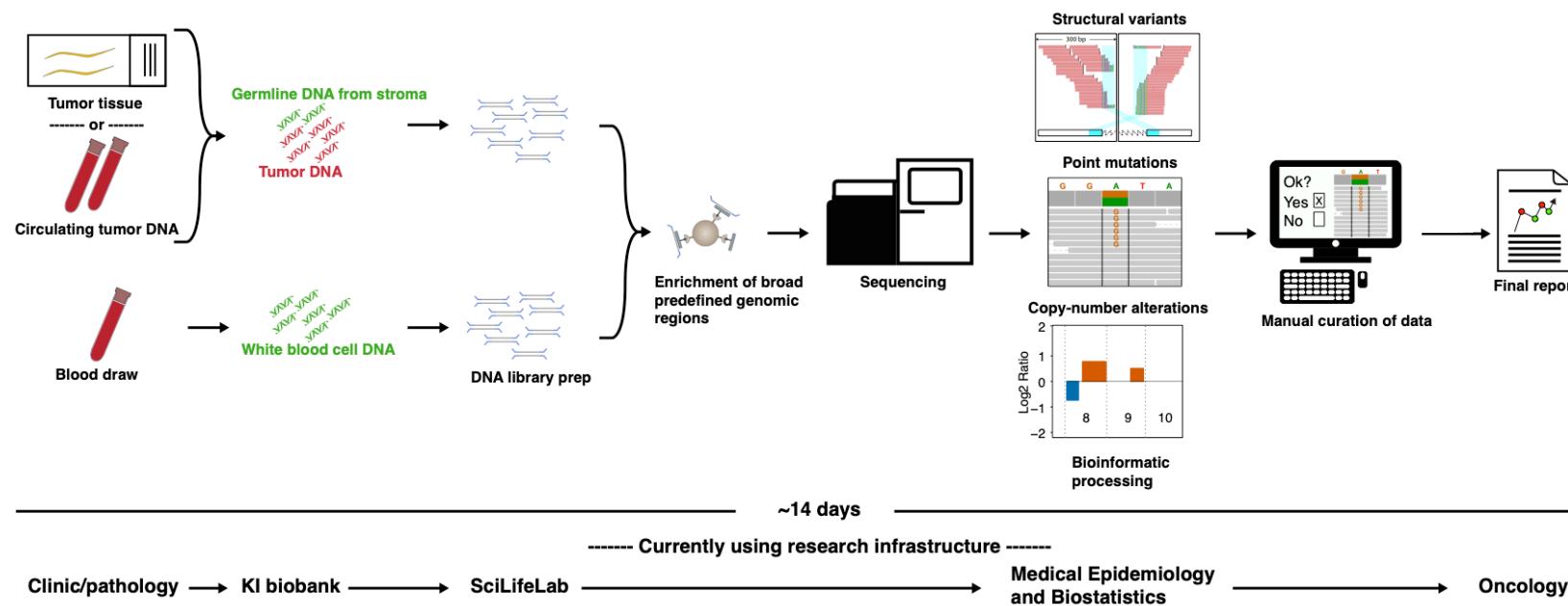


Practical considerations for performing cancer genomics



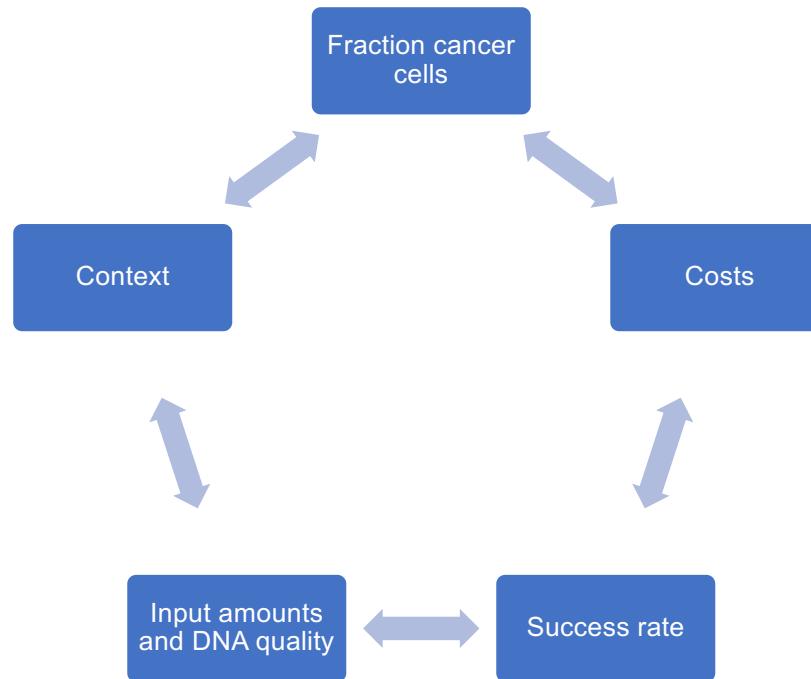
Karolinska
Institutet

Practical considerations for performing cancer genomics



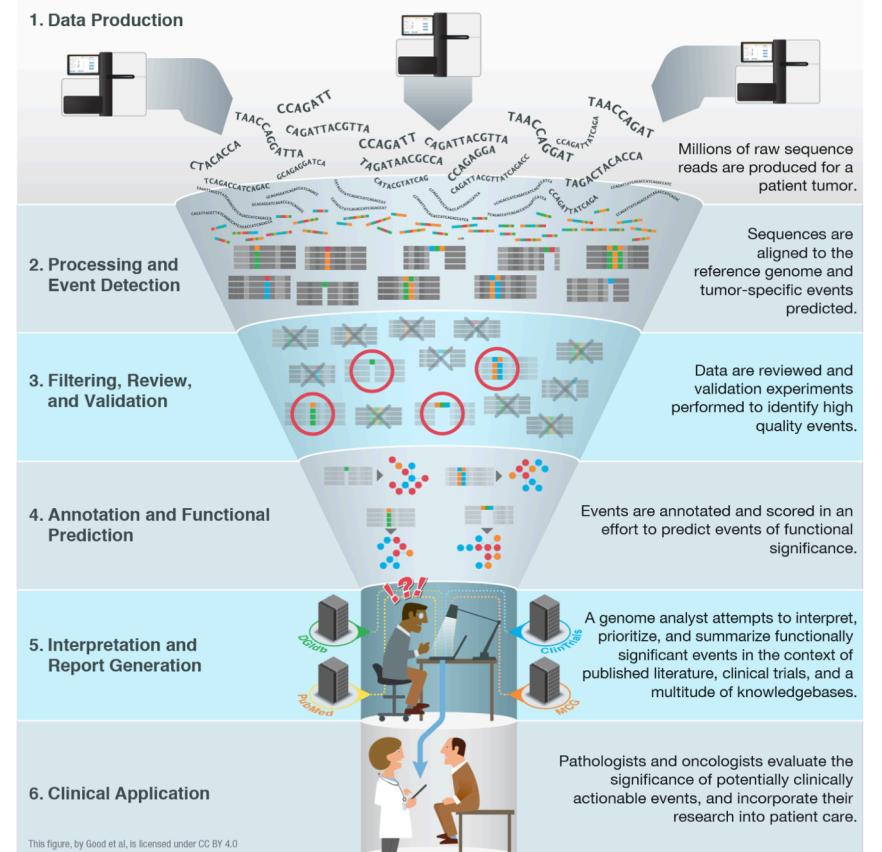
Examples given for DNA-seq but the same applies for RNA-seq

How to sequence depends on many things ..



