Somatic and germline variant calling

Outline

- Tumor purity and clonality
- Somatic mutation vs. germline mutation vs. germline polymorphism
- Germline variant calling
  - Methodology
  - Tools
  - Quiz
- Somatic variant calling
  - Methodology
  - Tools
  - Quiz
- Variant annotation (VEP)
- File format
- Manual curation
- Mutational signatures
  - General about signatures
  - Microsatellite instability/hypermutation
  - Quiz

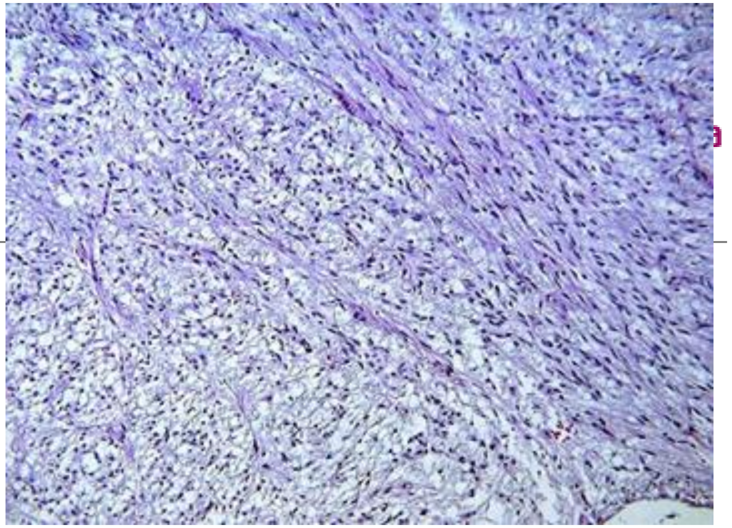
Learning outcomes and course content

- Learning outcomes:
  - Understand how to apply technology to obtain relevant information from the cancer genome.
  - Understand how to apply technology to obtain relevant information from the cancer genome.
  - Call somatic- and germline variation.
  - Annotate somatic- and germline variation.

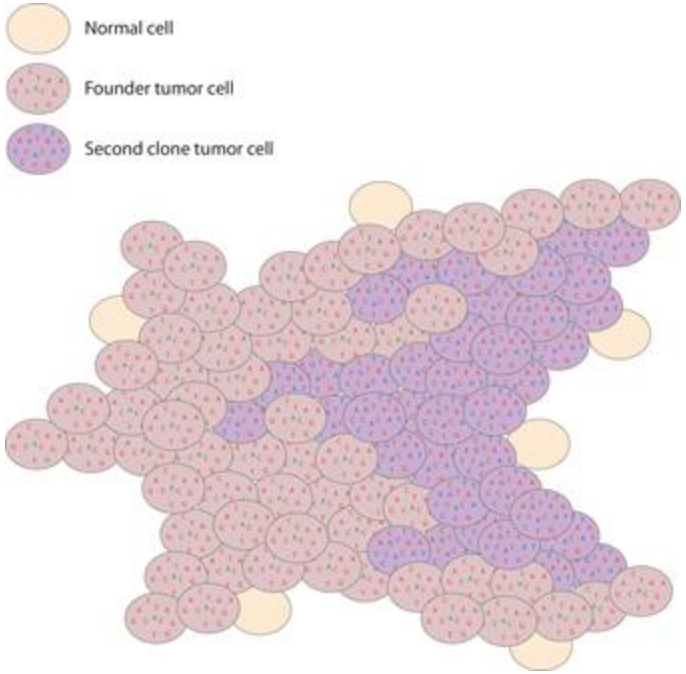
- Course content:
  - Calling somatic- and germline variation:
    - Point mutations and indels.
  - File formats for variant calling.
  - Annotating somatic- and germline variation.

Tumor purity and clonality

- Tumors are often impure
  - Mix of tumor cells and normal (germline) cells
  - Tumor purity expressed as fraction or percentage
  - Cancer DNA fraction is a very closely related term
- Tumors often contain multiple clones
  - Diverse collections of cells harboring different mutations
  - Often one original clone with initial mutations
  - Subclones containing additional mutations may form
  - Treatments may favor one subclone which has resistance mutation, causing it to take over
    - → clonal evolution

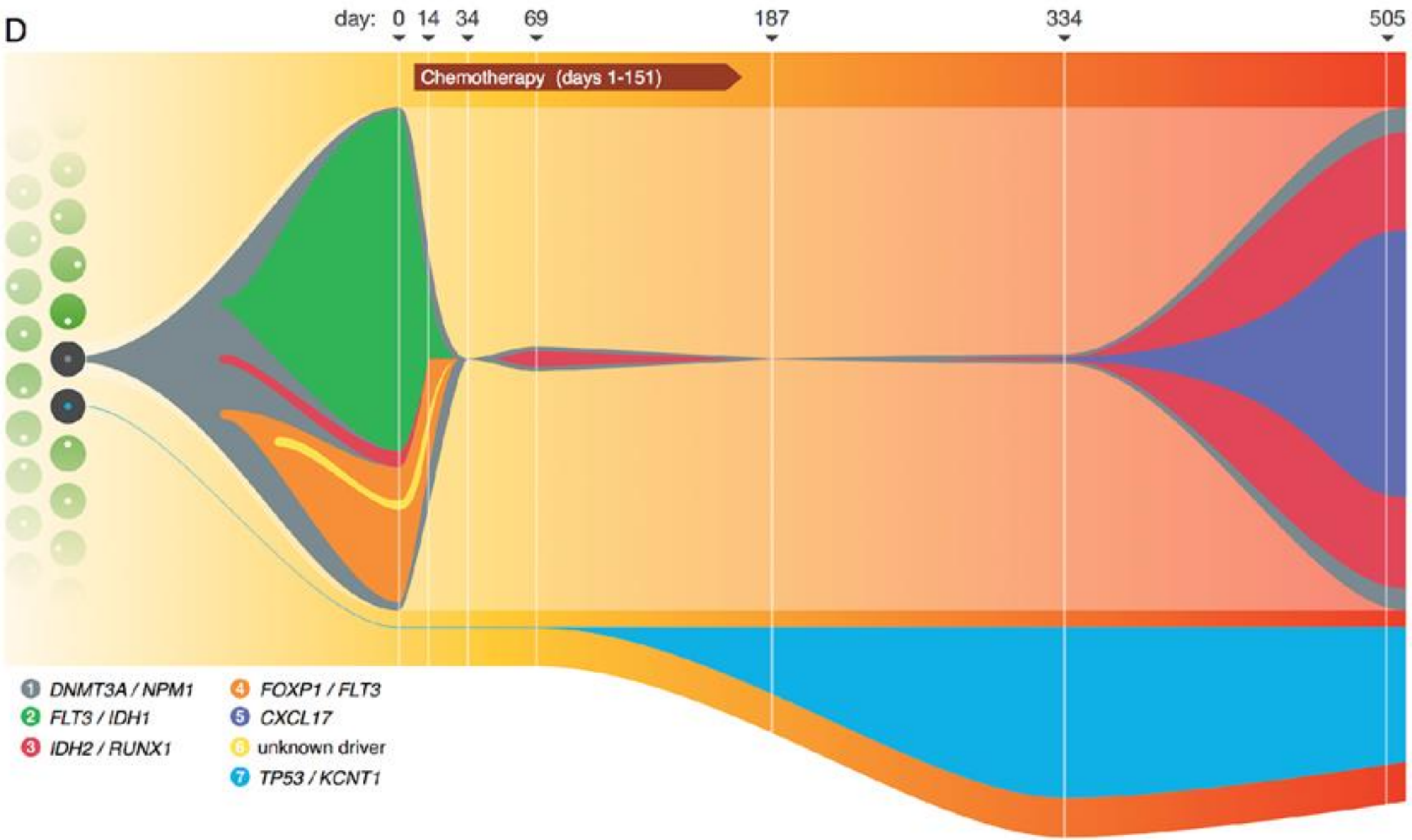


Credit Griffith brothers’ lecture #5



Credit Griffith brothers’ lecture #5

Clonal evolution



Credit Griffith brothers’ lecture #5

Somatic mutation vs. germline mutation vs. germline polymorphism

- Germline mutations
  - Present in egg or sperm
    - All cells of affected offspring
  - Heritable
  - Cause of familial cancers

- Germline polymorphisms
  - Present in egg or sperm
    - All cells of affected offspring
  - Heritable
  - Common in the population, > 1%
  - Generally not impacting disease
  - SNP – Single Nucleotide Polymorphism

- Somatic mutations
  - Occur in non-germline tissues
    - Only tumor cells (breast, lung, blood, etc.)
  - Non-heritable
  - Various reasons
    - Smoking, UV light, oxidation in cells, etc.
  - Cause of sporadic cancers
    - And of familial cancers – in combination with germline mutations

Small variants

- SNVs – single nucleotide variants
  - Change from one base to another at one single position of the genome
- Indels – Small insertions and deletions
  - 1 – ~30 bases added or removed

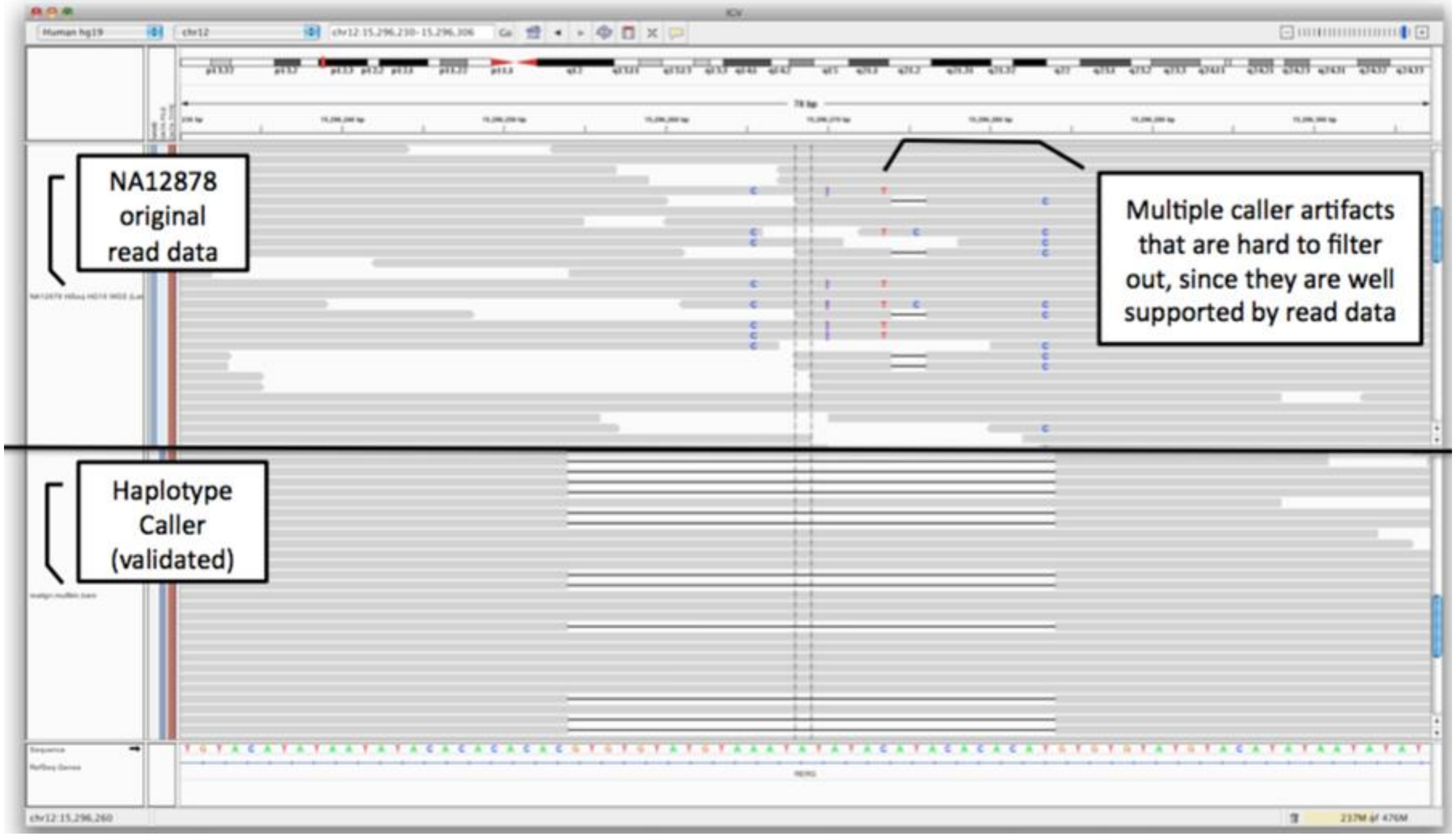
Germline mutation calling – methodology

- Mutation/variant calling – to identify mutations from e.g. sequencing data
  - Focus: DNA data
- Germline sample: often white blood cells, WBCs
- SNVs and small indels
- VAFs (variant allele frequencies): ~50% or 100%
  - One or two mutated alleles out of totally two alleles
- General method:
  - Find positions where a fraction of mapped reads have base deviating from the reference genome – alternate allele
  - If significant difference from the reference – call a germline variant

