wellcome trust sanger institute

Targeted capture and resequencing of whole exomes in breast cancer

Patrick Tarpey
Cancer Genome Project

Breast Cancer

Breast cancer is the most common cancer in the UK

In 2007 in the UK almost 45,700 women were diagnosed with breast cancer (125 women/day)

8 in 10 breast cancers are diagnosed in women aged 50 and over.

Table Two: Estimated risk of developing breast cancer by age, females, UK, 2001–2005[f]

Estimated risk at birth up to and including:	UK (2001–2005)
age 24	1 in 15,300
age 29	1 in 2,300
age 39	1 in 200
age 49	1 in 52
age 59	1 in 22
age 69	1 in 14
age 79	1 in 10
age 84	1 in 9
Lifetime risk	1 in 9

Breast Cancer

Almost 2 out of 3 women with breast cancer now survive their disease beyond 20 years

In the 1970s around 5 out of 10 breast cancer patients survived the disease beyond five years. Now it's 8 out of 10

9 out of 10 of women diagnosed with stage I breast cancer survive the disease beyond five years.

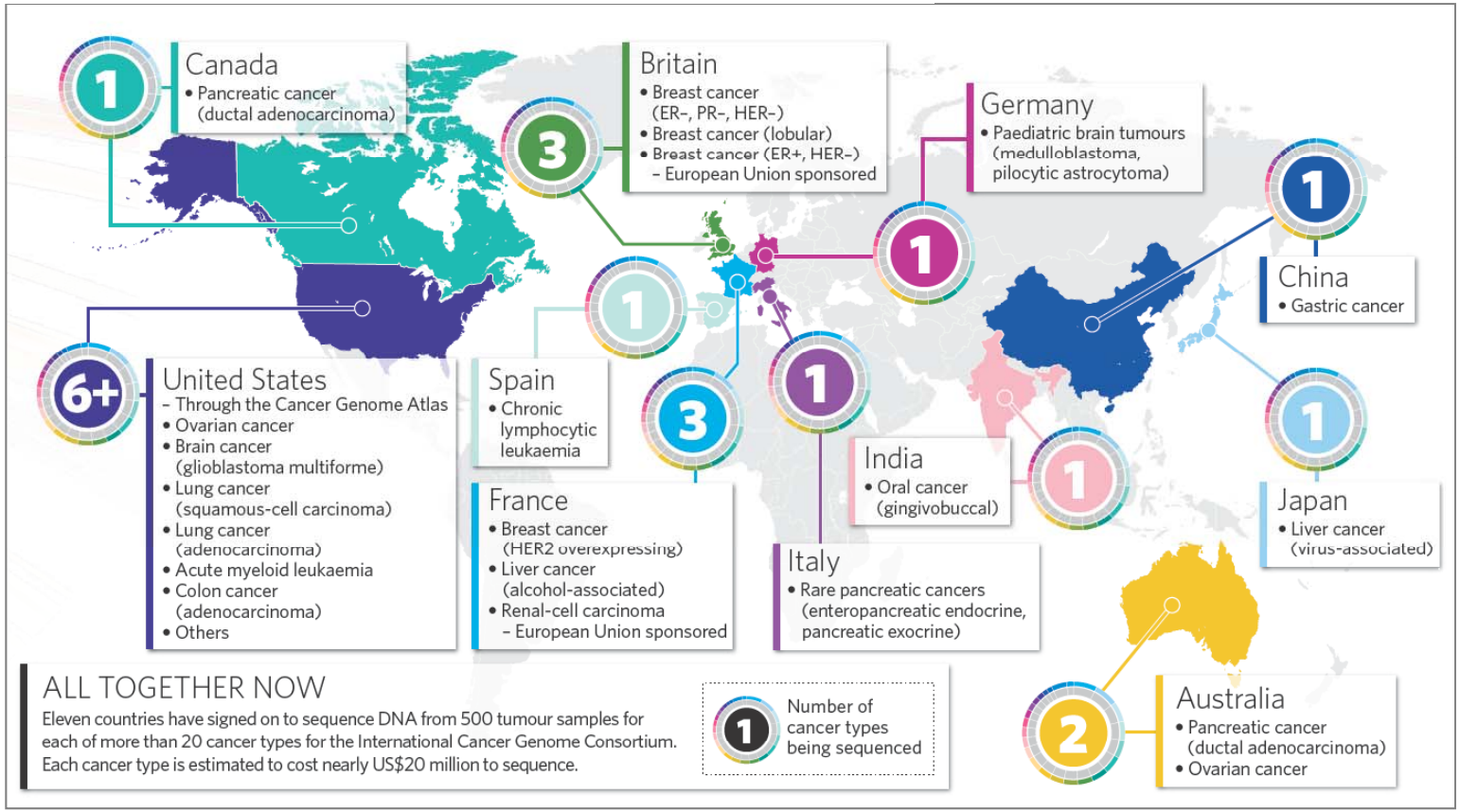
This drops to around 1 out of 10 diagnosed with stage IV

Breast Cancer

Breast cancer subtypes

- ER-, HER2-, PR- (triple negative)
- ER-, HER2+ tumors
- ER+, HER2+ tumors
- **ER+, HER2-, PR+ tumors; ductal-type**
- **ER+, HER2-, PR- tumors; ductal-type**
- Invasive lobular carcinoma
- “Rare” histological special types

International Cancer Genome Consortium (ICGC)



International Cancer Genome Consortium (ICGC)

Variants under investigation

base substitutions
insertions, deletions
copy number changes
Translocations
other chromosomal rearrangements

Intragenic Variants

Sanger sequencing

Candidate genes, small nos. of samples

BRIEF COMMUNICATIONS

nature genetics

A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer

Philip Stephens[1], Sarah Edkins[1], Helen Davies[1], Chris Greenman[1], Charles Cox[1], Chris Hunter[1], Graham Bignell[1], Jon Teague[1], Raffaella Smith[1], Claire Stevens[1], Sarah O'Meara[1], Adrian Parker[1], Patrick Tarpey[1], Tim Avis[1], Andy Barthorpe[1], Lisa Brackenbury[1], Gemma Buck[1], Adam Butler[1], Jody Clements[1], Jennifer Cole[1], Ed Dicks[1], Ken Edwards[1], Simon Forbes[1], Matthew Gorton[1], Kristian Gray[1], Kelly Halliday[1], Rachel Harrison[1], Katy Hills[1], Jonathon Hinton[1], David Jones[1], Vivienne Kosmidou[1], Ross Laman[1], Richard Lugg[1], Andrew Menzies[1], Janet Perry[1], Robert Petty[1], Keiran Raine[1], Rebecca Shepherd[1], Alexandra Small[1], Helen Solomon[1], Yvonne Stephens[1], Calli Tofts[1], Jennifer Varian[1], Anthony Webb[1], Sofie West[1], Sara Widaa[1], Andrew Yates[1], Francis Brasseur[2], Colin S Cooper[3], Adrienne M Flanagan[4], Anthony Green[5], Maggie Knowles[6], Suet Y Leung[7], Leendert H J Looijenga[8], Bruce Malkowicz[9], Marco A Pierotti[10], Bin Teh[11], Siu T Yuen[7], Andrew G Nicholson[12], Sunil Lakhani[13], Douglas F Easton[14], Barbara L Weber[15], Michael R Stratton[1,3], P Andrew Futreal[1] & Richard Wooster[1]

We examined the coding sequence of 518 protein kinases, ~1.3 Mb of DNA per sample, in 25 breast cancers. In many tumors, we detected no somatic mutations. But a few had numerous somatic mutations with distinctive patterns indicative of either a mutator phenotype or a past exposure.

We detected 92 somatic mutations, 90 base substitutions and two in-frame deletions (The Catalogue of Somatic Mutations in Cancer; Supplementary Table 1 online) in the complete protein kinase gene family (Supplementary Table 2 online)[1]. Of these, 58 base substitutions caused missense amino acid changes, 12 caused translational termination codons, 6 were at conserved positions in splice sites and 14 were silent (synonymous) changes.

The somatic mutations were distributed unevenly among the breast cancers that we examined. Twelve primary breast cancers had no somatic mutations; two had a single mutation each and one had two mutations. The remaining primary breast cancer (PD0119, an invasive pleomorphic lobular cancer in an 84-year-old woman, which we found by immunohistochemistry to be estrogen receptor–positive and E-cadherin–, *ERBB2*- and *TP53*-negative; Supplementary Table 3 online) had 52 somatic mutations, all base substitutions. The mutations in this cancer had a distinctive pattern (Fig. 1). Mutations occurred with a high frequency at C:G base pairs (96%, 50 of 52), with many C:G→G:C transversions (44%, 22 of 50). Mutations occurred in a specific sequence context ($P = 0.047$); specifically, at C:G base pairs, 94% (47 of 50) were 3′ to a T:A base pair, compared with an expected frequency of 26% (Fig. 1b).

In contrast to the primary breast cancers, eight of the nine immortal breast cancer cell lines had at least one somatic mutation. The cell line with the largest number of somatic mutations, HCC2218 (derived from an invasive ductal carcinoma from a 38-year-old female, which was estrogen receptor–negative and *ERBB2*- and *TP53*-positive), had a mutational spectrum similar to that of PD0119 (Fig. 1a). All eleven of the somatic mutations detected in HCC2218 were at C:G base pairs; most were C:G→G:C transversions (56%, 6 of 11), and all arose at C:G base pairs 3′ to T:A base pairs. The mutational spectra of HCC2218 and PD0119 differed from those of other cancers in the set ($P = 0.003$ for heterogeneity) and from that of *TP53* obtained from several hundred breast cancers (Fig. 1 and Supplementary Tables 4 and 5 online). The results indicate that a previously hidden diversity of mutational processes exists in breast cancer.

The high frequency and distinctive pattern of somatic mutations observed in PD0119 and HCC2218 are probably due to a mutator phenotype (*e.g.*, due to a DNA-repair defect), although mutagenic exposure cannot be excluded. *BRCA1* and *BRCA2* have been implicated in DNA repair. We did not find a truncating mutation in *BRCA1* or *BRCA2* in either PD0119 or HCC2218, but we did find one in a breast cancer cell line (HCC1395) that does not seem to have their characteristic mutational features. Neither PD0119 nor HCC2218 shows microsatellite instability (Supplementary Methods online). Therefore, it is unlikely that abnormalities in *BRCA1*, *BRCA2* or mismatch-repair genes account for the mutational spectra observed in PD00119 and HCC2218. These mutational spectra are also not readily attributable to other known DNA-repair defects. Yeast mutants

[1]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. [2]Ludwig Institute for Cancer Research, 74 Avenue Hippocrate, B-1200 Brussels, Belgium. [3]Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK. [4]The Institute of Orthopaedics, University College London, Stanmore, Middlesex, UK. [5]Cambridge Institute for Medical Research, Addenbrooke's Hospital, Cambridge CB2 2XY, UK. [6]Cancer Research UK Clinical Centre, St. James University Hospital, Leeds, LS9 7TF, UK. [7]Department of Pathology, The University of Hong Kong, Hong Kong. [8]Josephine Nefkens Institute, Erasmus University, Rotterdam, 3015 GE Rotterdam, The Netherlands. [9]Division of Urology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [10]Nazionale Tumori and FIRC Institute of Molecular Oncology, Milan, Italy. [11]Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. [12]Royal Brompton Hospital, Sydney Street, London, SW3 6NP, UK. [13]Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, Chester Beatty Laboratories, London SW3 6JB, UK. [14]Cancer Research UK Genetic Epidemiology Group, Strangeways Research Laboratories, Cambridge, CB1 4RN, UK. [15]University of Pennsylvania Cancer Centre, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. Correspondence should be addressed to M.R.S. (mrs@sanger.ac.uk).