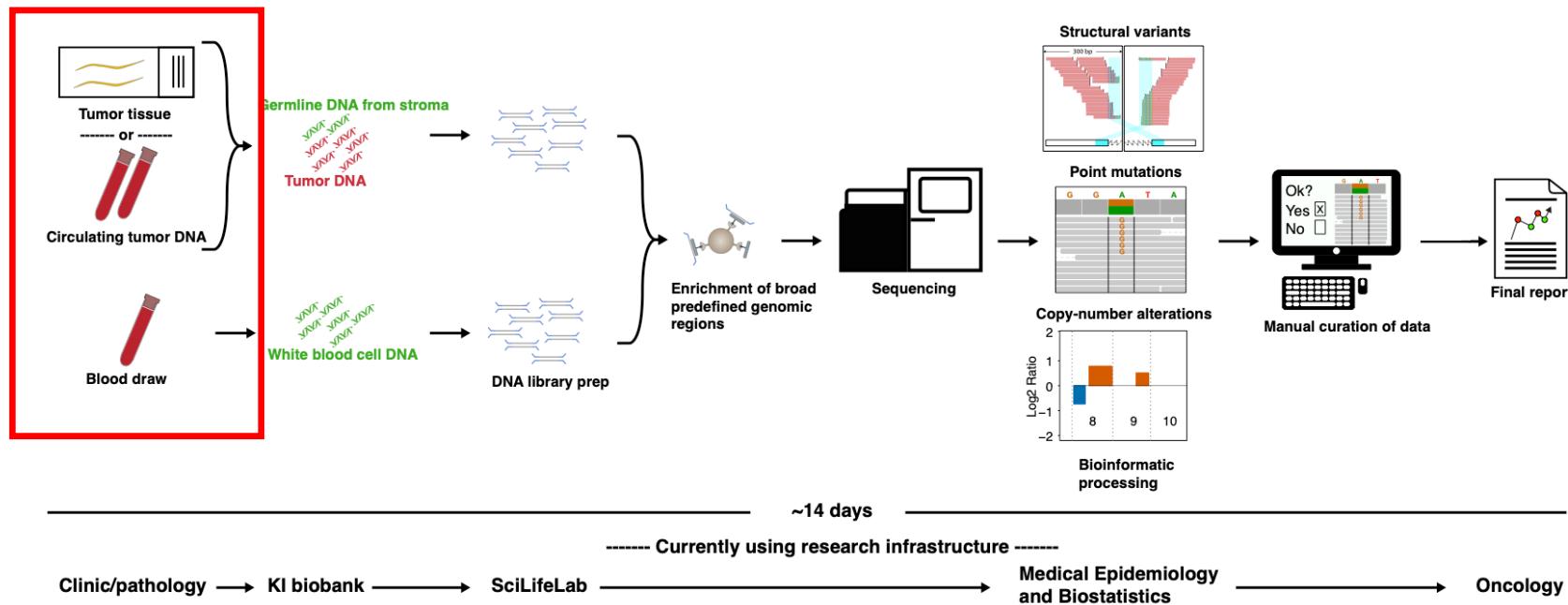
Context

- **Research**
 - Find new genomic features/associations in “large” cohorts
 - Accept a relatively high false positive rate
 - Turnover is flexible
- **Clinical**
 - Deliver on time – as high success rate as possible.
 - Predefined variants/events of interest
 - Not acceptable with false positives
 - Need to be honest about a potentially false negative situation
 - If cost-efficient, “research” data can be generated simultaneously

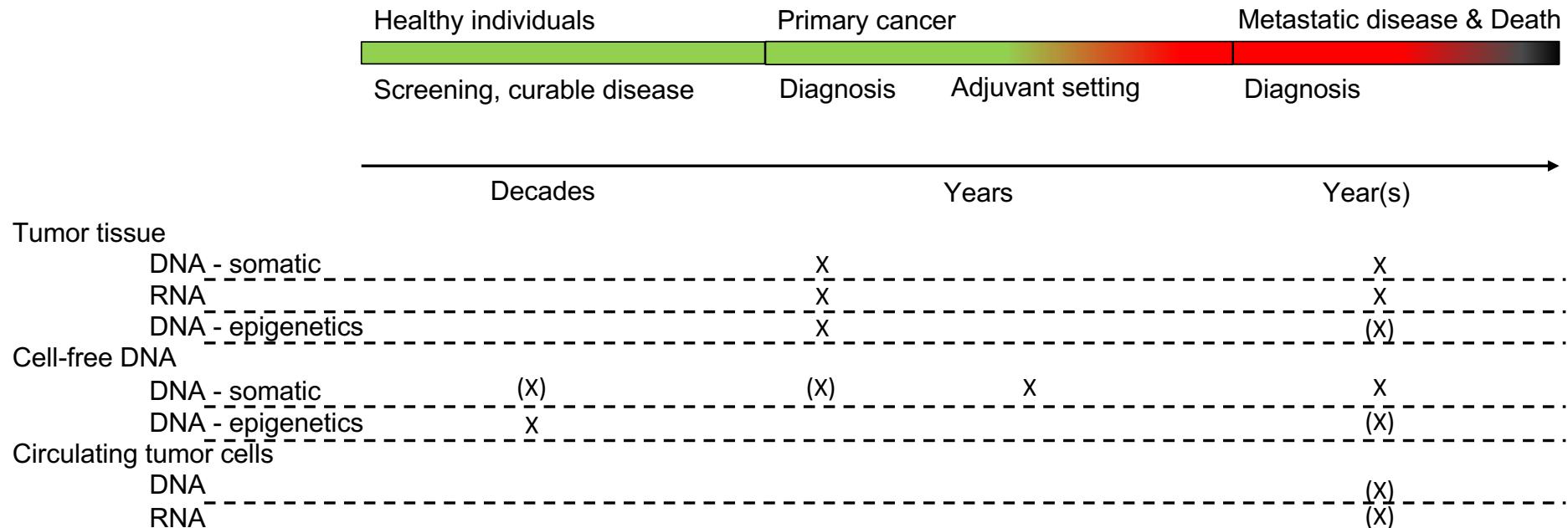


Good BM et al. Genome Biology 2014.

Practical considerations for performing cancer genomics

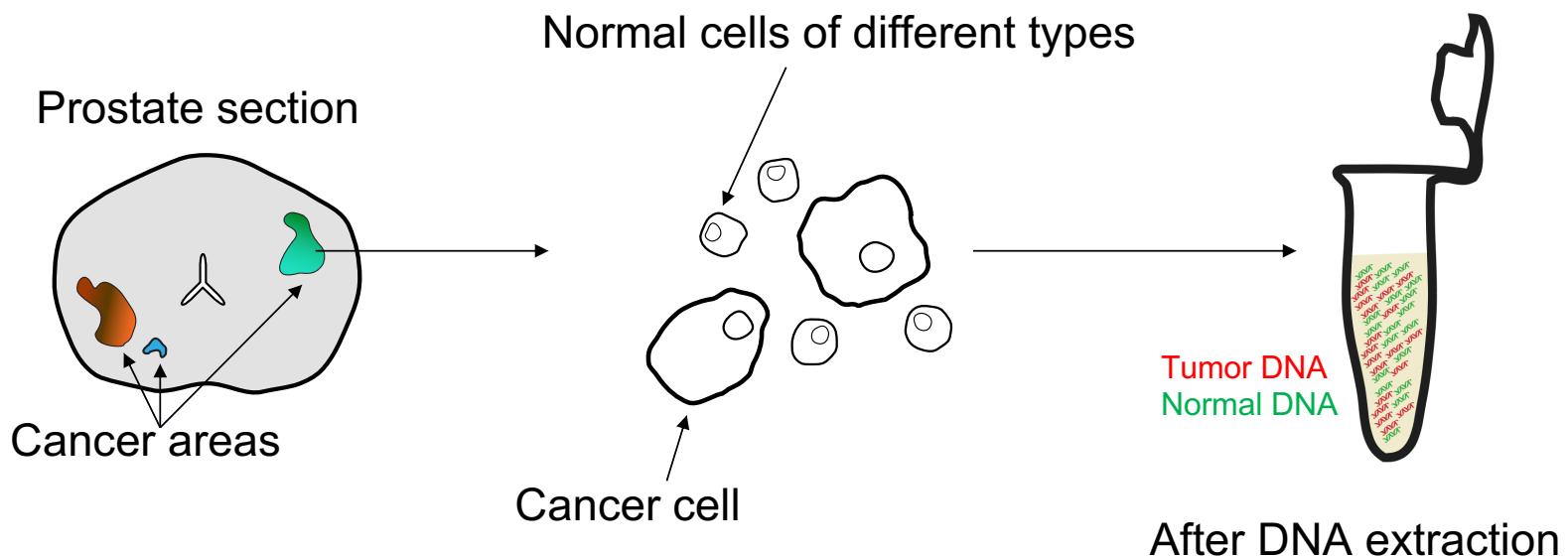


Tissue, analyte and context ...



Starting material – heterogeneity and purity

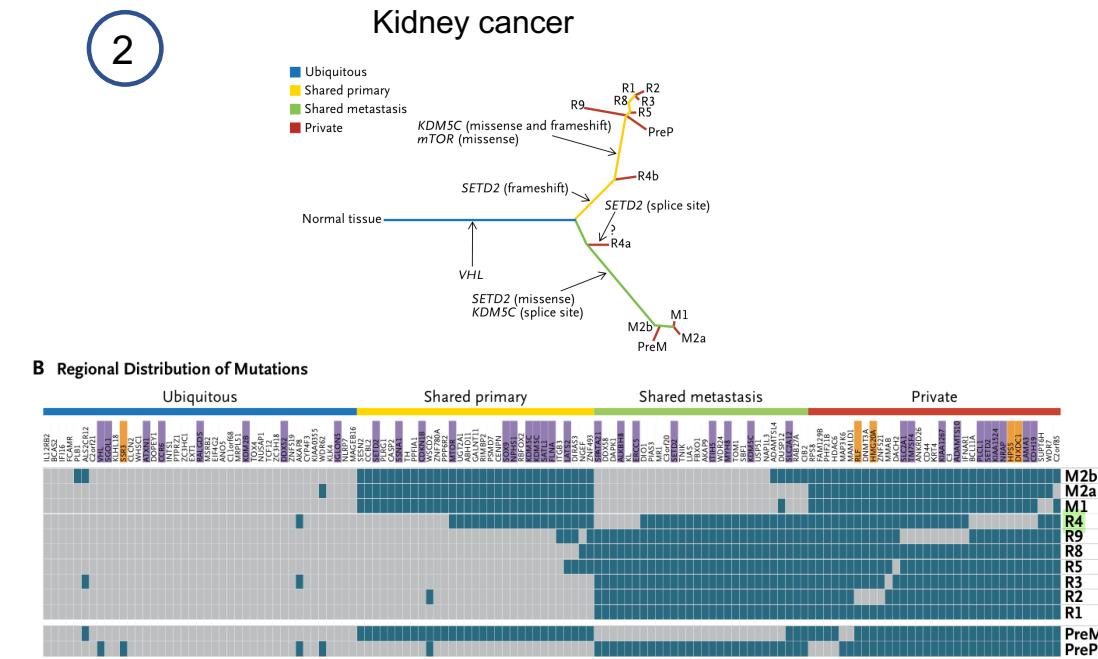
- Never 100% pure cancer cells
- Vary from 0.0001 – 0.7



Tumor heterogeneity – primary cancer

- 1
 - Pronounced heterogeneity
 - Vary with stage and grade
 - Want to treat the stem!

