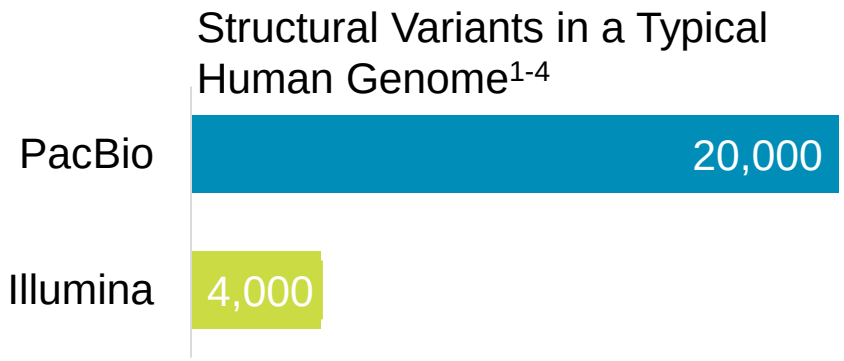
PACIFIC BIOSCIENCES®

Structural variation detection with long read sequencing in a melanoma cell line

Towards a method for robust discovery of patterns of structural variation in cancer

DO WE HAVE THE COMPLETE PICTURE OF ALL THE VARIANTS THAT DRIVE CANCER?

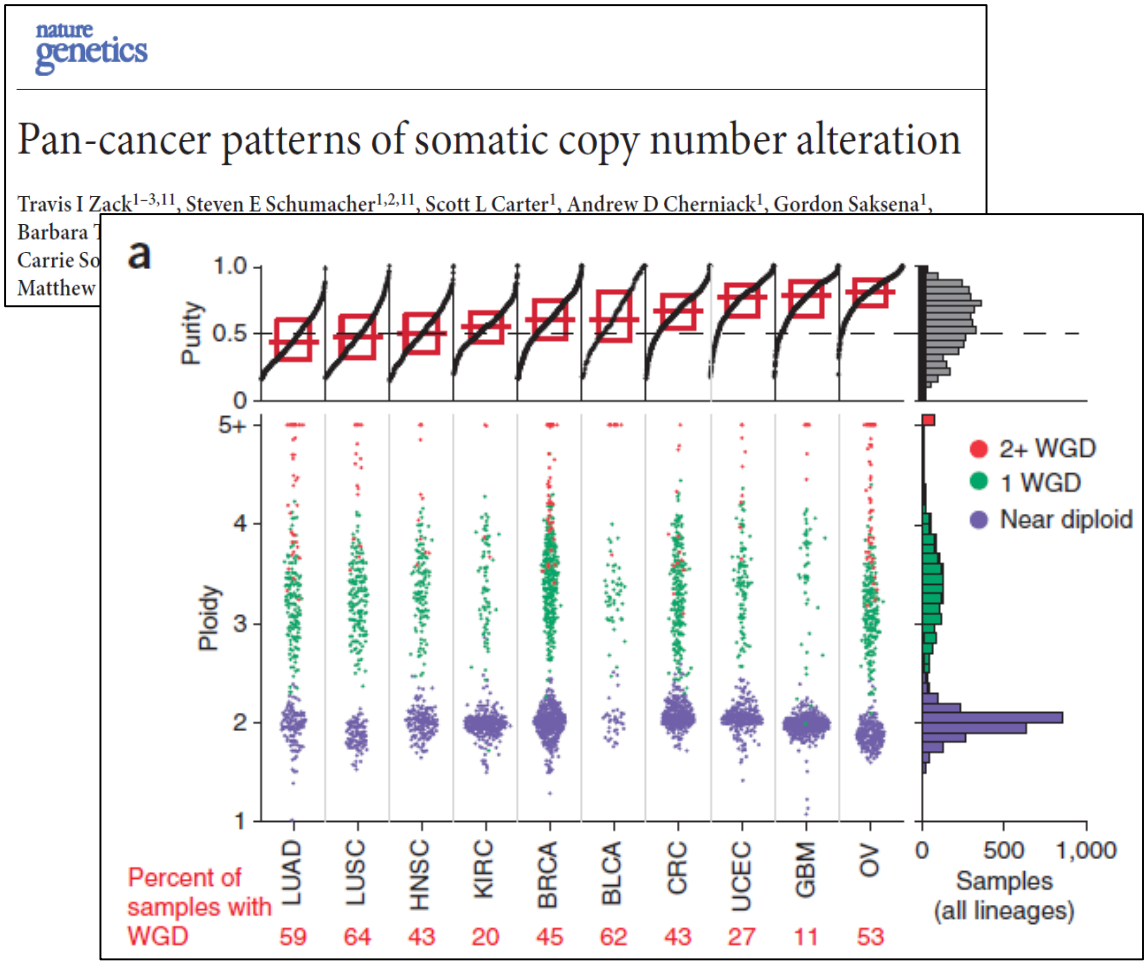
Personal Genome	PacBio Coverage	Deletions ≥ 50 bp	Insertions ≥ 50 bp
CHM1 (haploid)[1]	41-fold	6,111	9,638
HX1[2]	103-fold	9,891	10,284
AK1[3]	101-fold	7,358	10,077



PacBio reference genome projects revealed short reads *significantly underestimate* the abundance of structural variants even in normal, healthy genomes.

[1]Chaisson et al. (2015) *Nature* 517:608-11.
[2]Shi et al. (2016) *Nat Commun* 7:12065.
[3]Seo et al. (2016) *Nature* 538:243-7.
[4]Sudmant et al. (2015) *Nature* 526:75-81.