Germline mutation calling – methodology and tools

- Call SNVs and Indels separately by considering each variant locus
  - Very fast
  - Assumes bases are independent
- Call SNVs and indels simultaneously via Bayesian genotype likelihood model
  - More computationally intensive
- Call SNVs, indels and SVs simultaneously by performing a local de novo assembly
  - More computationally intensive
  - More accurate – gets rid of many false positives especially indels
  - GATK HaplotypeCaller
  - Strelka

Germline mutation calling – methodology



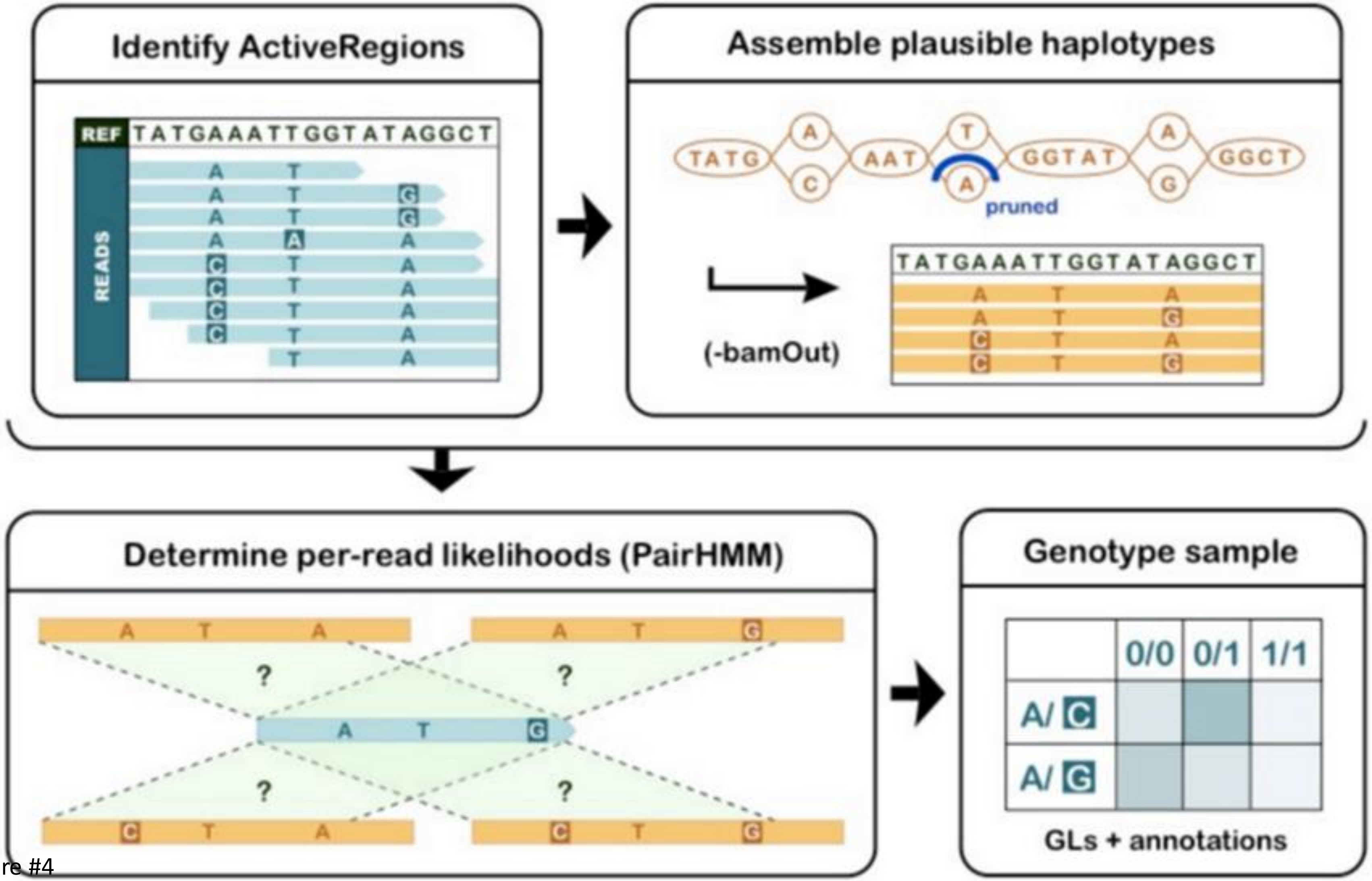
NA12878 original read data

Multiple caller artifacts that are hard to filter out, since they are well supported by read data

Haplotype Caller (validated)

Credit Griffith brothers’ lecture #4

Germline mutation calling – HaplotypeCaller

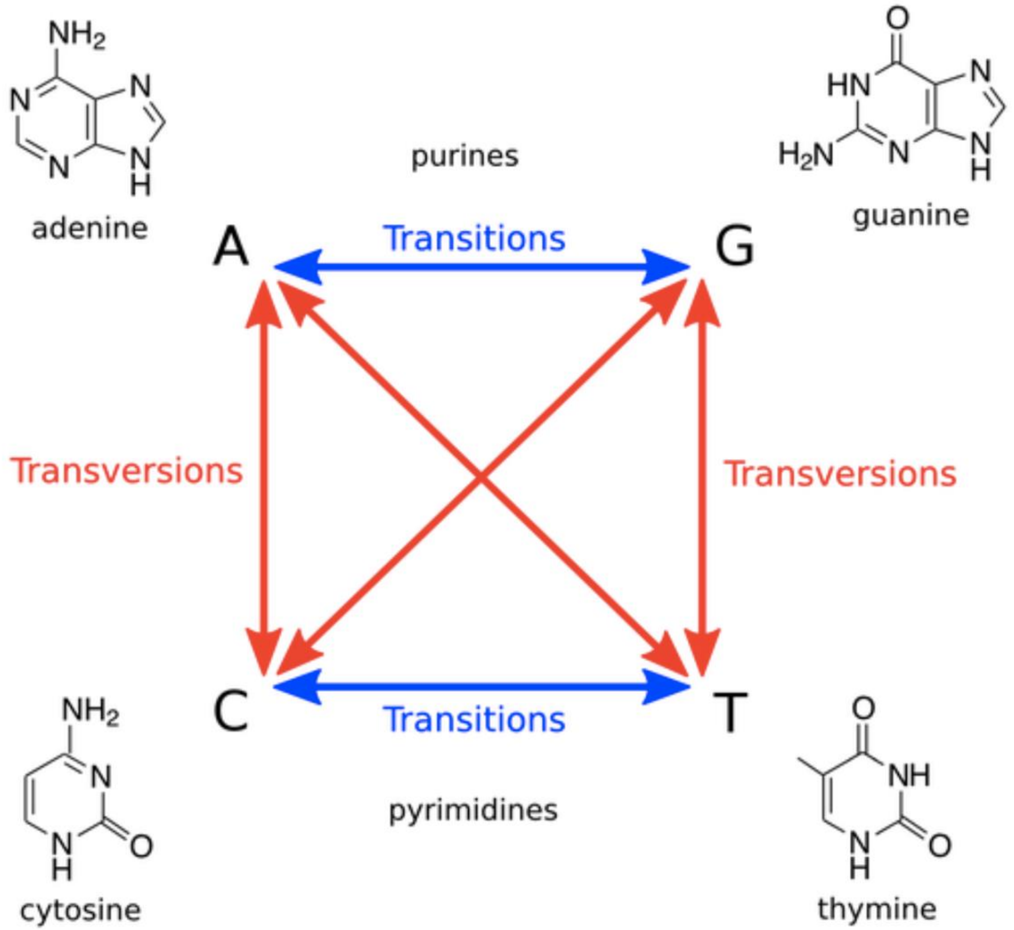


Credit Griffith brothers’ lecture #4

GATK recommended filters

- SNPs
  - QD < 2.0 (variant quality/depth of non-ref samples)
  - MQ < 40.0 (Mapping quality)
  - FS > 60.0 (Phred score Fisher's test pvalue for strand bias)
  - SOR > 3.0 (Strand odds ratio, aims to evaluate whether there is strand bias in the data)
  - MQRankSum < -12.5 (mapping quality of reference reads vs alt reads)
  - ReadPosRankSum < -8.0 (distance of alt reads from end of the read)
- Indels
  - QD < 2.0
  - FS > 200.0
  - SOR > 10.0
  - ReadPosRankSum < -20.0
  - InbreedingCoeff < -0.8

Transition/Transversion ratio (Ti/Tv)



Ratios:

Random = 0.5

WGS = 2.0-2.1

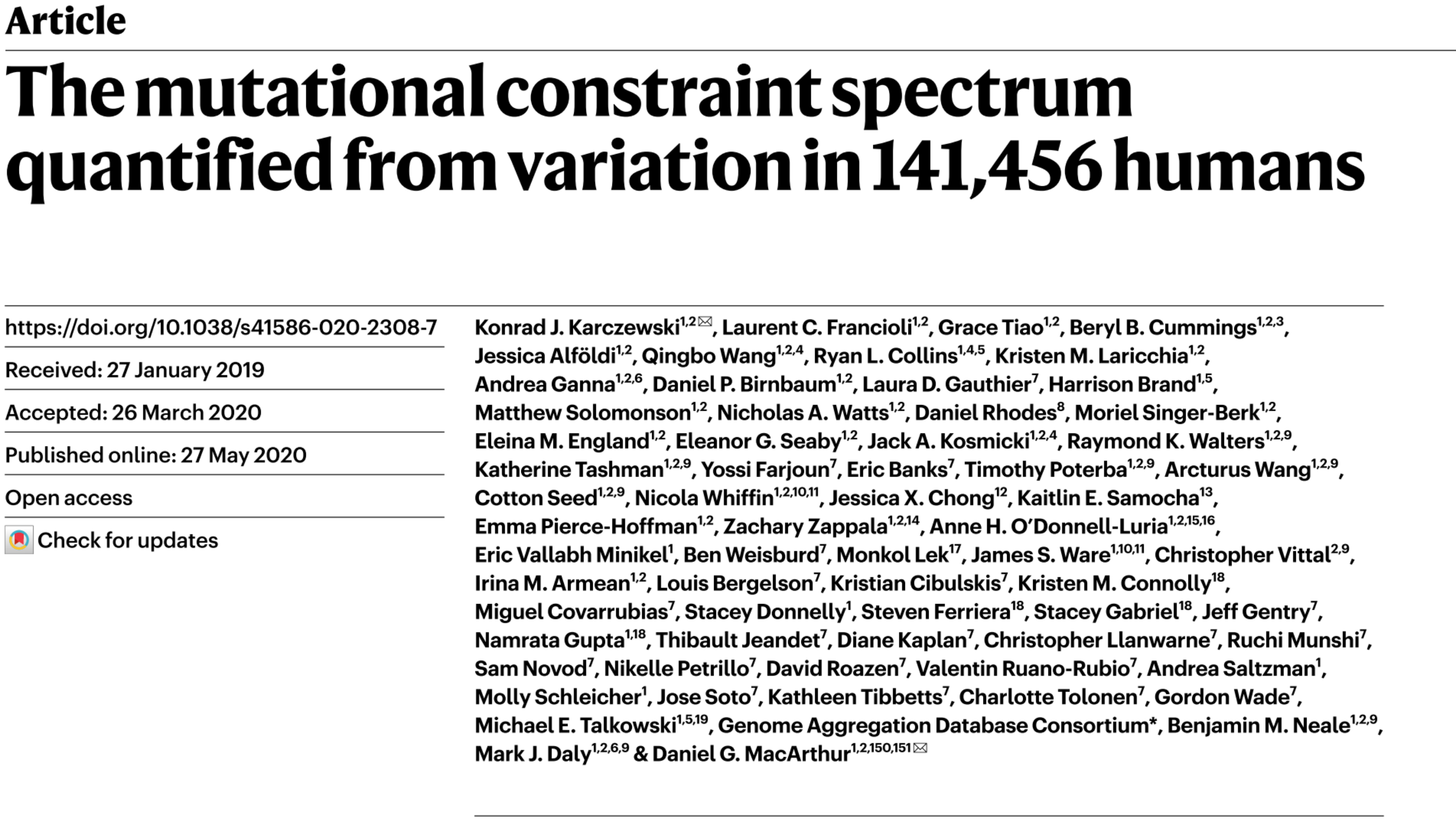
Exome = 3-3.5

Watch for major deviation from typical ratio

Credit Griffith brothers' lecture #4

A good paper

- The “gnomAD” paper
- Published in Nature
- Supplementary Information:
  Many details on how to properly call and filter germline variants