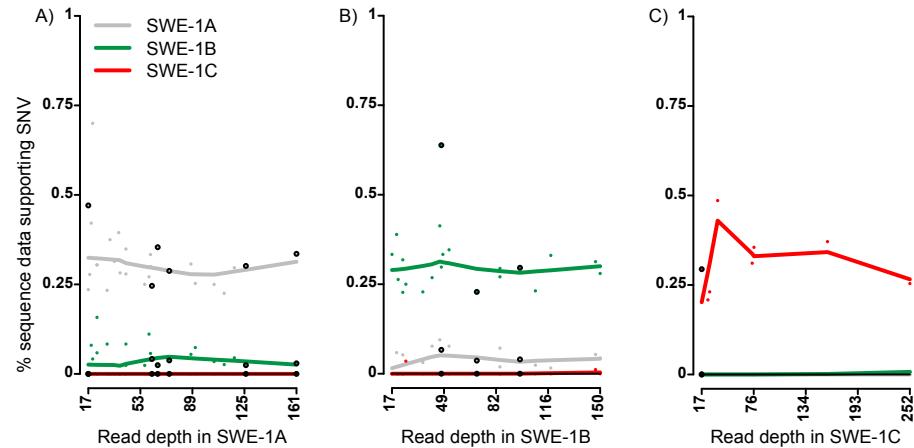
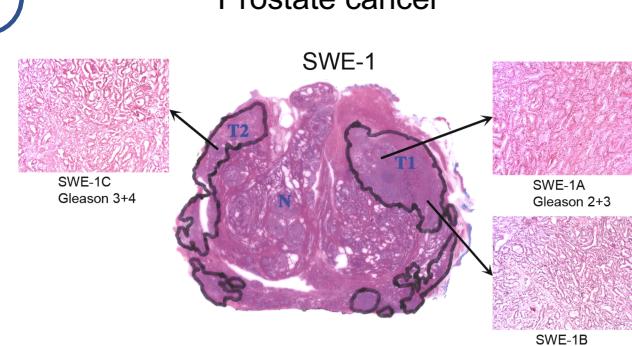
2



Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing, NEJM 2012

3

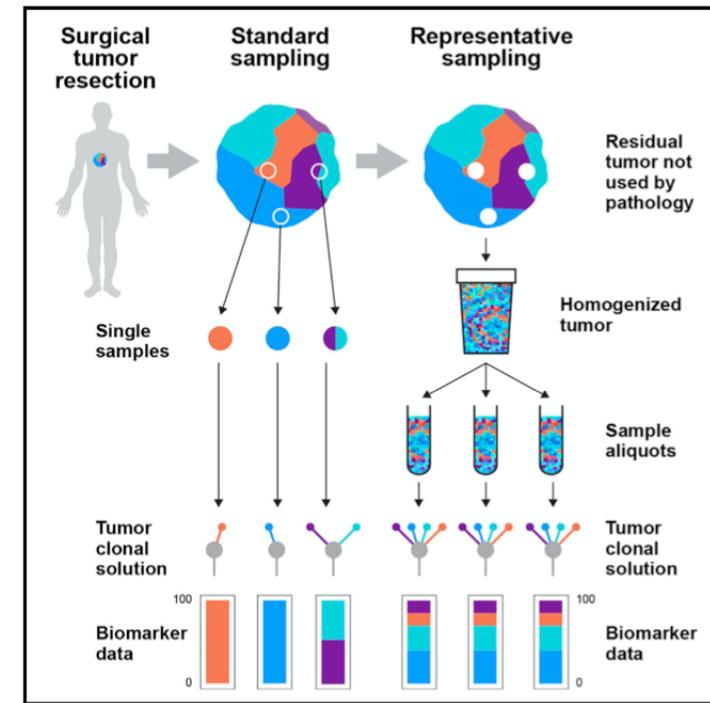
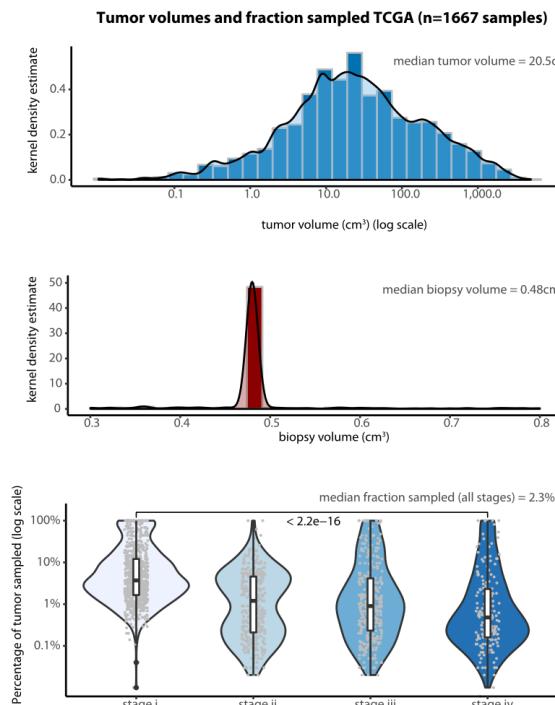
Prostate cancer



Exome Sequencing of Prostate Cancer Supports the Hypothesis of Independent Tumour Origins, Euro Uro 2012

Tumor heterogeneity – primary cancer

- Potential remedy – sequence a homogenized multiregional sample
- If not possible – only regarding "clonal" variants improves the situation
 - Difficult if cancer DNA fraction is low

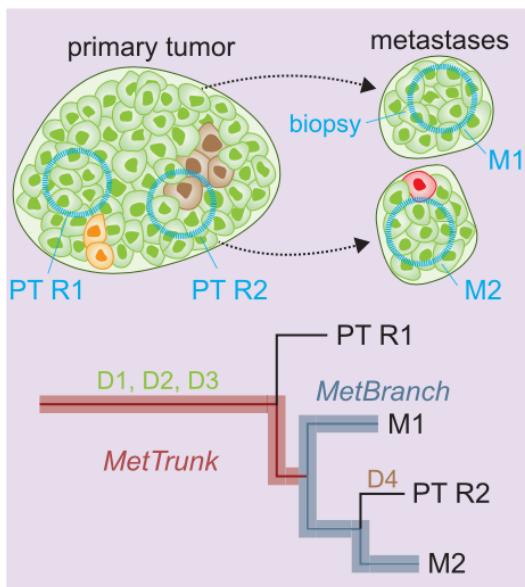


Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue, Cell Reports
2020

Tumor heterogeneity – metastatic cancer

- De novo metastatic disease

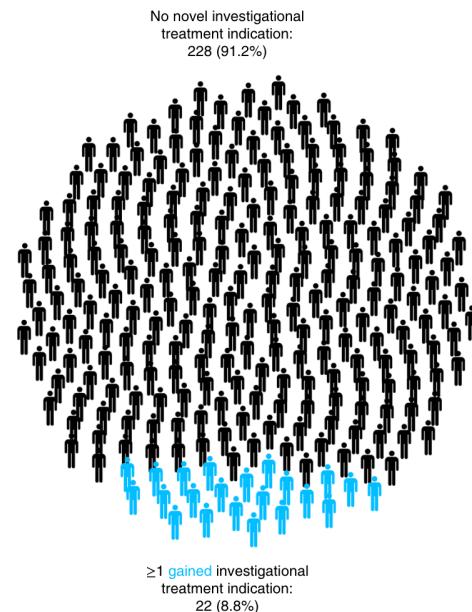
Limited heterogeneity



Minimal functional driver gene heterogeneity among untreated metastases, Science 2020

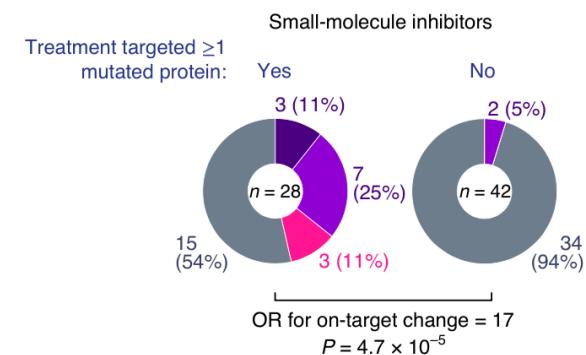
- Treated metastatic disease

The drive genes basically remain the same after treatment.



Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. Nat Med, 2021

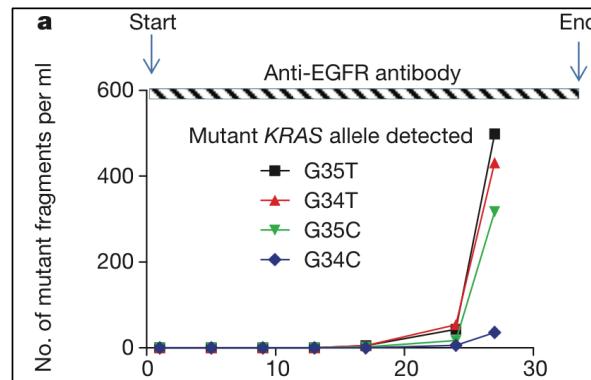
Genomic evolution occur in the targeted genes



Likely underestimate due to single-biopsy sequencing

Tumor heterogeneity – metastatic cancer

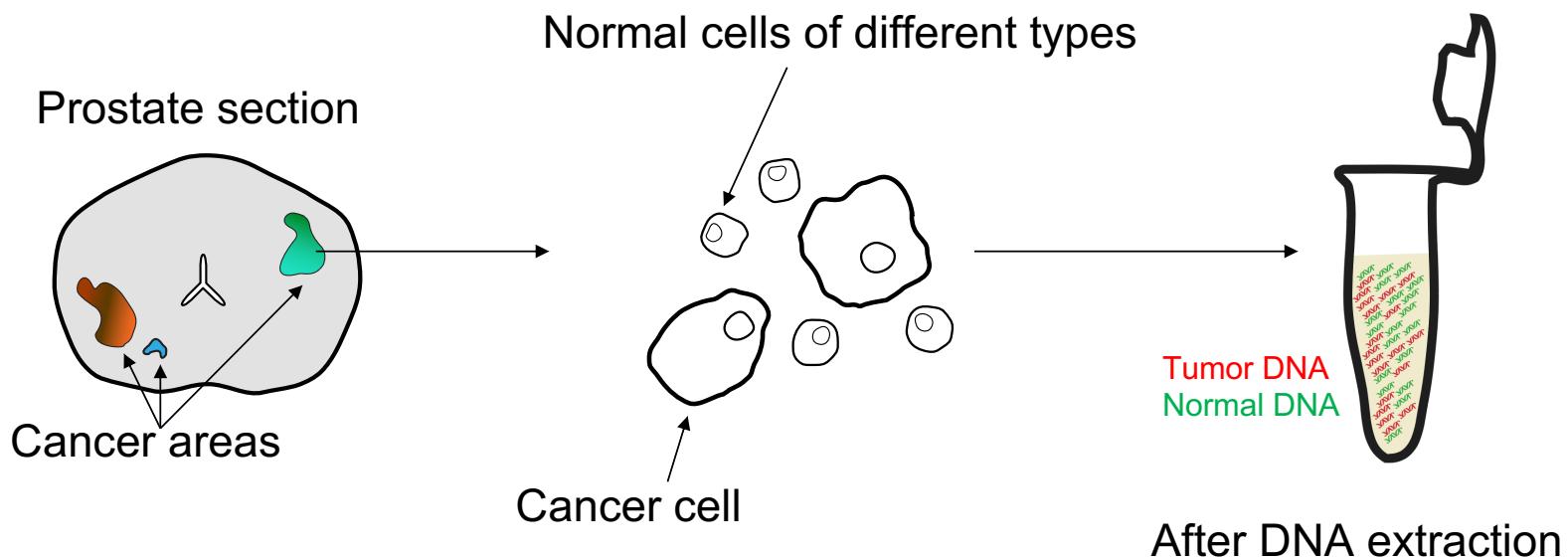
- Summary
 - Localised cancer
 - Pronounced heterogeneity
 - Remedy: sequenced homogenized multiregional samples
 - Newly diagnosed advanced cancer
 - Limited heterogeneity
 - One metastatic sample likely provides an adequate representation
 - Treated advanced cancer
 - As for newly diagnosed BUT clonal evolution will take place in the targeted genes/genes associated with resistance.



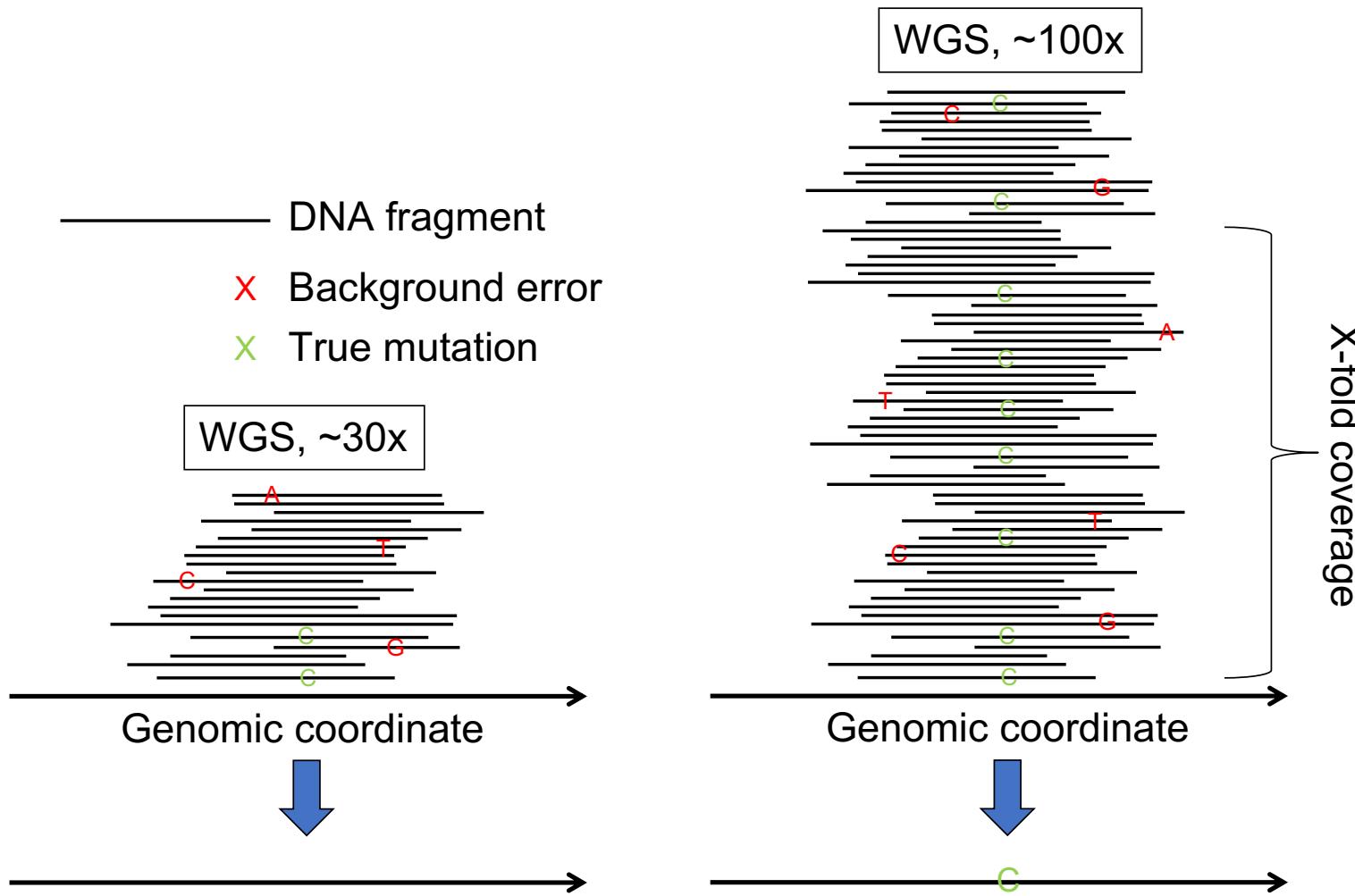
The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers, Nature, 2012

Starting material – heterogeneity and purity

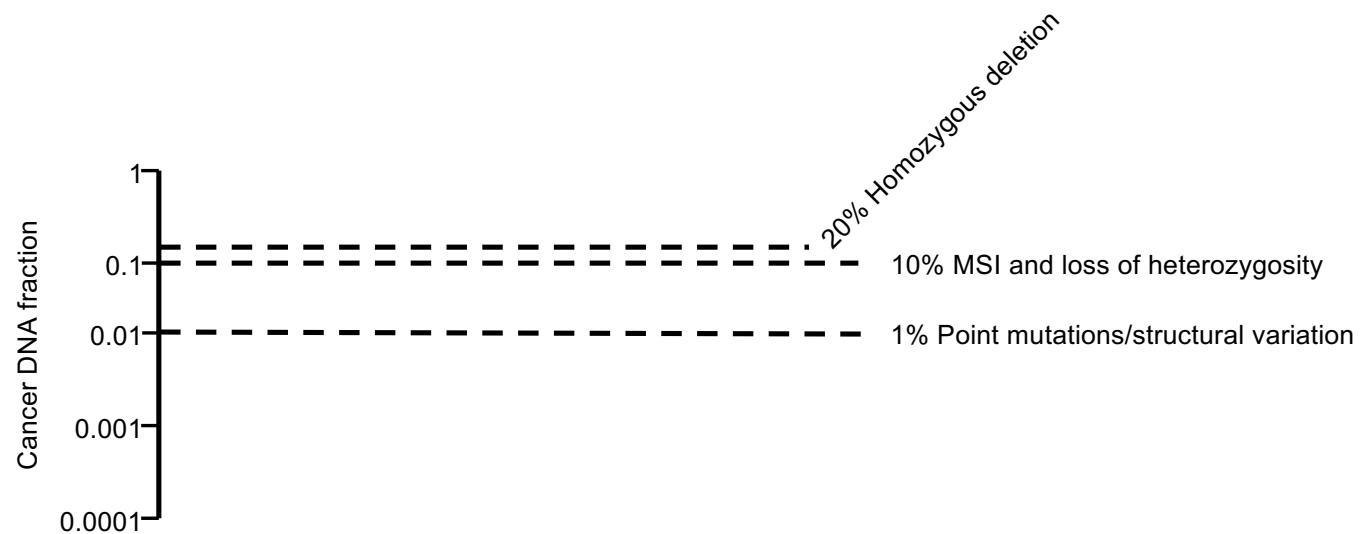
- Never 100% pure cancer cells
- Vary from 0.0001 – 0.7