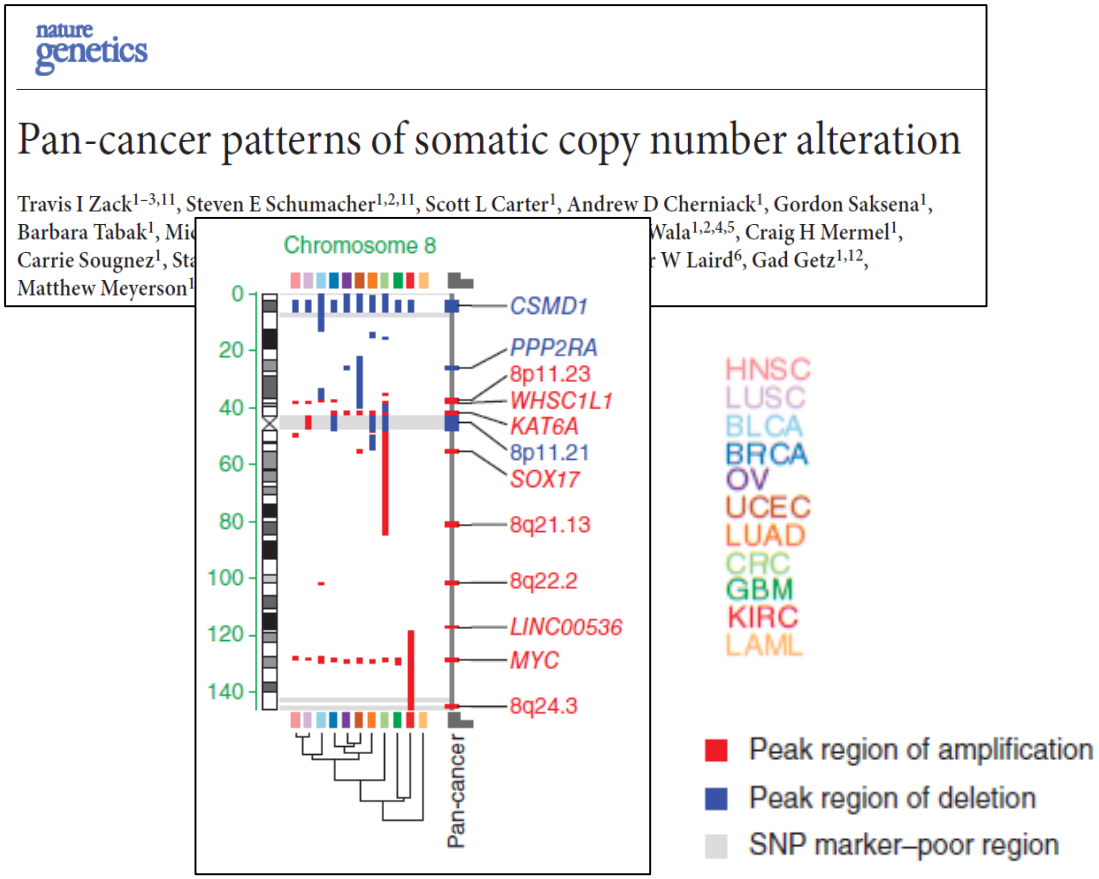
ANALYSIS OF THE TCGA PAN-CANCER DATA SET INDICATES THAT STRUCTURAL VARIATION IS COMMON IN MOST CANCERS



- SNP arrays were used to establish genome-wide copy number variation
- 3,847 samples studied
- Whole genome duplication events were common

Zack et. al., (2013) Nature Genet 45:1134-1143

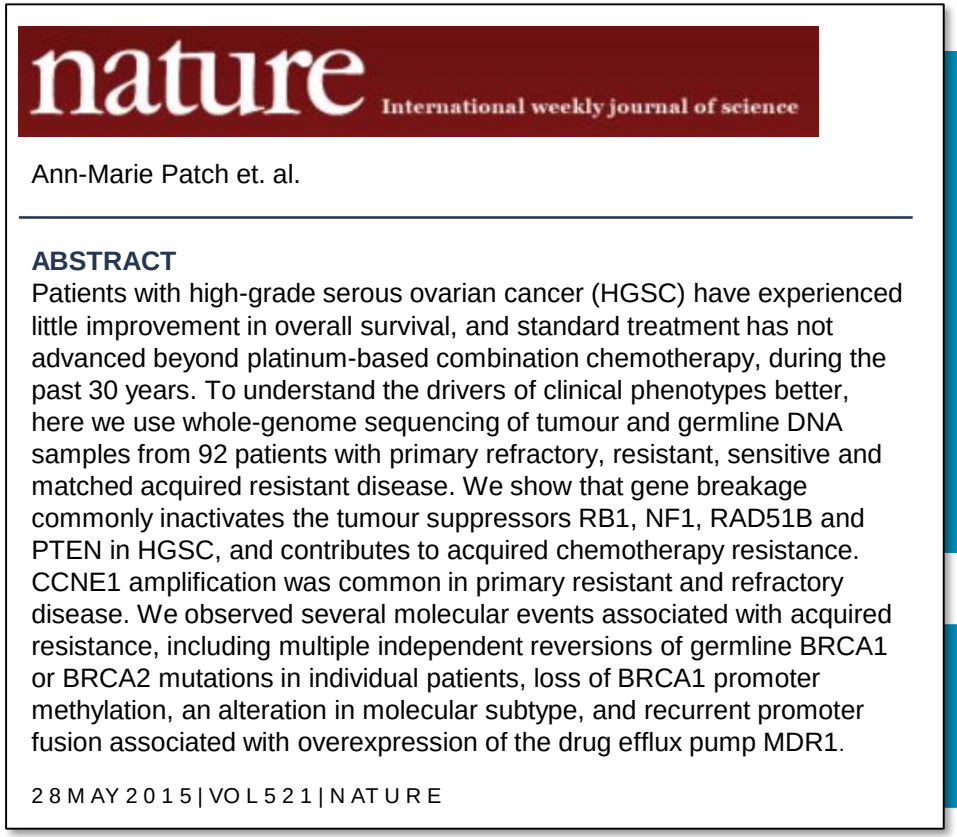
…AND SUGGESTS THAT SOME OF THESE STRUCTURAL VARIANTS ARE DRIVER MUTATIONS



- SNP arrays were used to establish genome-wide copy number variation
- 3,847 samples studied
- Whole genome duplication events were common
- However, recurrent, focal hotspots for CNVs were also identified in each chromosome, indicating that like SNPs, some SVs are driver mutations and under strong selective pressure

Zack et. al., (2013) Nature Genet 45:1134-1143

SOME TYPES OF CANCER ARE KNOWN TO HAVE VERY HIGH LEVELS OF STRUCTURAL REARRANGEMENT



nature International weekly journal of science

Ann-Marie Patch et. al.

ABSTRACT

Patients with high-grade serous ovarian cancer (HGSC) have experienced little improvement in overall survival, and standard treatment has not advanced beyond platinum-based combination chemotherapy, during the past 30 years. To understand the drivers of clinical phenotypes better, here we use whole-genome sequencing of tumour and germline DNA samples from 92 patients with primary refractory, resistant, sensitive and matched acquired resistant disease. We show that gene breakage commonly inactivates the tumour suppressors RB1, NF1, RAD51B and PTEN in HGSC, and contributes to acquired chemotherapy resistance. CCNE1 amplification was common in primary resistant and refractory disease. We observed several molecular events associated with acquired resistance, including multiple independent reversions of germline BRCA1 or BRCA2 mutations in individual patients, loss of BRCA1 promoter methylation, an alteration in molecular subtype, and recurrent promoter fusion associated with overexpression of the drug efflux pump MDR1.

