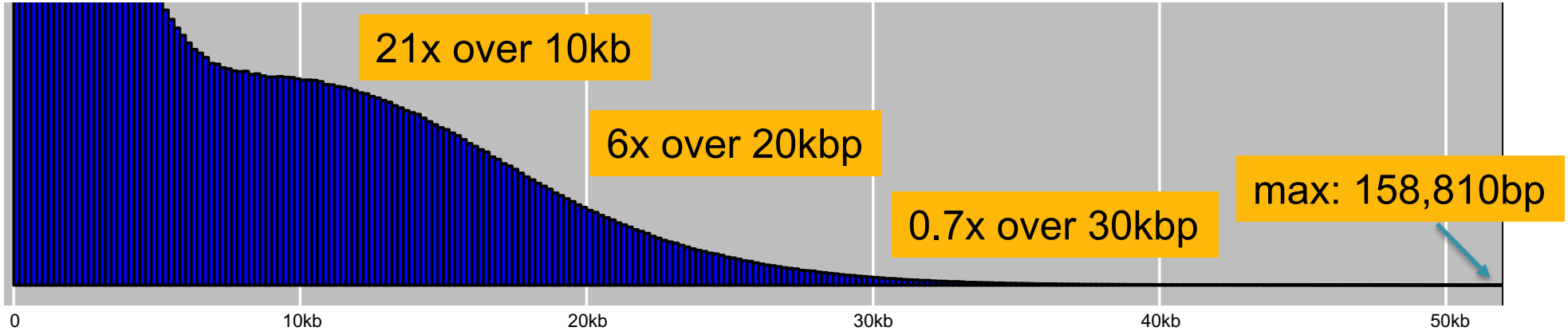


Long Read Sequencing of SKBR3

PacBio RSII: 26.3M reads, 72.6X coverage, n50=13,336 bp



Oxford Nanopore GridION: 13.6M reads, 31.8X coverage, n50=13,350 bp

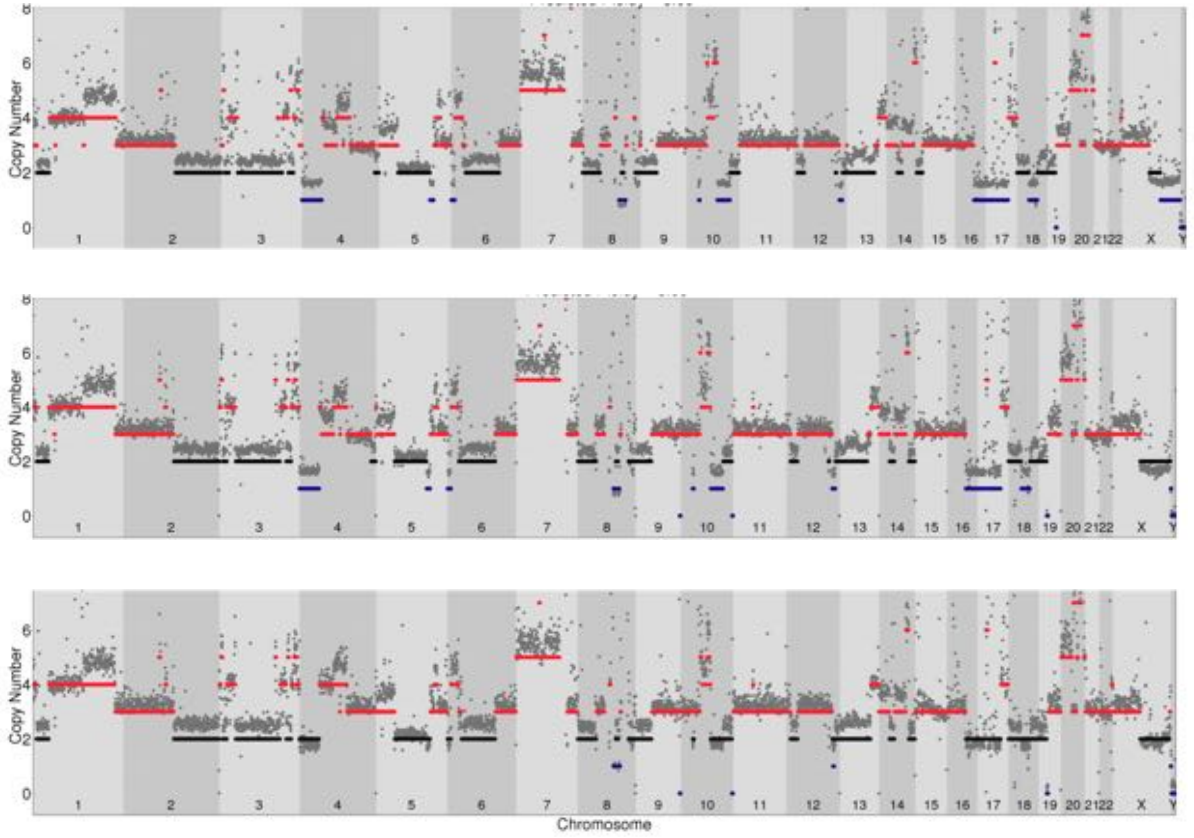


Consistent Profiles of Megabase CNVs

illumina®

PACBIO®

Oxford NANOPORE Technologies

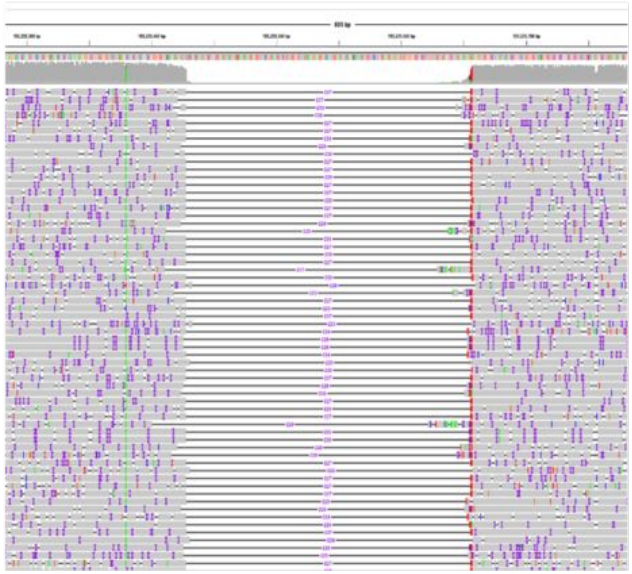


Interactive analysis and assessment of single-cell copy-number variations ("Ginkgo")
Garvin, Aboukhalil, *et al.* (2015) Nature Methods doi:10.1038/nmeth.3578

Structural Variation Identification with Long Reads

BWA-MEM:

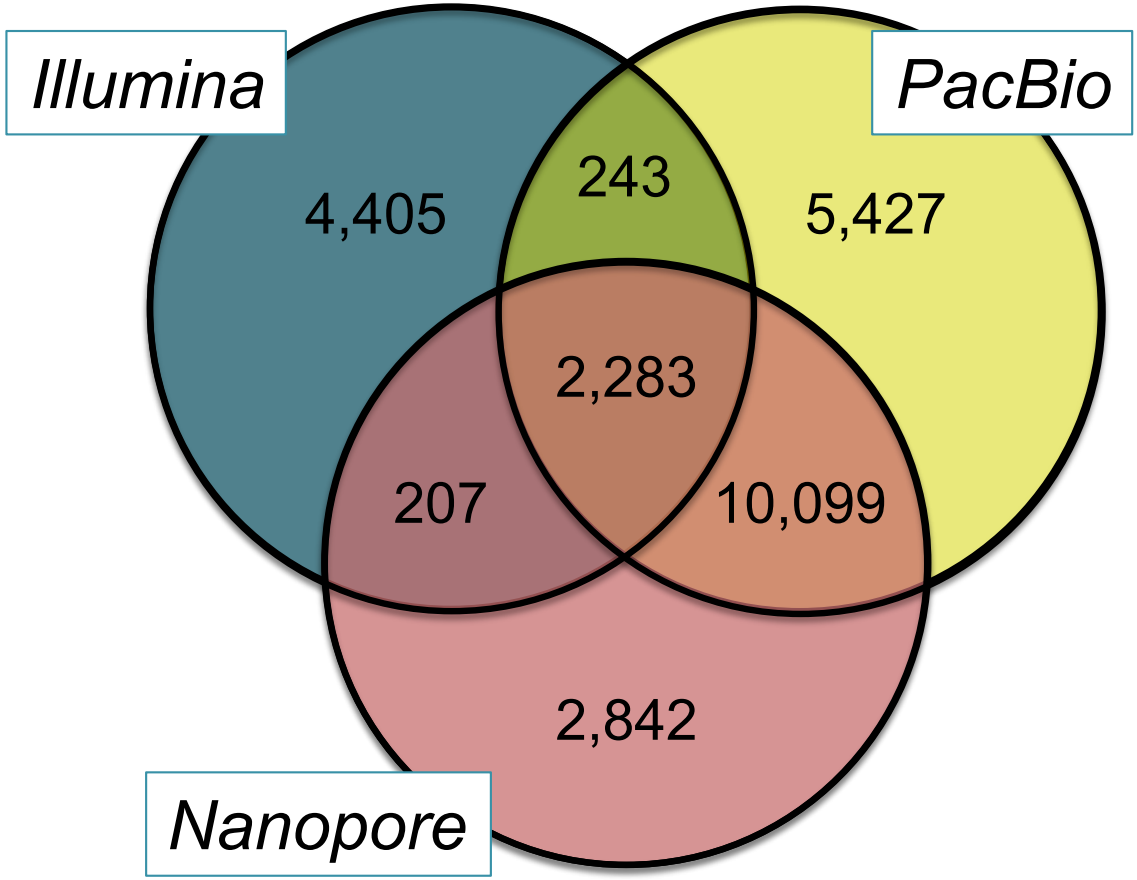
NGMLR:



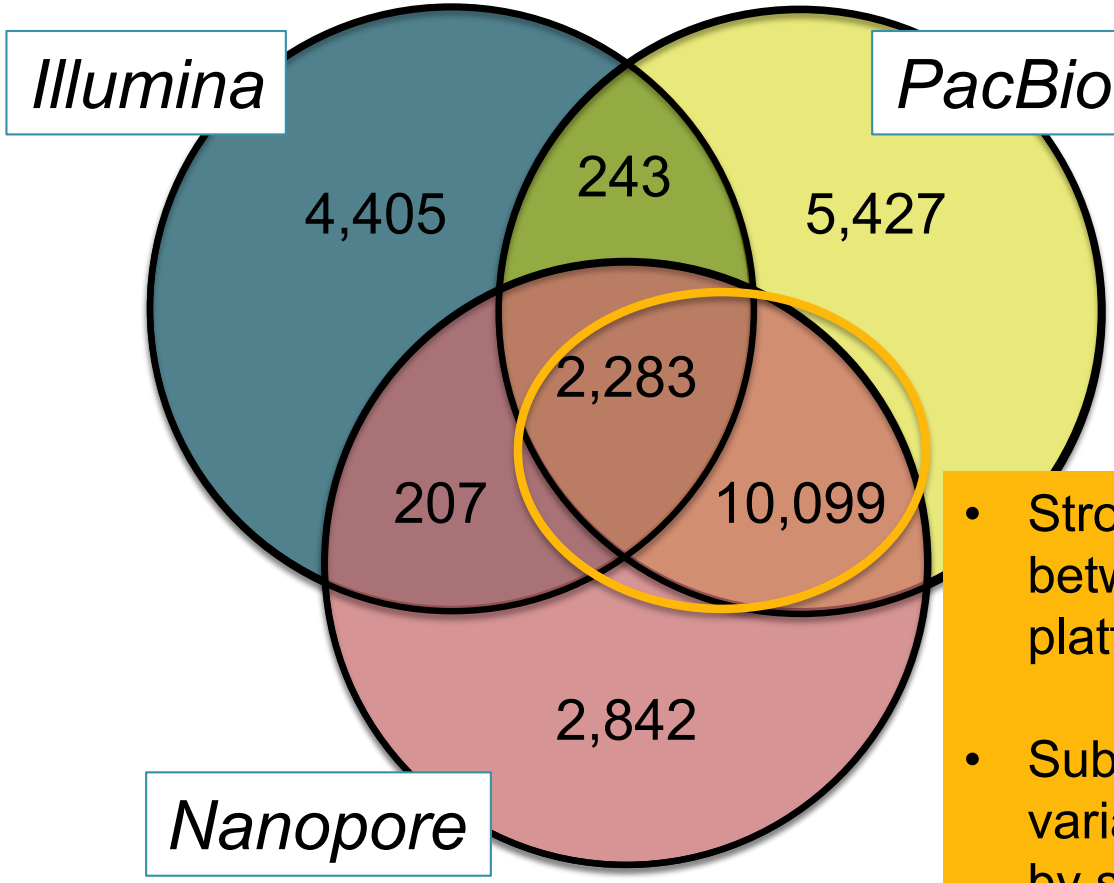
NGMLR: Dual mode scoring to accommodate many small gaps from sequencing errors along with less frequent but larger SVs

Accurate detection of complex structural variations using single molecule sequencing
Sedlazeck, Rescheneder, *et al* (2018) Nature Methods. *In Press.* (*bioRxiv https://doi.org/10.1101/169557*)