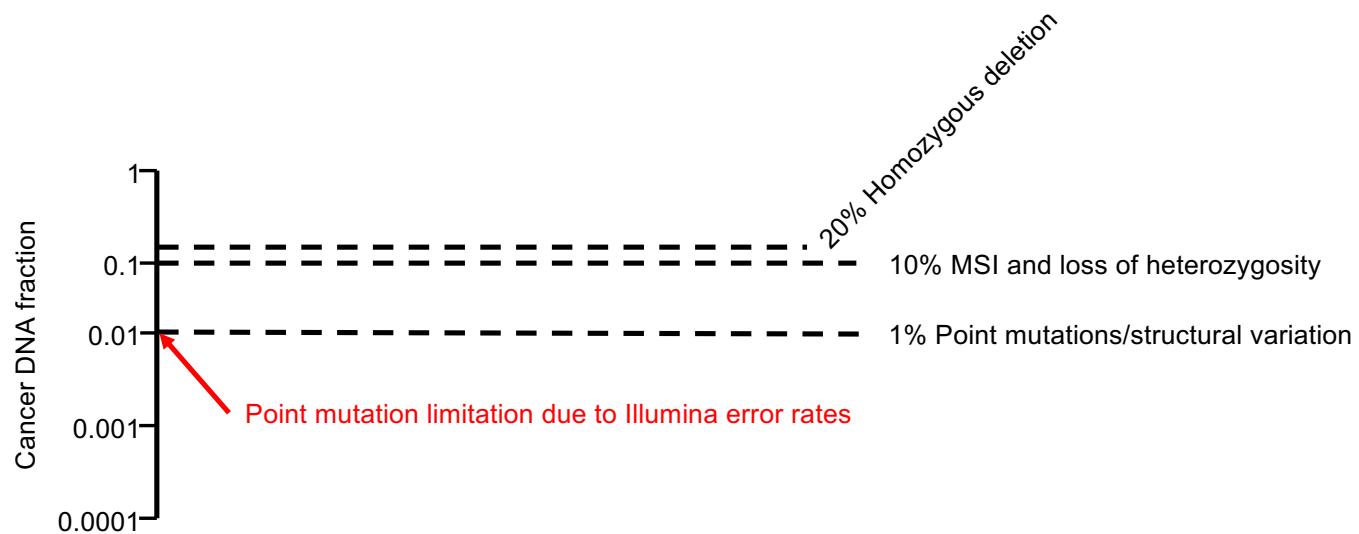
Coverage is king ... but not remedy for everything



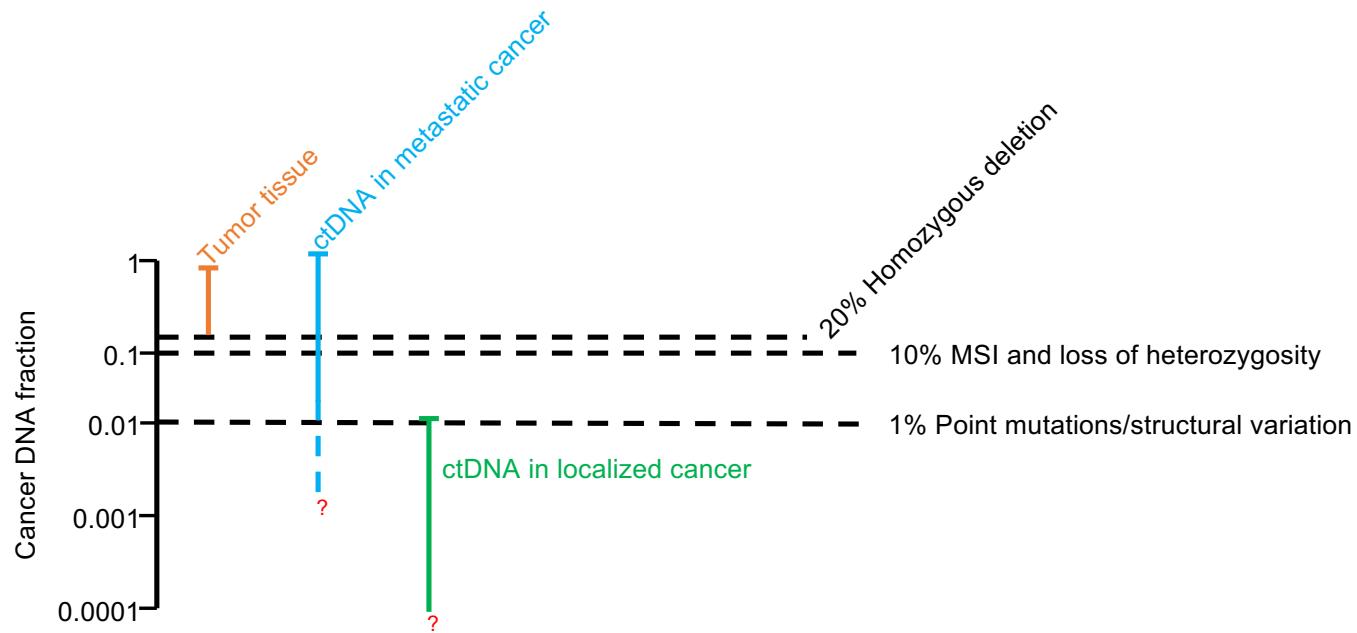
Cancer DNA fraction and consequence for genomic profiling using sequencing



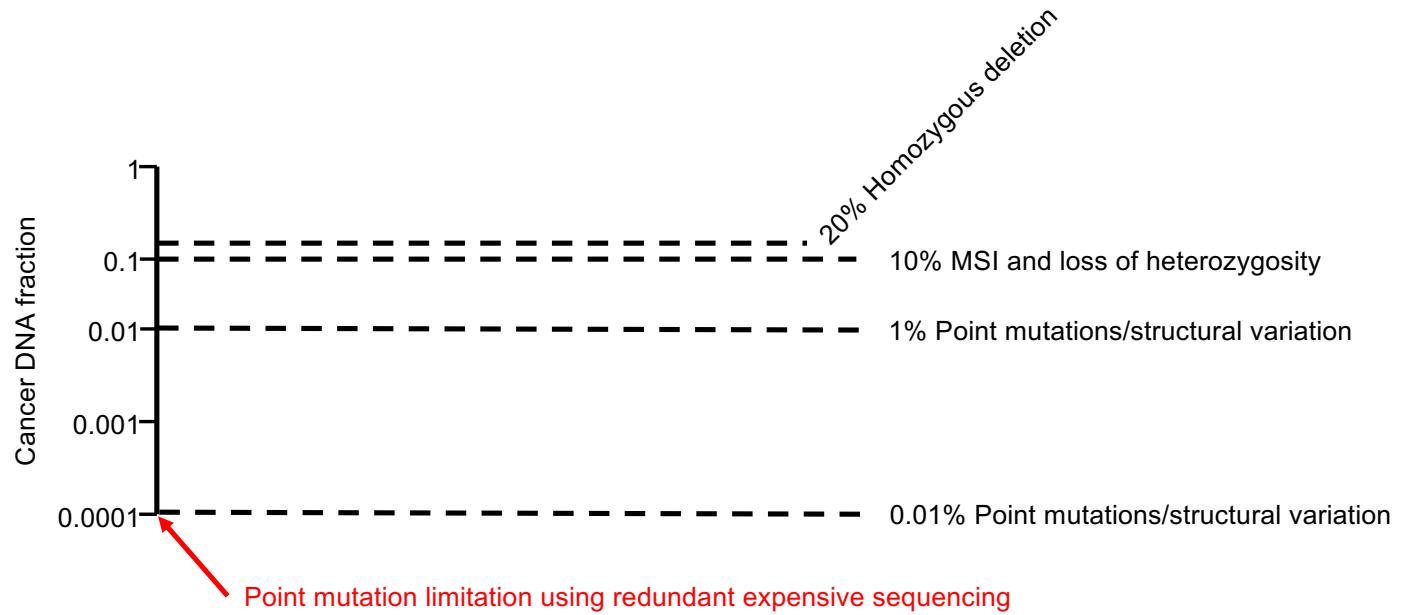
Cancer DNA fraction and consequence for genomic profiling using sequencing



