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

W. Richard McCombie

Disclosures

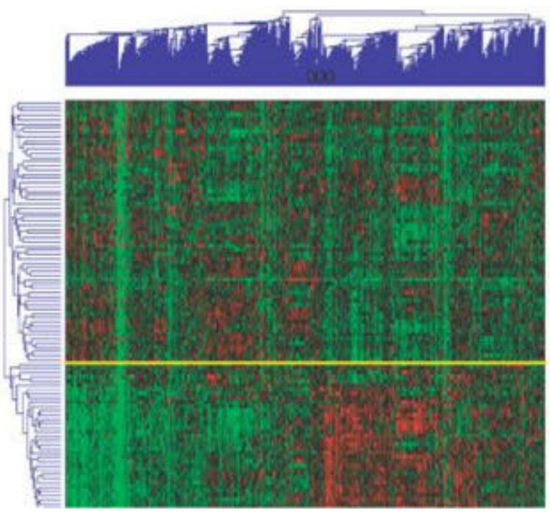
Orion Genomics – Founder and Shareholder
Cancer epigenetics and plant genomics

Previously Compensated Speaker for Illumina, Inc.

Previously Compensated Speaker for Pacific Biosciences, Inc.

Evolution of Cancer Genomics

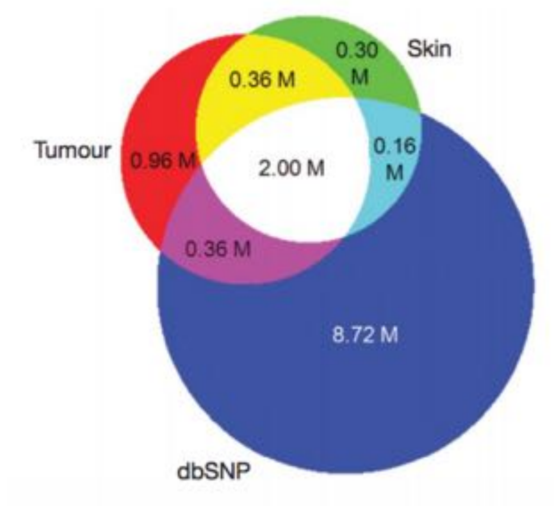
Microarray Profiling



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

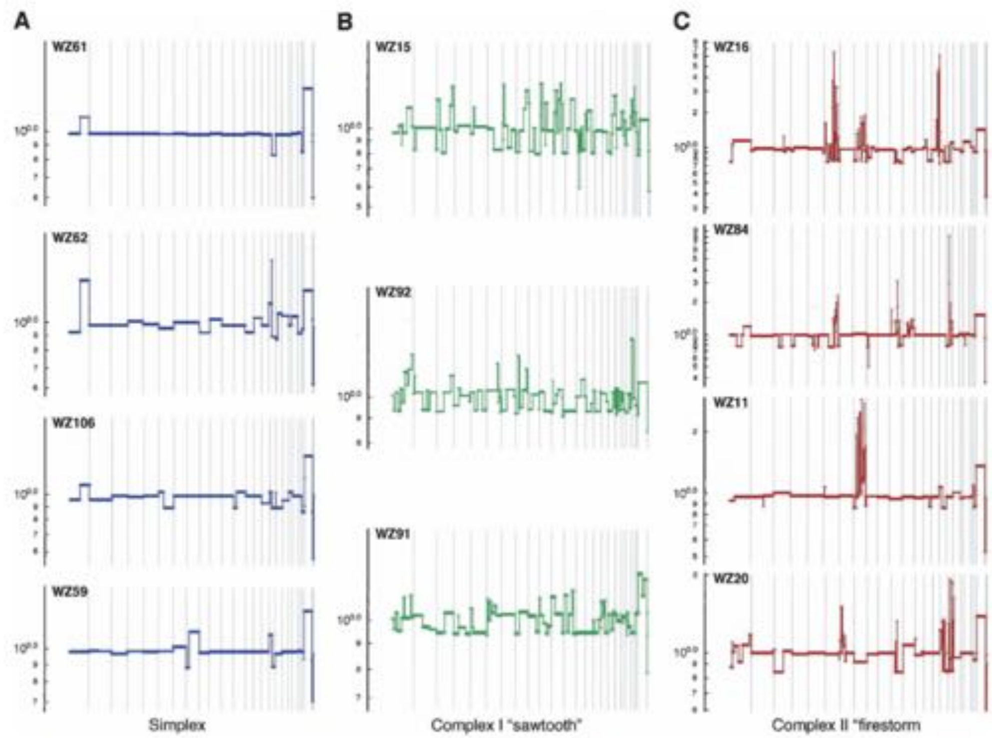


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The *y*-axis displays 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

Prognostic indicator

Greater genome instability generally leads to worse patient outcomes

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 sequencing but misses the vast majority of SVs

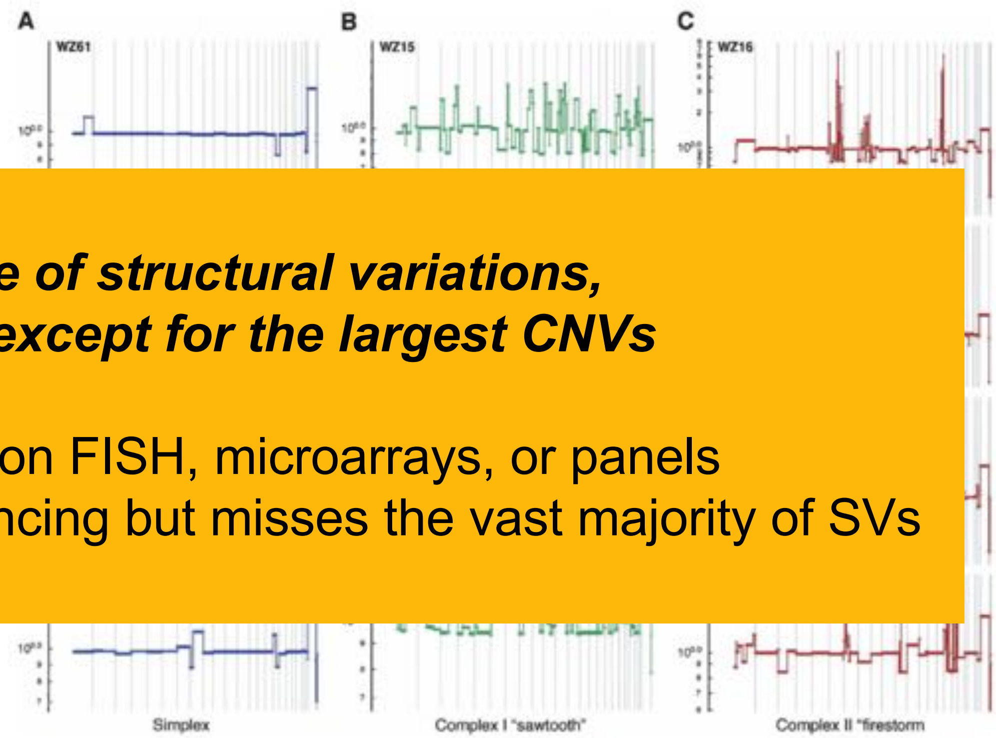
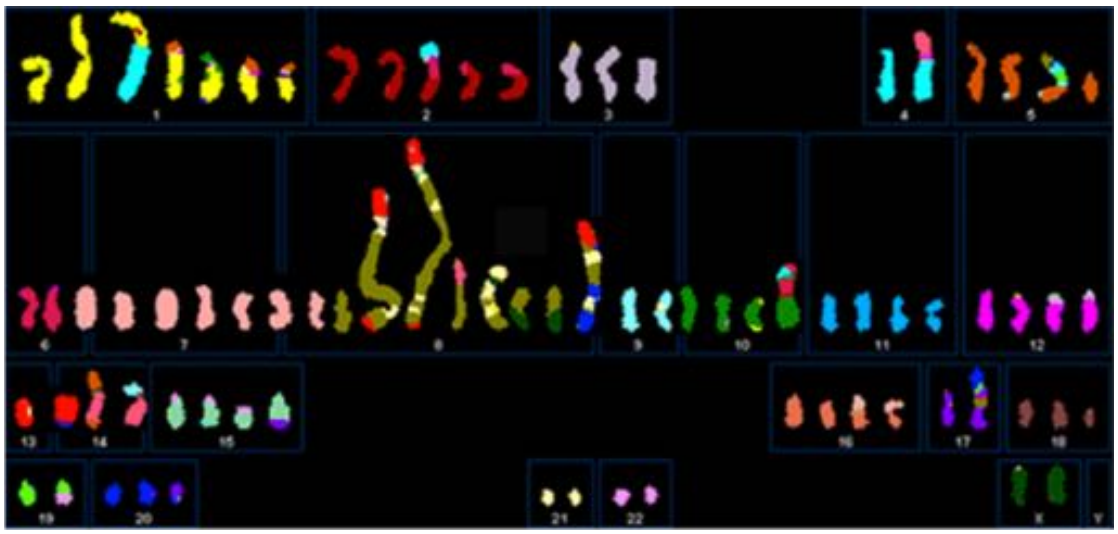


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The y-axis displays 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.