Published online 22 May 2005; doi:10.1038/ng1571

Intragenic Variants

Illumina exome sequencing

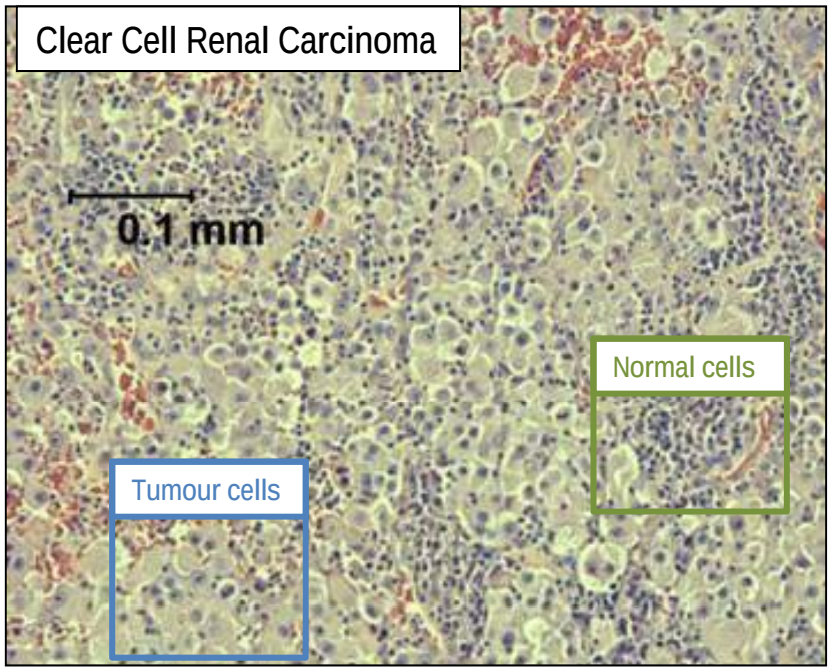
All genes, **large nos**. of samples (recurrence)

Intragenic Variants

Illumina exome sequencing

All genes, large nos. of samples

Higher sensitivity

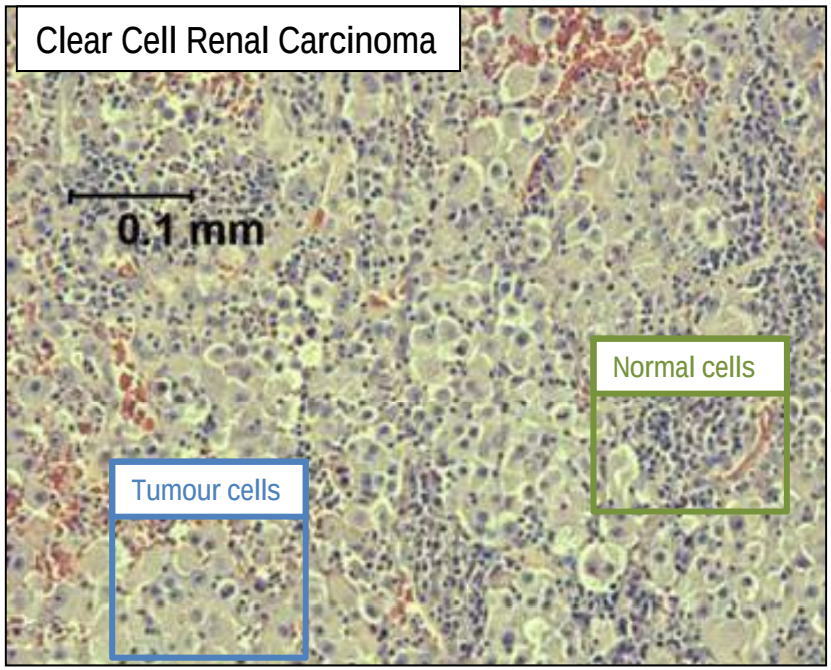


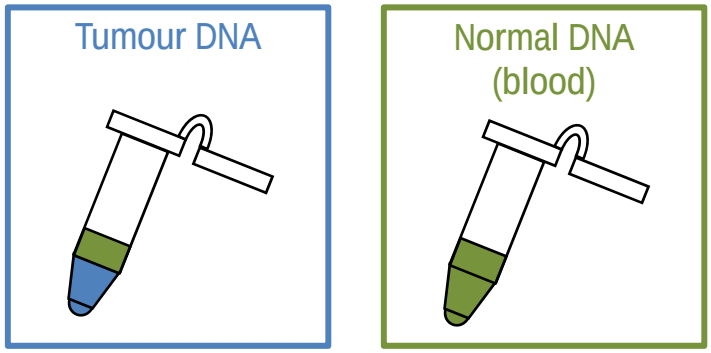
Intragenic Variants

Illumina exome sequencing

All genes, large nos. of samples

Higher sensitivity



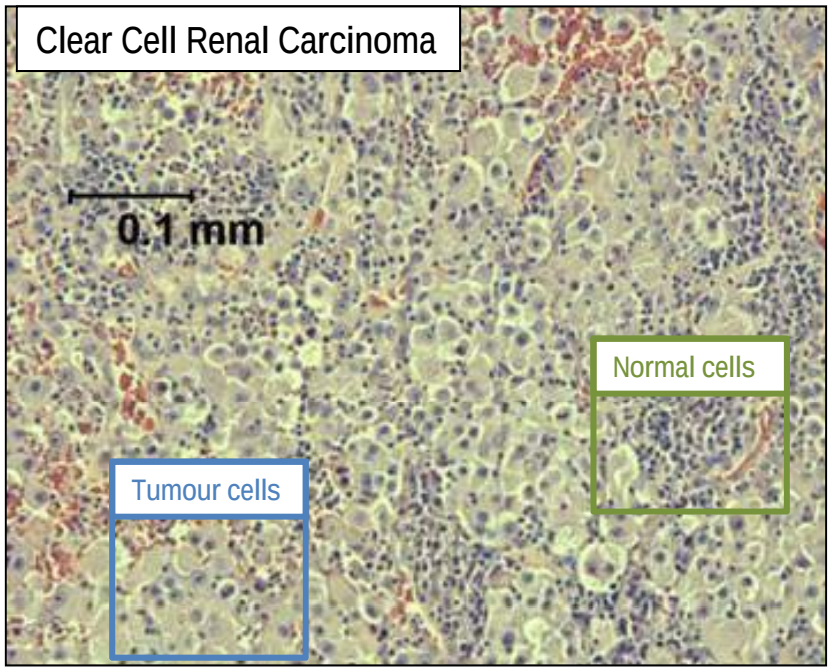


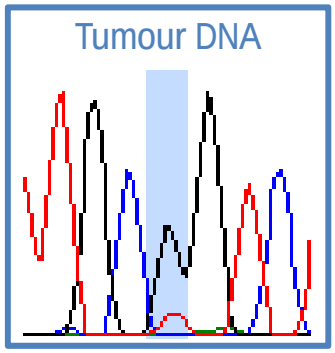
Intragenic Variants

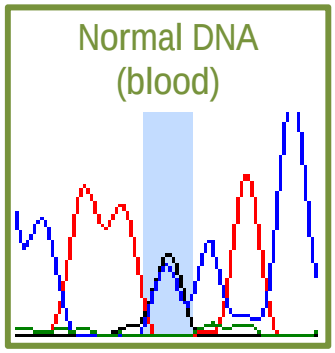
Illumina exome sequencing

All genes, large nos. of samples

Higher sensitivity







Intragenic Variants

Illumina exome sequencing

All genes, large nos. of samples

Higher sensitivity

Acquire all variants simultaneously

Samples screened

Sporadic breast cancer

28 breast cancer primary tumours

3 breast cancer cell lines

Samples screened

Sporadic breast cancer

28 breast cancer primary tumours [SOMATIC]

3 breast cancer cell lines [SOMATIC]

Samples screened

Sporadic breast cancer

28 breast cancer primary tumours [SOMATIC] [GERMLINE]

3 breast cancer cell lines [SOMATIC]

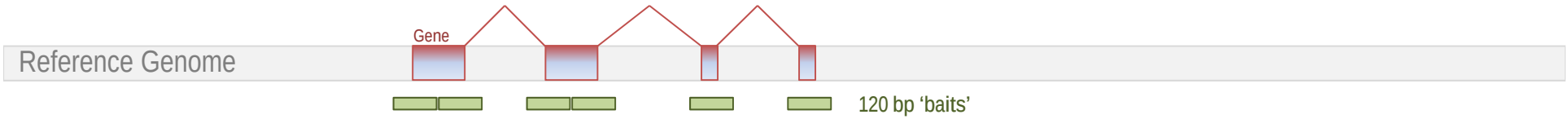
Exome Resequencing

Agilent Technologies

SureSelect
Solution Capture

illumina

Sanger SureSelect Exome



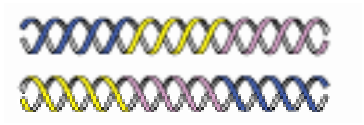
21,416 Protein Coding Genes
1664 miRNA's

50.3 Mb target

SureSelect Protocol

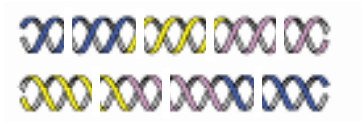
Genomic DNA

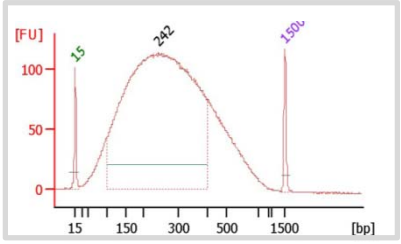
5ug



Fragment
(Covaris)

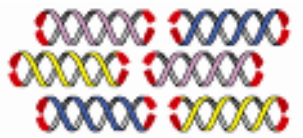
5ug

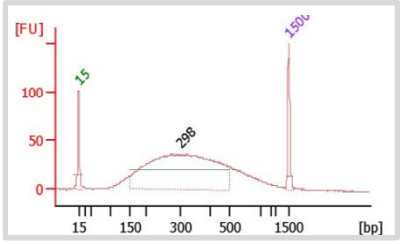




Ligate adapters
(Library)