Article

The mutational constraint spectrum quantified from variation in 141,456 humans

https://doi.org/10.1038/s41586-020-2308-7

Received: 27 January 2019

Accepted: 26 March 2020

Published online: 27 May 2020

Open access

Check for updates

Konrad J. Karczewski[1,2✉], Laurent C. Francioli[1,2], Grace Tiao[1,2], Beryl B. Cummings[1,2,3], Jessica Alföldi[1,2], Qingbo Wang[1,2,4], Ryan L. Collins[1,4,5], Kristen M. Laricchia[1,2], Andrea Ganna[1,2,6], Daniel P. Birnbaum[1,2], Laura D. Gauthier[7], Harrison Brand[1,5], Matthew Solomonson[1,2], Nicholas A. Watts[1,2], Daniel Rhodes[8], Moriel Singer-Berk[1,2], Eleina M. England[1,2], Eleanor G. Seaby[1,2], Jack A. Kosmicki[1,2,4], Raymond K. Walters[1,2,9], Katherine Tashman[1,2,9], Yossi Farjoun[7], Eric Banks[7], Timothy Poterba[1,2,9], Arcturus Wang[1,2,9], Cotton Seed[1,2,9], Nicola Whiffin[1,2,10,11], Jessica X. Chong[12], Kaitlin E. Samocha[13], Emma Pierce-Hoffman[1,2], Zachary Zappala[1,2,14], Anne H. O’Donnell-Luria[1,2,15,16], Eric Vallabh Minikel[1], Ben Weisburd[7], Monkol Lek[17], James S. Ware[1,10,11], Christopher Vittal[2,9], Irina M. Armean[1,2], Louis Bergelson[7], Kristian Cibulskis[7], Kristen M. Connolly[18], Miguel Covarrubias[7], Stacey Donnelly[1], Steven Ferriera[18], Stacey Gabriel[18], Jeff Gentry[7], Namrata Gupta[1,18], Thibault Jeandet[7], Diane Kaplan[7], Christopher Llanwarne[7], Ruchi Munshi[7], Sam Novod[7], Nikelle Petrillo[7], David Roazen[7], Valentin Ruano-Rubio[7], Andrea Saltzman[1], Molly Schleicher[1], Jose Soto[7], Kathleen Tibbetts[7], Charlotte Tolonen[7], Gordon Wade[7], Michael E. Talkowski[1,5,19], Genome Aggregation Database Consortium*, Benjamin M. Neale[1,2,9], Mark J. Daly[1,2,6,9] & Daniel G. MacArthur[1,2,150,151✉]

Germline mutation calling – quiz

1. Are germline mutations heritable?
   a) Yes
   b) No
   c) Only sometimes
2. Which of these variant allele frequencies is most likely for a germline variant in a germline sample?
   a) 25%
   b) 50%
   c) 75%
3. What is a haplotype?
   a) A germline variant caller
   b) A very common germline variant
   c) A possible combination of nearby germline variants

Germline mutation calling – quiz answers

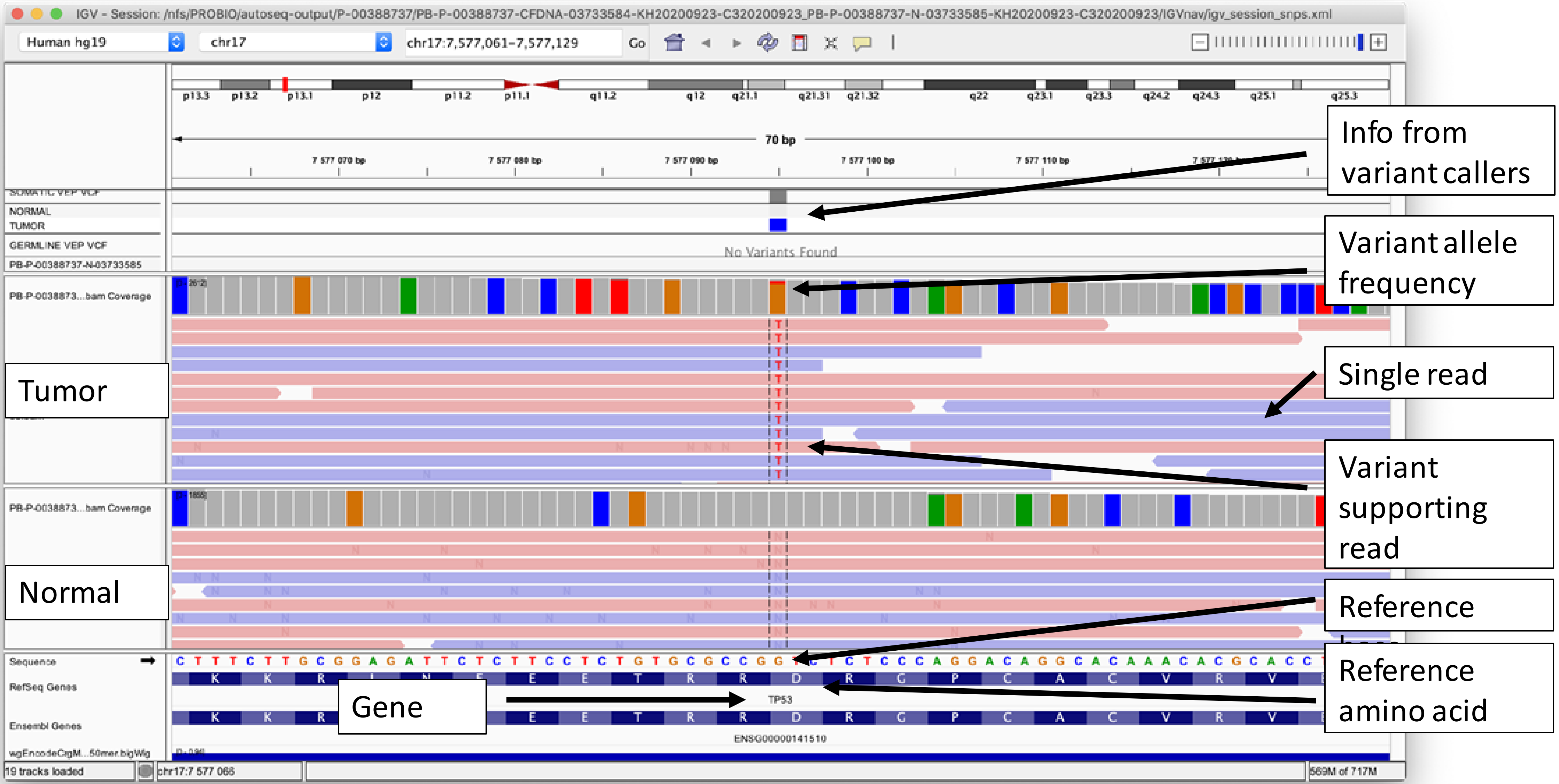
1. Are germline mutations heritable?
   a) Yes – these are the mutations we are born with, and they can be inherited by our children – they are in our germ cells
2. Which of these variant allele frequencies is most likely for a germline variant in a germline sample?
   b) 50% - 1 mutated out of 2 alleles
3. What is a haplotype?
   c) A possible combination of nearby germline variants
      - HaplotypeCaller is a germline variant caller that calls variants by testing the likelihood of possible haplotypes based on the DNA read support

Germline variant calling – Questions?

Somatic mutation calling – methodology

- Mutation/variant calling – to identify mutations from e.g. sequencing data
  - Focus: DNA data
- Somatic mutations are best distinguished by comparison of tumor to a matched normal
- Normal sample: Germline sample from the same individual
  - If not available: use healthy donor sample → requires additional filtering
- SNVs and small indels
- VAFs: ~1% - 100%, depending on purity, ploidy, clonality
  - Limited detection capacity for VAF < 1%
- General method:
  - Find positions where a fraction of mapped reads have base deviating from the reference genome – alternate allele
  - Compare with germline sample
  - If significant difference to germline – call a somatic variant