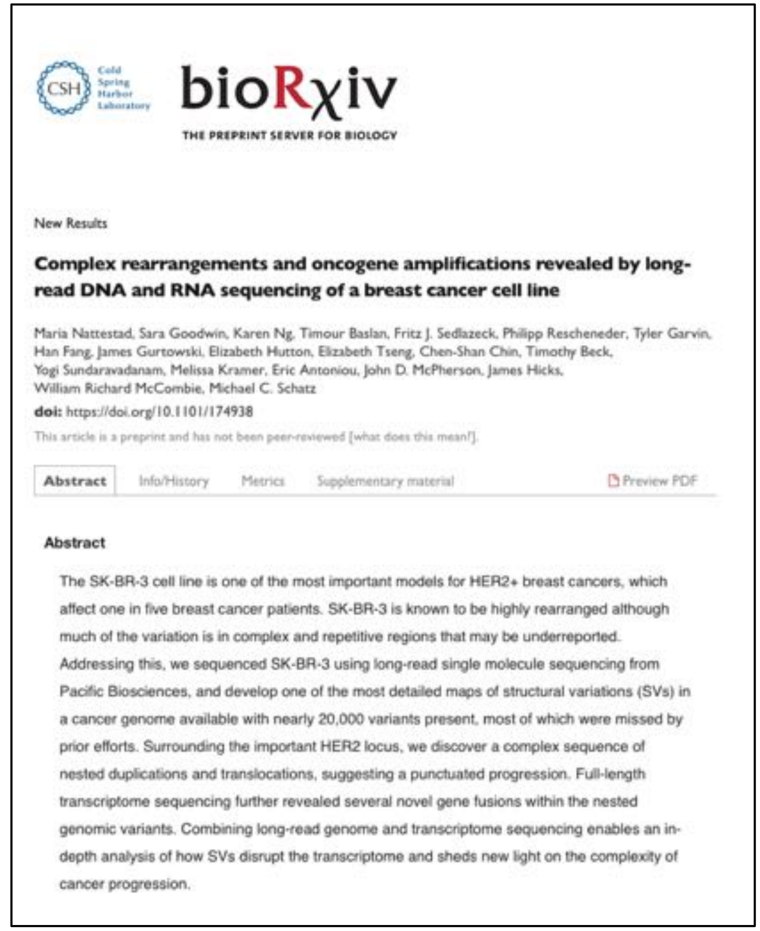
(Hicks *et al,* 2006, Genome Research)

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)



CSH Cold Spring Harbor Laboratory

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

This article is a preprint and has not been peer-reviewed [what does this mean?].

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

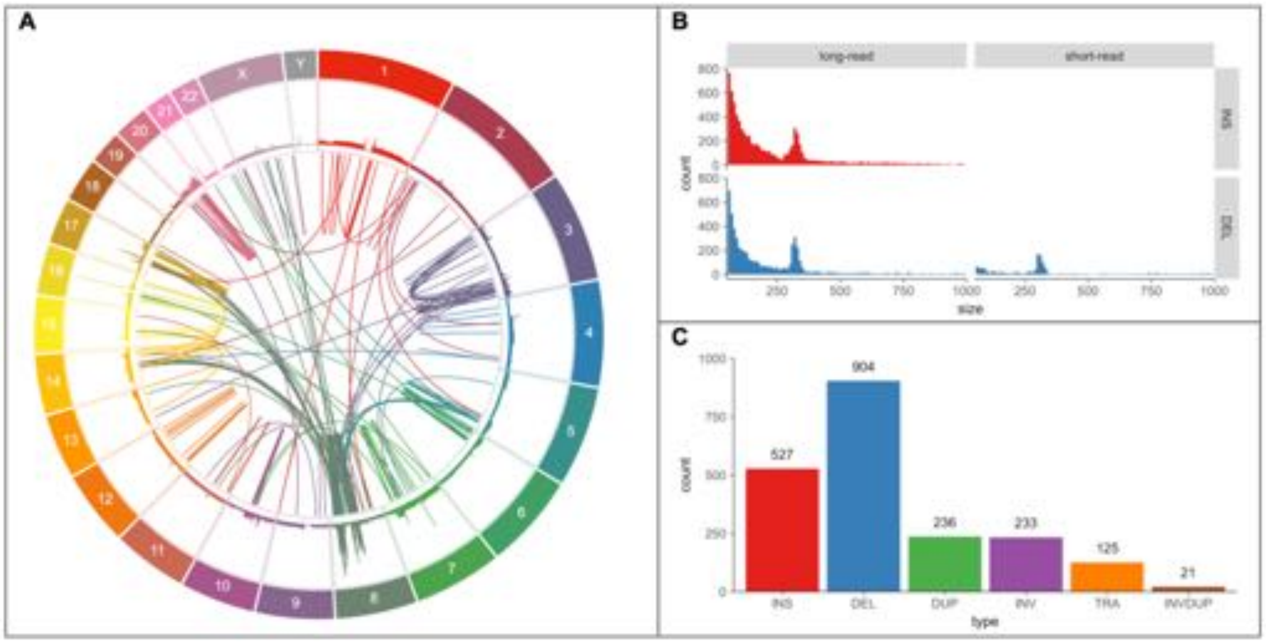


Figure 1 | Variants found in SK-BR-3 with PacBio long-read sequencing. (A) Circos plot showing long-range (larger than 10 kbp or interchromosomal) variants found by Sniffles from split-read alignments, with read coverage shown in the outer track. (B) Variant size histogram of deletions and insertions 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.

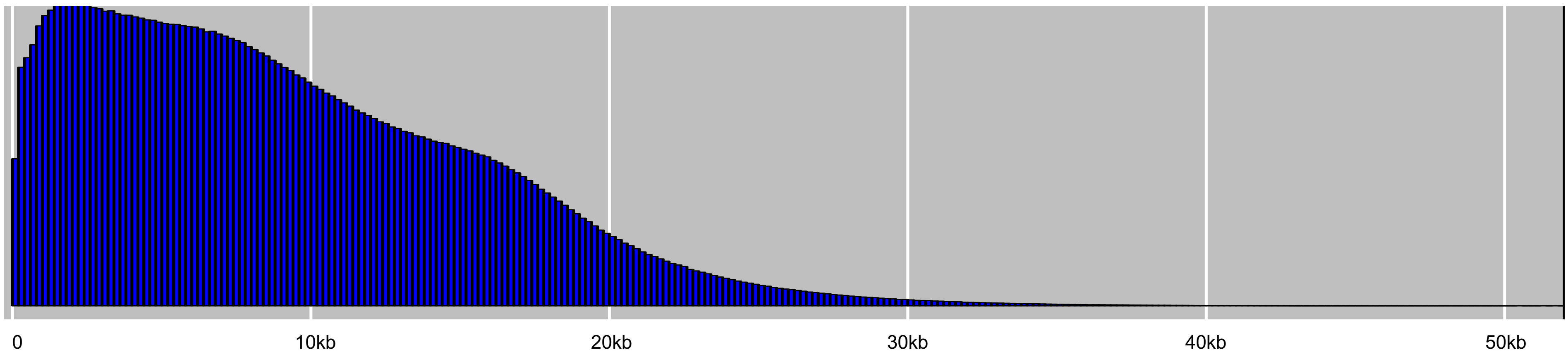
- Finding 10s of thousands of additional variants in the cancer
- PCR validation confirms high accuracy of long read calls
- With improved SV analysis, can infer the progression of the cancer
- Detect many novel gene fusions