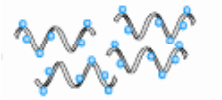
1.5 ug

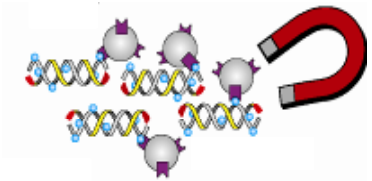




Hybridise

0.5 ug

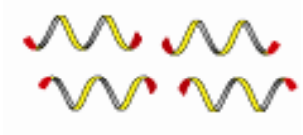


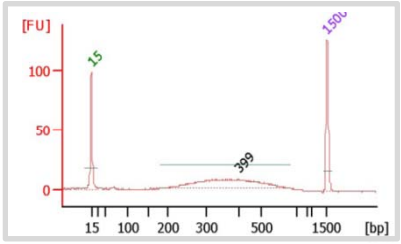


PCR

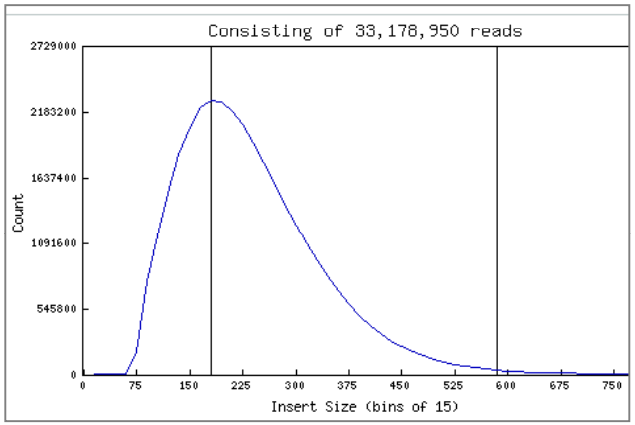
14 cycles

Elute and amplify



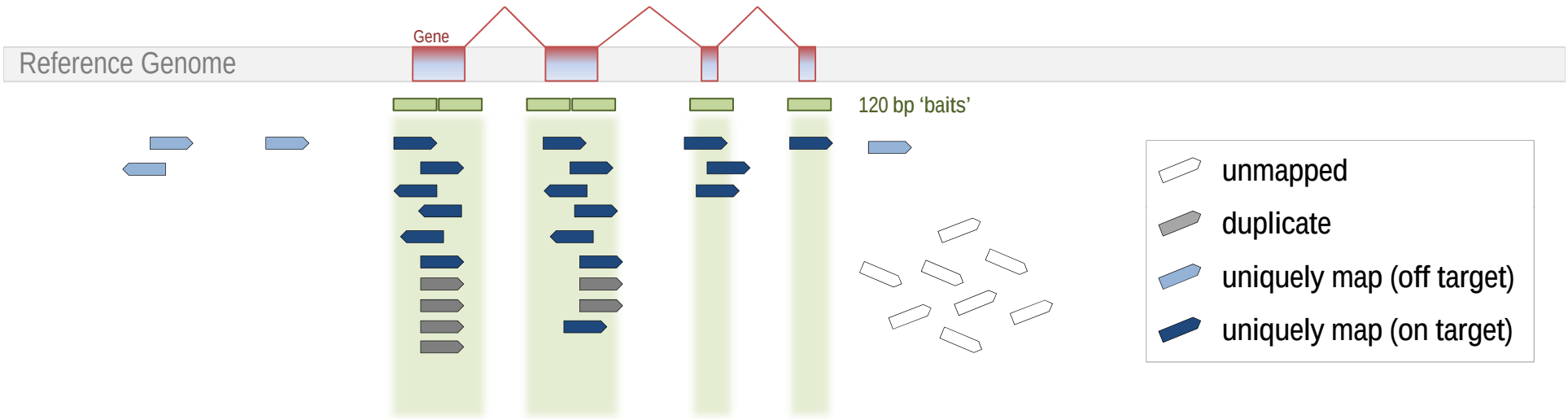


Paired-end sequencing

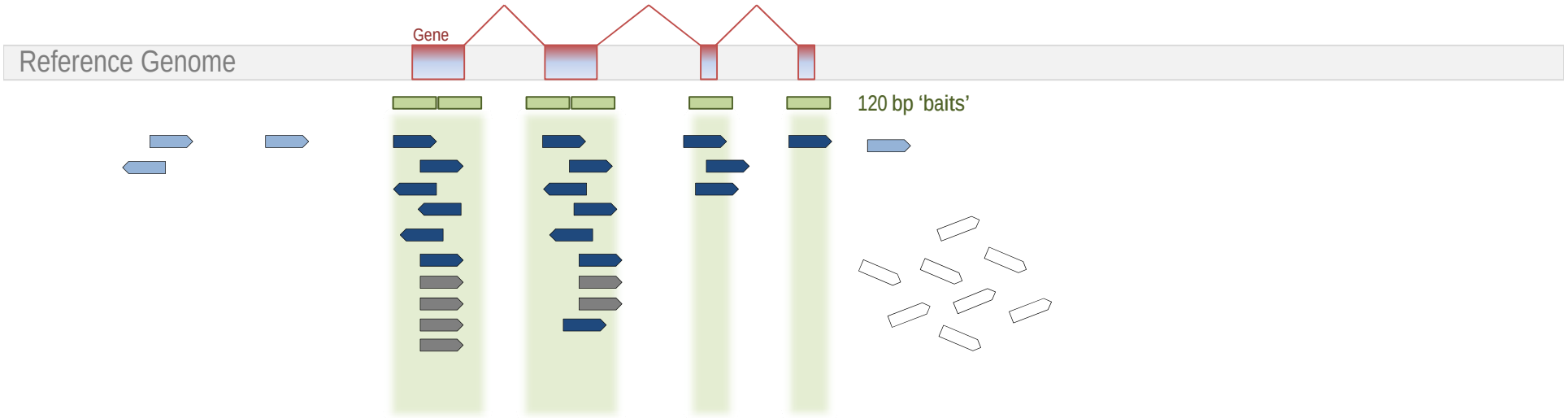


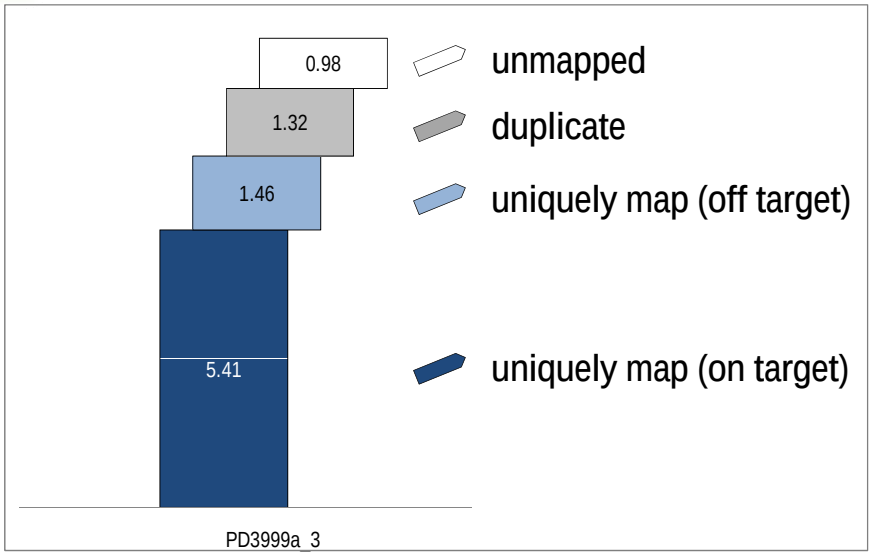
Performance Metrics