Somatic mutation calling – SNV



Info from variant callers

Variant allele frequency

Single read

Variant supporting read

Reference

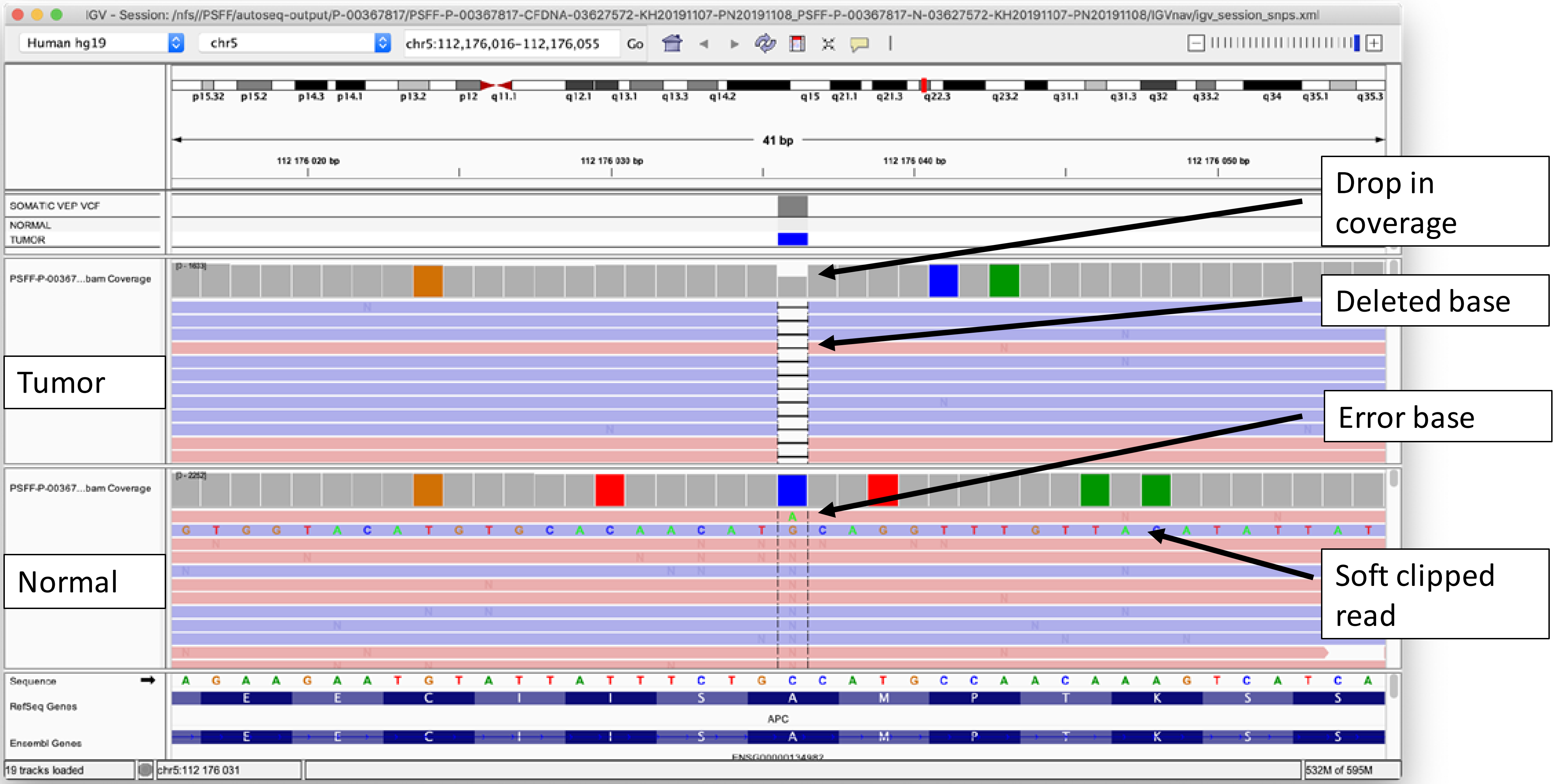
Reference amino acid

Tumor

Normal

Gene

Somatic mutation calling – deletion


IGV - Session: /nfs//PSFF/autoseq-output/P-00367817/PSFF-P-00367817-CFDNA-03627572-KH20191107-PN20191108_PSFF-P-00367817-N-03627572-KH20191107-PN20191108/IGVnav/igv_session_snps.xml
Human hg19
chr5
chr5:112,176,016-112,176,055
Go
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q12.1 q13.1 q13.3 q14.2 q15 q21.1 q21.3 q22.3 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.3
41 bp
112 176 020 bp
112 176 030 bp
112 176 040 bp
112 176 050 bp
SOMATIC VEP VCF
NORMAL
TUMOR
PSFF-P-00367...bam Coverage
[0 - 1633]
PSFF-P-00367...bam Coverage
[0 - 2252]
Sequence
A G A A G A A T G T A T T A T T T C T G C C A T G C C A A C A A A G T C A T C A
RefSeq Genes
E E C I I S A M P T K S S
APC
Ensembl Genes
ENSG00000134982
19 tracks loaded
chr5:112 176 031
532M of 595M

Tumor

Normal

Drop in coverage

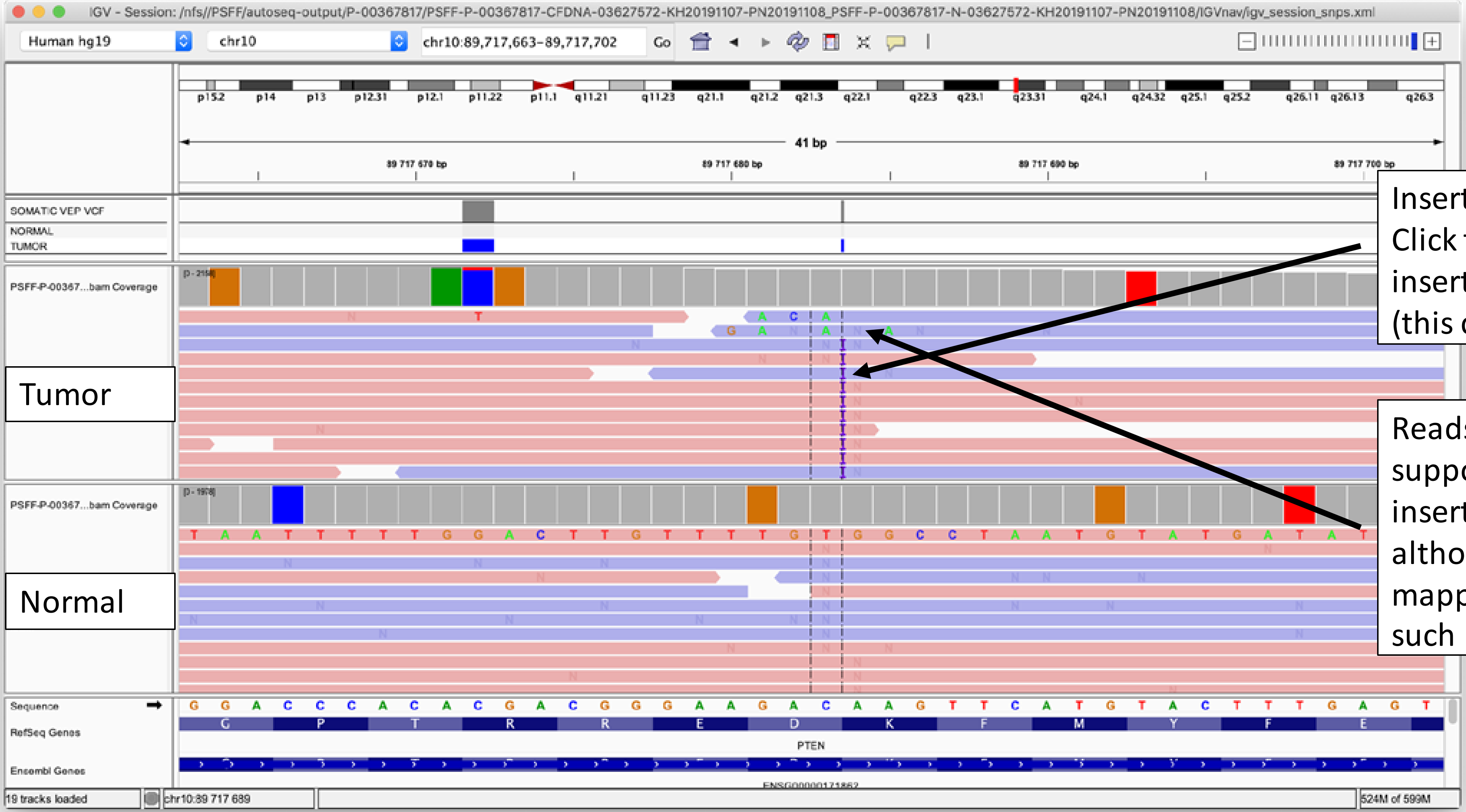
Deleted base

Error base

Soft clipped read

Somatic mutation calling – insertion

Karolinska Institutet



IGV - Session: /nfs//PSFF/autoseq-output/P-00367817/PSFF-P-00367817-CFDNA-03627572-KH20191107-PN20191108_PSFF-P-00367817-N-03627572-KH20191107-PN20191108/IGVnav/igv_session_snps.xml

Human hg19 | chr10 | chr10:89,717,663-89,717,702 | Go

19 tracks loaded | chr10:89 717 689 | 524M of 599M

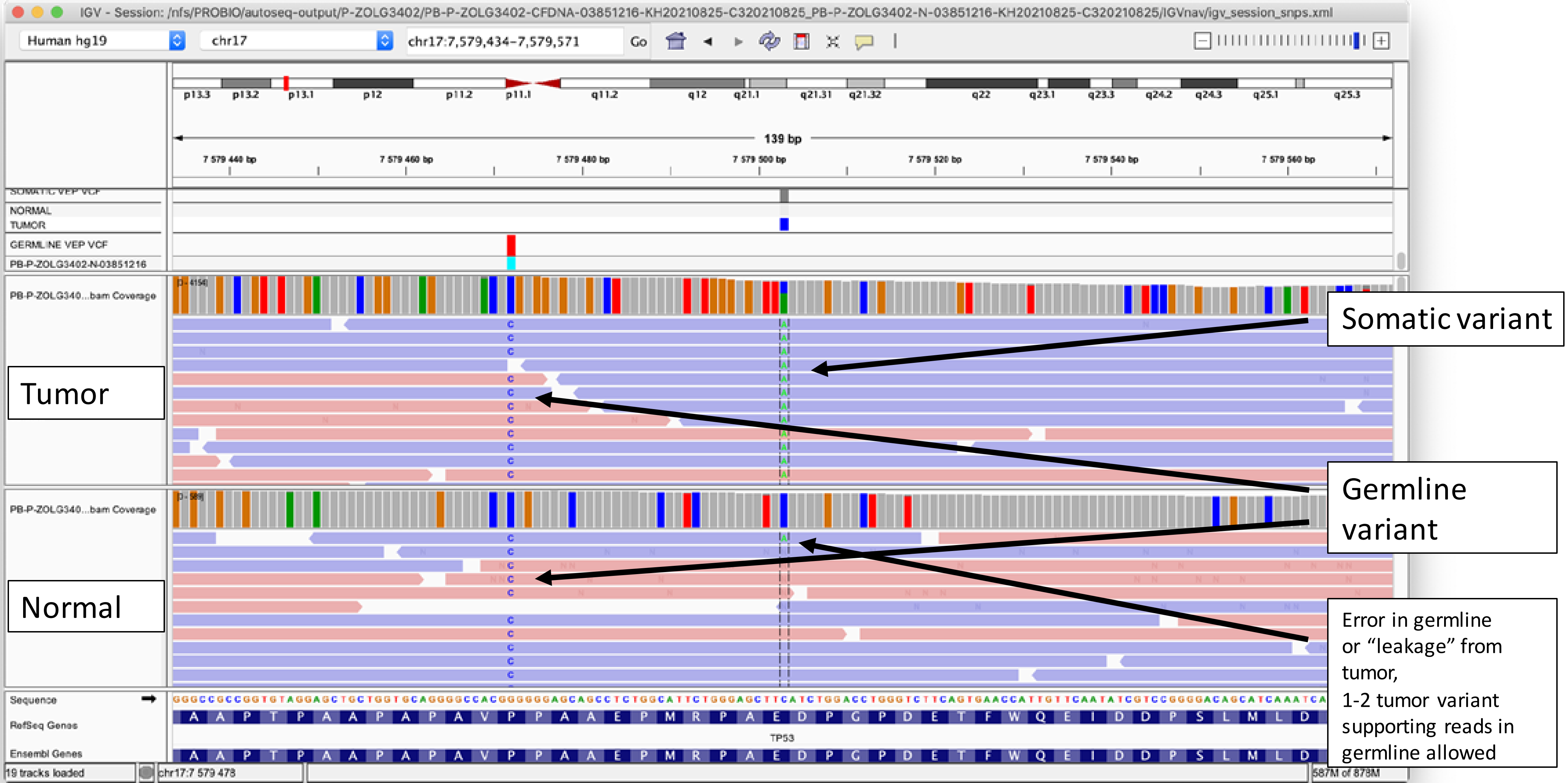
Tumor

Normal

Insertion
Click to see
inserted bases
(this case A)

Reads also
supporting
insertion
although not
mapped as
such

Germline vs somatic variant


IGV - Session: /nfs/PROBIO/autoseq-output/P-ZOLG3402/PB-P-ZOLG3402-CFDNA-03851216-KH20210825-C320210825_PB-P-ZOLG3402-N-03851216-KH20210825-C320210825/IGVnav/igv_session_snps.xml
Human hg19
chr17
chr17:7,579,434-7,579,571
Go
p13.3
p13.2
p13.1
p12
p11.2
p11.1
q11.2
q12
q21.1
q21.31
q21.32
q22
q23.1
q23.3
q24.2
q24.3
q25.1
q25.3
139 bp
7 579 440 bp
7 579 460 bp
7 579 480 bp
7 579 500 bp
7 579 520 bp
7 579 540 bp
7 579 560 bp
SOMATIC VEP VCF
NORMAL
TUMOR
GERMLINE VEP VCF
PB-P-ZOLG3402-N-03851216
PB-P-ZOLG340...bam Coverage
[0 - 4154]
PB-P-ZOLG340...bam Coverage
[0 - 589]
Sequence
GGGCCGCCGGTGTAGGAGCTGCTGGTGCAGGGGCCACGGGGGGAGCAGCCTCTGGCATTCTGGGAGCTTCATCTGGACCTGGGTCTTCAGTGAACCATTGTTCAATATCGTCCGGGGACAGCATCAAATCA
RefSeq Genes
A A P T P A A P A P A V P P A A E P M R P A E D P G P D E T F W Q E I D D P S L M L D
TP53
Ensembl Genes
A A P T P A A P A P A V P P A A E P M R P A E D P G P D E T F W Q E I D D P S L M L D
19 tracks loaded
chr17:7 579 473
587M of 878M