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

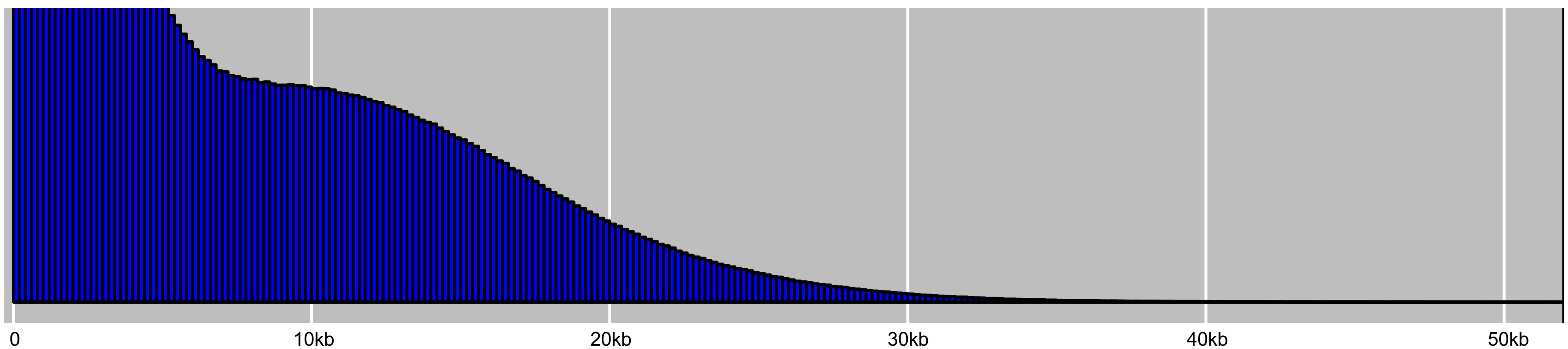
Nattestad, M et al (2017) bioRxiv https://doi.org/10.1101/174938