Structural Variant Comparison of SKBR3

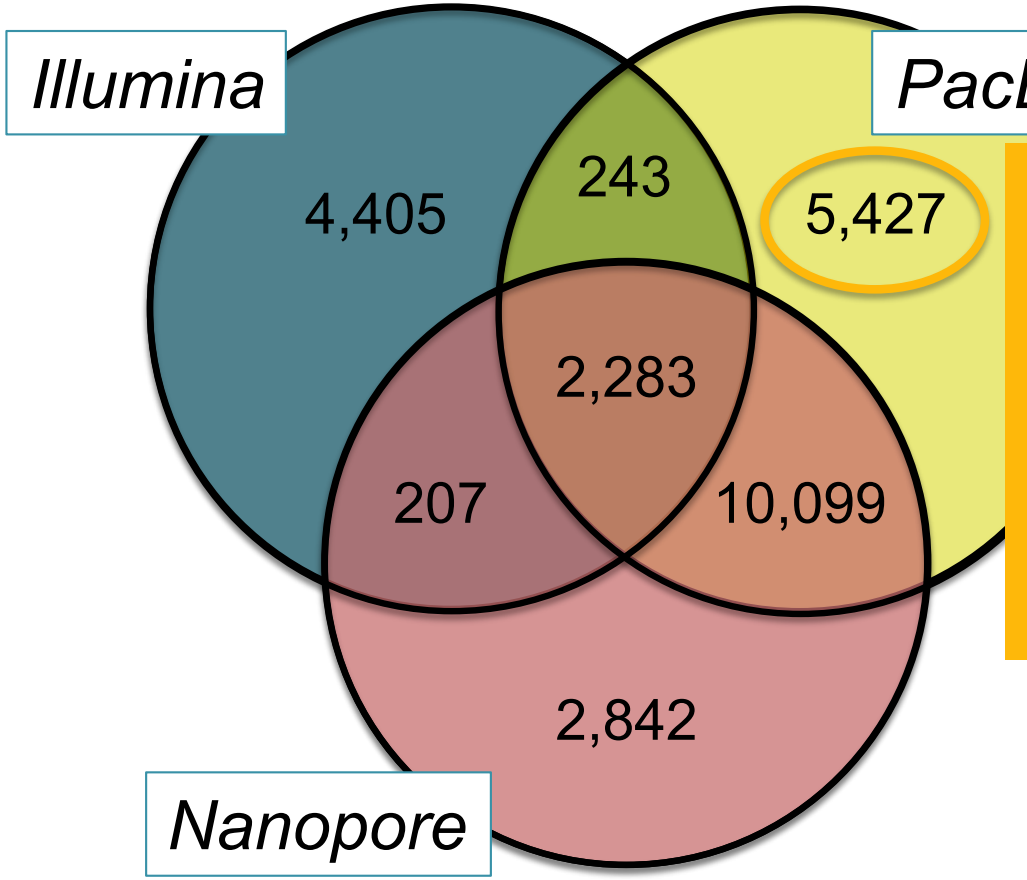


Structural Variant Comparison of SKBR3



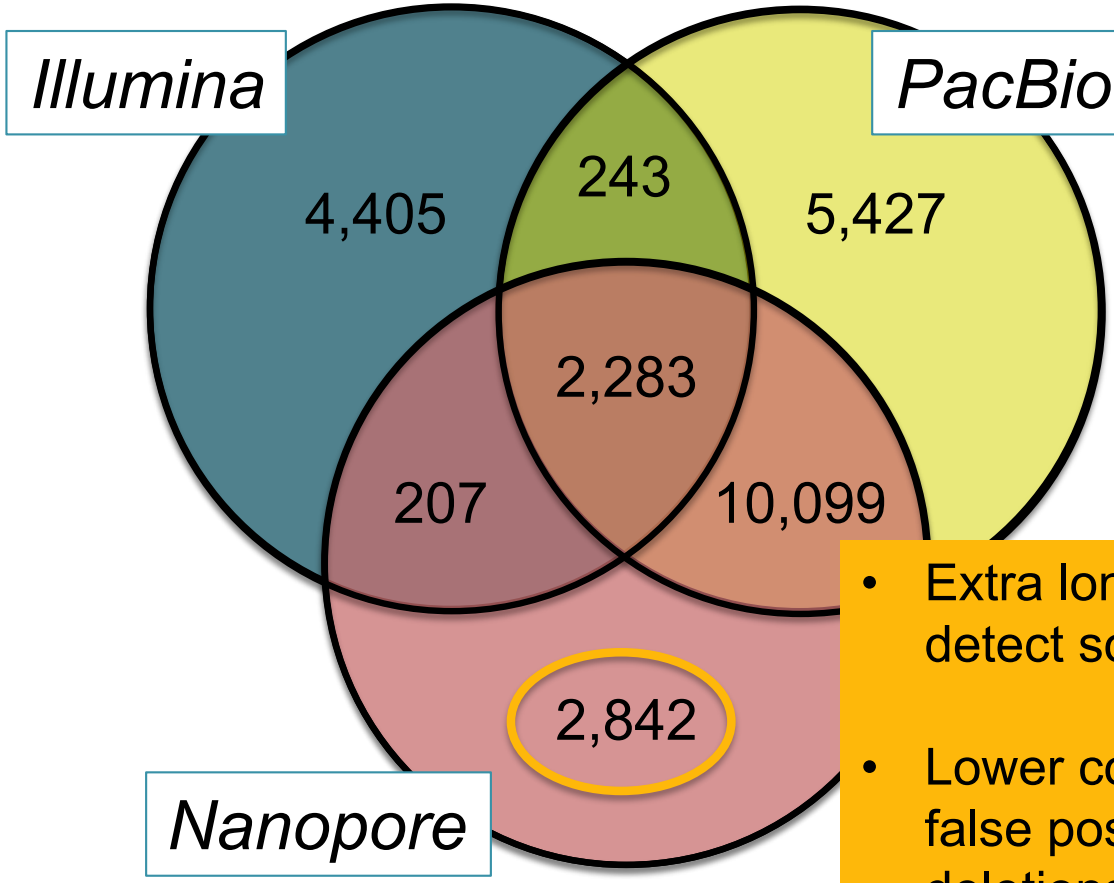
- Strong concordance between long read platforms
- Substantially more variants than detected by short reads