Somatic variant

Tumor

Germline variant

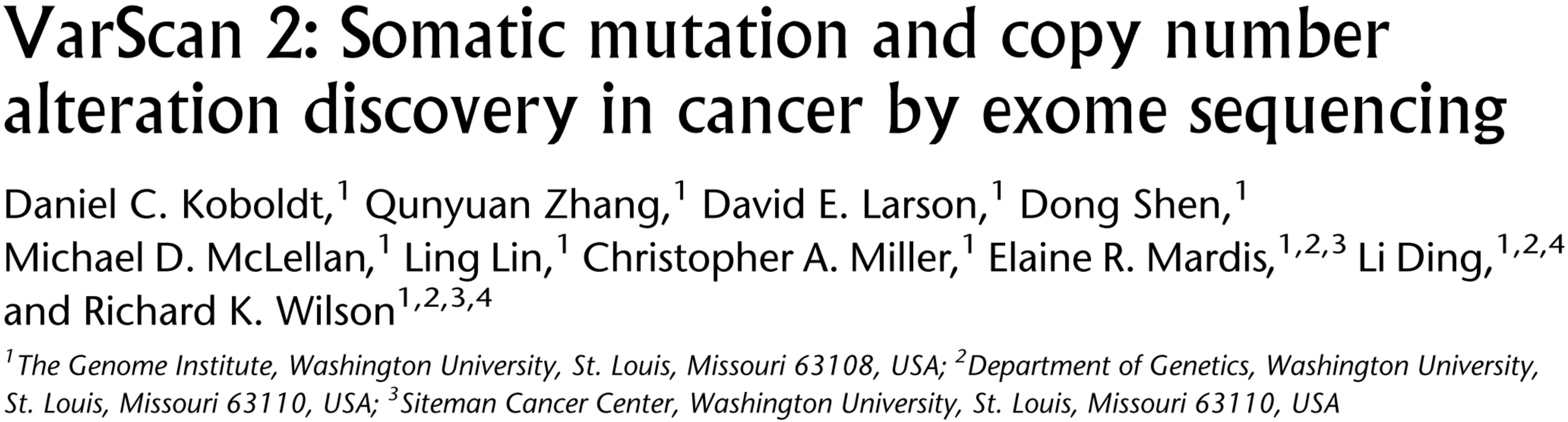
Normal

Error in germline or "leakage" from tumor,
1-2 tumor variant supporting reads in germline allowed

Somatic mutation calling – tools

- Tools (variant callers):
  - Vardict, Varscan – Counting the reads with ref bases and non-ref bases (alt allele) in tumor and normal, calculating significance test, if significant call the variant
  - Strelka, Mutect2 – Local de novo assembly of sites with non-ref bases
  - SAGE – Identifying sites with non-ref bases and nearby "read context", weighting reads based on the various quality parameters and summing the weights, if high enough weighted support in tumor (and <4% in germline) call the variant
  - Different properties – better at different types of variants

VarScan 2



VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing

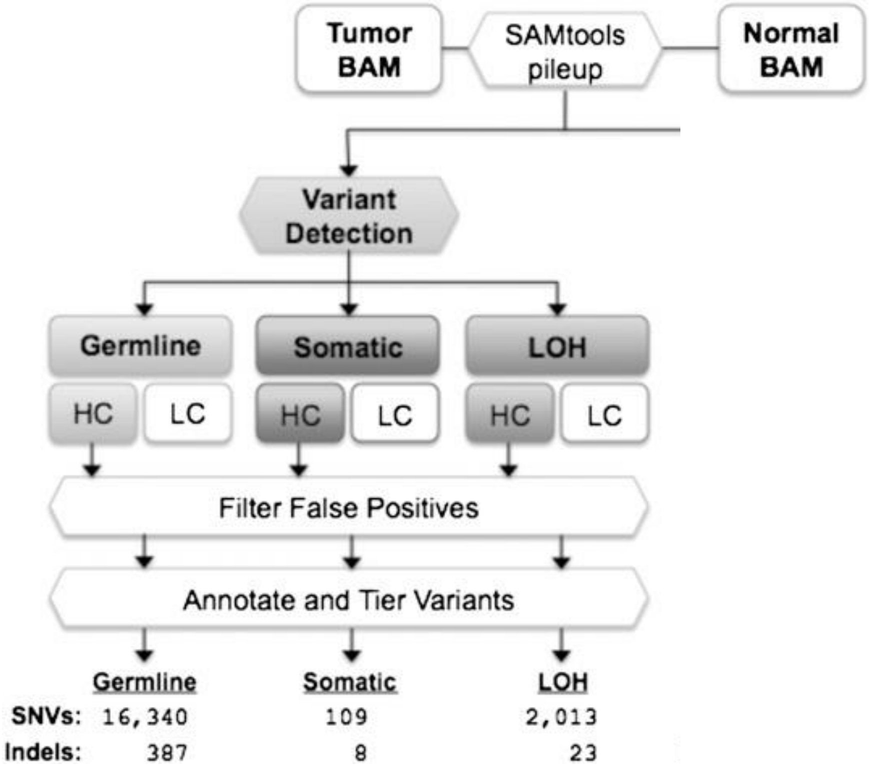
Daniel C. Koboldt,[1] Qunyuan Zhang,[1] David E. Larson,[1] Dong Shen,[1] Michael D. McLellan,[1] Ling Lin,[1] Christopher A. Miller,[1] Elaine R. Mardis,[1,2,3] Li Ding,[1,2,4] and Richard K. Wilson[1,2,3,4]

[1] The Genome Institute, Washington University, St. Louis, Missouri 63108, USA; [2] Department of Genetics, Washington University, St. Louis, Missouri 63110, USA; [3] Siteman Cancer Center, Washington University, St. Louis, Missouri 63110, USA

Published in Genome Research

VarScan 2

- Samtools pileup
  - Checks every position in given regions, which base every read covering it has (A, C, G, T, ins, del) and the total depth
- Variant detection:
  - Thresholds for coverage, base quality, variant allele frequency
  - Statistically significant variant: Fisher's exact test of ref and alt read counts, compared to expected sequencing error distribution
- Variant classification:
  - Comparison between variants in germline and in tumor
  - Fisher's exact test of ref and alt read counts in tumor and in normal
  - If variant only in tumor → somatic variant
  - If heterozygous (VAF < 75%) in germline but homozygous in tumor (VAF > 75%) → LOH variant (loss of heterozygosity, germline variant where the ref allele is somehow lost in the tumor)
  - If variant has same genotype (heterozygous/homozygous) in both tumor and germline → germline variant
  - Filters on read position, strandedness, variant reads, variant frequency, distance to 3', homopolymer, mapping quality difference, read length difference, MMQS (mismatch quality sum) difference

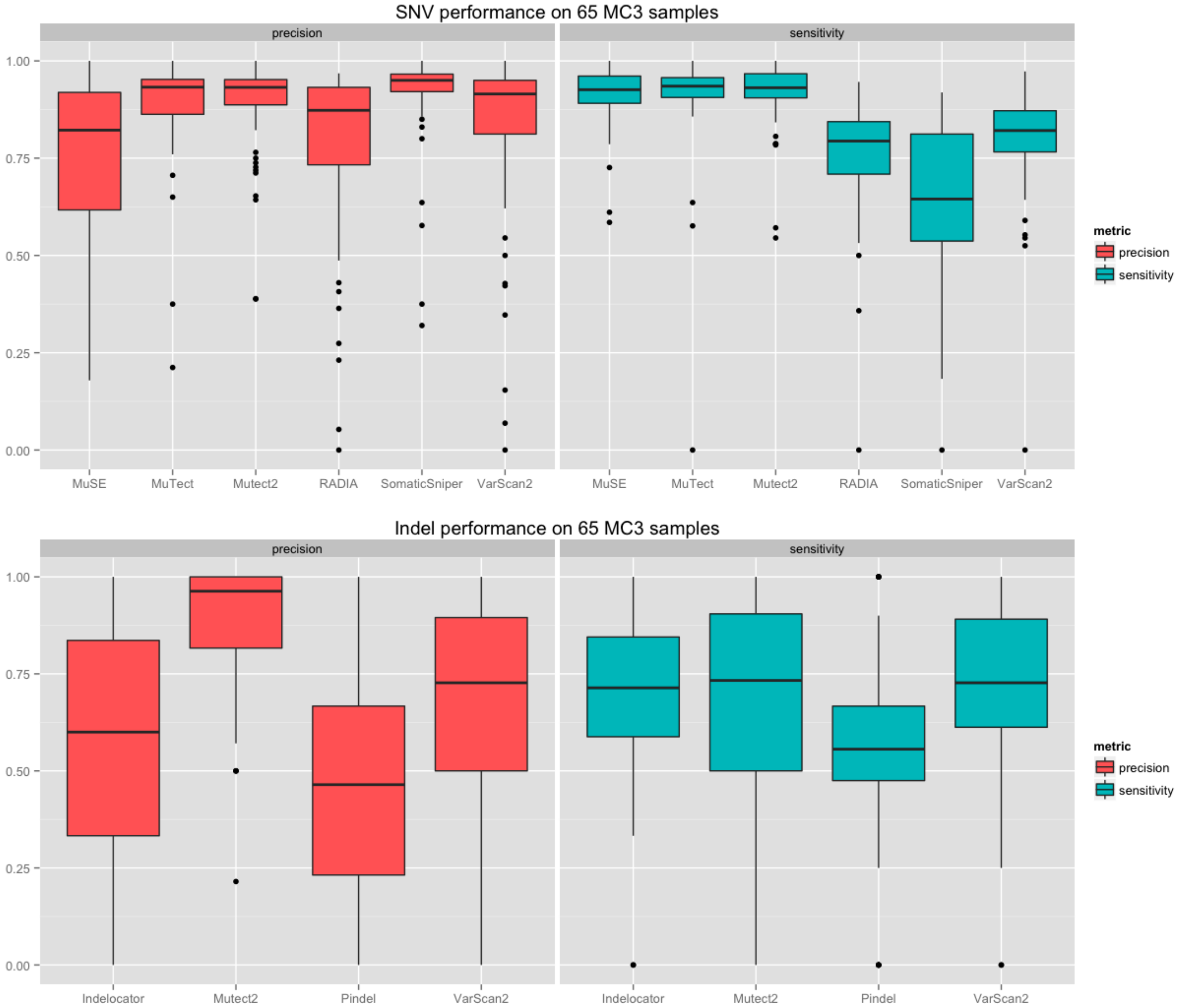


Fisher's test 2x2	Germline sample	Tumor sample
Reference allele	572	585
Alternate allele	1	213

Comparison of callers

Karolinska Institutet

- Callers have different performance on different variant types
- Comparison from Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., & Lichtenstein, L. (2019). Calling Somatic SNVs and Indels with Mutect2. *BioRxiv*. https://doi.org/10.1101/861054