Long Read Sequencing of SKBR3

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

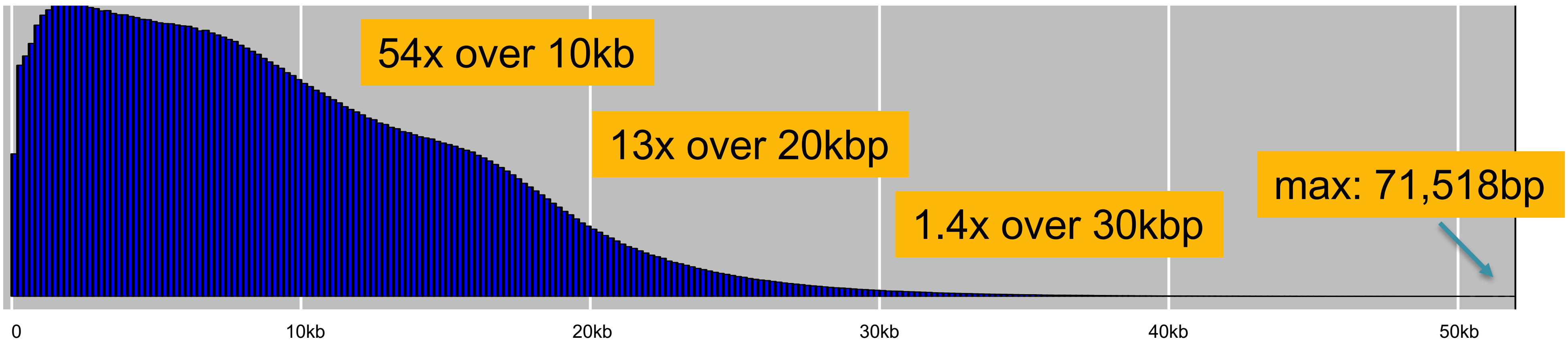


Oxford Nanopore GridION: 13.6M reads, 31.8X coverage, n50=13,350 bp

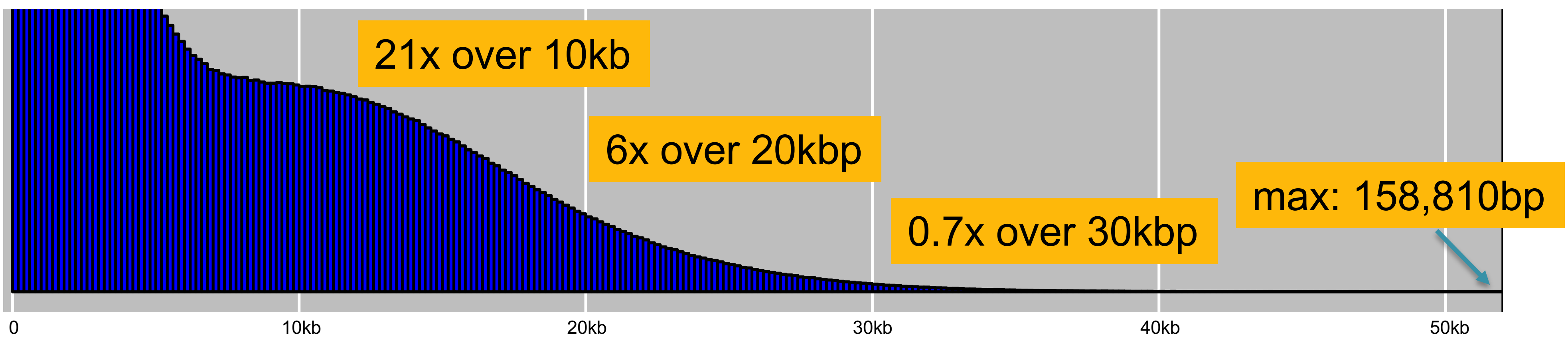


Long Read Sequencing of SKBR3

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

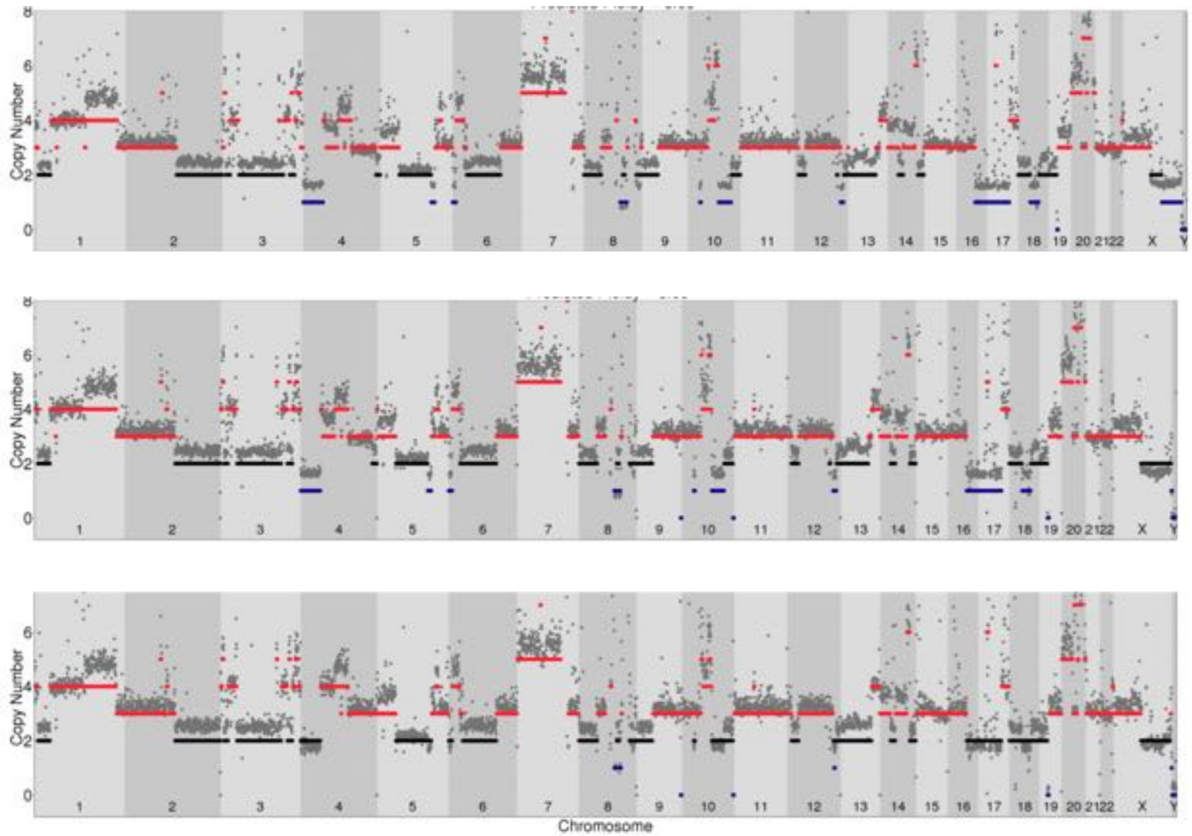


Consistent Profiles of Megabase CNVs

illumina®

PACBIO®

Oxford NANOPORE Technologies

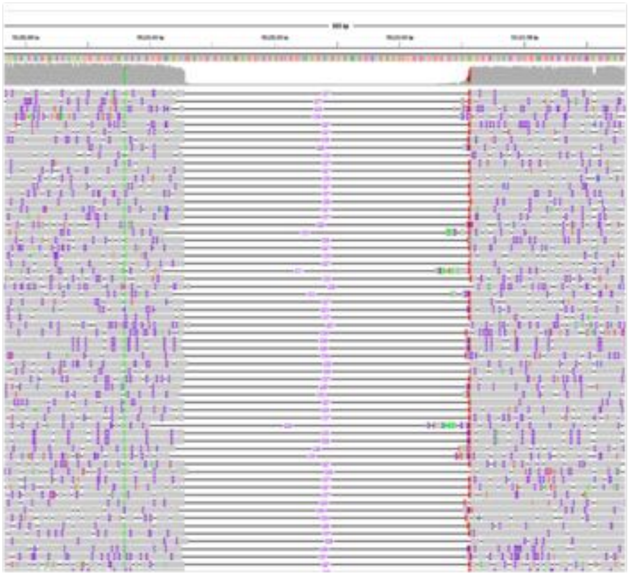


Interactive analysis and assessment of single-cell copy-number variations ("Ginkgo")
Garvin, Aboukhalil, *et al.* (2015) Nature Methods doi:10.1038/nmeth.3578

Structural Variation Identification with Long Reads

BWA-MEM:

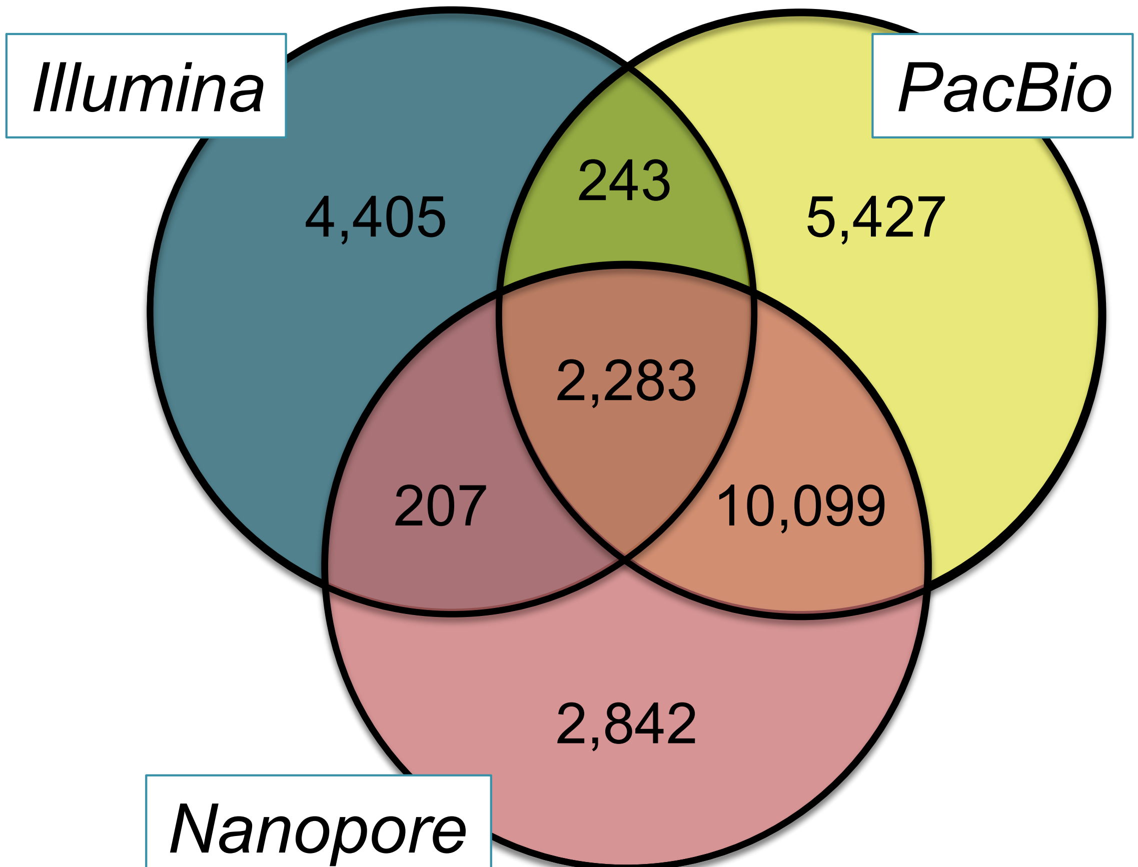
NGMLR:



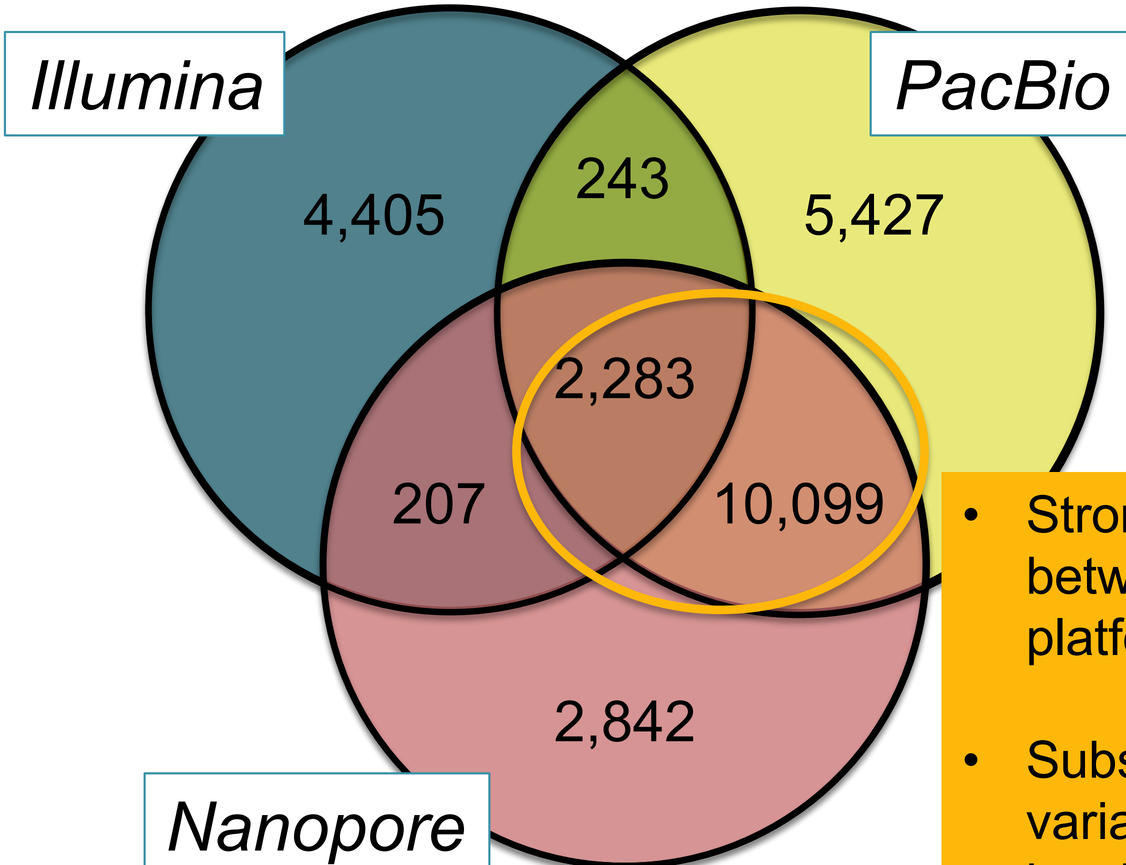
NGMLR: Dual mode scoring to accommodate many small gaps from sequencing errors along with less frequent but larger SVs