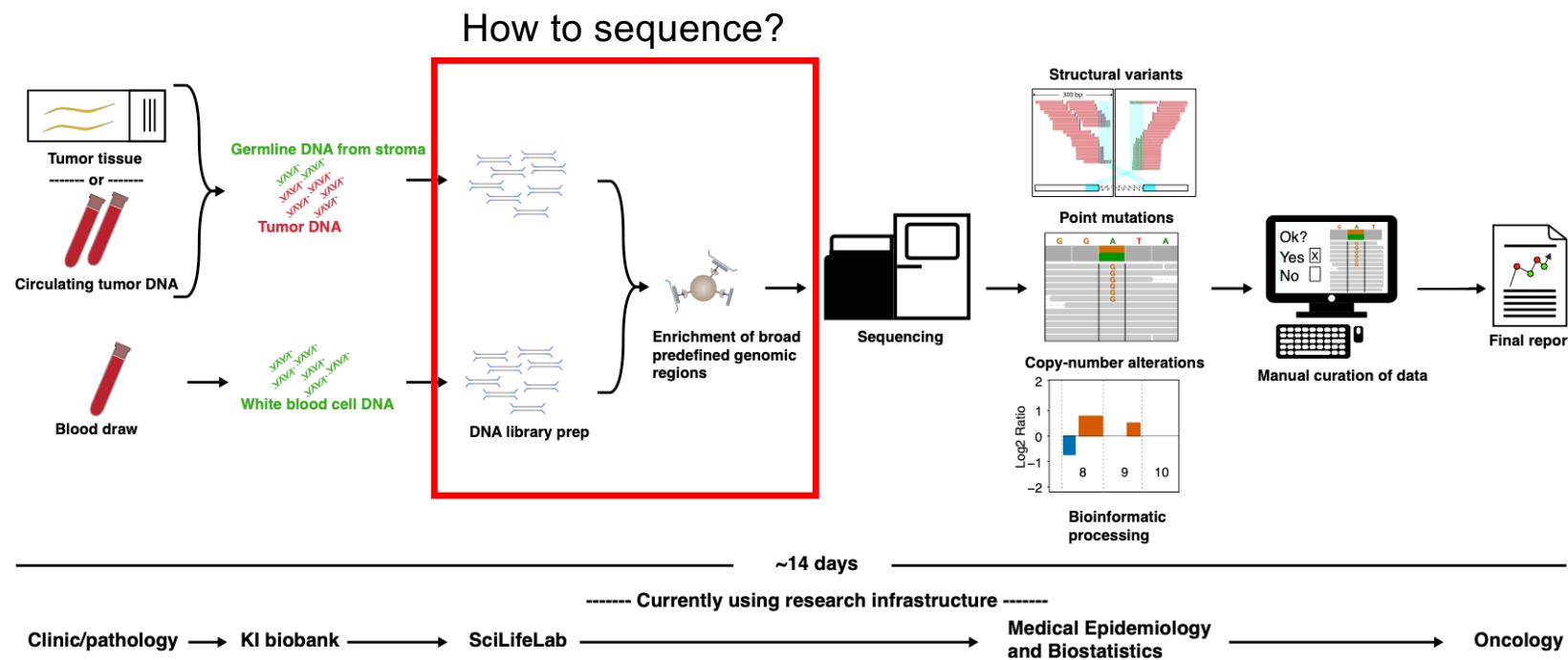
Cancer DNA fraction and consequence for genomic profiling using sequencing



Cancer DNA fraction and consequence for genomic profiling using sequencing

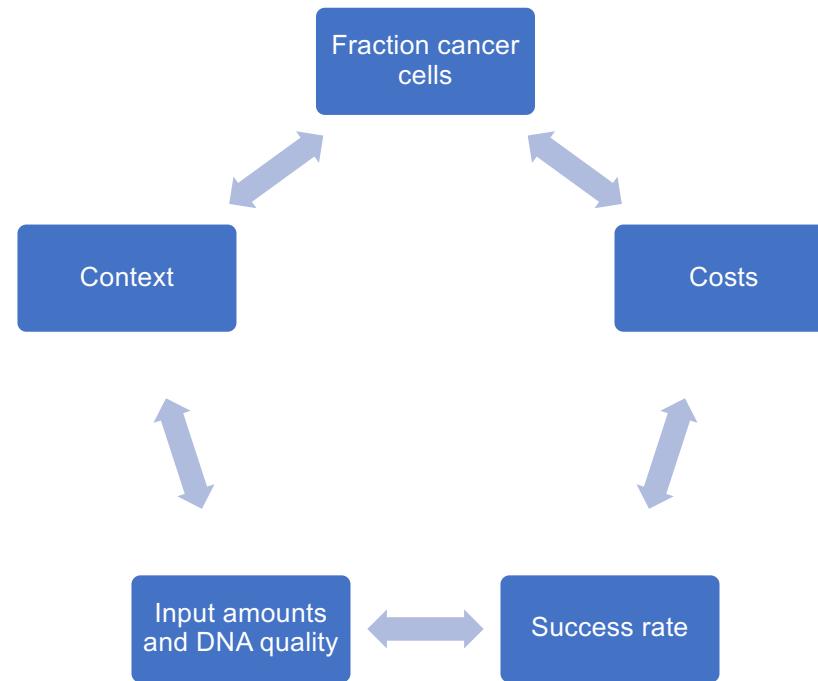


Practical considerations for performing cancer genomics



How to sequence depends on many things ..

- Research vs clinical.
- Clinical trial
 - High success rate
 - Fast turnover
 - Not acceptable with false positives
 - Allow for retrospective research

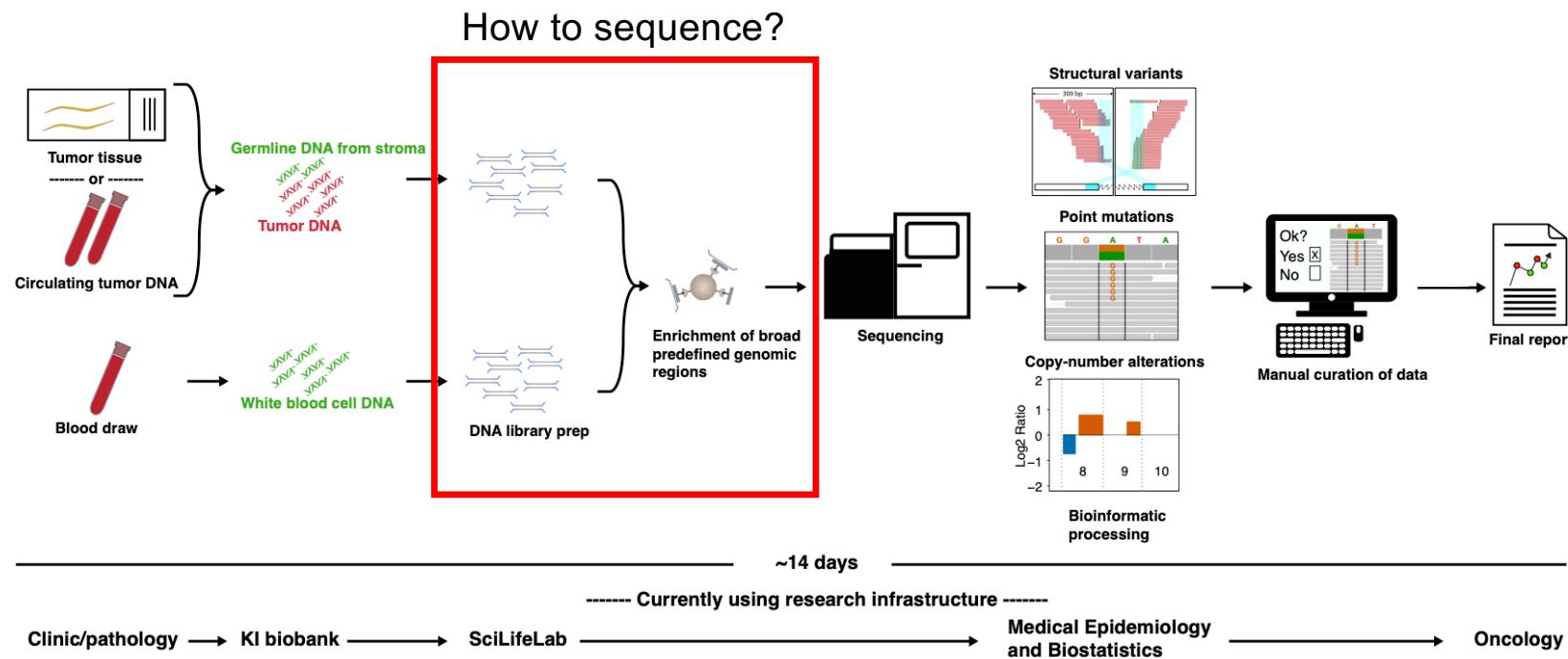


Why not deep whole genome sequencing?