28 MAY 2015 | VOL 521 | NATURE

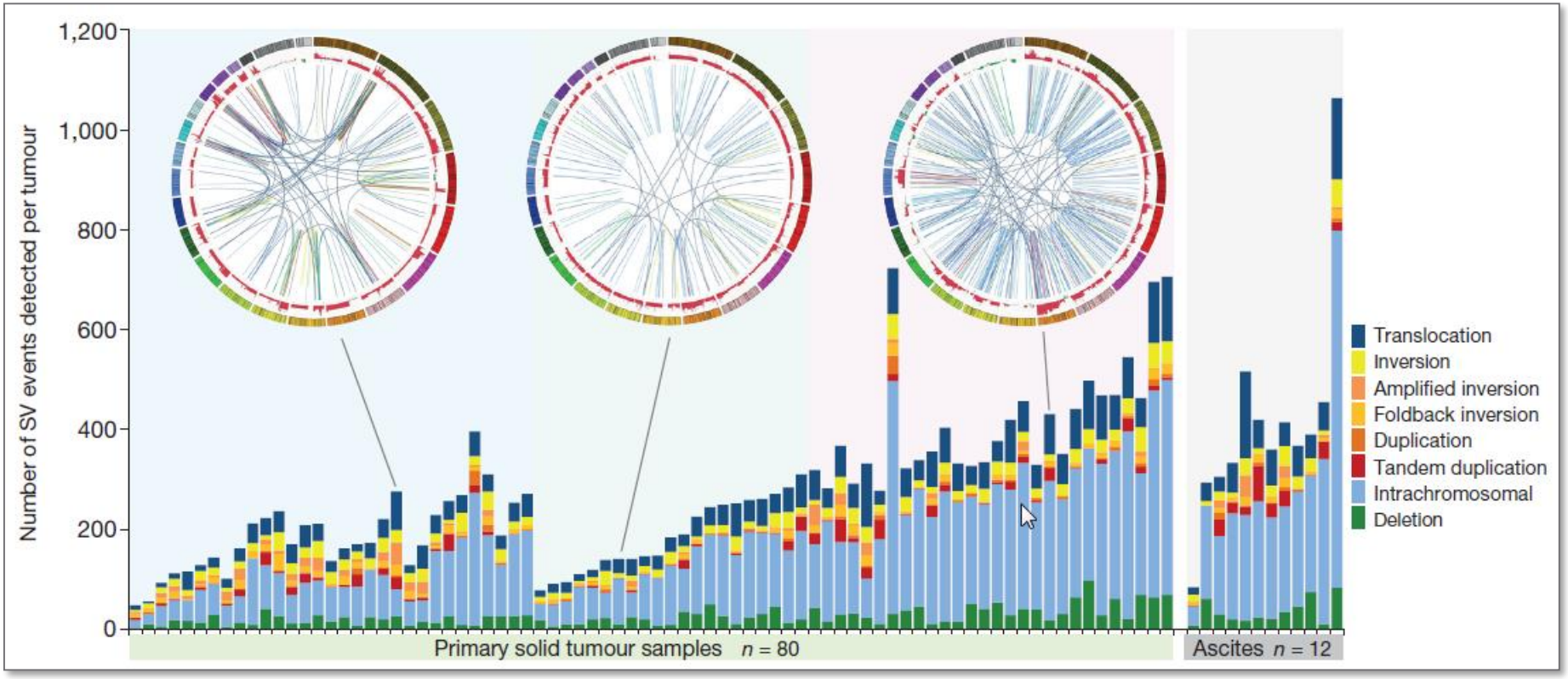
"Patients with high-grade serous ovarian cancer (HGSC) have experienced little improvement in overall survival, and

standard treatment has not advanced beyond platinum-based combination chemotherapy, during the past 30 years."

The authors performed WGS and RNA-seq on 92 individual cases.

Ann-Marie Patch, et al . (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. *521: 489–494.*

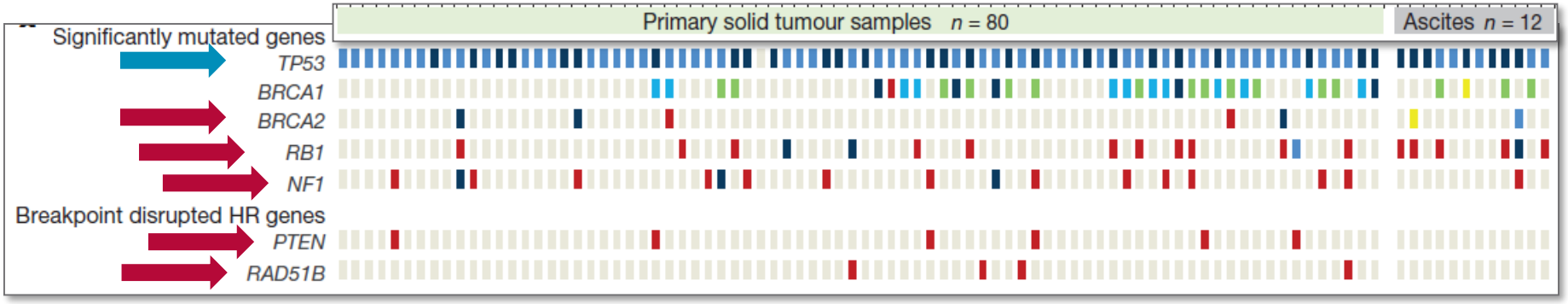
HIGH GRADE SERIOUS OVARIAN CANCER IS CHARACTERIZED BY EXTENSIVE GENOMIC STRUCTURAL CHANGES



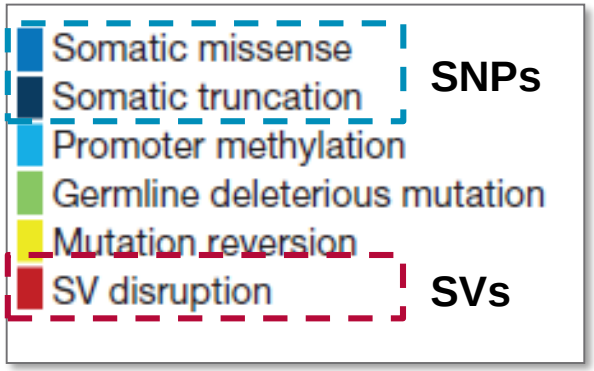
- A total of 36,561 somatic structural variants were detected in primary and recurrence samples by WGS

Ann-Marie Patch, et al . (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. *521: 489–494.*

APART FROM TP53, SOMATIC POINT MUTATIONS IN DRIVER GENES ARE INFREQUENT IN PRIMARY HGSC



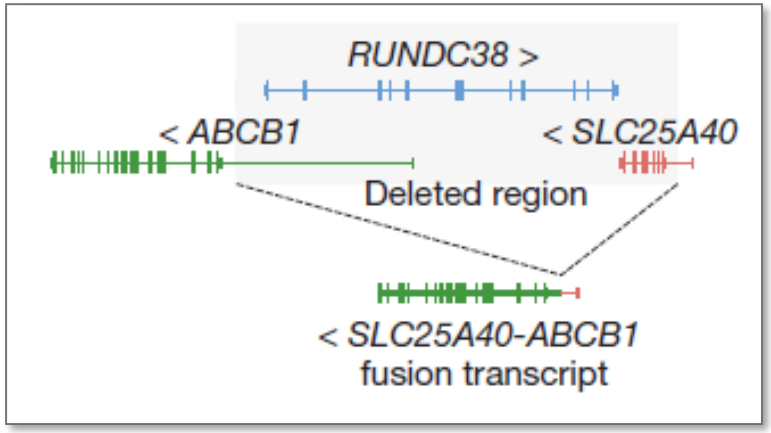
- NF1 and RB1 were inactivated by point mutations and indels in only 6% of primary samples (blue ticks)
- Inclusion of SVs raised the frequency of inactivating mutations to 20% for NF1 and 17.5% for RB1 (red ticks)
- Gene inactivation of homologous repair genes PTEN and RAD51B was by breakage