Accurate detection of complex structural variations using single molecule sequencing
Sedlazeck, Rescheneder *et al* (2017) *bioRxiv https://doi.org/10.1101/169557*

Structural Variant Comparison of SKBR3

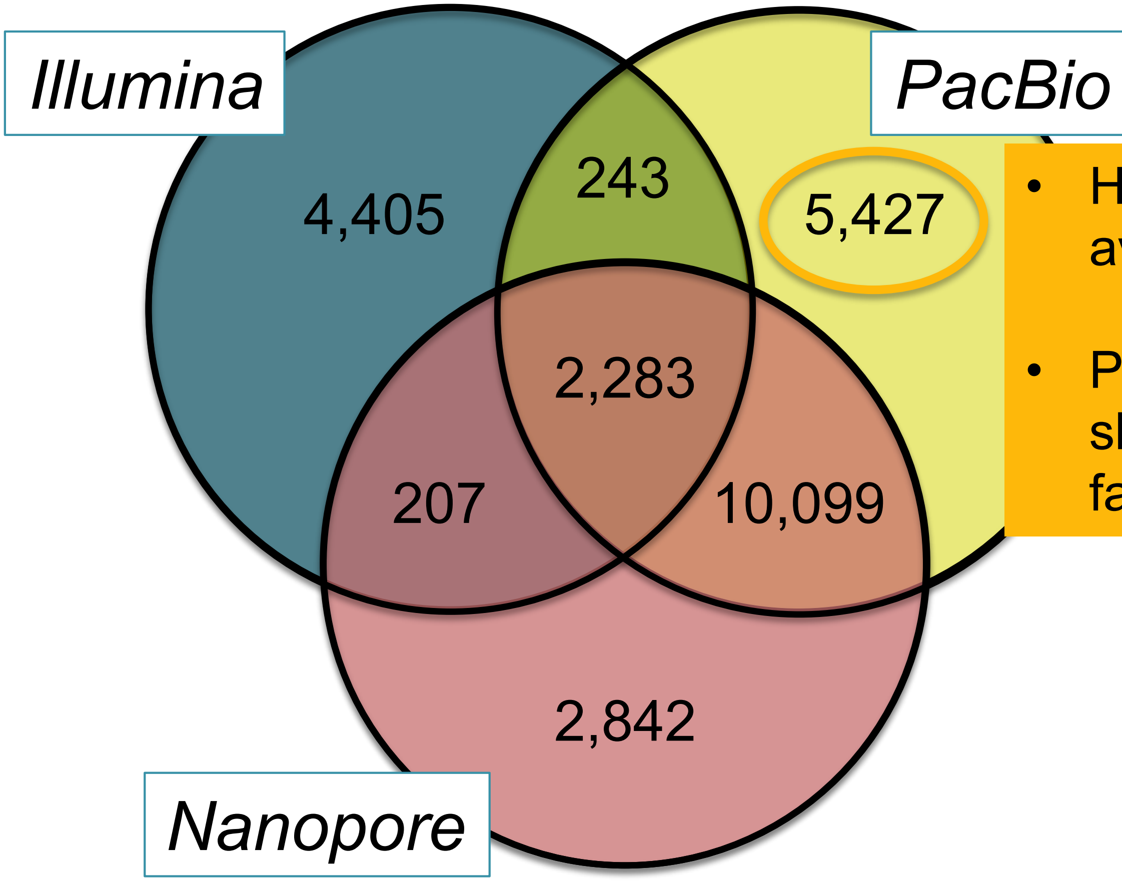


Structural Variant Comparison of SKBR3



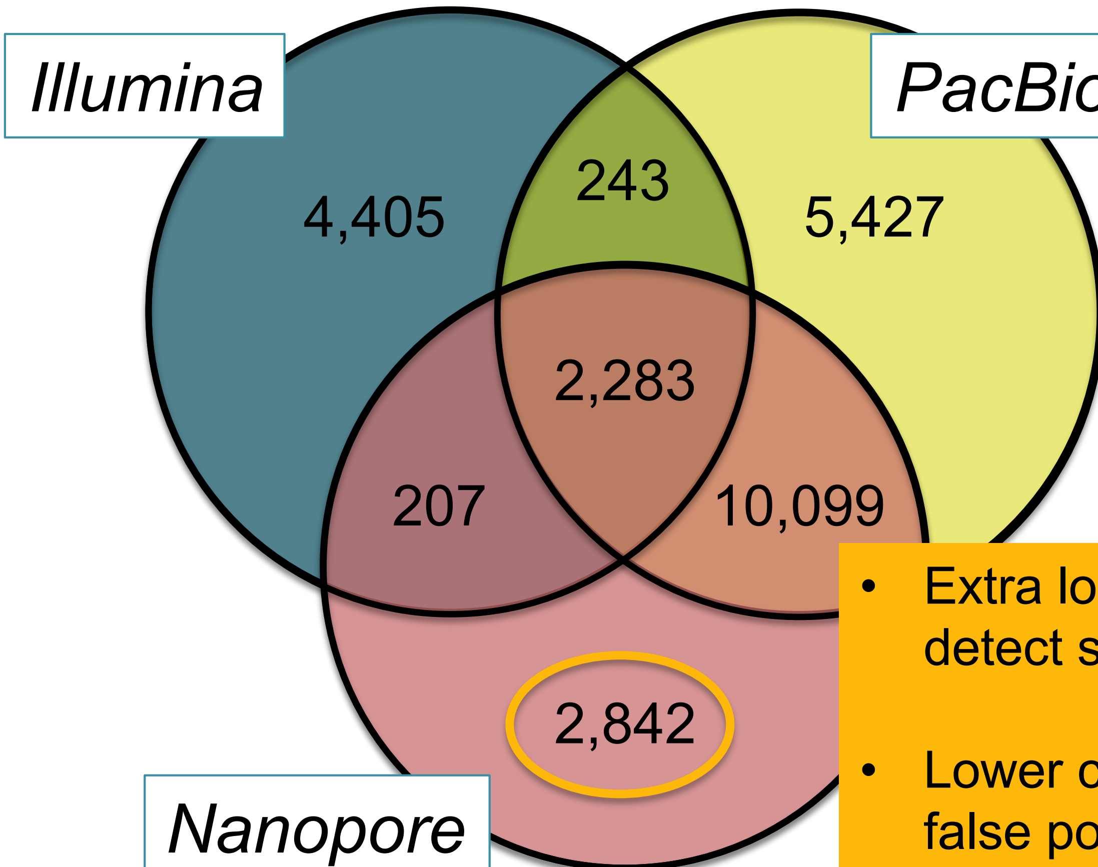
- Strong concordance between long read platforms
- Substantially more variants than detected by short reads