Structural Variant Comparison of SKBR3



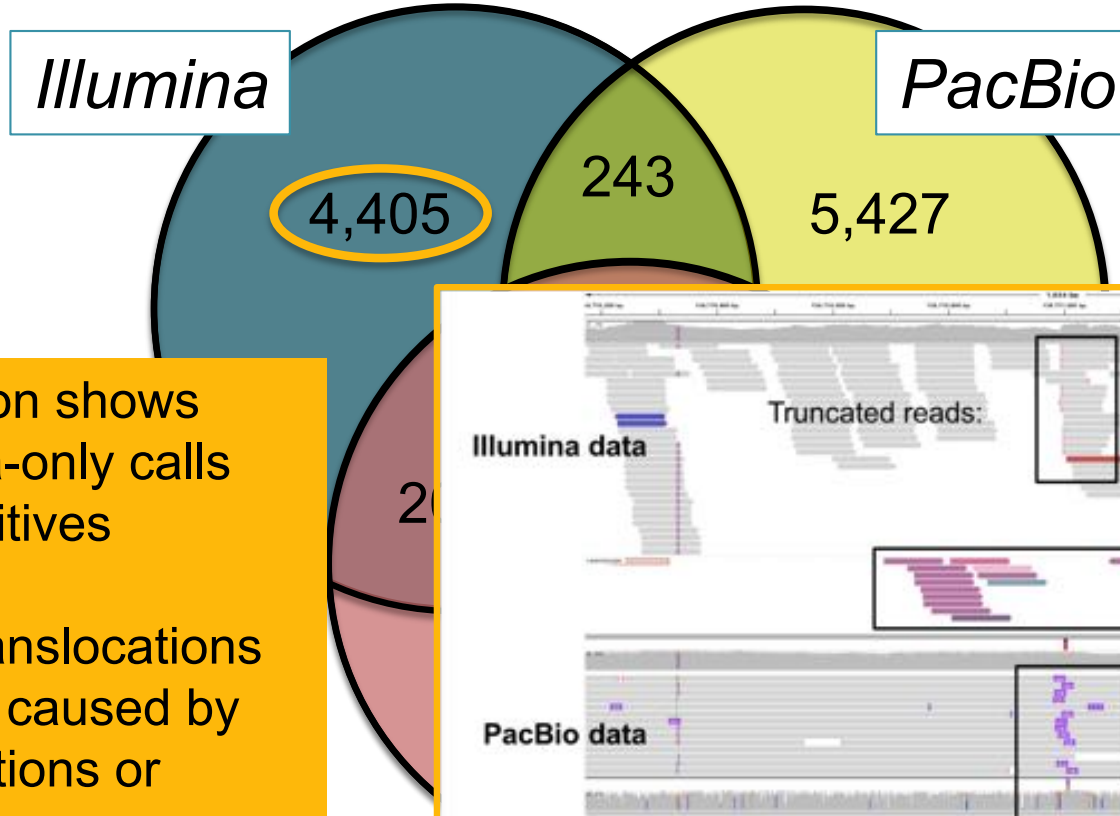
- Higher coverage available
- PCR validation shows low rate of false positives compared to Illumina

Structural Variant Comparison of SKBR3

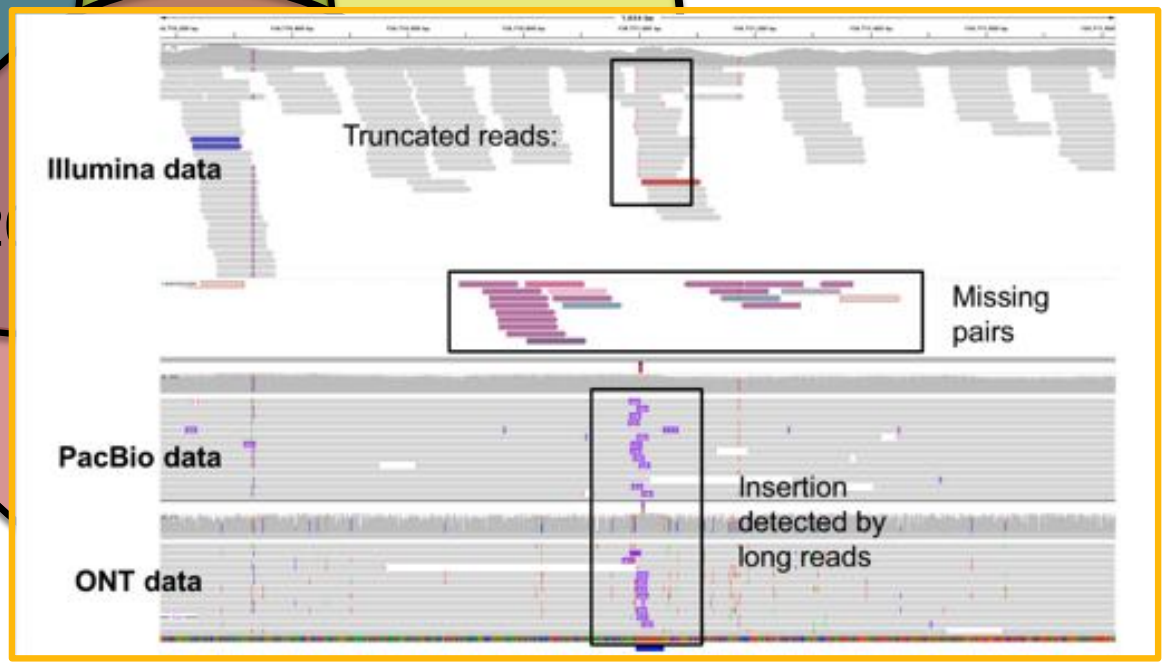


- Extra long read able to detect some novel variants
- Lower coverage increases false positives, especially deletions

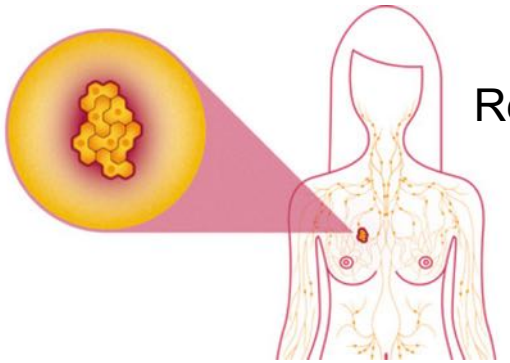
Structural Variant Comparison of SKBR3



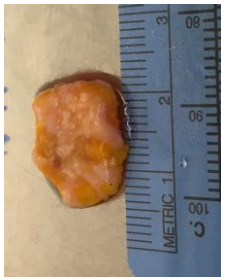
- PCR validation shows most Illumina-only calls are false positives
- Especially translocations or inversions caused by smaller insertions or deletions



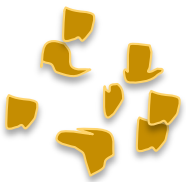
Taking Nanopore Sequencing into the Clinic