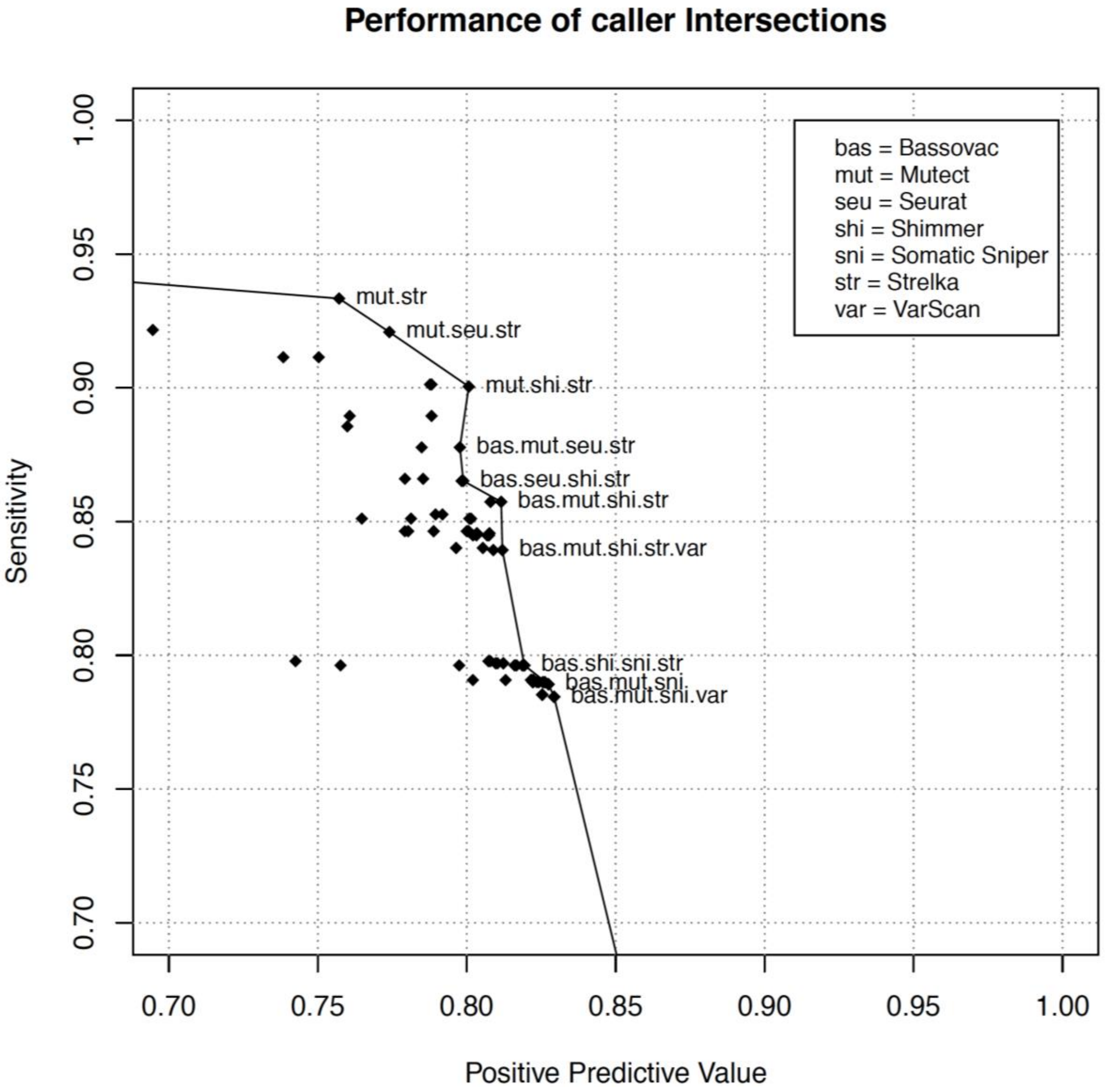
SNV performance on 65 MC3 samples

Indel performance on 65 MC3 samples

Somatic mutation calling – tools

- Consensus:
  - Not all called variants are true
  - Take consensus from the four callers to improve specificity
  - Literature review in 2018 showed that these (Mutect, Strelka, Varscan, Vardict) were the most commonly used and give a good combination of properties
    - Evaluation ongoing for update of which callers we will use in future
  - Tool somaticseq
  - If ≥ 2 callers call the same mutation, include it in table for manual curation

Somatic mutation calling – tools



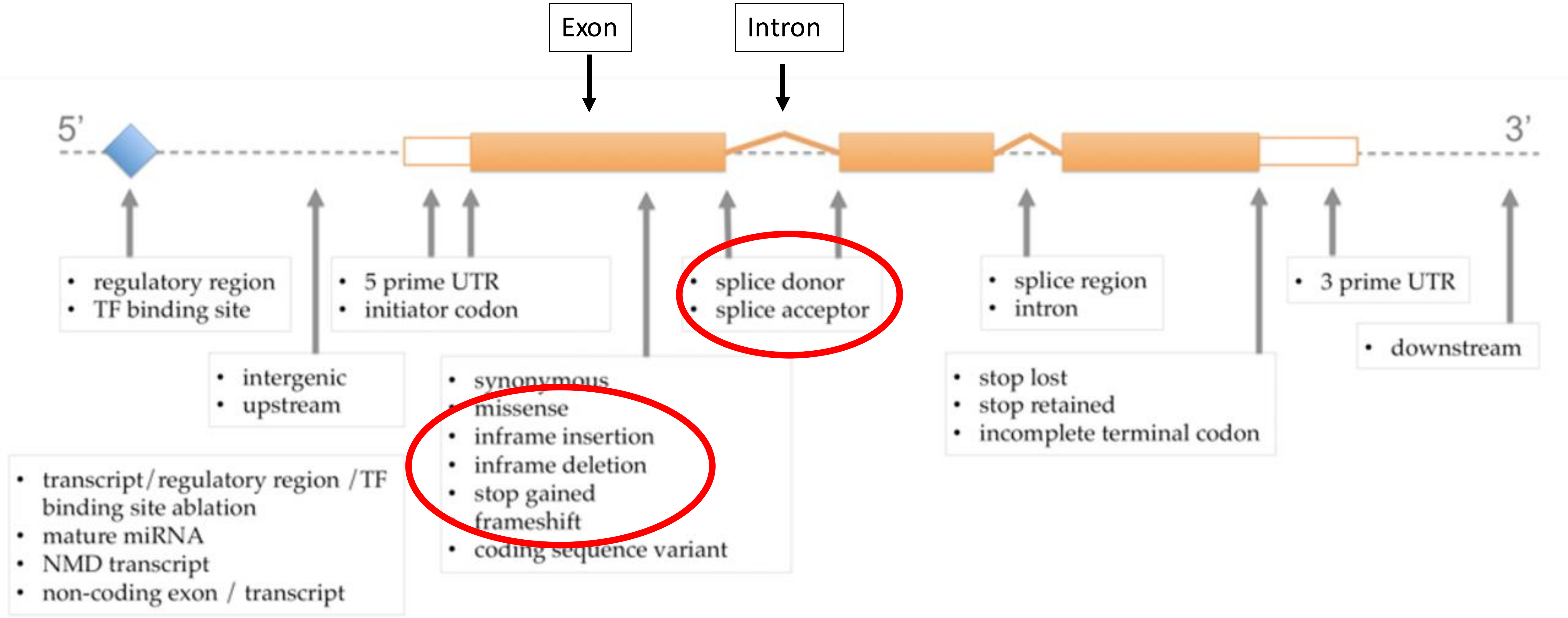
Credit Griffith brothers' lecture #5

Somatic mutation calling – annotation

- What effects do the mutations have?
  - On proteins
  - On disease
  - On treatment alternatives

- VEP – Variant Effect Predictor
  - From Ensembl
  - “VEP determines the effect of your variants … on genes, transcripts, and **protein sequence**, as well as regulatory regions.” (https://www.ensembl.org/info/docs/tools/vep/index.html)

Somatic mutation calling – VEP annotation categories



- regulatory region
- TF binding site

- intergenic
- upstream

- transcript/regulatory region /TF binding site ablation
- mature miRNA
- NMD transcript
- non-coding exon / transcript

- 5 prime UTR
- initiator codon

- synonymous
- missense
- inframe insertion
- inframe deletion
- stop gained
- frameshift
- coding sequence variant

- splice donor
- splice acceptor

- splice region
- intron

- stop lost
- stop retained
- incomplete terminal codon

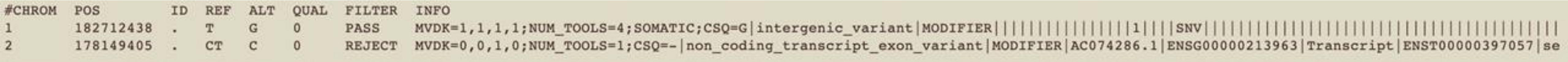
- 3 prime UTR

- downstream

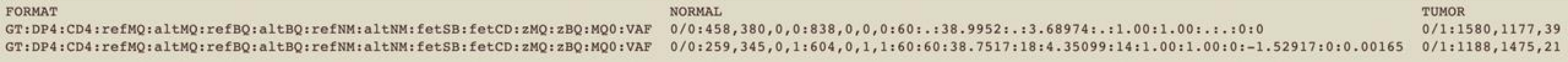
Credit Griffith brothers' lecture #4

Mutation calling – file format

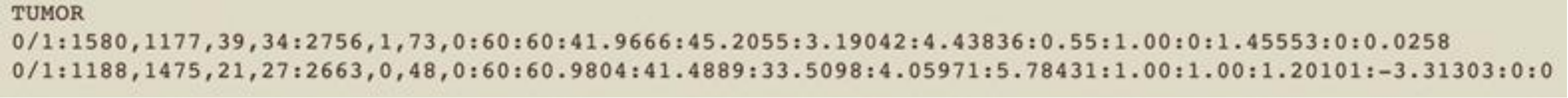
- VCF – variant call format
- Header – metadata describing different fields
- Main



```
#CHROM  POS        ID  REF  ALT  QUAL  FILTER  INFO
1       182712438  .   T    G    0     PASS    MVDK=1,1,1,1;NUM_TOOLS=4;SOMATIC;CSQ=G|intergenic_variant|MODIFIER|||||||||||||||1||||SNV|||||||||||||||||||||||||||||||||||||||||
2       178149405  .   CT   C    0     REJECT  MVDK=0,0,1,0;NUM_TOOLS=1;CSQ=-|non_coding_transcript_exon_variant|MODIFIER|AC074286.1|ENSG00000213963|Transcript|ENST00000397057|se
```



```
FORMAT                                                        NORMAL                                                                  TUMOR
GT:DP4:CD4:refMQ:altMQ:refBQ:altBQ:refNM:altNM:fetSB:fetCD:zMQ:zBQ:MQ0:VAF  0/0:458,380,0,0:838,0,0,0:60:.:38.9952:.:3.68974:.:1.00:1.00:.:.:0:0             0/1:1580,1177,39
GT:DP4:CD4:refMQ:altMQ:refBQ:altBQ:refNM:altNM:fetSB:fetCD:zMQ:zBQ:MQ0:VAF  0/0:259,345,0,1:604,0,1,1:60:60:38.7517:18:4.35099:14:1.00:1.00:0:-1.52917:0:0.00165  0/1:1188,1475,21
```



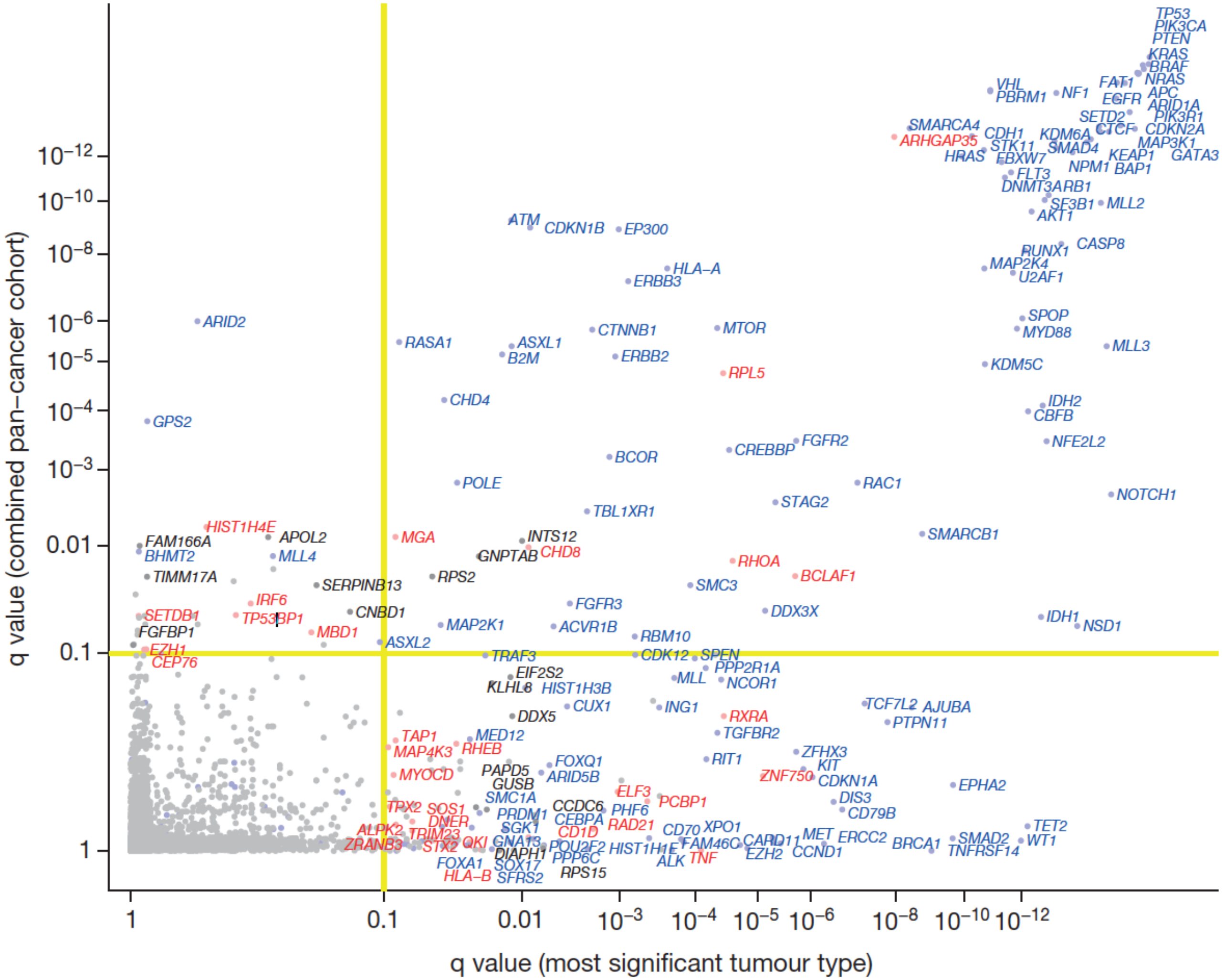
```
TUMOR
0/1:1580,1177,39,34:2756,1,73,0:60:60:41.9666:45.2055:3.19042:4.43836:0.55:1.00:0:1.45553:0:0.0258
0/1:1188,1475,21,27:2663,0,48,0:60:60.9804:41.4889:33.5098:4.05971:5.78431:1.00:1.00:1.20101:-3.31303:0:0
```