Structural Variant Comparison of SKBR3



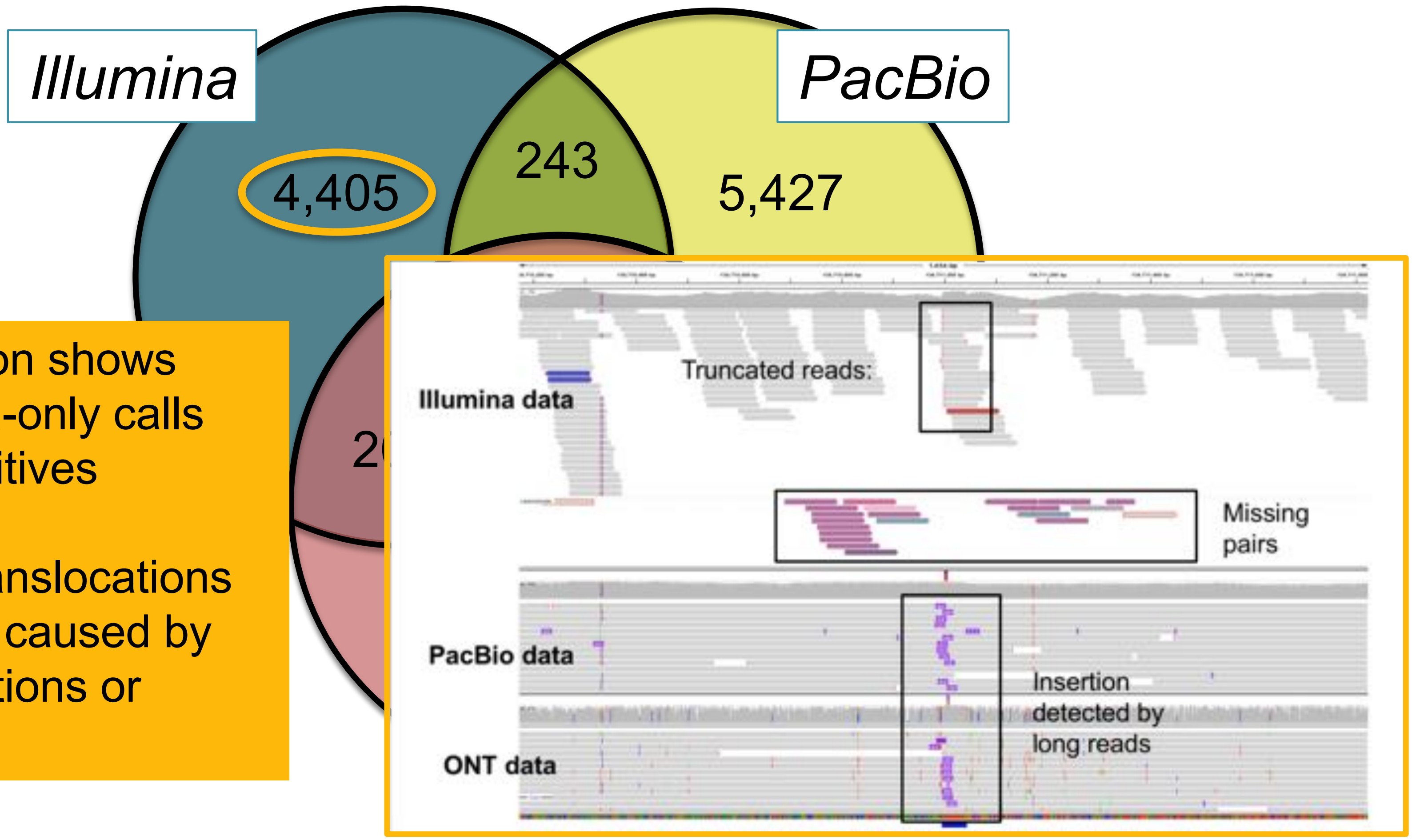
- Higher coverage available
- PCR validation shows low rate of false positives

Structural Variant Comparison of SKBR3



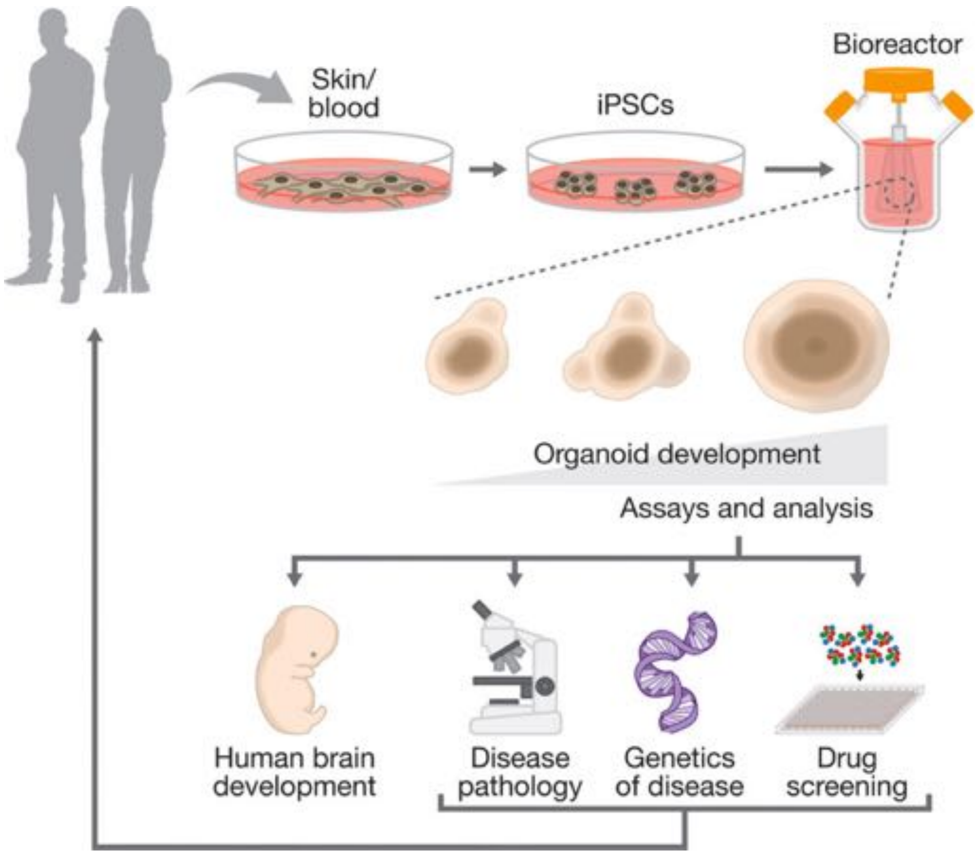
- Extra long read able to detect some novel variants
- Lower coverage increases false positives, especially deletions

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

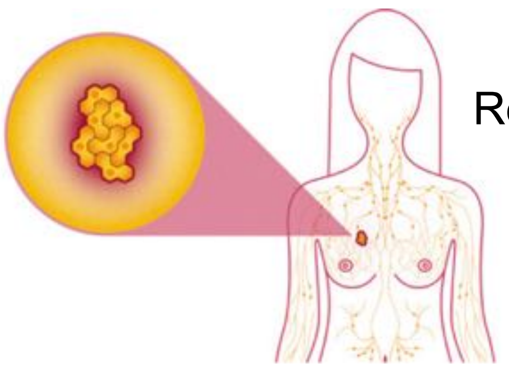
Organoids are an improved model for cancer



(Arlotta, Nature Methods, 2018)

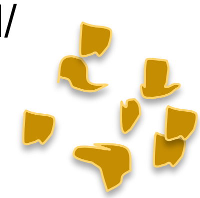
“Method of the Year” 2017

Generation of patient derived organoids

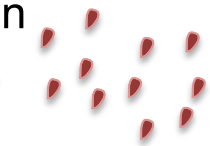


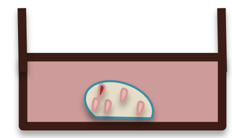
Resected primary tumor

Mechanical/ enzymatic



Purification

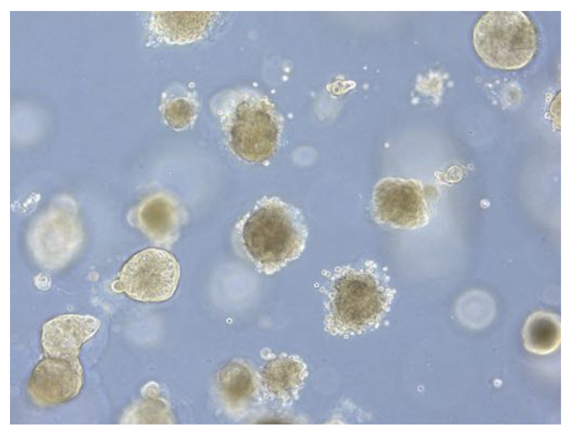




Plating on Matrigel

Add growth factors

Tumor organoids in culture

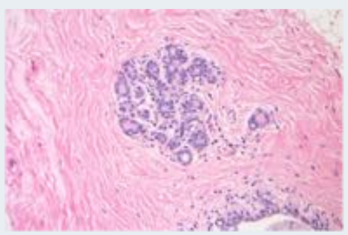
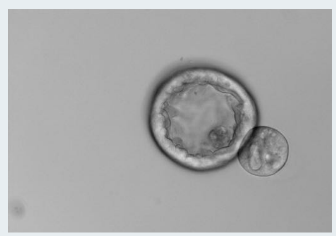
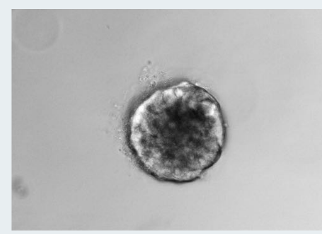
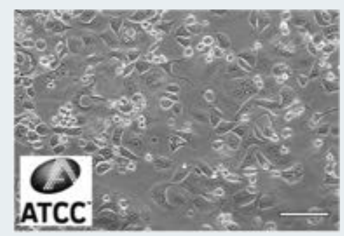