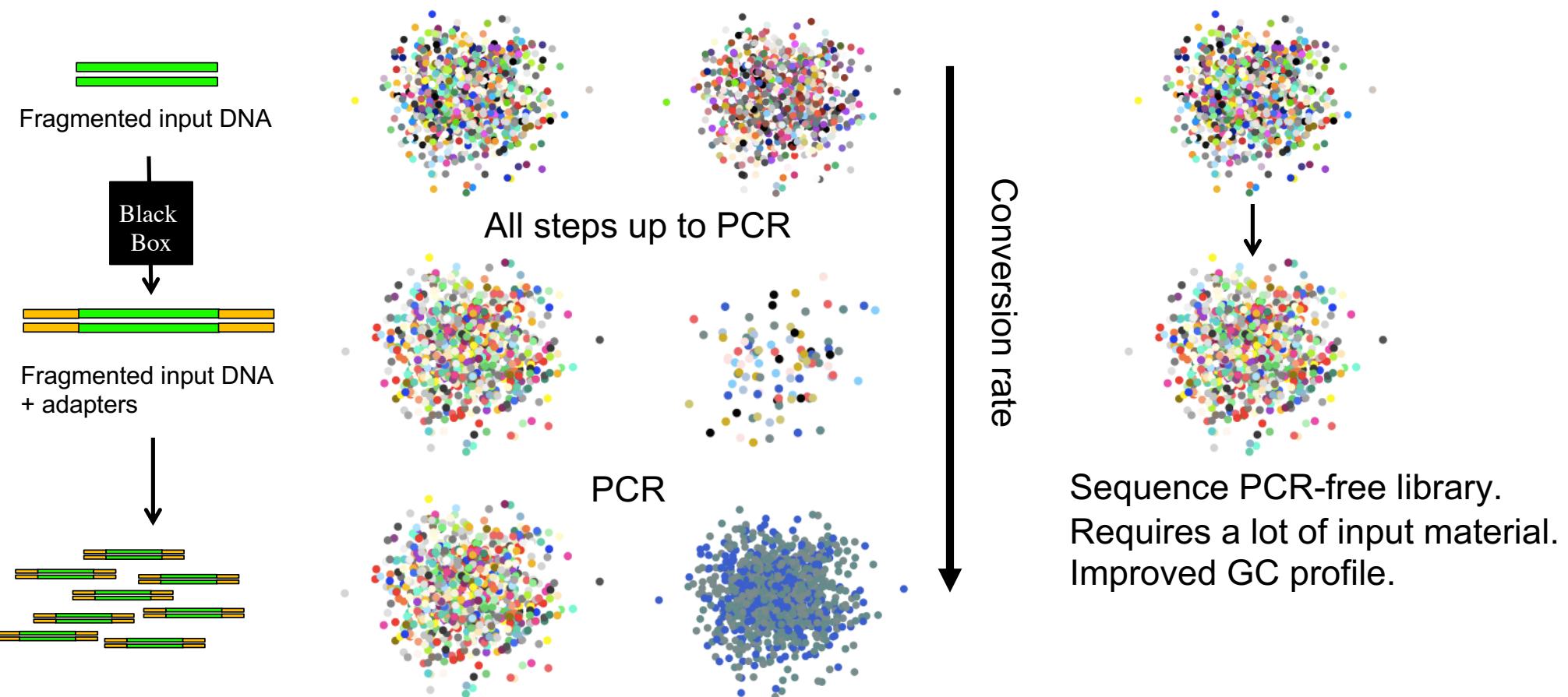
	Purity					
	100%	50%	20%	10%	1%	0.1%
Required coverage	100x	100x	500x	>1000x	>1500x	>5000x
Cost with WGS	€ 5 000	€ 5 000	-	-	-	-
Cost with targeted sequencing	€ 1 500	€ 1 500	€ 1 500	€ 1 500	€ 1 500	-
Minimal target and tailored bioinformatics	-	-	-	-	-	€1 500 - €2 000

Paired tumor and germline DNA analysis to enable identification of somatic- and germline alterations with good performance

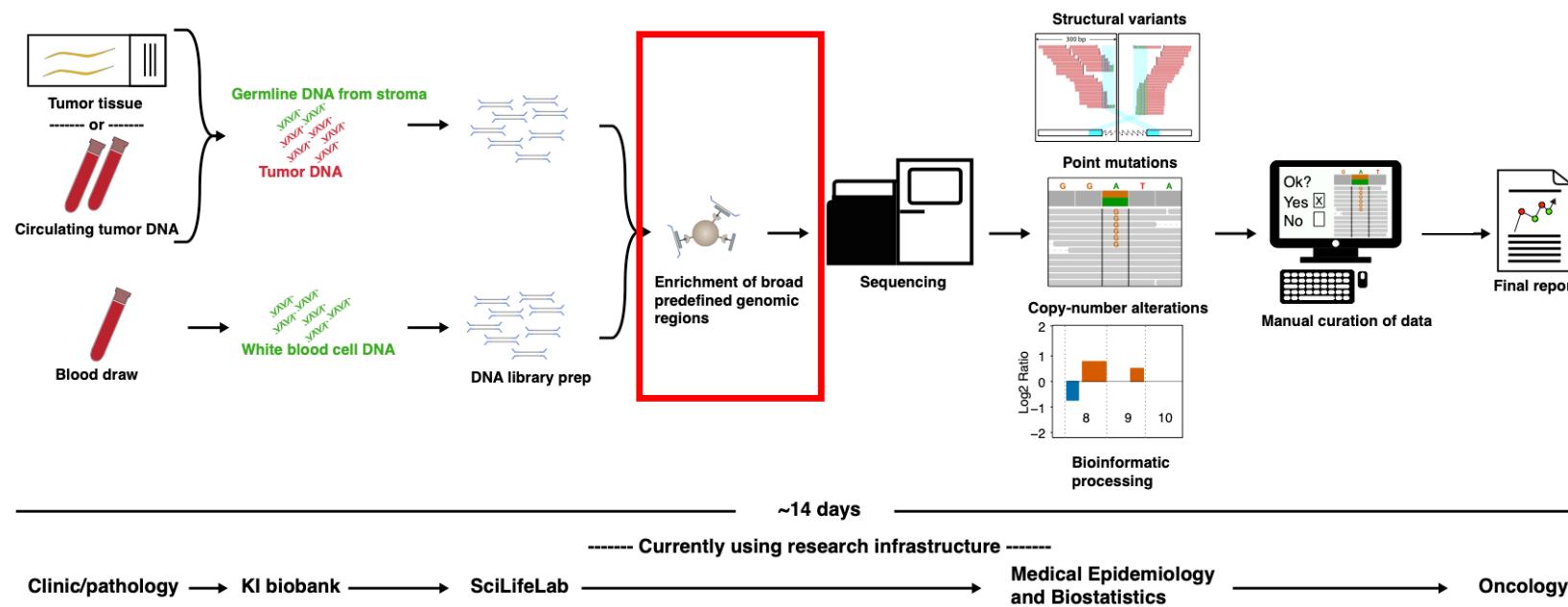
Practical considerations for performing cancer genomics