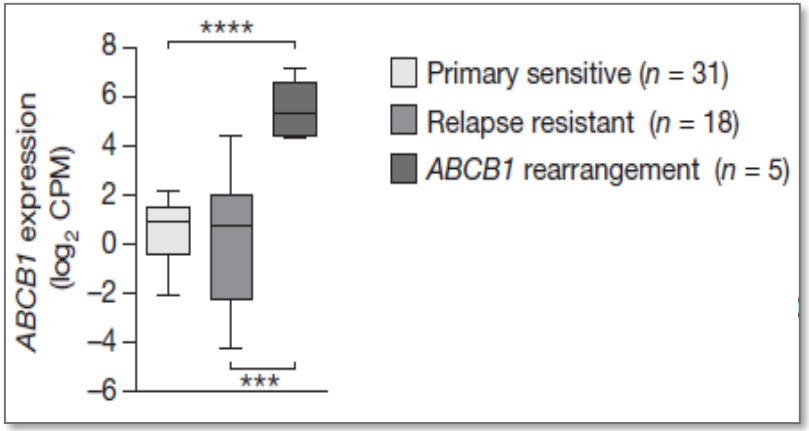


Ann-Marie Patch, et al . (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. *521: 489–494.*

MUTATIONS SPECIFIC TO RESISTANCE WERE OFTEN NON-SNV IN NATURE

On average there were 1.6 X more structural variants in recurrence vs. primary samples

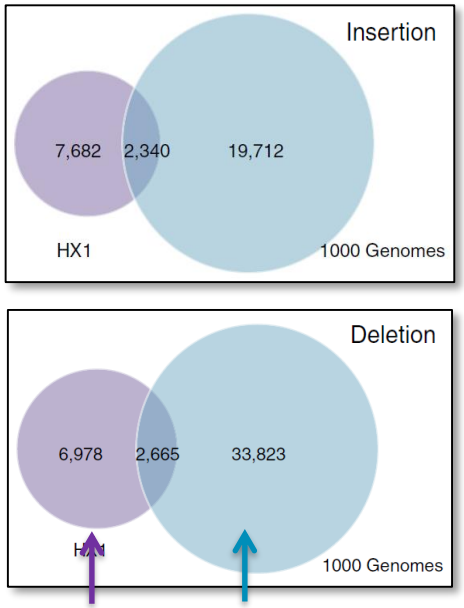




- ABCB1 encodes the drug efflux pump MDR1
- The promoter and non-coding exon 1 of SLC25A40 was fused with exon 2 of ABCB1
- …driving increased gene expression

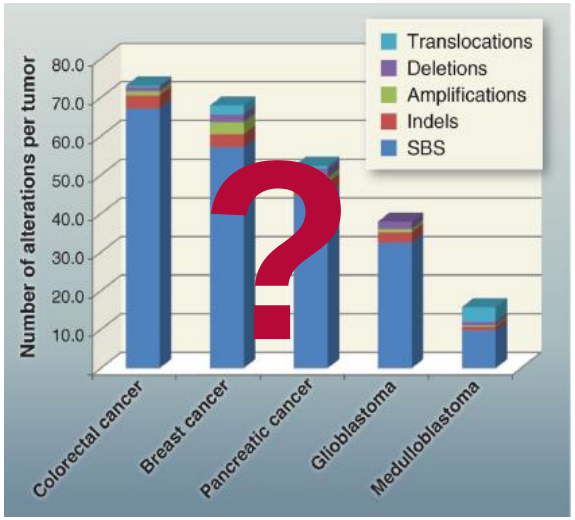
Ann-Marie Patch, et al . (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. *521: 489–494.*

WHAT MIGHT WE LEARN BY REVEALING THE HIDDEN LANDSCAPE OF SVs IN CANCER GENOMES?



1 Individual with PacBio

2500 Individuals with Illumina



Vogelstein, et. al.. (2013). Cancer Genome Landscapes. Science. 339(6127):1546–1558.

A TRUTH SET IS A FIRST STEP TOWARDS DEVELOPING A ROBUST METHOD FOR SV DETECTION

- What technology or technologies?
- How much coverage?
- What analysis pipeline? With what filtering?
- What samples are appropriate?
- What experimental design?

Proposed using COLO829 and matched normal lymphoblastoid cell line (COLO829BL) to develop a validated truth set of variant calls

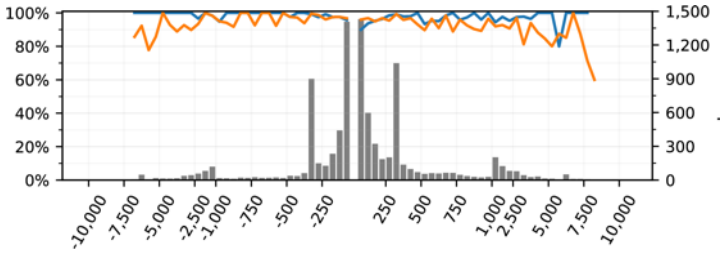
Technologies

- Illumina
- PacBio
- Oxford Nanopore
- BioNano
- 10X Genomics

PACBIO LONG READ SEQUENCING IS THE MOST RELIABLE METHOD FOR DETECTION STRUCTURAL VARIANTS

Technology	Precision	Recall
PacBio	96.13%	95.99%
Oxford Nanopore	83.23%	87.46%
Illumina	85.35%	55.88%
10X Genomics	83.79%	39.83%

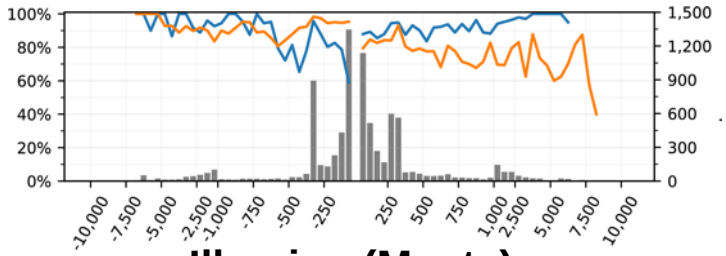
PacBio (pbsv)



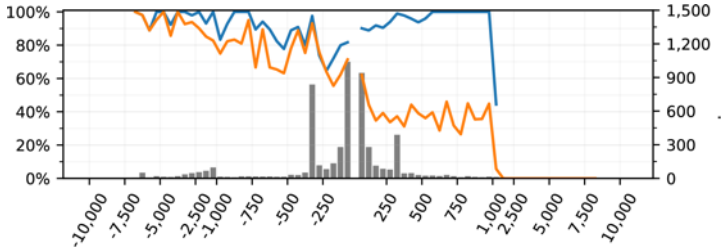
deletions insertions

Structural variant length (bp)