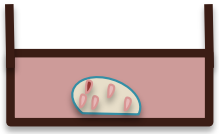
Resected primary tumor



Mechanical/ enzymatic



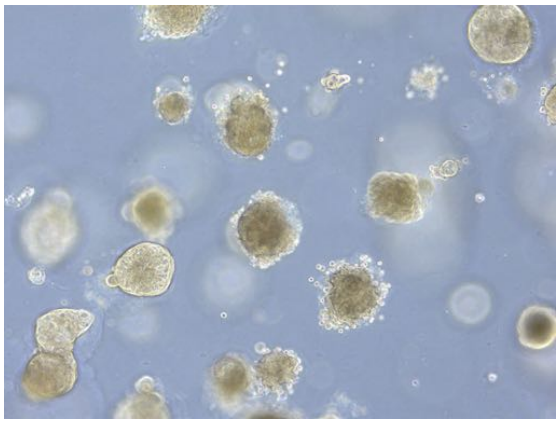
Purification



Plating on Matrigel

Add growth factors

Tumor organoids in culture



- ✓ Stable Growth in 3D
- ✓ Recapitulate tumor pathology & treatment response
- ✓ Maintenance of tissue/tumor heterogeneity
- ✓ "2017 Method of the Year" - Nature Methods

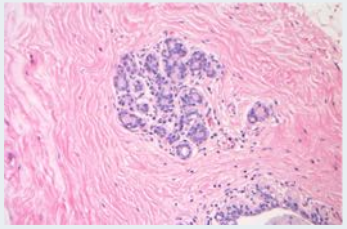
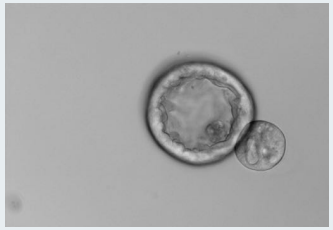
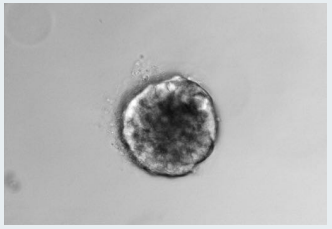
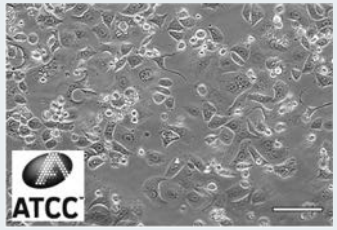
CSH

David Spector

Northwell Health

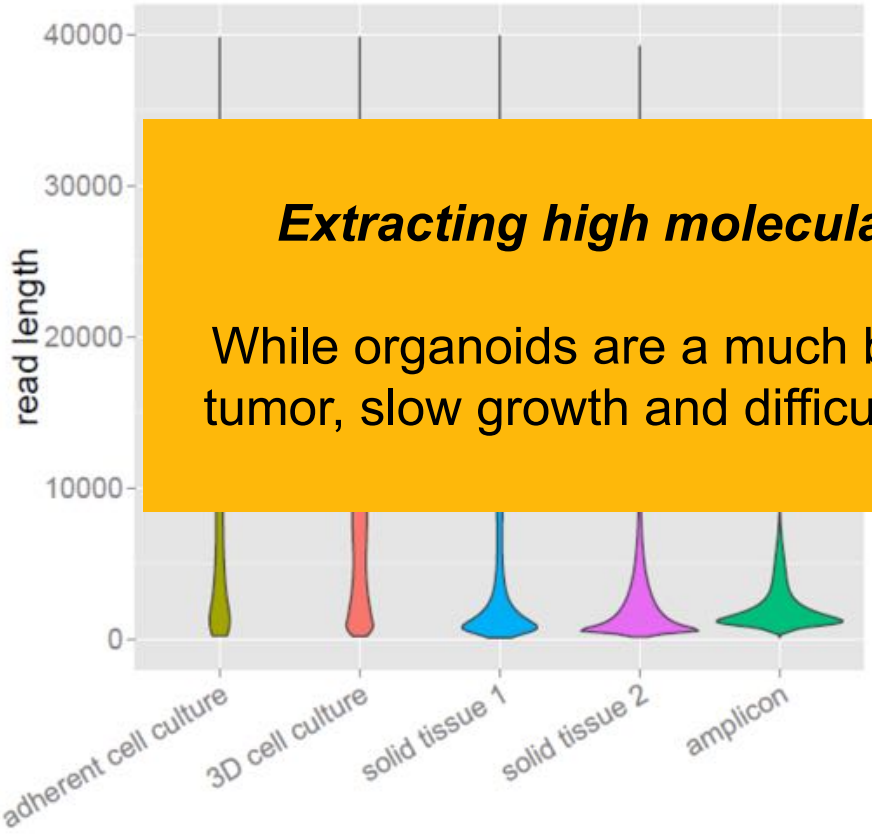
Karen Kostroff

Multi-omics Long Read Analysis of Cancer

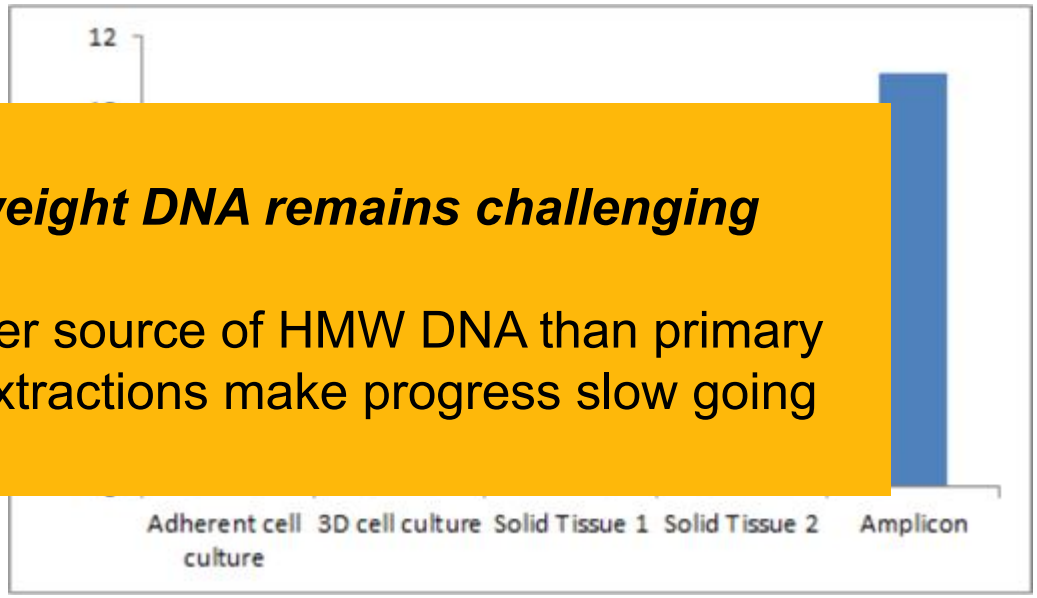
	Normal Breast Tissue	Normal Breast Organoid	Tumor Breast Organoid	SK-BR-3 Breast Cancer Cell Line
Oxford Nanopore WGS	**Y**	**N**	**Y**	**Y**
PacBio WGS	**N**	**N**	**N**	**Y**
ONT Methylation	**Y**	**N**	**Y**	**Y**
Illumina Methylation	**Y**	**N**	**Y**	**Y**
Illumina RNA-seq	**N**	**Y**	**Y**	**Y**
PacBio RNA-seq	**N**	**N**	**N**	**Y**
Pathology	**NA**	**NA**	**ER+, PR+, Her2-**	**ER-, PR-, Her2+**
Histology				
Image Source	Digital Atlas of Breast Pathology	David Spector, CSHL	David Spector, CSHL	ATCC