Sanger SureSelect Exome

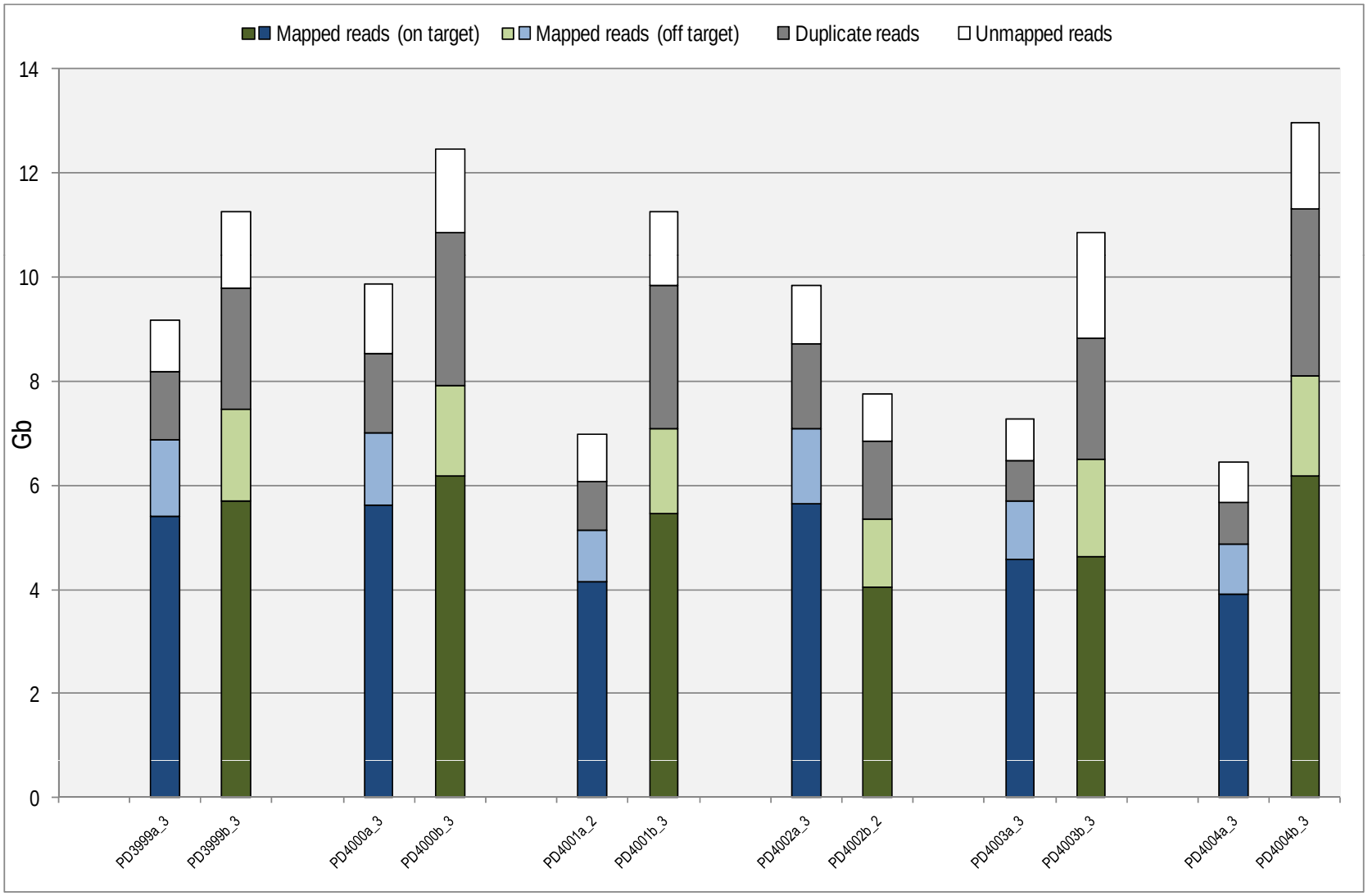


Performance Metrics

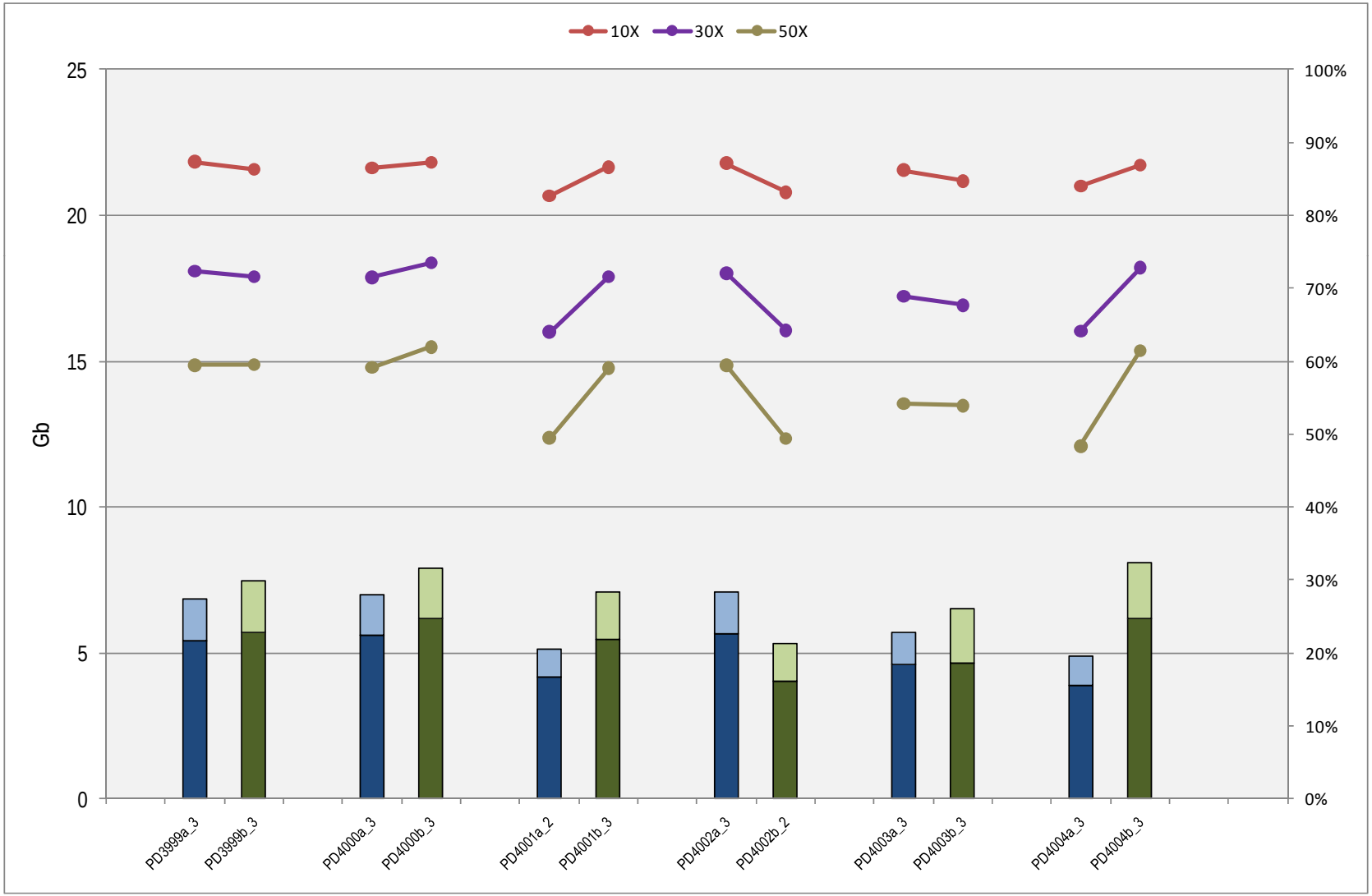




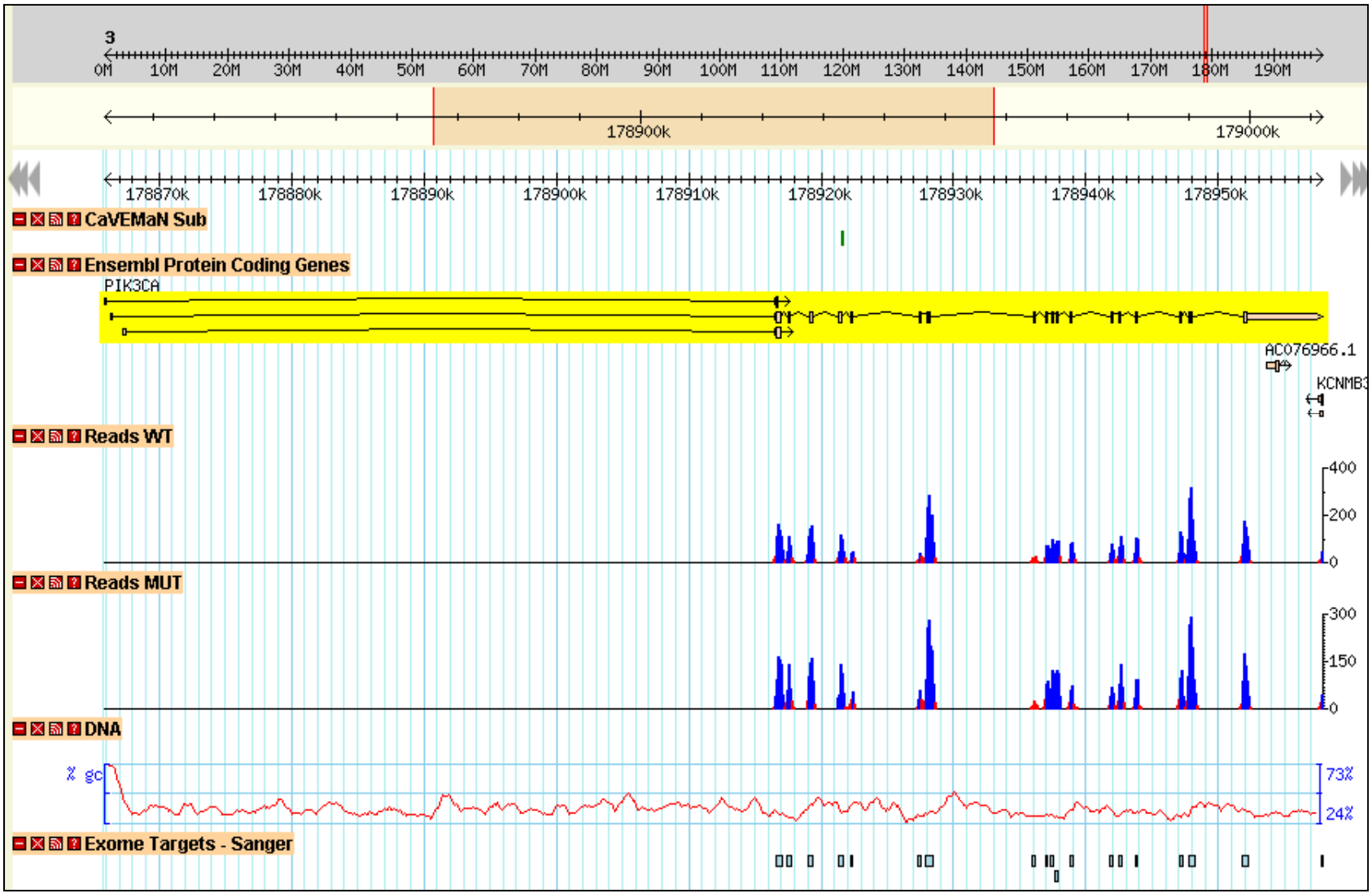
Sequence Metrics



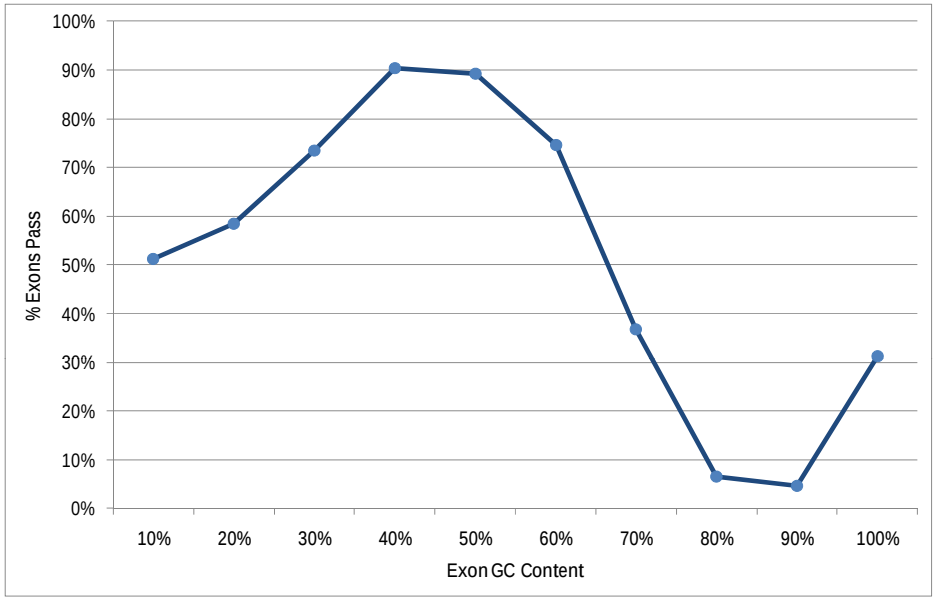
Coverage Metrics

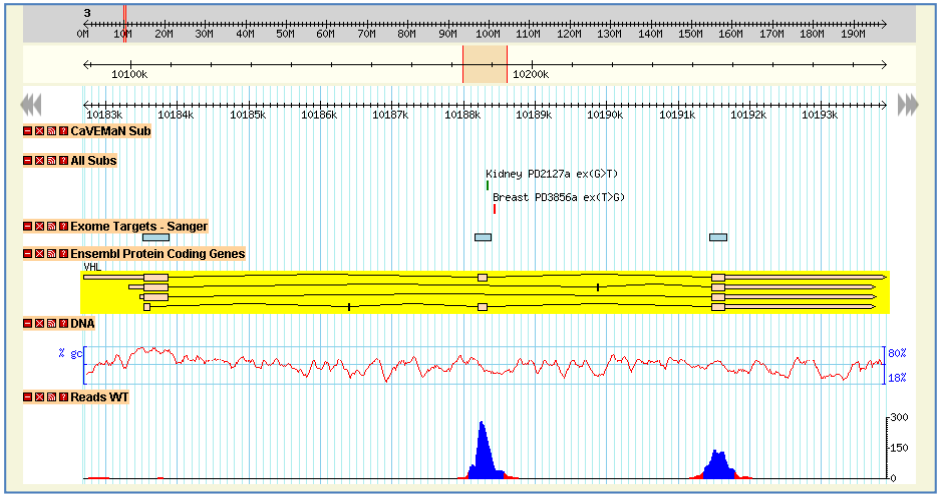


Gbrowse:PIK3CA



GC Effect


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% Exons Pass
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Exon GC Content