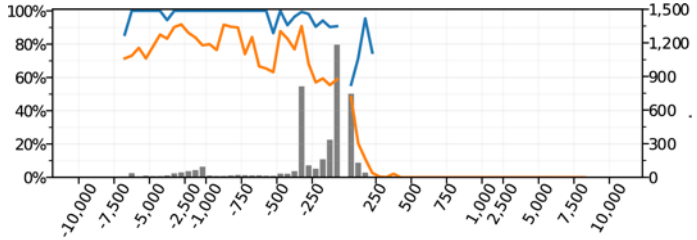
Oxford Nanopore (pbsv)



Illumina (Manta)



10X Genomics (LongRanger)



deletions insertions

Structural variant length (bp)

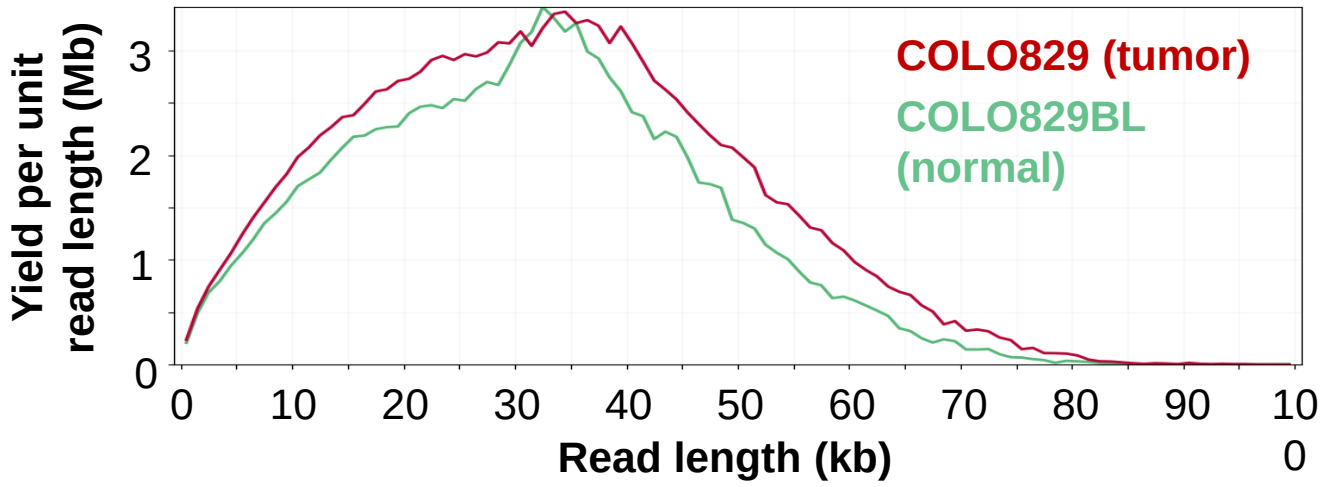
Variant calling performance against the GIAB HG002 v0.6 benchmark.

LONG-READ SEQUENCING OF COLO829 TUMOR AND NORMAL ON PACBIO SEQUEL SYSTEM

Metric	COLO829	COLO829BL
Yield	146.6 Gb	154.9 Gb
Reads	9,673,612	10,500,345
Read Length N50	29,161 bp	26,950 bp

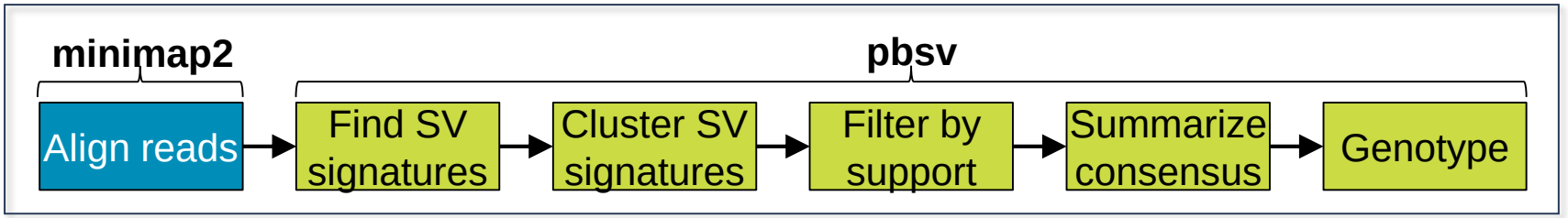
High-coverage Illumina sequencing identified:

- >35,000 tumor-specific SNVs
- 446 indel variants



Marcel Nelen, Radboud University Medical Center; Wigard Kloosterman, University Medical Center Utrecht

STRUCTURAL VARIANT CALLING WITH PBSV



- For both cell lines, reads were aligned against GCRh37 and structural variants were called with pbsv

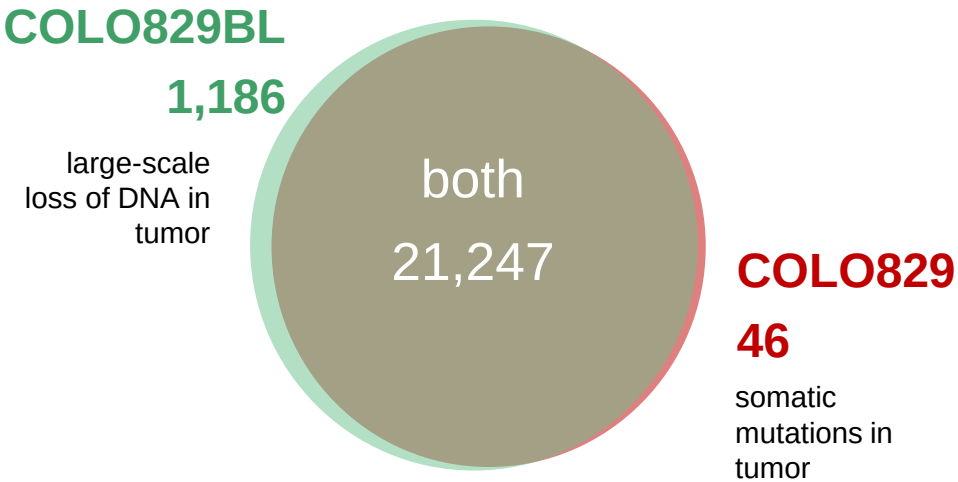
- The genotypes of the tumor and normal cells lines were then compared to assess insertions, deletions, inversions, and translocations, and CNVs.