Oxford Nanopore Sequencing Results

Tissue source impacts read length



Tissue source impacts yield per flow cell



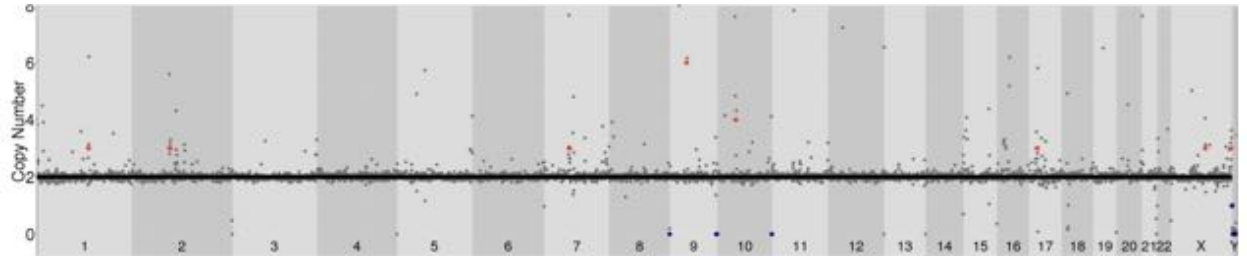
Extracting high molecular weight DNA remains challenging

While organoids are a much better source of HMW DNA than primary tumor, slow growth and difficult extractions make progress slow going

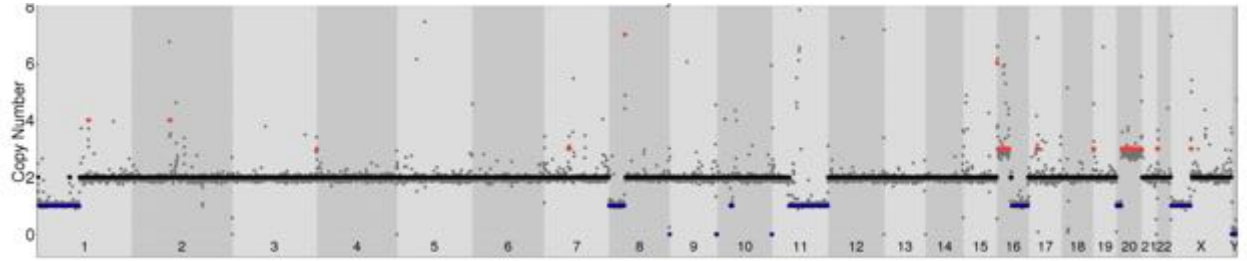
Sequencing by W. Richard McCombie @ CSHL
Winston Timp @ JHU

Copy Number Profiling with Long Reads

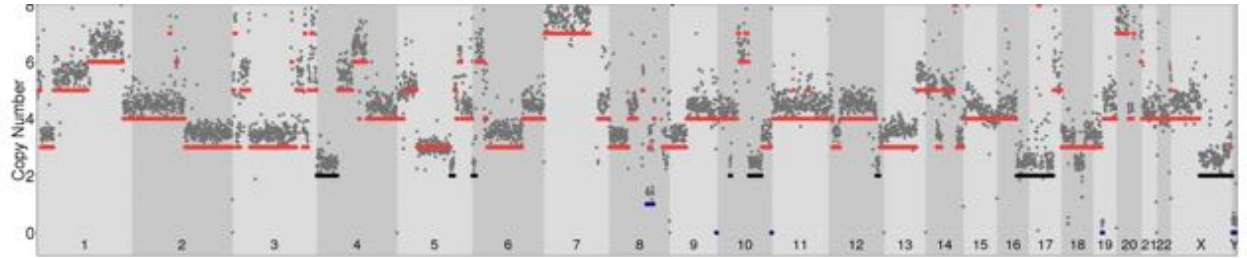
Normal Tissue (10x)



Tumor Organoid (25x)



SKBR3 Cell Line (30x)



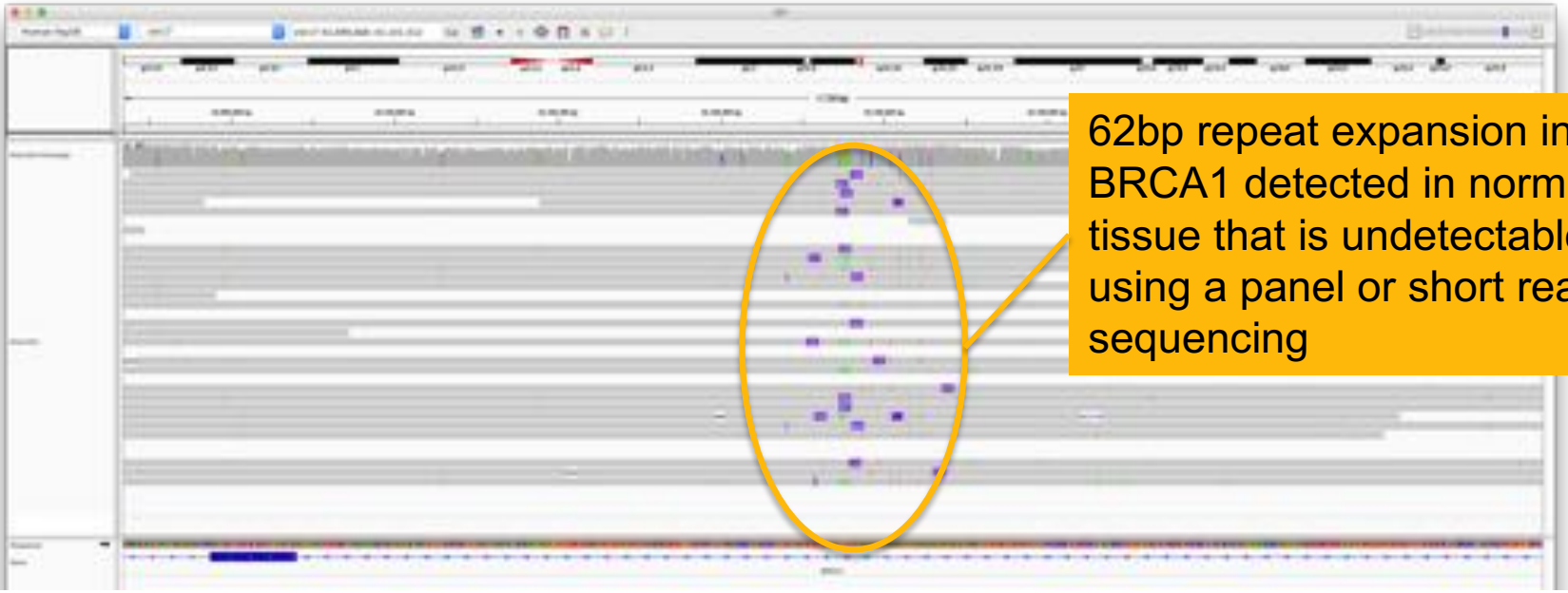
Preliminary Structural Variations Analysis