3
10100k
10200k
CaVEMaN Sub
All Subs
Kidney PD2127a ex(G>T)
Breast PD3856a ex(T>G)
Exome Targets - Sanger
Ensembl Protein Coding Genes
VHL
DNA
% gc
80%
18%
Reads WT
300
150
0

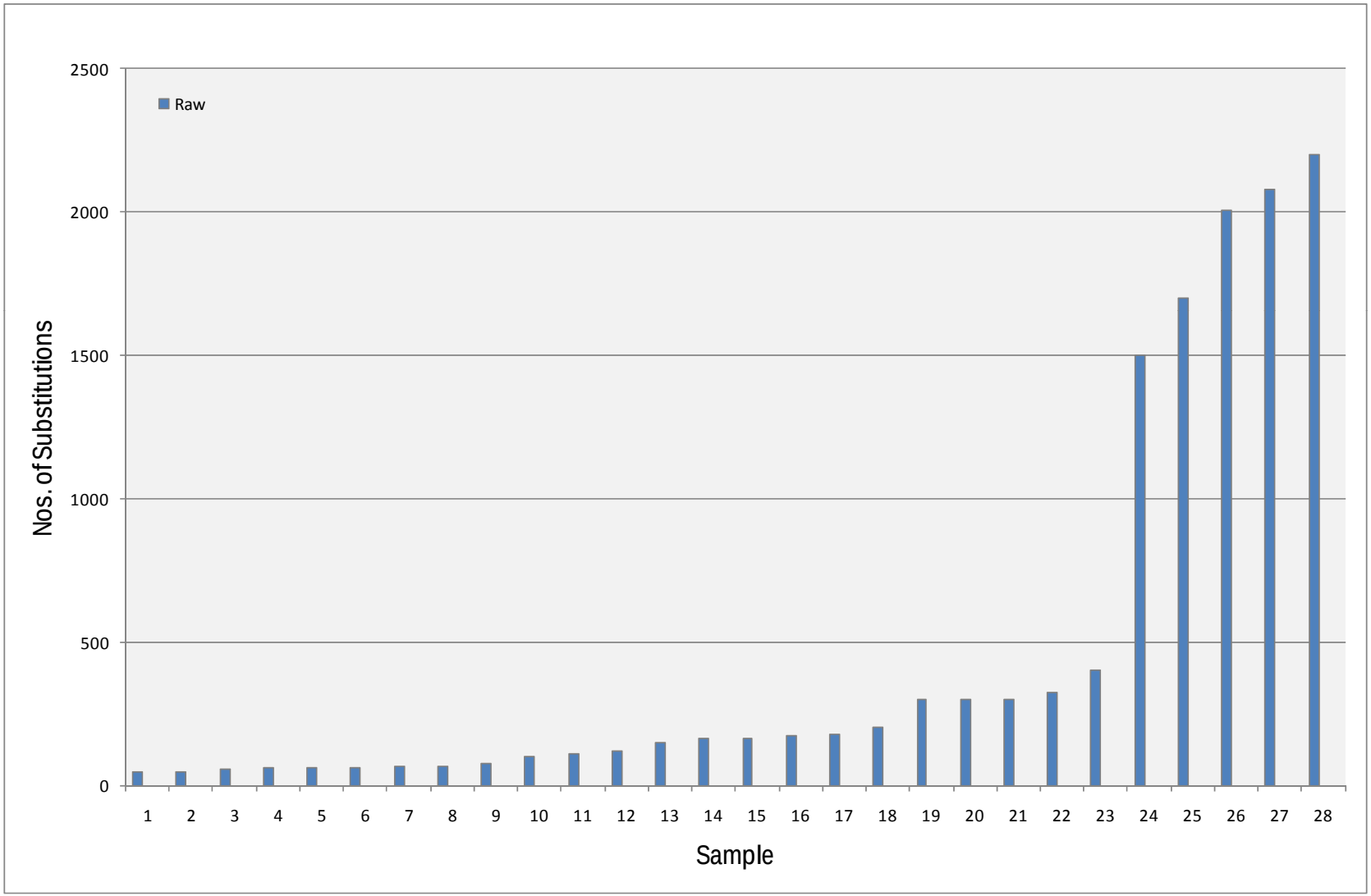
Variant Detection

Caveman: Substitutions

Pindel: Insetions/Deletions

Variant Detection: Specificity

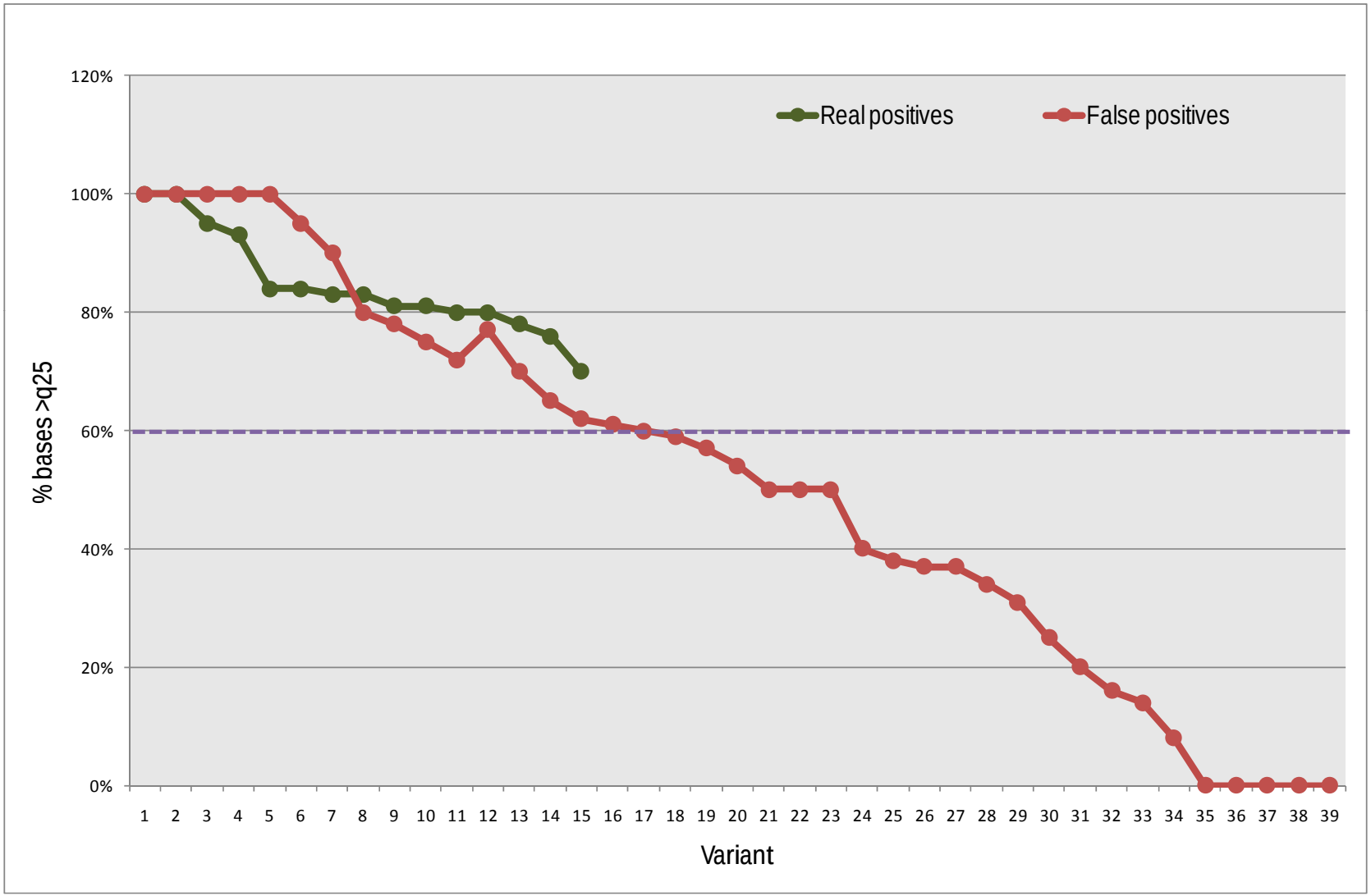
Caveman: Specificity



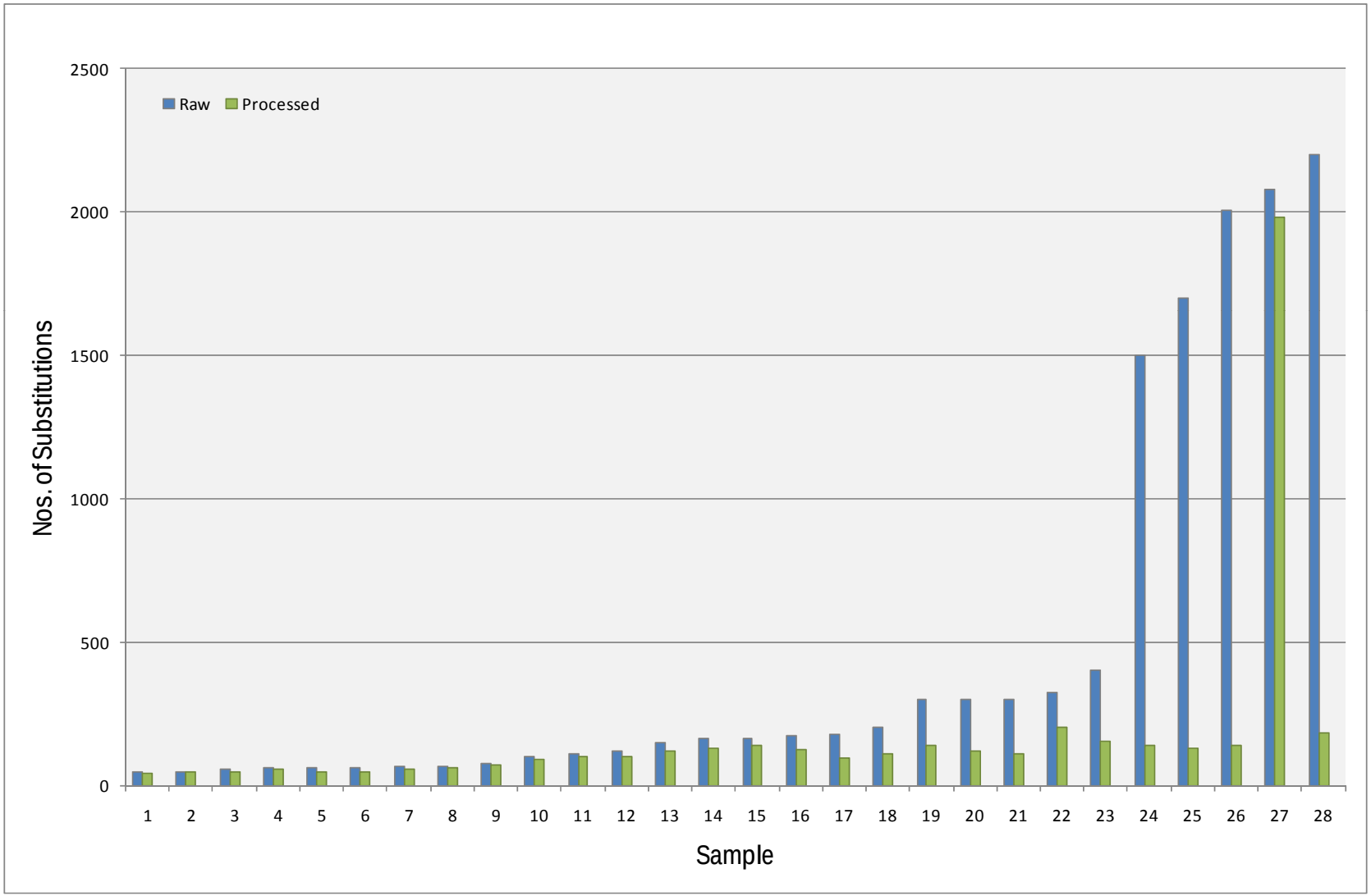
Caveman: Specificity

- Quality score of variant bases
- Position of variant bases within 76bp read
- Representation of variant bases on both strands