Mutation calling – manual curation

- Not all called variants are true – even after consensus
- If filters were sharpened to decrease number of false positives, we would instead miss some variants
- Some properties are hard to account for/discover programmatically
- Manual curation necessary
- Purpose: to identify true variants with impact on the protein structure

Different genes mutated in different cancers

- x-axis: how common in the tumor type where it’s most common?
- y-axis: how common in all tumor types together?
- Above yellow line → candidate cancer gene
- Many genes that are candidate cancer genes when looking at a specific tumor type are not significantly mutated when looking at all cancer types together
- Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, *505*(7484), 495–501. https://doi.org/10.1038/nature12912



Somatic mutation calling – quiz

1. Are somatic mutations heritable?
   a) Yes
   b) No
   c) Only sometimes
2. In which sample can you find the somatic mutations?
   a) Tumor sample
   b) Germline sample
   c) Both
3. What is a frameshift variant?
   a) A single position base change, that gives a different codon and amino acid
   b) Insertion of a new stop codon before the actual stop of the gene, causing premature stop of the protein
   c) Insertion or deletion that changes the open reading frame, and thus all downstream amino acids
4. What is a VCF?
   a) A file listing the regions of interest, used as input to variant caller
   b) A file containing variants, e.g. as output of variant caller
   c) A file with positions that don't have any variant, so called wild type positions

Somatic mutation calling – quiz answers

1. Are somatic mutations heritable?
   b) No
2. In which sample can you find the somatic mutations?
   a) Tumor sample
3. What is a frameshift variant?
   c) Insertion or deletion that changes the open reading frame, and thus all downstream amino acids
      - One codon is 3 bases, if indel size is not multiple of 3 the downstream codons won’t be read correctly, but under a different reading frame
4. What is a VCF?
   b) A file containing variants, e.g. as output of variant caller

Somatic variant calling – Questions?

Mutational signatures

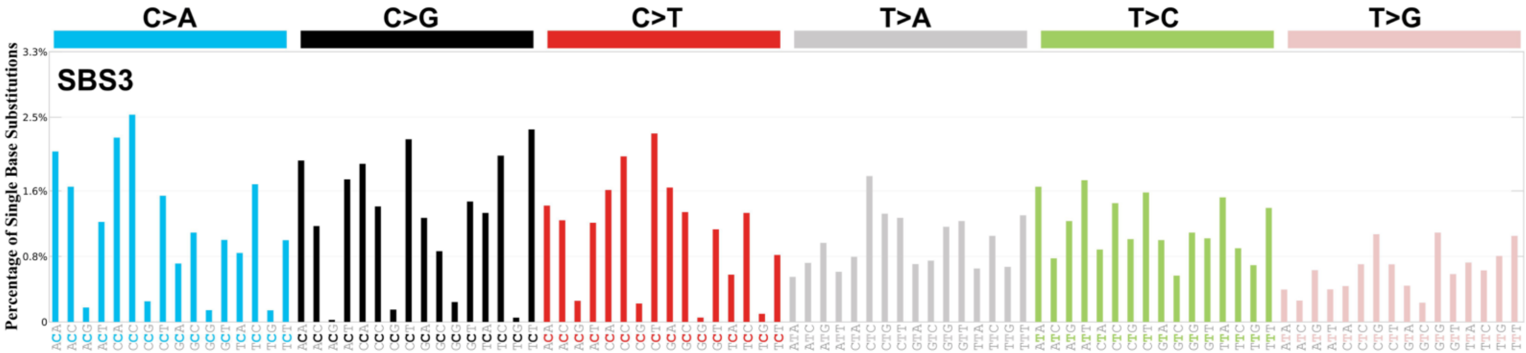
- Somatic mutations are caused by both exogenous and endogenous processes.
  - DNA repair, smoking, sunlight, ageing, chemotherapy etc
- Mathematical methods have been developed to investigate mutation data to determine the mutational signatures from a cancer genome and its origin.

Statistically significant mutational signature x Cancer type association x Known explaining mutagenic process?
Associated with treatment outcome?

The repertoire of mutational signatures in human cancer, Alexandrov et al., Nature 2020

Mutational signatures

- Single base substitutions (SBS), an example
    - Input data, possible mutations: C>A, C>G, C>T, T>A, T>C and T>G
- Account for 5’ and 3’ base, leads to 96 possibilities.
- Account for transcribed or untranscribed strand
    - 192 possibilities

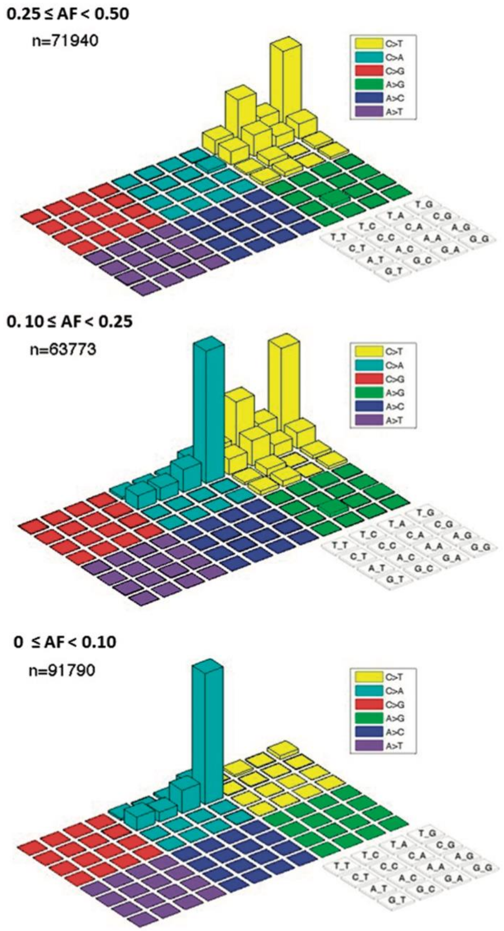


SBS3: signature associated with homologous recombination deficiency and sensitivity to parp inhibitors and carboplatin.

The repertoire of mutational signatures in human cancer, Alexandrov et al., Nature 2020

Mutational signatures and artefacts

- At the Broad Institute noise variants were detected in the TCGA whole exome data
- Another way of displaying the 96 options
- Similarly done for
  - Indels
  - Double base substitutions
  - Vary flanking bases
  - Etc …



Low frequency artefactual mutations increase

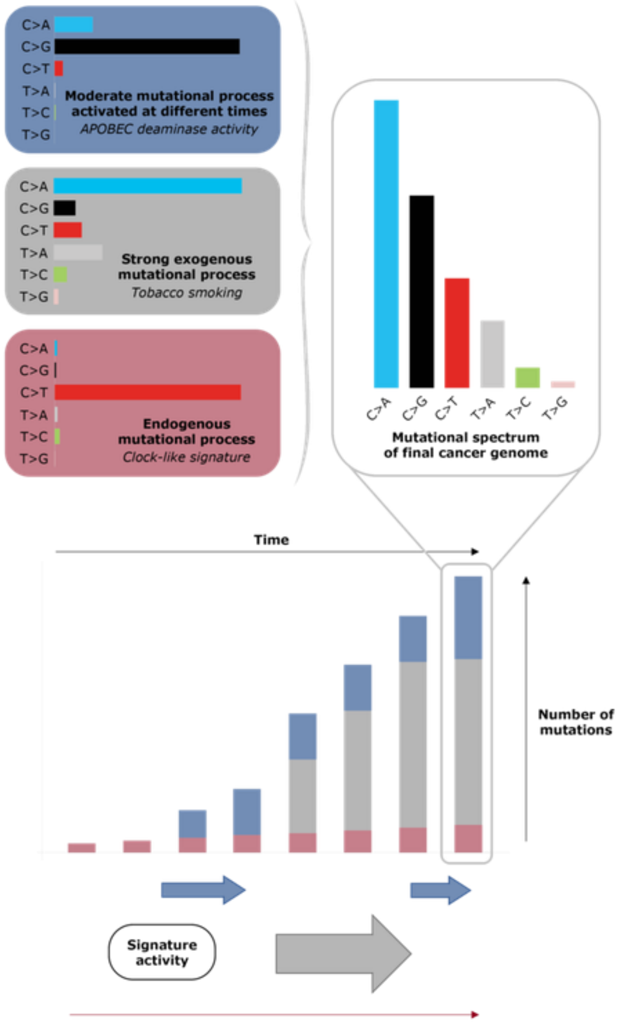
True signature

Artefact signature

Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation, Costello et al., NAR 2013

Mutational signatures

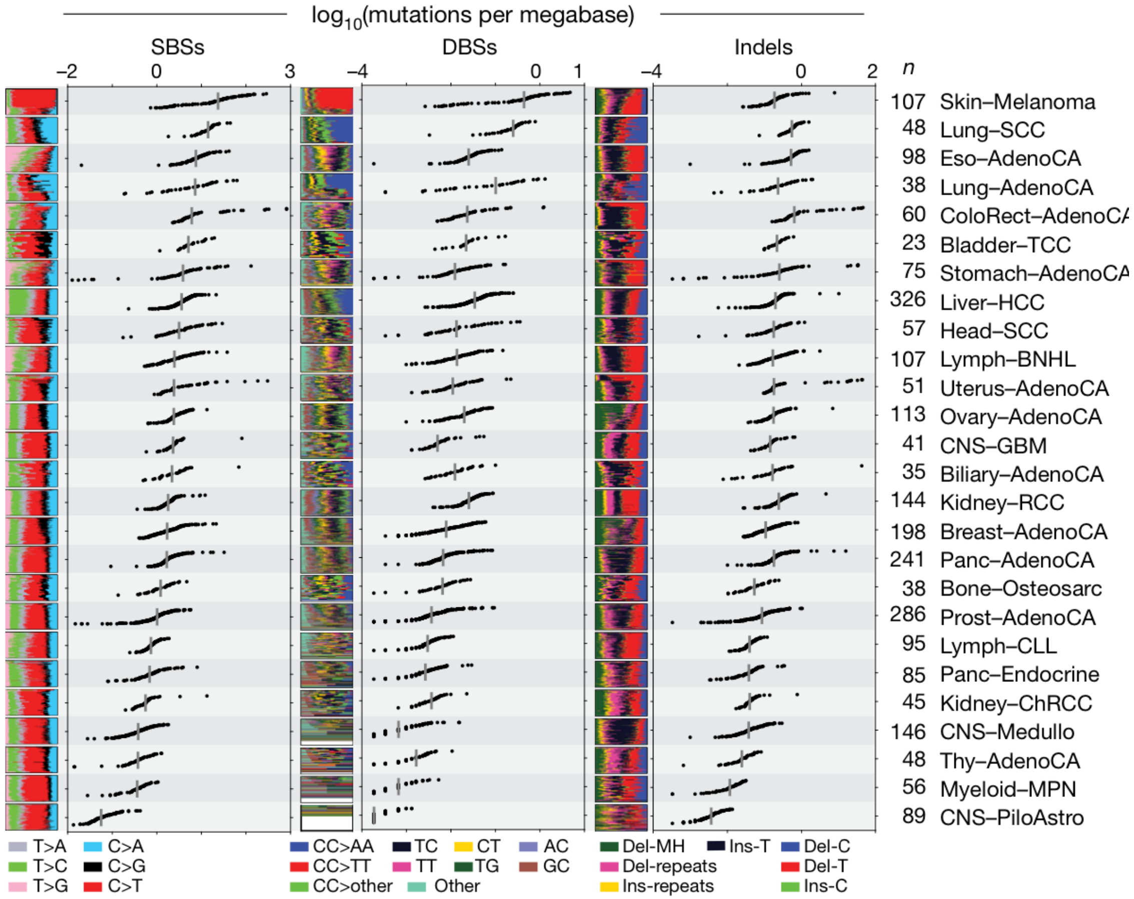
Identify the mutational causing processes for an individual cancer