	Total	Deletions	Duplications	Insertions	Inversions	Translocations
All SVs in normal	9816	5225	578	3727	130	156
All SVs in tumor	13737	7020	988	5292	202	235
SVs only in tumor (Also exclude NA12878)	3662	1805	420	1250	98	89

3,662 SVs specific to the tumor, most are undetectable using short read or 10X sequencing

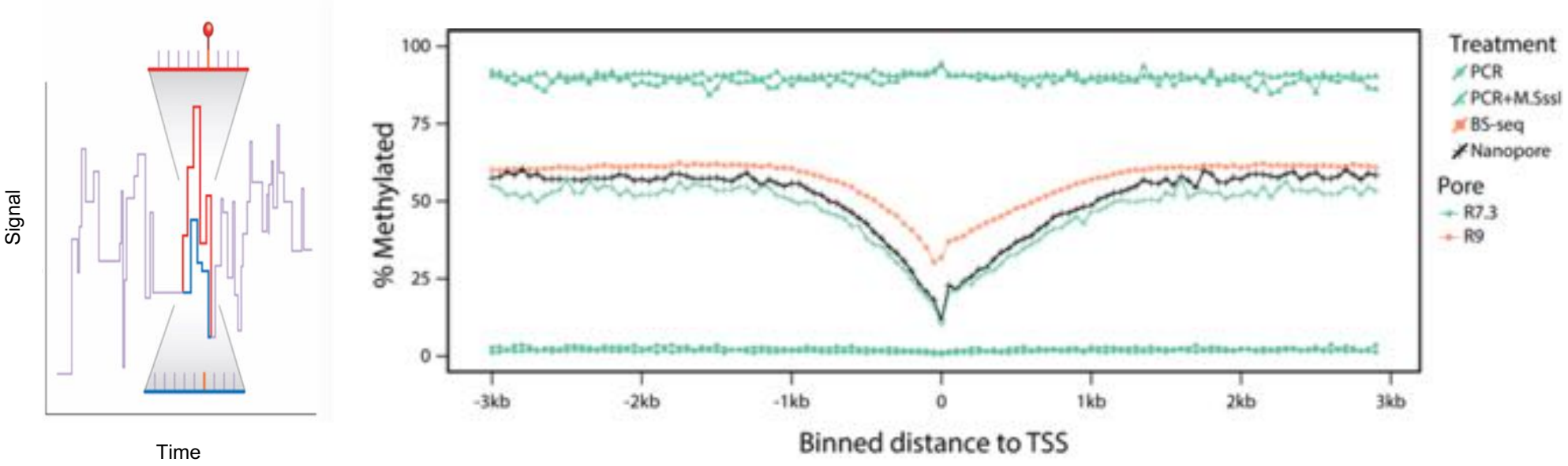
Preliminary Structural Variations Analysis

	Total	Deletions	Duplications	Insertions	Inversions	Translocations
All SVs in normal	9816	5225	578	3727	130	156
All SVs in tumor	13737	7020	988	5292	202	235
SVs only in tumor (Also exclude NA12878)	3662	1805	420	1250	98	89



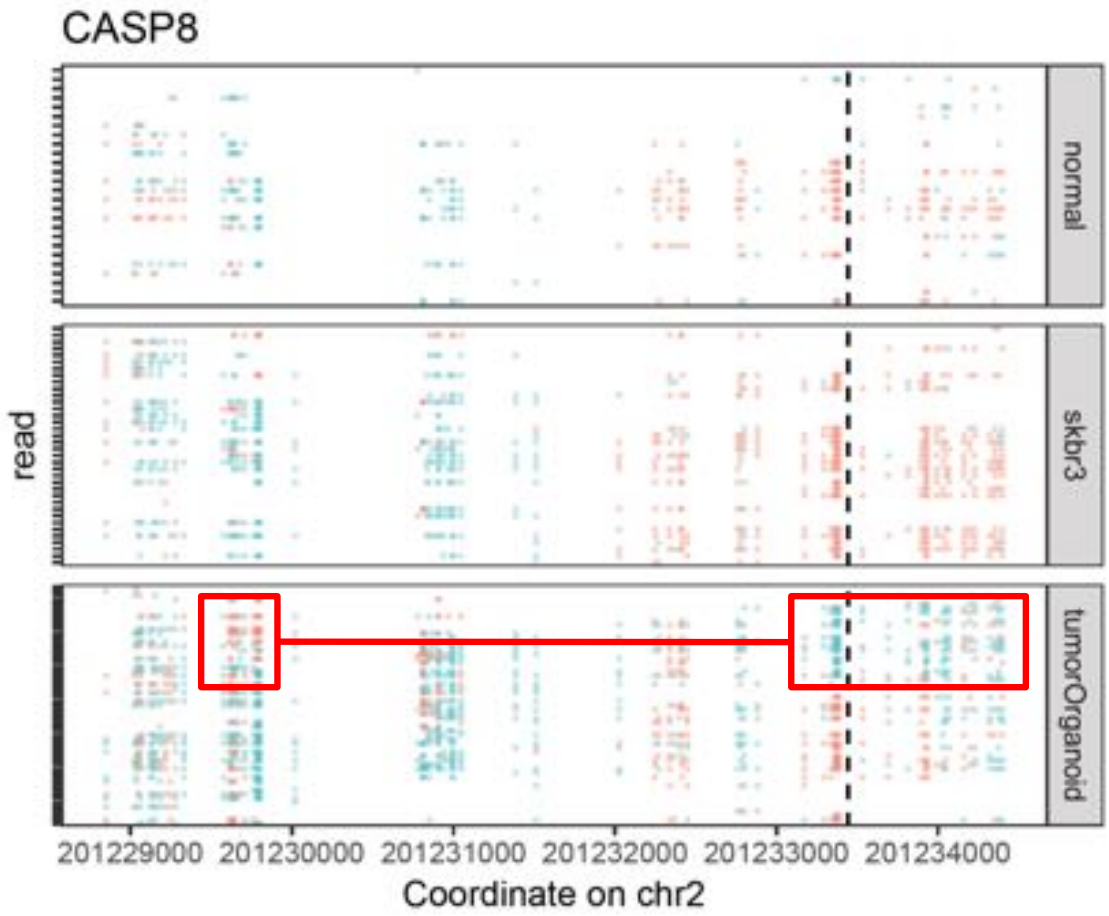
62bp repeat expansion in BRCA1 detected in normal tissue that is undetectable using a panel or short read sequencing