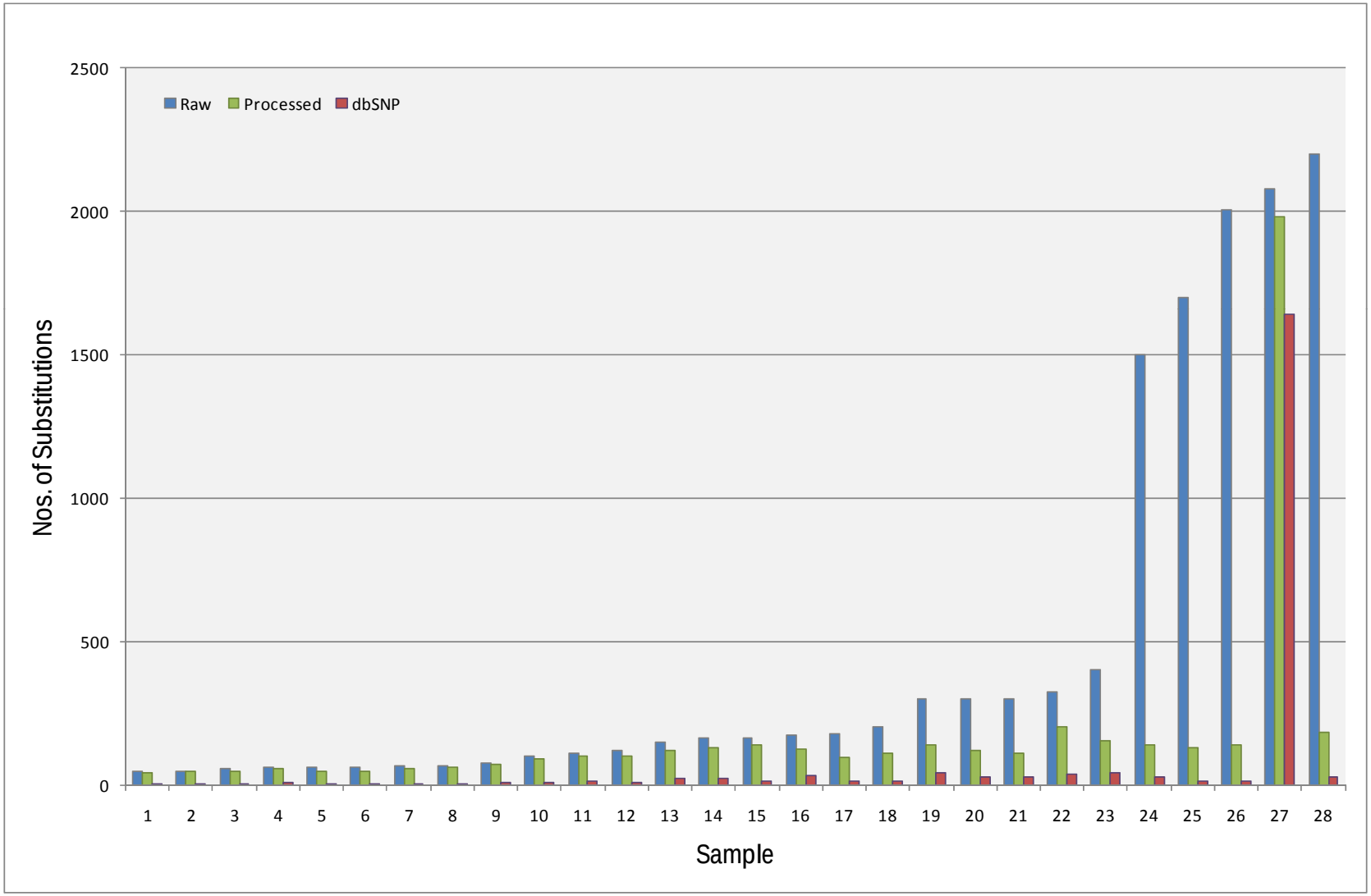
Post-Processing: Base q Score



Caveman: Specificity



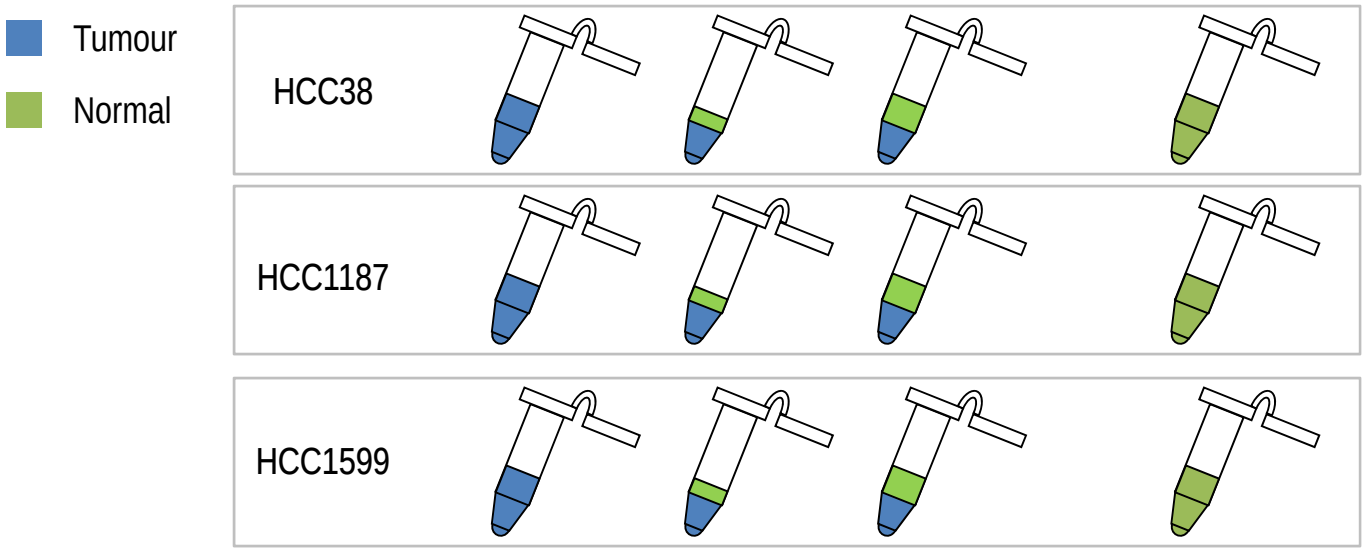
Caveman: Specificity

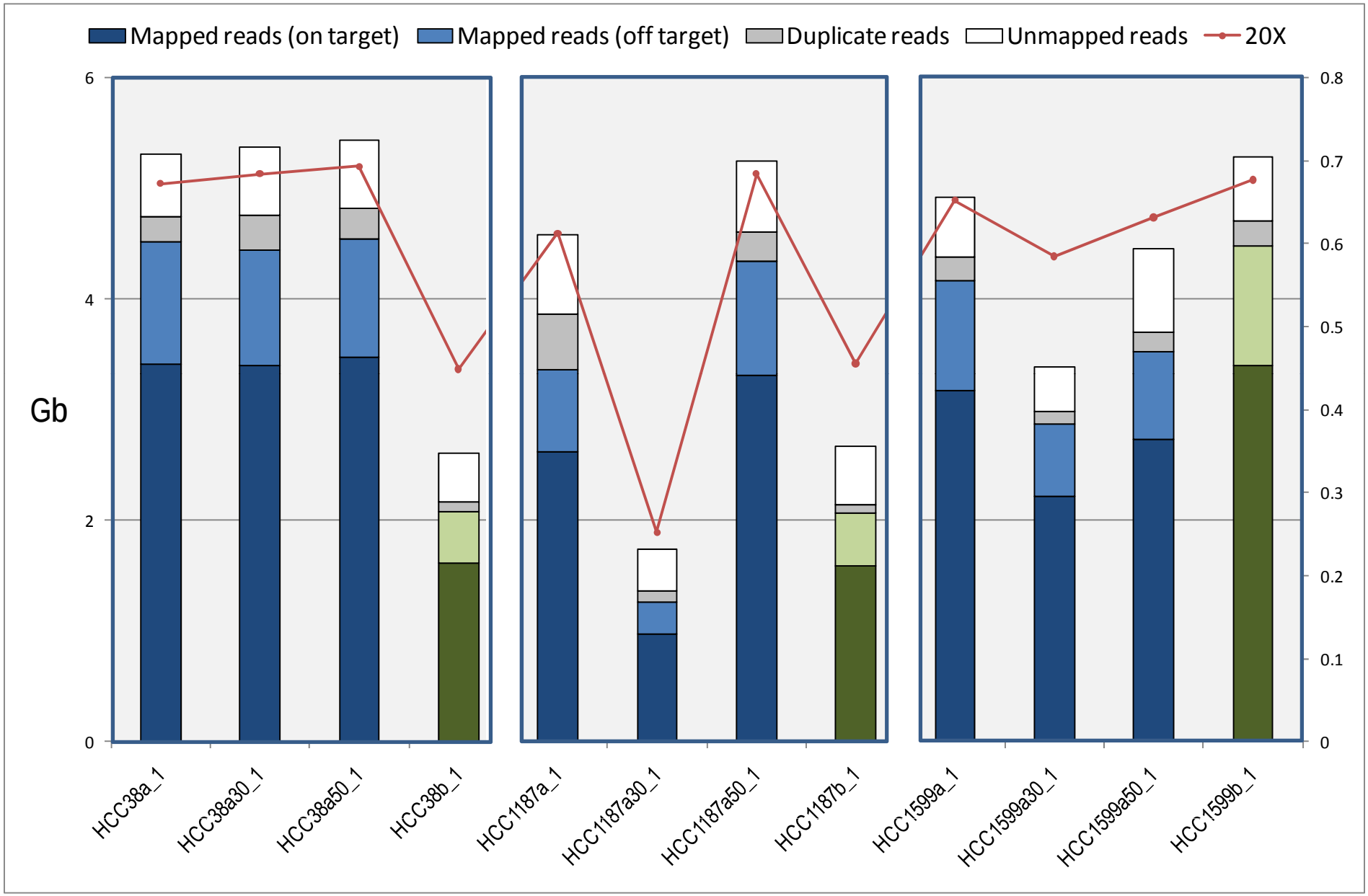


Variant Detection: Sensitivity

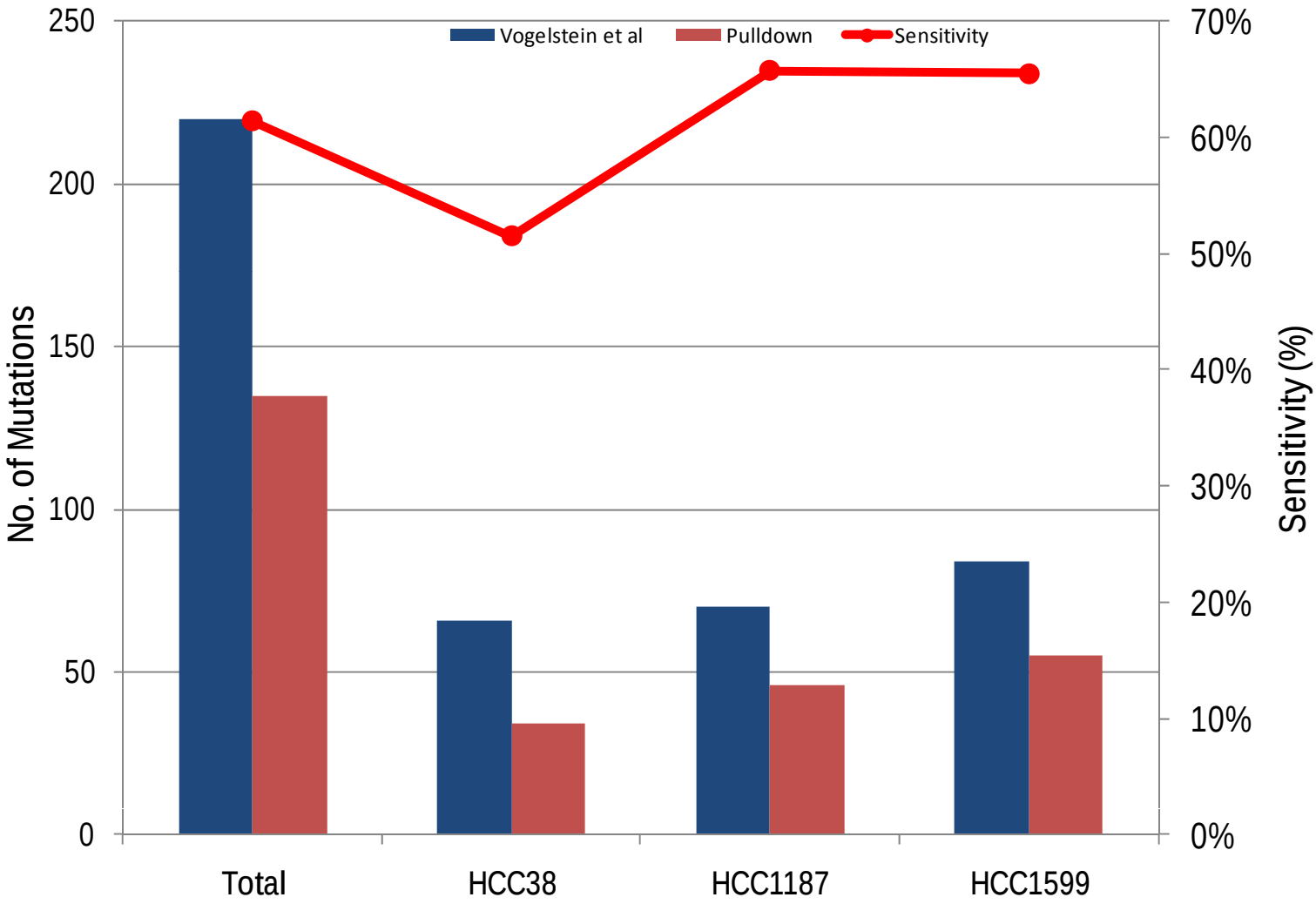
The Genomic Landscapes of Human Breast and Colorectal Cancers

Laura D. Wood,[1*] D. Williams Parsons,[1*] Siân Jones,[1*] Jimmy Lin,[1*] Tobias Sjöblom,[1*†] Rebecca J. Leary,[1] Dong Shen,[1] Simina M. Boca,[1,2] Thomas Barber,[1‡] Janine Ptak,[1] Natalie Silliman,[1] Steve Szabo,[1] Zoltan Dezso,[3] Vadim Ustyanksky,[3] Tatiana Nikolskaya,[3,4] Yuri Nikolsky,[3] Rachel Karchin,[5] Paul A. Wilson,[5] Joshua S. Kaminker,[6] Zemin Zhang,[6] Randal Croshaw,[7] Joseph Willis,[8] Dawn Dawson,[8] Michail Shipitsin,[9] James K. V. Willson,[10] Saraswati Sukumar,[11] Kornelia Polyak,[9] Ben Ho Park,[11] Charit L. Pethiyagoda,[12] P. V. Krishna Pant,[12] Dennis G. Ballinger,[12] Andrew