Analysis by Aaron Wenger, PacBio

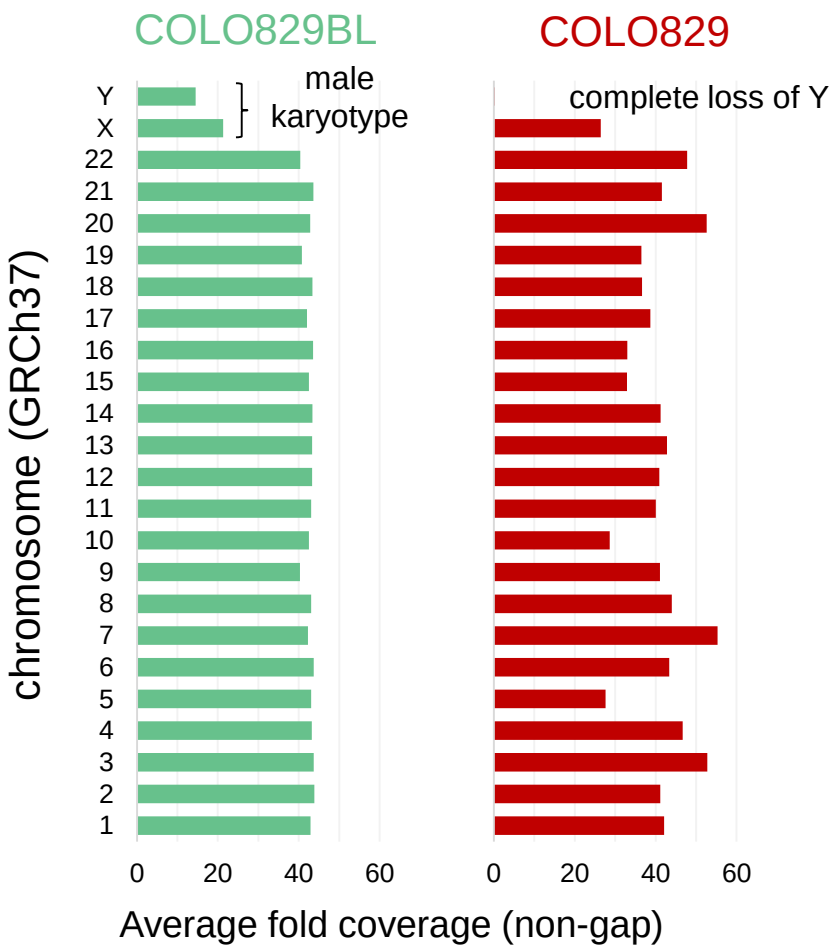
THE PBSV PIPELINE IDENTIFIED SIGNIFICANT DIFFERENCES IN STRUCTURAL VARIANTS BETWEEN GRCH37 AND EACH CELL LINE

- The relatively low number of SVs in COLO289 makes it a useful cell line for exhaustive validation of variant calls
- Variants present in the control but absent from the tumor (green) reflect large scale loss of DNA in the tumor