C>A
C>G
C>T
T>A
T>C
T>G
Moderate mutational process activated at different times
APOBEC deaminase activity
Strong exogenous mutational process
Tobacco smoking
Endogenous mutational process
Clock-like signature
C>A
C>G
C>T
T>A
T>C
T>G
Mutational spectrum of final cancer genome
Time
Number of mutations
Signature activity

https://cancer.sanger.ac.uk/signatures/

Reflected in the mutation rates of different cancers



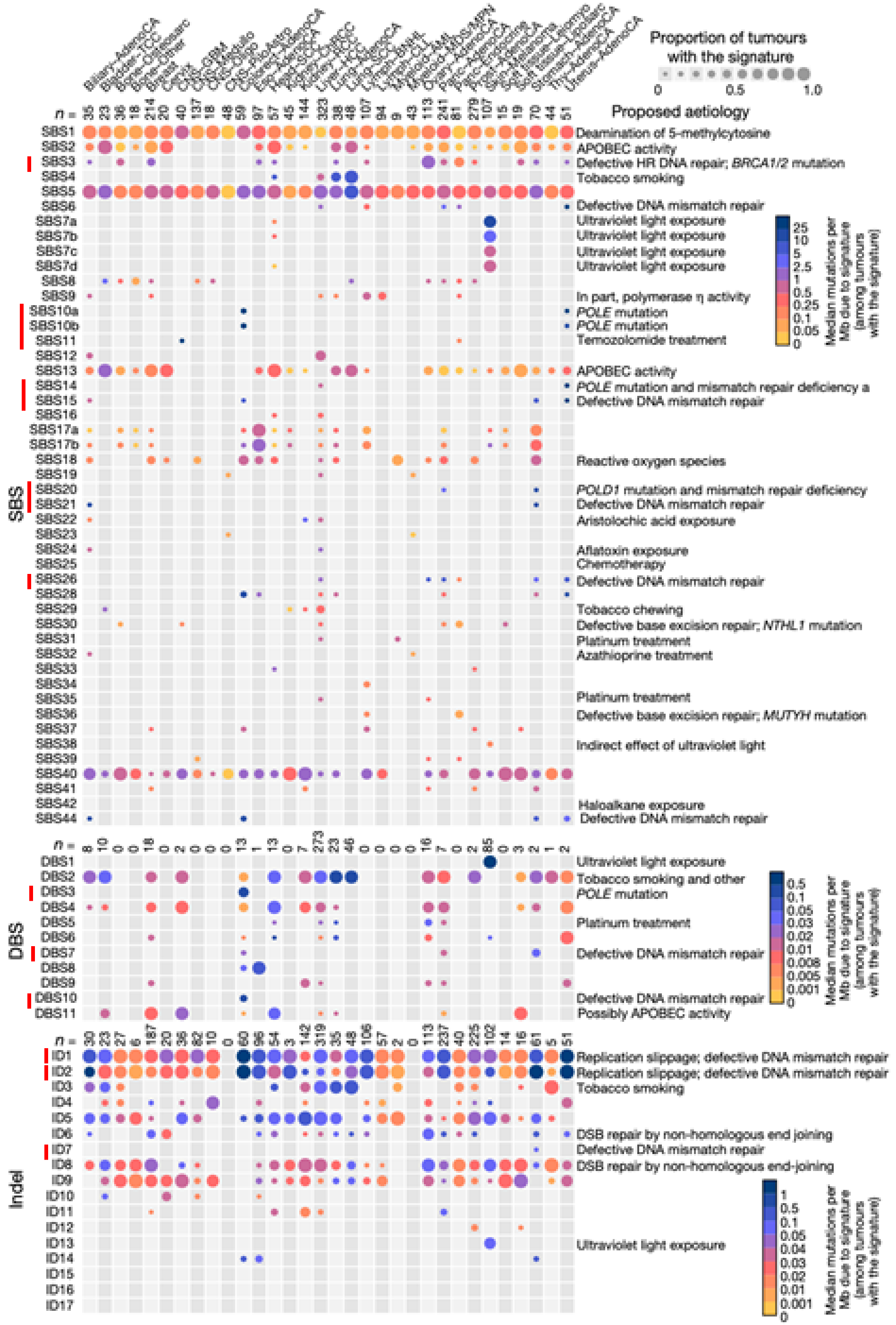
The repertoire of mutational signatures in human cancer, Alexandrov et al., Nature 2020

Mutational signatures

Treatment relevant mutational signatures, e.g. homologous recombination deficiency (SBS3).

An argument for WGS/WES.
Panel sequencing is often 5x – 50x smaller than WES
Hard to robustly assess mutational signatures with fewer positions sequenced.

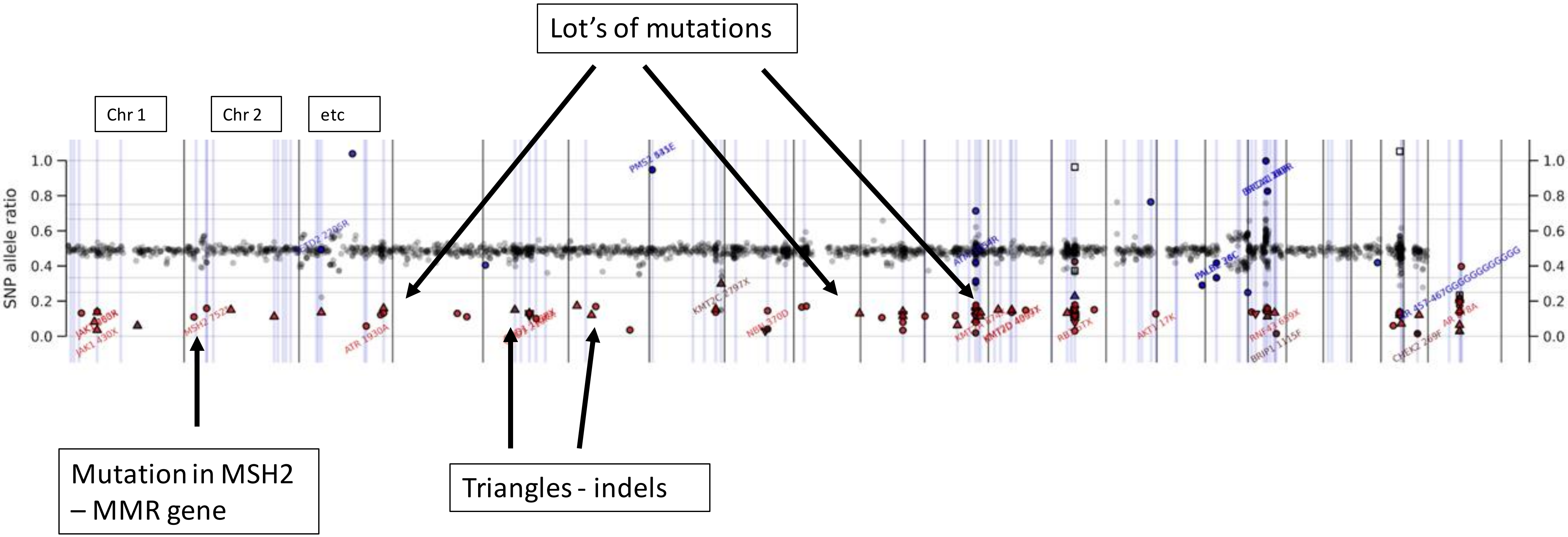
If time allows you will get to upload mutations to an online portal to see if you can determine the relevant signatures for some cancers.



MSI – microsatellite instability

- Microsatellite: small repetitive sequence of the genome, e.g. TTTTT or TATATATA
- MSI: increased rate of insertion or deletion of repeated segments in tumor
  - → plenty of insertions and deletions
- Caused by dysfunctional mismatch repair (MMR) mechanism
  - Indels of repeated segments not corrected
- Multiple different mutational signatures recognized as MSI
- Levels:
  - MSI-H: microsatellite instability high
  - (MSI-L: microsatellite instability low)
  - MSS: microsatellite stable
- Many mutations → tumor cells may express lots of weird proteins on their surfaces
  - Immunotherapy may be effective

MSI – an example



Lot’s of mutations

Chr 1

Chr 2

etc

Mutation in MSH2 – MMR gene

Triangles - indels

MSI - tool

- mSINGS - MicroSatellite Instability detection by Next Generation Sequencing
- Given list of microsatellites, 63 in our designs
- Background control created by ~20 healthy donor samples
- Comparing number of repeats in each microsatellite locus in tumor sample to the same sites in background control
- Locus called as unstable if significantly different from control samples
- Fraction of unstable loci gives mSINGS score
  - Threshold for MSI-H: > ~0.2
  - Confirm by visual inspection of mutation plot

Hypermutation

- MSI is a type of hypermutation – phenotypes with highly increased levels of mutation frequency
- Another type is caused by defective DNA replication repair (mismatch repair), due to mutations in DNA polymerases
- This gives different mutational signature, with more SNVs than indels
  - SBS14
- Other types of hypermutation can be caused by environmental factors (e.g. UV light, smoking, chemotherapy), and are associated with other specific mutational signatures

CHIP - Clonal Hematopoiesis of Indeterminate Potential

- Sub-population of blood cells carrying the same mutation(s)
- Age-related
- Increased risk of blood cancer and cardiovascular disease
- Shows in germline DNA, often at VAFs ≠ 50% or 100%
- In germline from WBC
  - May show in cfDNA but not tissue tumor samples

Mutational signatures – quiz

1. What is a mutational signature?
   a) All the mutations one patient have
   b) A pattern of specific mutation types, caused by a specific process
   c) The mutation types which are most typical for a certain tumor type
2. What factors can cause mutational signatures?
   a) Environmental factors (smoking, UV light etc.)
   b) Previous cancer treatments
   c) Ageing
   d) Artefacts from sequencing
3. What is MSI (microsatellite instability)?
   a) Increased level of single nucleotide variants (SNVs) in microtubule encoding genes
   b) A well-defined mutational signature common in breast cancer
   c) Increased level of insertions and deletions in repetitive sequences

Mutational signatures – quiz answers

1. What is a mutational signature?
   b) A pattern of specific mutation types, caused by a specific process
2. What factors can cause mutational signatures?
   a) Environmental factors (smoking, UV light etc.)
   b) Previous cancer treatments
   c) Ageing
   d) Artefacts from sequencing
3. What is MSI (microsatellite instability)?
   c) Increased level of insertions and deletions in repetitive sequences

Credits

- Malachi Griffith, Obi Griffith, Zachary Skidmore, Huiming Xia
  - Lecture notes from the course “Introduction to bioinformatics for DNA and RNA sequence analysis (IBDR01)”, 29 October – 2 November, 2018
  - McDonell Genome Institute, Washington University of St Louis School of Medicine

More questions?