B. Sparks,[12§] James Hartigan,[13] Douglas R. Smith,[13] Erick Suh,[13] Nickolas Papadopoulos,[1] Phillip Buckhaults,[7] Sanford D. Markowitz,[14] Giovanni Parmigiani,[1‖] Kenneth W. Kinzler,[1‖] Victor E. Velculescu,[1‖] Bert Vogelstein[1‖]

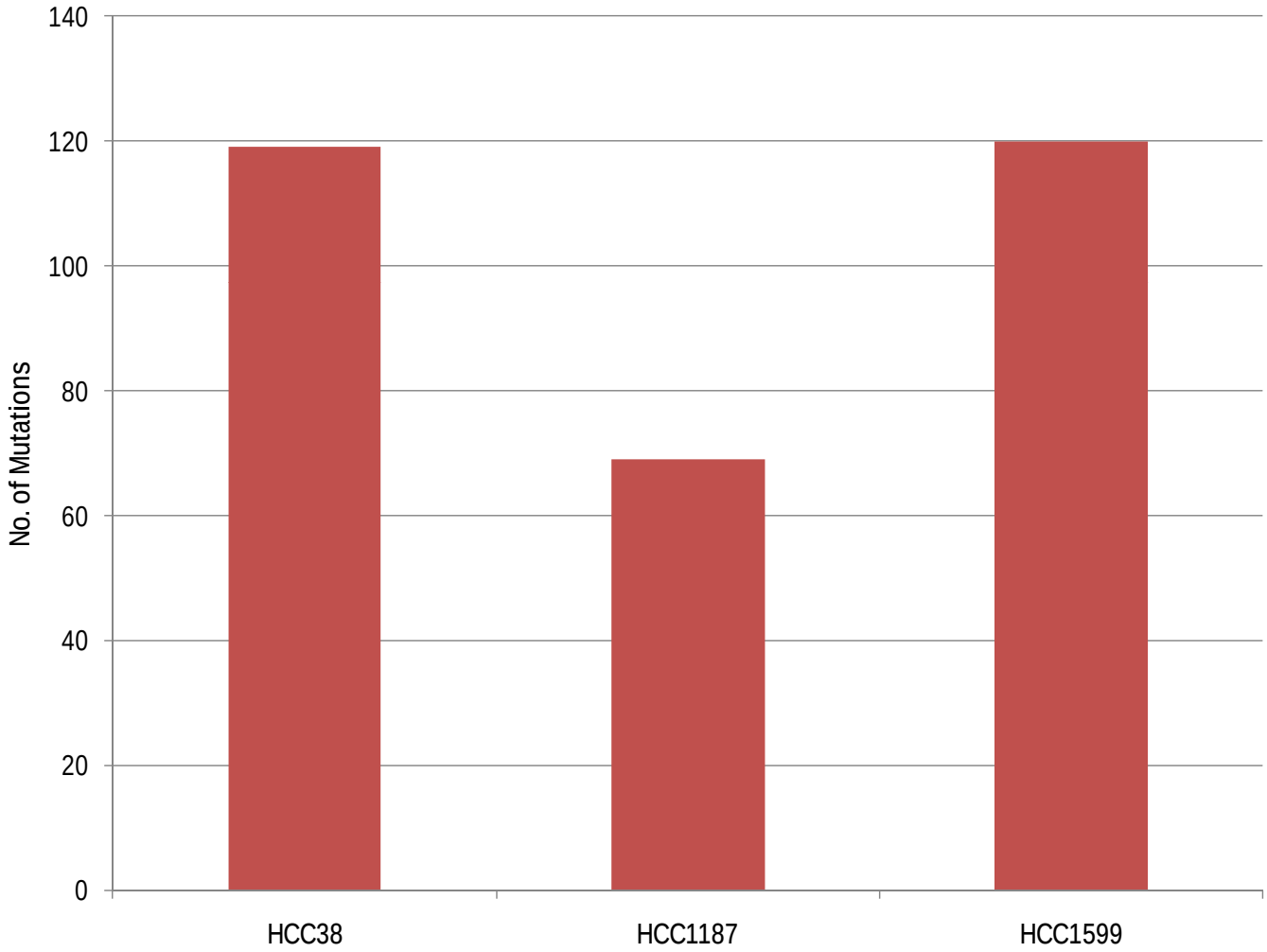



Mapped reads (on target)
Mapped reads (off target)
Duplicate reads
Unmapped reads
20X
Gb
6
4
2
0
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
HCC38a_1
HCC38a30_1
HCC38a50_1
HCC38b_1
HCC1187a_1
HCC1187a30_1
HCC1187a50_1
HCC1187b_1
HCC1599a_1
HCC1599a30_1
HCC1599a50_1
HCC1599b_1

Variant Detection: Sensitivity



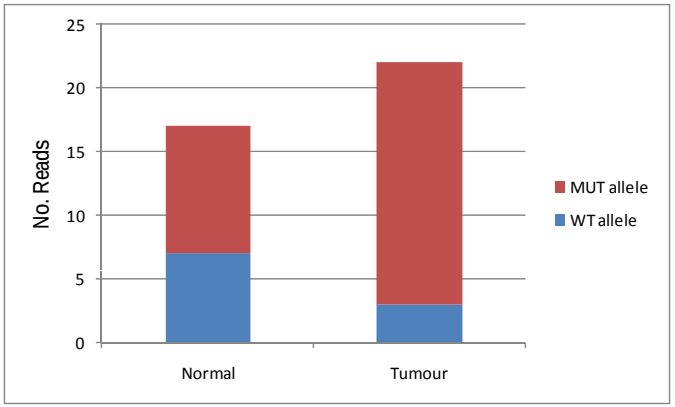
Novel Substituions



Germline Variants in Known Cancer Genes

28 Sporadic breast cancer cases

CHEK2: W411X
BRCA2: del4bp



Somatic Variants in Known Cancer Genes

Search

Tissue Overview for Breast

The genes displayed below are associated with samples that have the tissue type you have selected.