Structural variants



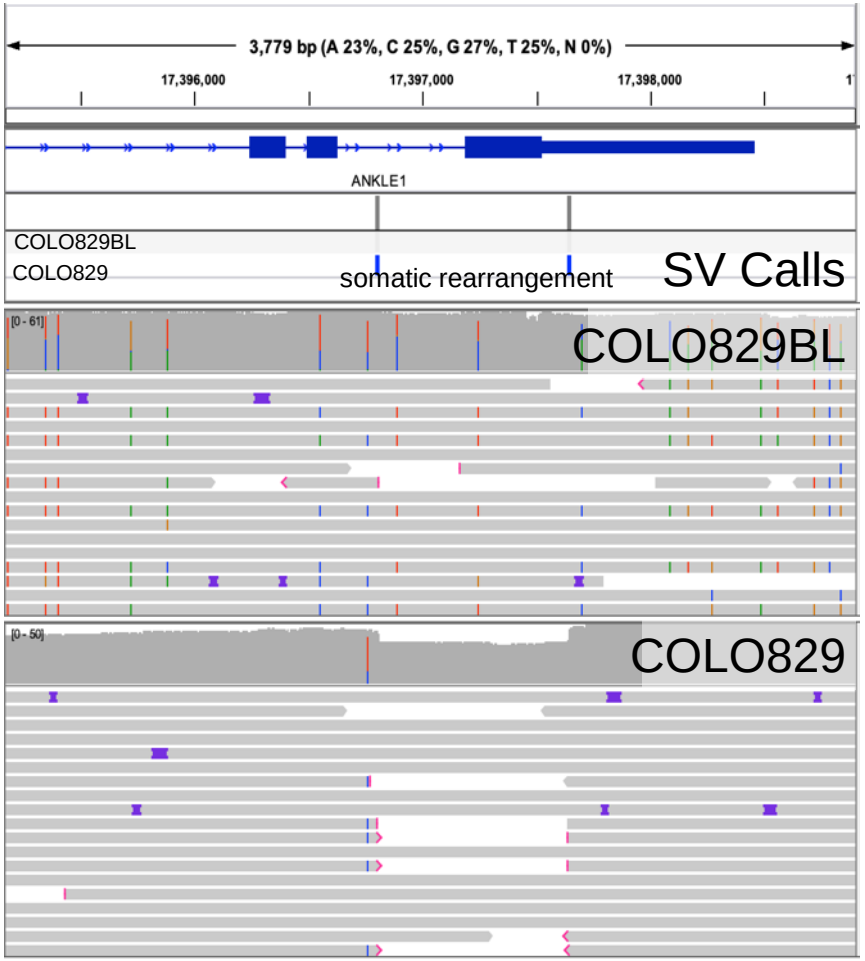
Copy number profile

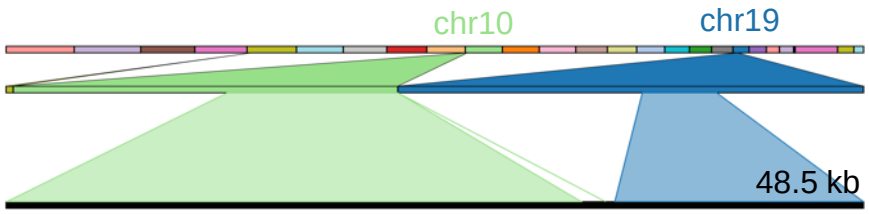


Analysis by Aaron Wenger, PacBio

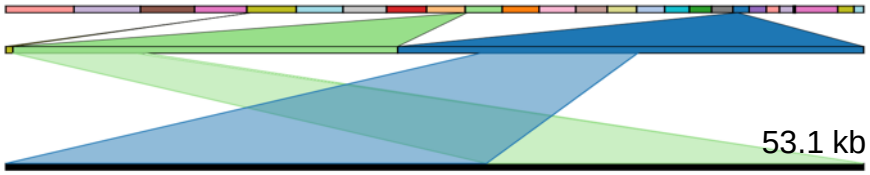
CHR 19 EXAMPLE: SOMATIC REARRANGEMENT IN ANKLE1

GRCh37 chr19:17,395,165-17,398,943





Read m54150_180626_101156/6160504/0_48531

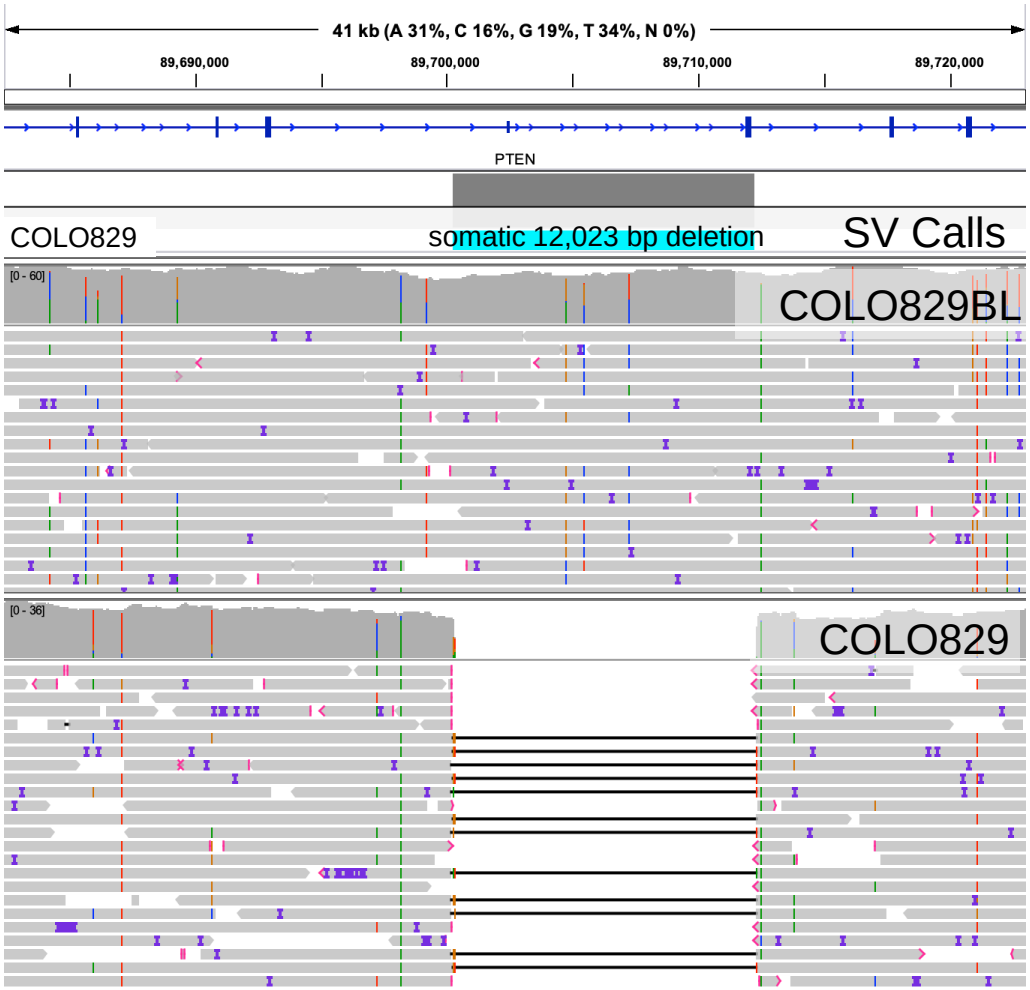


Read m54150_180604_100758/54920184/0_53139

- Ankle1 is a nuclease involved in DNA damage repair
- SNVs in the Ankle1 gene have been linked to increased breast and ovarian cancer risk

Analysis by Aaron Wenger, PacBio

CHR 10 EXAMPLE: SOMATIC DELETION IN PTEN



- PTEN is a tumor suppressor commonly mutated in cancers.
- COLO829 has a homozygous deletion within PTEN.

Analysis by Aaron Wenger, PacBio

HOW MUCH PACBIO COVERAGE IS NEEDED FOR SV DISCOVERY?

Rare disease

- PacBio reads map uniquely and rarely have indel errors between 20-50,000 bp
- In this size range, 2 reads is therefore sufficient to make a confident call

Tumor samples:

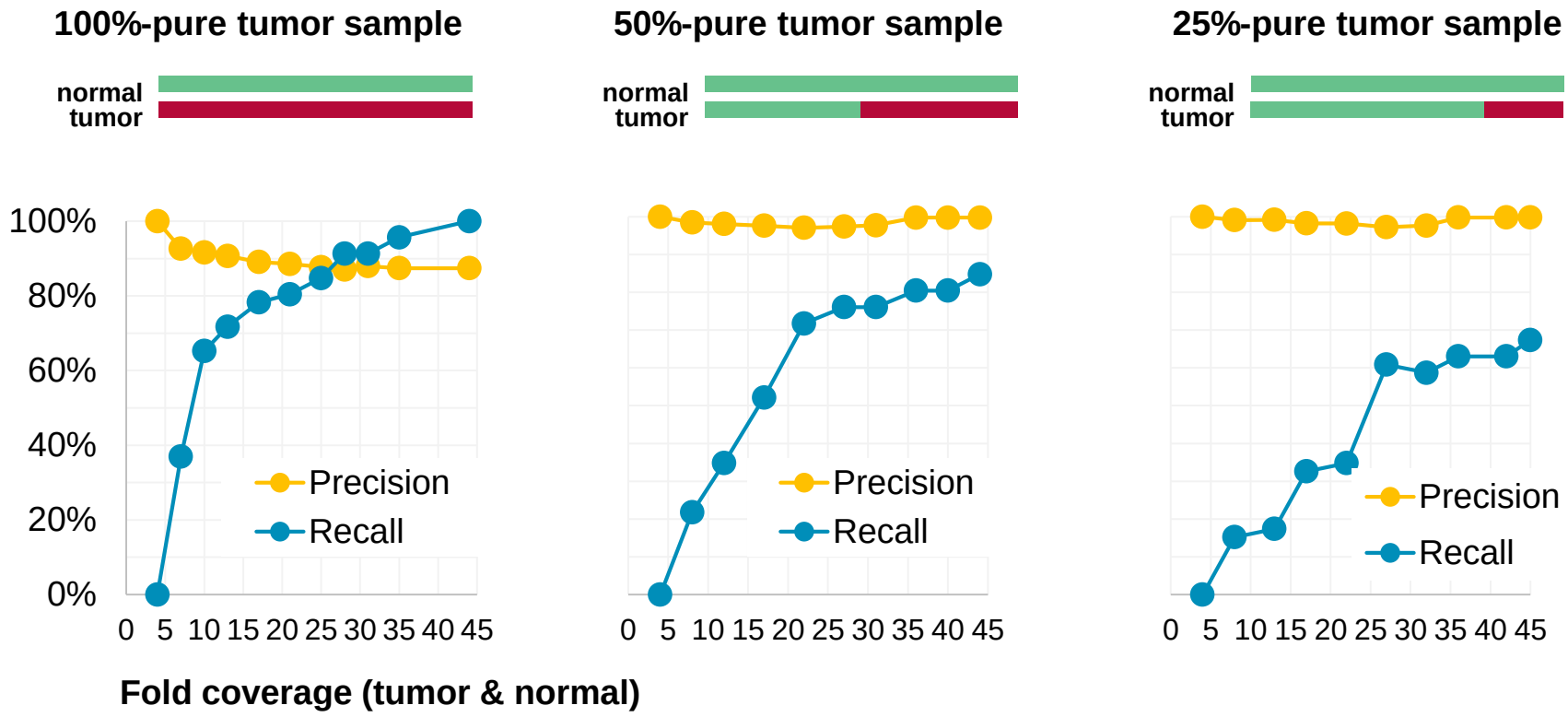
- **Challenge #1: Tumors are not clonal**