Differential Methylation of Cancer



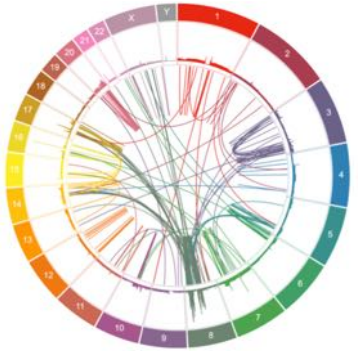
Detecting DNA cytosine methylation using nanopore sequencing ("Nanopolish")
Simpson *et al* (2017) *Nature Methods. doi:10.1038/nmeth.4184*

Long-range methylation correlation



Isac Lee

Summary and Future Work



Long reads are crucial for accurate SV calling

- Finding thousands to tens of thousands of additional SVs over short reads
- Resolves the false positives observed with short reads
- Detecting potential cancer risk factors that would otherwise go undetected

Long read platforms have matured significantly in the last few years

- PacBio and Oxford Nanopore producing similar length distributions
- Overcome error sequencing with improved informatics
- Oxford Nanopore exciting for methylation & direct RNA capabilities

Sample & DNA requirements one of the largest barriers for clinical application

- Continue to advance protocols for extracting, preparing samples
- Organoids (as opposed to primary tumors) enable large DNA amounts for long read sequencing, though it remains much more difficult then cell culture
- Organoids also enable application and profiling of other molecular and pharmaceutical assays

Moving quickly towards profiling many more patient samples!

Acknowledgements

Schatz Lab
Mike Alonge
Amelia Bateman
Charlotte Darby
Han Fang
Michael Kirsche
Sam Kovaka
Laurent Luo
Srividya Ramakrishnan
T. Rhyker Ranallo-Benavide
Your Name Here

Baylor Medicine
Fritz Sedlazeck

University of Vienna
Arndt von Haeseler
Philipp Rescheneder

DNAnexus
Maria Nattestad

CSHL
Gingeras Lab
Jackson Lab
Lippman Lab
Lyon Lab
Martienssen Lab
McCombie Lab
Spector Lab
Tuveson Lab
Ware Lab
Wigler Lab

SBU
Skiena Lab
Patro Lab

GRC
Roderic Guido
Alessandra Breschi
Anna Vlasova

Yale
Gerstein Lab

JHU
Battle Lab
Langmead Lab
Leek Lab
Salzberg Lab
Taylor Lab
Timp Lab
Wheelan Lab

Cornell
Susan McCouch
Lyza Maron
Mark Wright

OICR
John McPherson
Karen Ng
Timothy Beck
Yogi Sundaravadanam

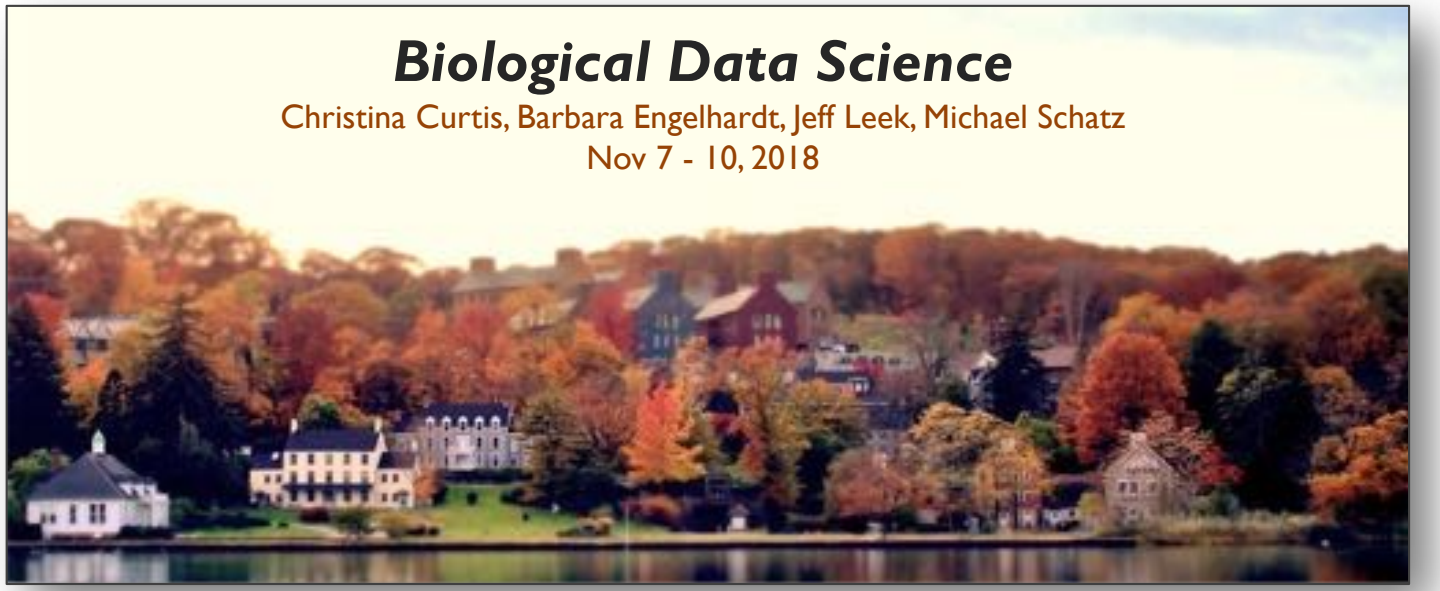
Northwell Health
Karen Kostroff

NSF

National Human Genome Research Institute

NIH NATIONAL CANCER INSTITUTE

ALFRED P. SLOAN FOUNDATION



Thank you

http://schatz-lab.org

@mike_schatz