» Top 5 Genes with Sample Data

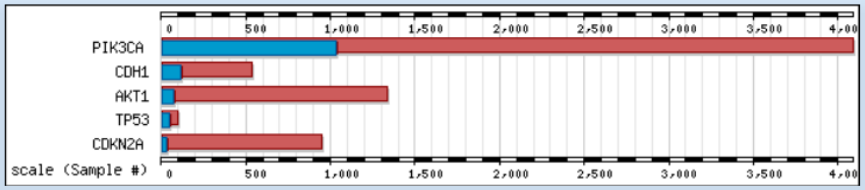


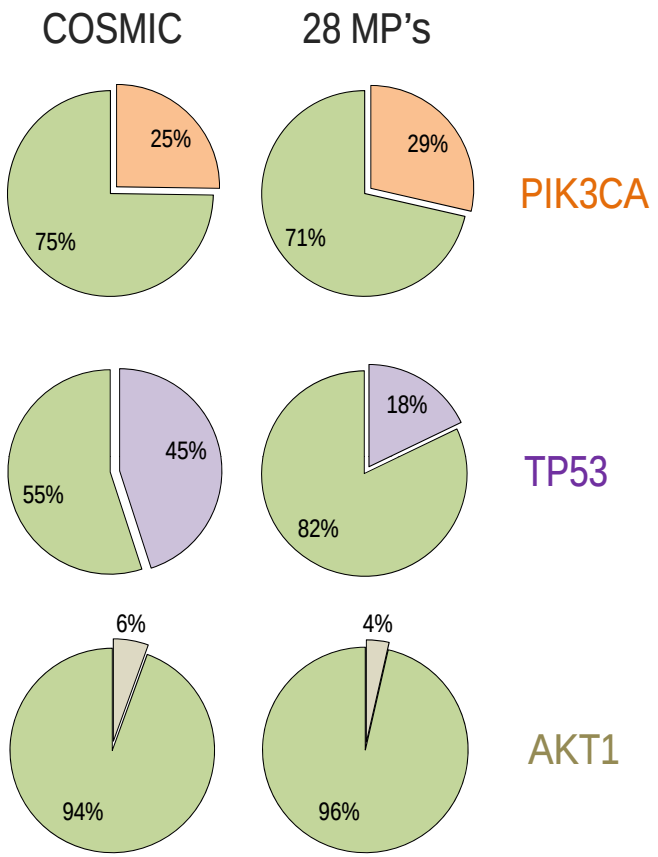
Image Key: Samples with mutations | All samples

Gene Name	Sample Number	Positive Samples	Percent Mutated
PIK3CA	4085	1031	25%
CDH1	539	116	22%
AKT1	1334	74	6%
TP53	100	45	45%
CDKN2A	948	32	3%

» Other genes with mutations for this tissue

PTEN (30), APC (19), KRAS (15), BRCA1 (13), RB1 (12), BRAF (8), CTNNB1 (7), MAP2K4 (7), BRCA2 (6), ERBB2 (5), OBSCN (5), SPTAN1 (5), SMAD4 (4), TTN (4), CTNNA1 (4), ATP8B1 (4), ZNF668 (4), DIP2C (4), CUBN (4), FLNB (4), TECTA (4), HRAS (3), NF2 (3), PRPF4B (3), ATR (3), SMG1 (3), SgK085 (3), SYNE2 (3), CDH20 (3), RFX2 (3), ICAM5 (3), CIC (3), GRIN2D (3), FSCB (3), SIX4 (3), ABCA3 (3), MYH1 (3), DNAH9 (3), MAMDC4 (3), SORL1 (3), COL7A1 (3), VEPH1 (3), LRBA (3), ODZ1 (3), MACF1 (3), GLI1 (3), TMPRSS6 (3), LRRFIP1 (3), APC2 (3), NUP214 (3), ADAM12 (3), SBNO1 (3), AKAP9 (3), HDLBP (3), CENTD3 (3), CHD5 (3), KIAA1632 (3), ATM (2), FBXW7 (2), NRAS (2), TAF1 (2), TEK (2), MAP3K6 (2), PRKAA2 (2), WNK1 (2), JAK3 (2), PTCH1 (2), NF1 (2), UTX (2), ROCK1 (2), EPHA10 (2), PFTK2 (2), NRK (2), HNF1A (2), SMARCA4 (2), KIAA0999 (2), PARP1 (2), MRE11A (2), SMARCAL1 (2), XDH (2), DYSF (2), KCNIP3 (2), CNNM4 (2), SLC9A2 (2),

Sample	PIK3CA	TP53	Other
PD3849a			
PD3852a			
PD3853a			
PD3854a			
PD3856a	PIK3CA: H1047R		
PD3857a	PIK3CA: H1047R		
PD3858a			
PD3983a	PIK3CA: H1047R		
PD3984a			
PD3985a	PIK3CA: H1047R	TP53: R306*	
PD3986a		TP53: Ess splice	
PD3987a		TP53: Y220C	
PD3988a	PIK3CA: H1047R		
PD3989a	PIK3CA: E545K		
PD3990a			
PD3991a		TP53:G245S	
PD3992a	PIK3CA: H1047R		
PD3993a			
PD3994a	PIK3CA: N354K		
PD3995a			AKT1: E17K
PD3996a			PTEN: Y27D NF1: -1G>T
PD3997a			
PD3998a			
PD3999a			
PD4000a			
PD4001a			
PD4002a		TP53: H179Y	
PD4003a			
PD4004a			

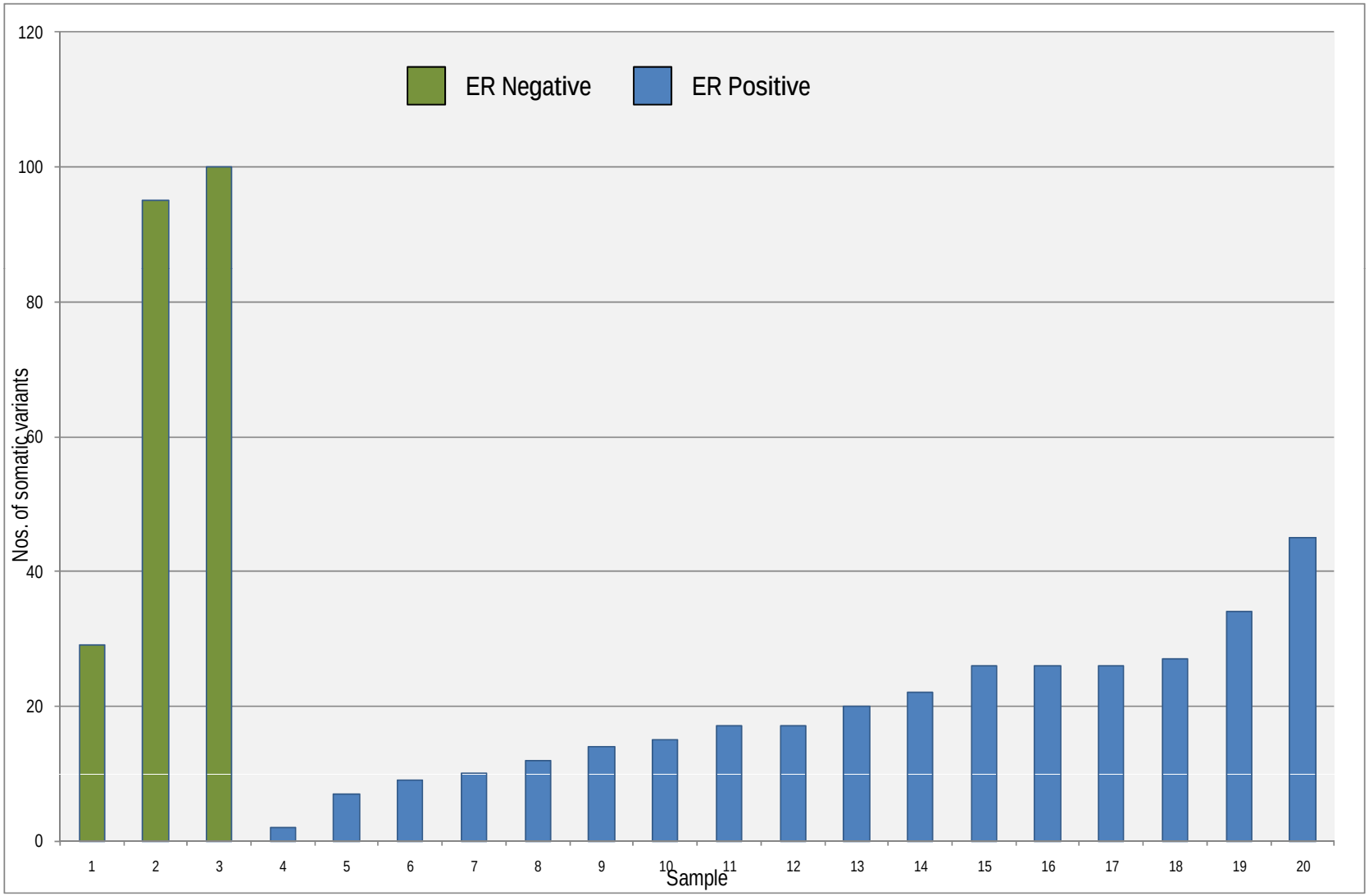


Novel Somatic Variants

638 **Confirmed Somatic Subs**

38 **Confirmed Somatic InDels**

Novel Somatic Variants



Summary

Exome resequencing has proven to be robust and effective for somatic variant detection

Candidate genes are being followed up in additional breast cancer cases

Many additional breast cancer whole exomes are being processed

Applying protocol to additional cancer types

Mike Stratton
Andy Futreal
Peter Campbell

Phil Stephens
Ignacio Varela
Calli Latimer
Serena Nik Zainal
Keiran Raine
David Jones
Mingming Jia
Andrew Menzies
David McBride
Graham Bignell
Helen Davies
Wendy Haynes
Catherine Leroy
Meng-Lay Lin
Rebecca Shepherd
Juok Cho

Adam Butler
Lucy Stebbings
Kelly Halliday
Simon Forbes
Elisabeth Dawson
Sally Bamford
David Beare
Gurpreet Bhamra
Nidhi Bindal
Claire Stevens
Laura Muddie
Andrew Barthorpe
Arjunan Rajasingham
Stuart Mclaren
Sancha Martin
King Wai Lau

Alison Coffey
Eleanor Howard

Dan Turner
Lira Mamanova

Felix Kokocinski
Carol Scott
Kai Ye

The Wellcome Trust