  "All happy families are alike; each unhappy family is unhappy in its own way."
  Tolstoy, Anna Karenina

  - But, if the goal is to find *driver* mutations, these should be clonally shared by most tumor cells (~100% frequency)

- **Challenge #2: Tumor samples are not purely tumor**

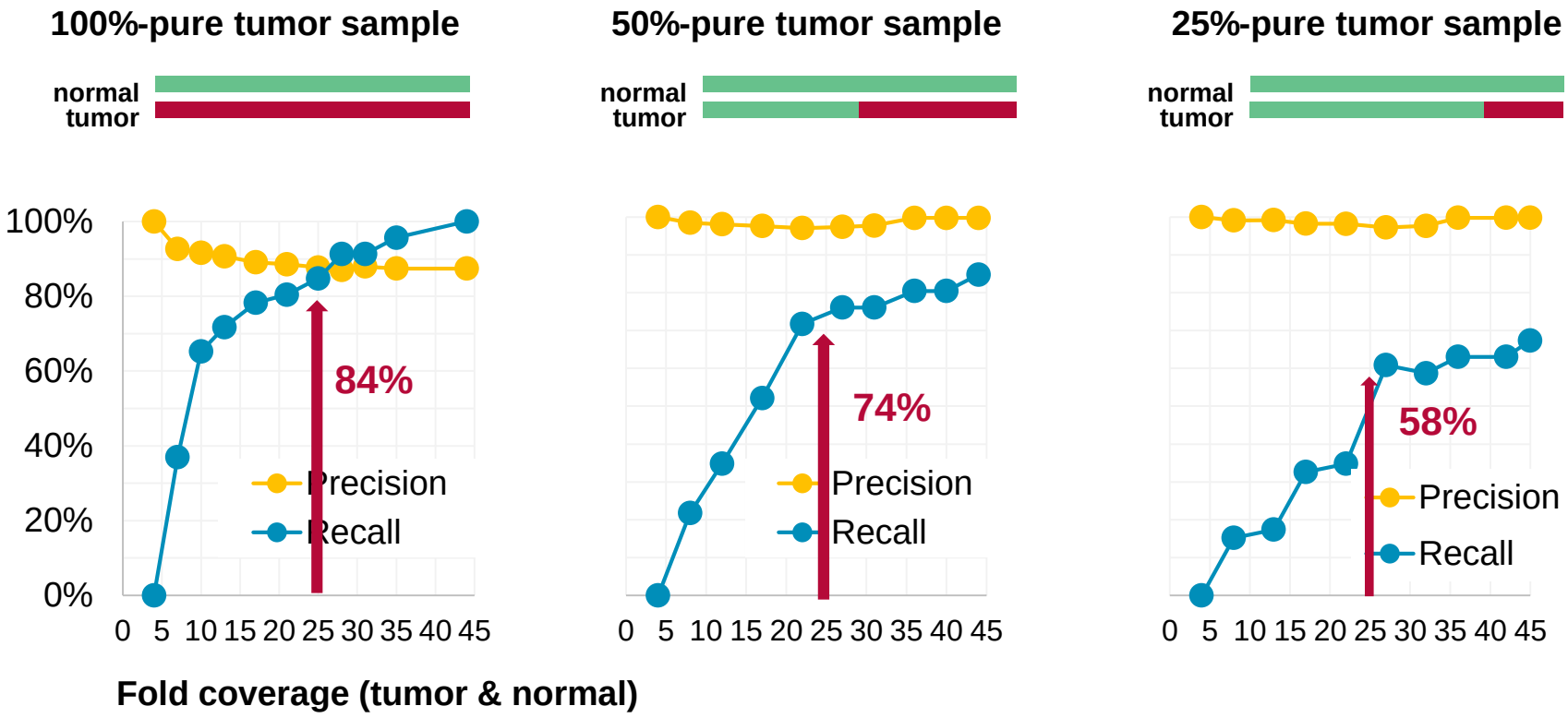
PRECISION AND RECALL FOR SOMATIC VARIANTS



- Truth set: 46 somatic variants from full-coverage data with ≥6 variant reads in tumor, 0 in normal
- Calling criteria: ≥2 variant reads in tumor, 0 in normal)

Analysis by Aaron Wenger, PacBio

PRECISION AND RECALL FOR SOMATIC VARIANTS



- Precision is high across a range of tumor purity and coverage.
- Recall remains high down to 20-fold coverage in samples that are at least half-tumor, and nearly saturates by 30-fold coverage at all purity levels.

Analysis by Aaron Wenger, PacBio

REVEALING THE DARK MATTER OF CANCER GENOMES

How can we reveal the full spectrum of variants that impact cancer development, progression, treatment response, and relapse?

Sequencing:

- Low-fold PacBio coverage to reveal uncharacterized structural variants and indels
  - ~80 Gb of CLR data per SMRT Cell 8M
  - ~25-fold coverage per SMRT Cell at a price similar to the cost of WGS with Illumina
- High coverage Illumina sequencing with gene panels to catalog SNPs in already identified hotspots (TCGA, ICGC)

Analysis:

- Minimap2 and pbsv
- Require 2 reads per variant for SVs between 20-50,000 bp

HOW LARGE A COHORT IS NEEDED TO FIND PREDISPOSITION ALLELES OR DRIVER MUTATIONS?

Lower ‘N’…

- *Technology has high recall*
- *Tumor purity* is high
- Variant is a *driver mutation* / is relatively clonal

Higher ‘N’…

- *Technology has low recall*
- *Incorrect clustering*: Cancer cohort is composed of similar but biologically distinct tumors
- *Biological complexity*: distinct variants produce the same biological result
  - Gene ontology?
  - Familial predisposition cases?

"We basically can see the entire picture. We're not looking under a lamppost for the keys. It's daylight, and we can see the whole neighborhood. **So we're gonna find the keys**."

- Dan Geraghty, Fred Hutchinson Cancer Center

http://mendelspod.com/podcast/what-projects-will-you-pursue-long-read-technology/

ACKNOWLEDGEMENTS

Aaron Wenger, PhD
PacBio
Principal Scientist, Bioinformatics

Wigard Kloosterman, PhD
Principal Investigator – Assoc. Professor
UMC Utrecht
(CSO, Frame Cancer Therapeutics)

Marcel Nelen, PhD
Deputy Head Genome Diagnostics,
Clinical Molecular Geneticist,
Radboud UMC, Nijmegen

Hartwig Medical Foundation

Pacific Biosciences

PACBIO®

www.pacb.com