- ✓ Growth in 3D
- ✓ Stable genotype
- ✓ Recapitulate tumor pathology
- ✓ Maintenance of tissue/tumor heterogeneity

Northwell Health

Karen Kostroff

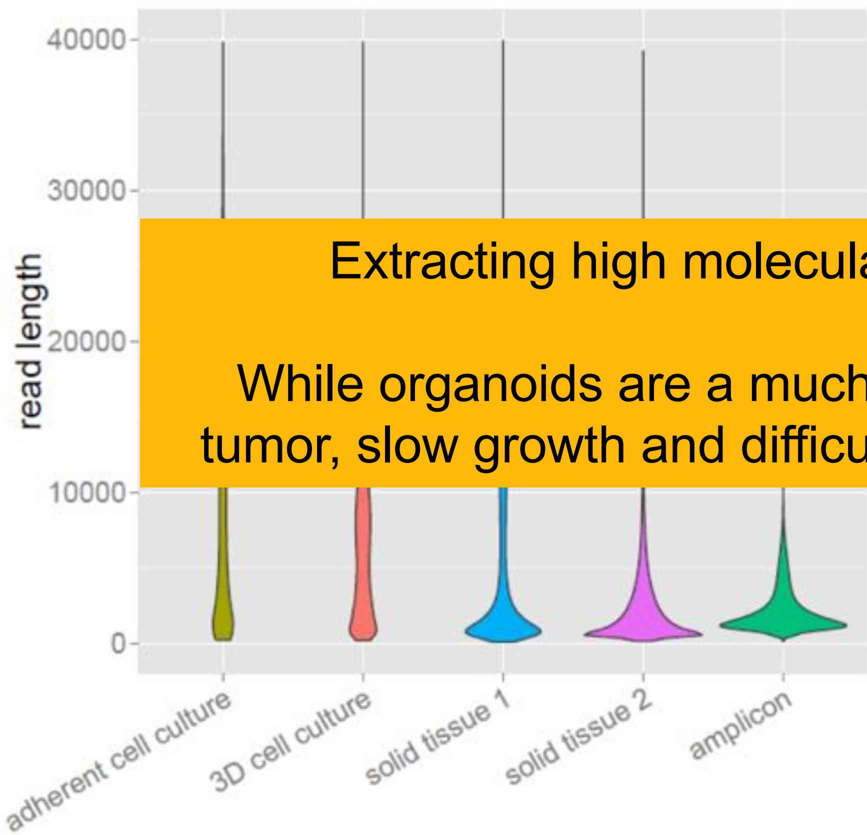
Multi-omics Long Read Analysis of Cancer

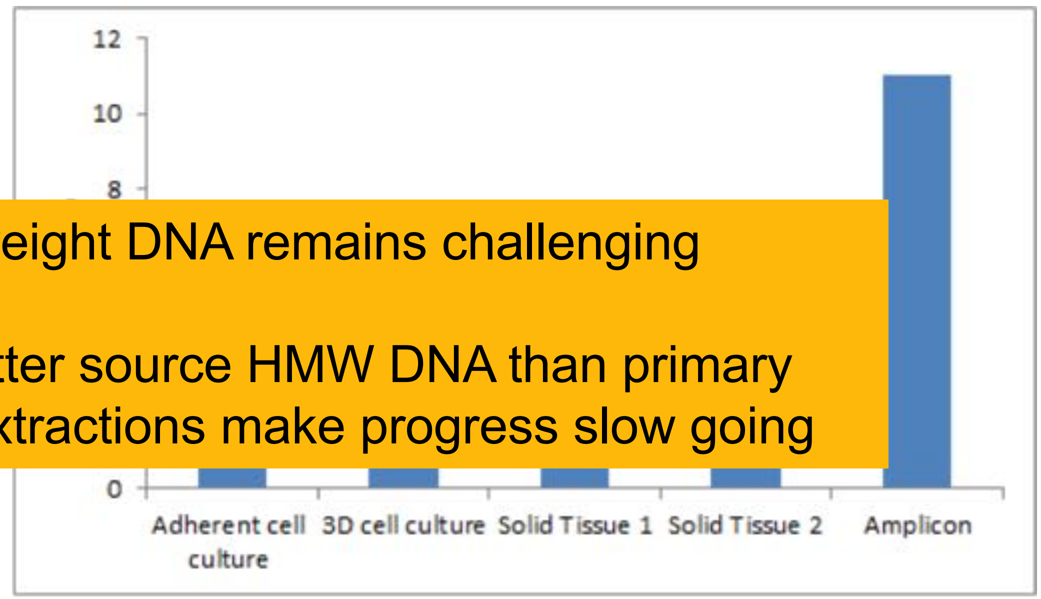
	Normal Breast Tissue	Normal Breast Organoid	Tumor Breast Organoid	SK-BR-3 Breast Cancer Cell Line
Oxford Nanopore WGS	**Y**	**N**	**Y**	**Y**
PacBio WGS	**N**	**N**	**N**	**Y**
ONT Methylation	**Y**	**N**	**Y**	**Y**
Illumina Methylation	**Y**	**N**	**Y**	**Y**
Illumina RNA-seq	**N**	**Y**	**Y**	**Y**
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



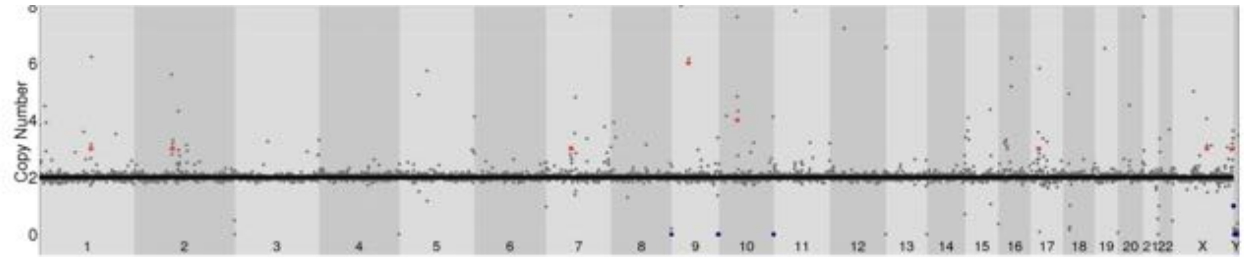


Extracting high molecular weight DNA remains challenging

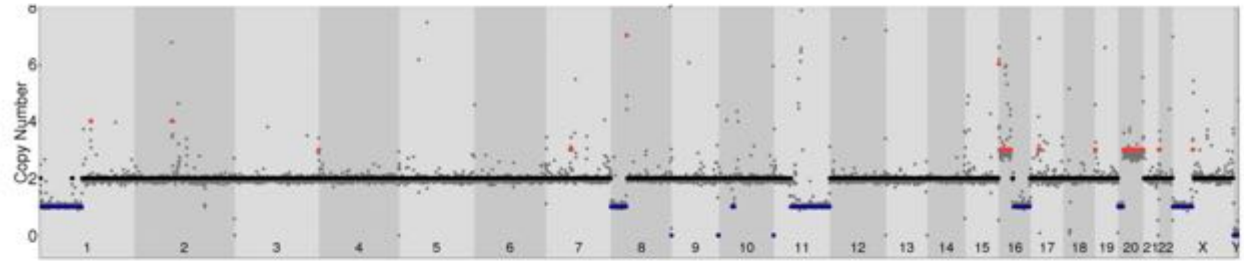
While organoids are a much better source HMW DNA than primary tumor, slow growth and difficult extractions make progress slow going

Copy Number Profiling with Long Reads

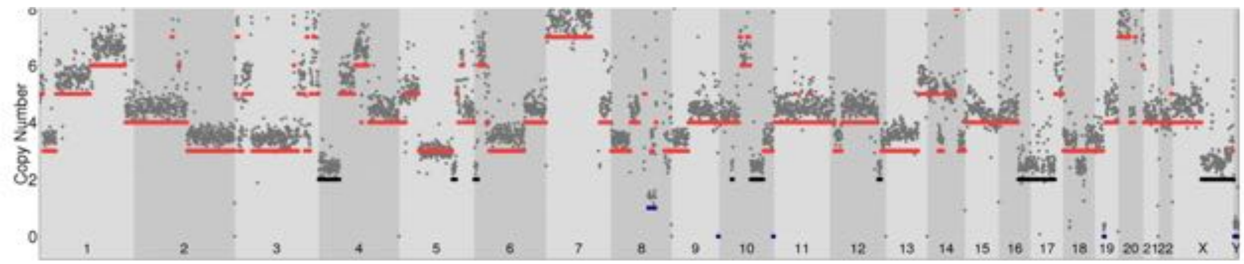
Normal Tissue



Tumor Organoid



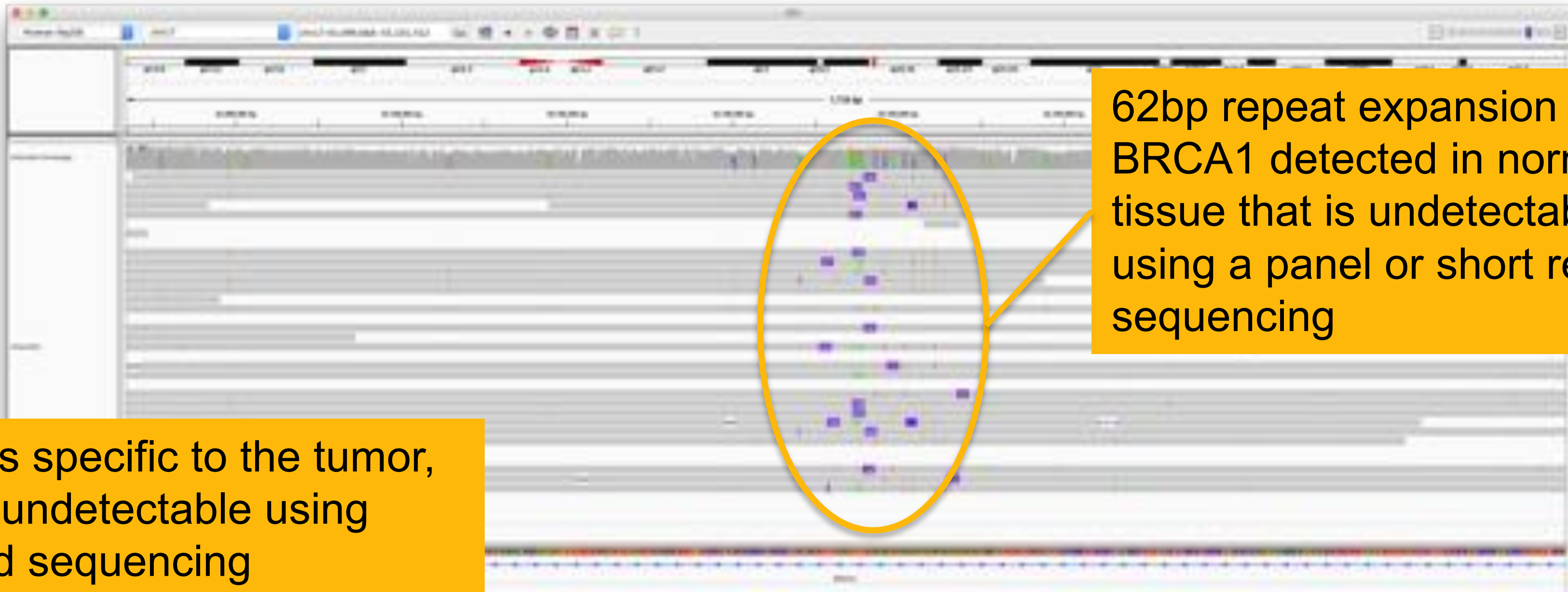
SKBR3 Cell Line



Preliminary Structural Variations Analysis

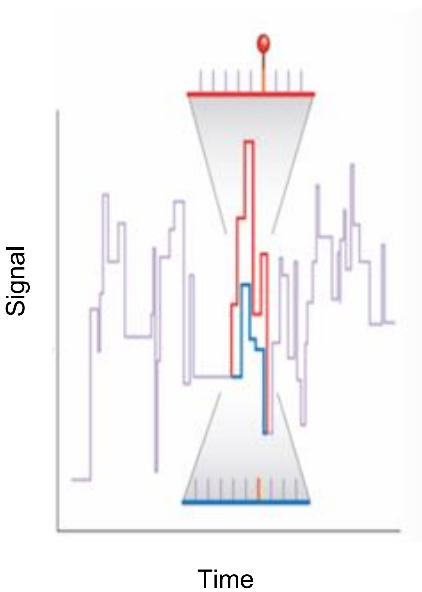
	Total	Deletions	Duplications	Insertions	Inversions	Translocations
All SVs in normal	9816	5225	578	3727	130	156
All SVs in tumor	13737	7020	988	5292	202	235
SVs only in tumor (Also exclude NA12878)	3662	1805	420	1250	98	89

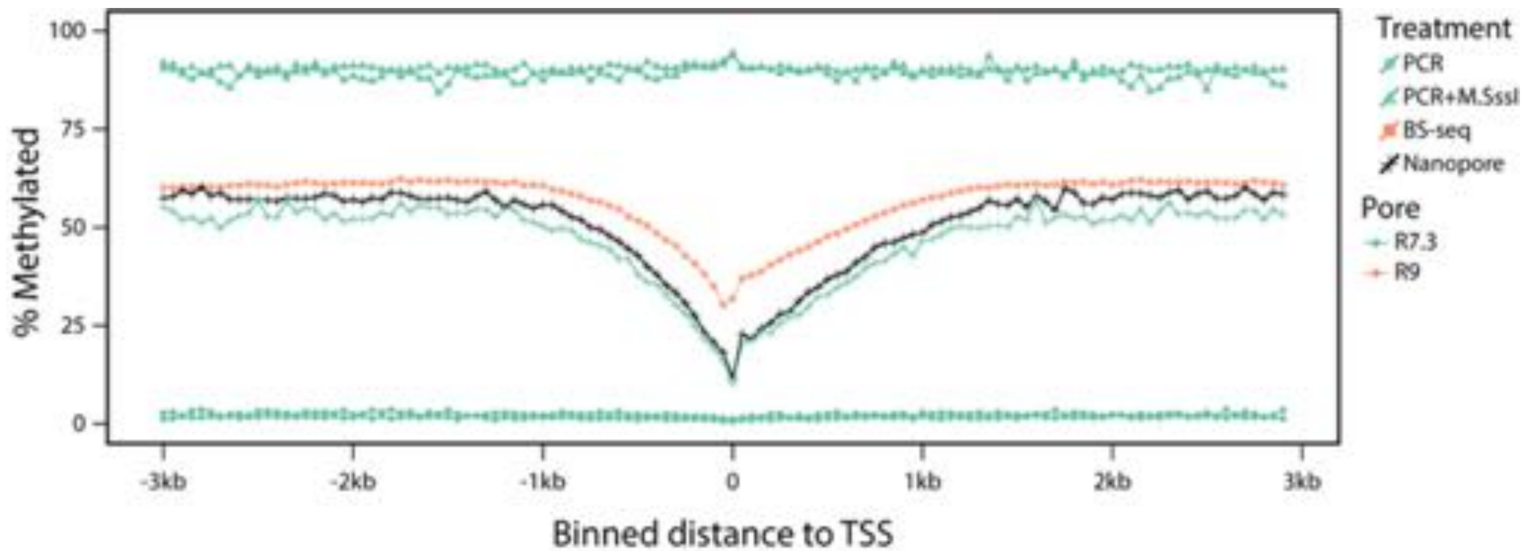
Preliminary Structural Variations Analysis



	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

Differential Methylation of Cancer





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

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 unnoticed

Methylation data can be derived from raw Oxford Nanaopore reads
- There is good concordance between Illumina and ONT methods for modified base detection
- Several oncogenes, including GATA3, show differential methylation patterns between tumor and Organoid
- Several genes with differential expression levels between tumor and normal, including WNT5B and BLC11B, have been identified

Long read platforms have matured significantly in the last few years
- PacBio and Oxford Nanopore producing similar length distributions
- Overcome high 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, including normal organoids

Acknowledgements

CSH

Baylor College of Medicine

McCombie Lab
Sara Goodwin
Melissa Kramer
Robert Wappel

Spector Lab
Gayatri Arun

Schatz Lab
Sam Kovaka
Michael Kirsche

Timp Lab
Isac Lee

Fritz Sedlazeck

Karen Kostroff

NIH NATIONAL CANCER INSTITUTE

National Human Genome Research Institute

NSF

Thank you!