Library preparation in a nutshell

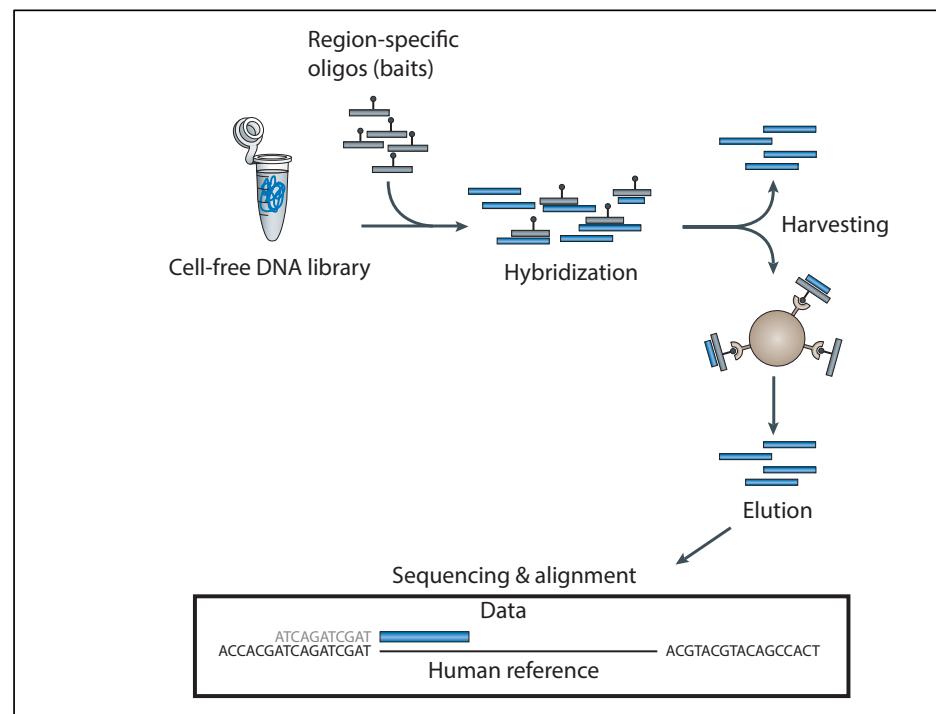


Key steps when performing DNA/RNA sequencing of cancer



Key steps when performing DNA/RNA sequencing of cancer

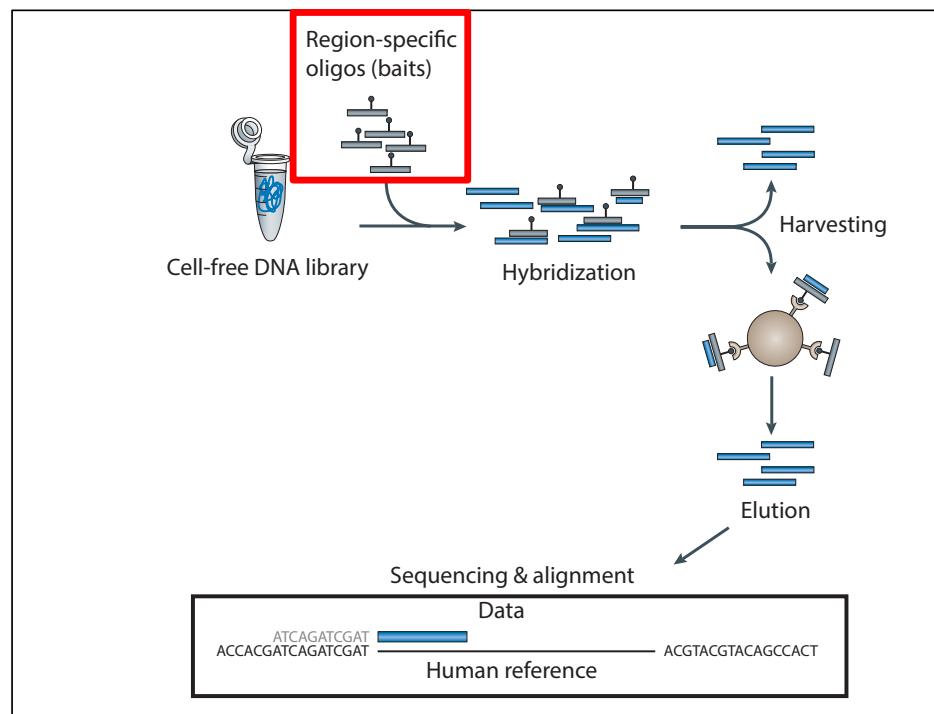
- In solution hybridisation based capture
- Flexibility in both breadth/depth



Adopted from: Advances in understanding cancer genomes through second-generation sequencing, Nat Rev Gen 2010

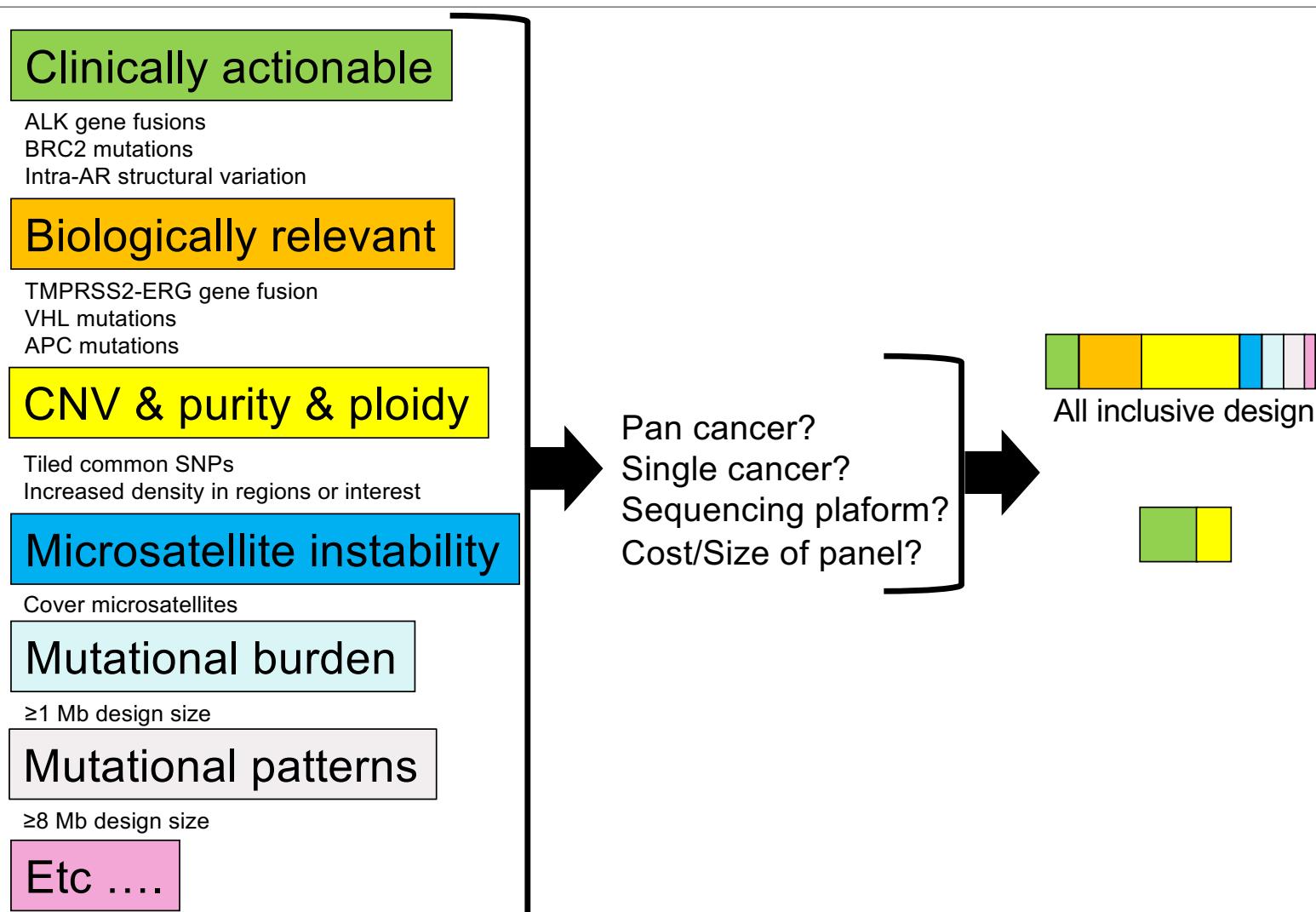
Key steps when performing DNA/RNA sequencing of cancer

- The baits determine the assay properties

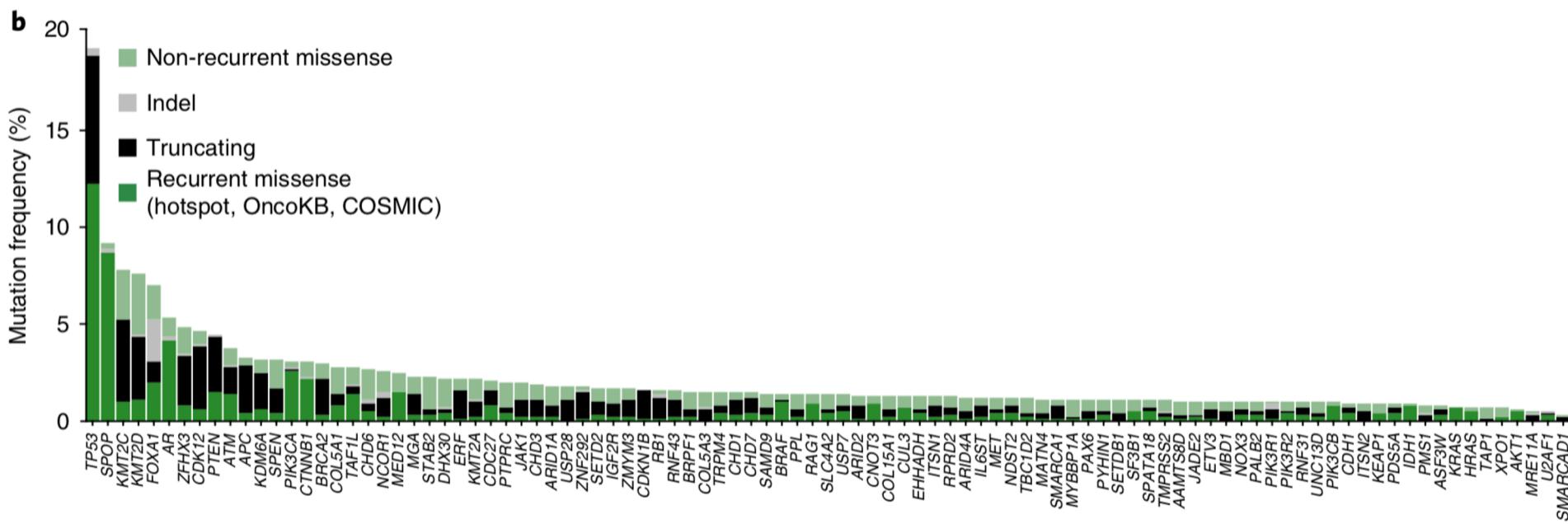


Adopted from: Advances in understanding cancer genomes through second-generation sequencing, Nat Rev Gen 2010

Defining a panel/bait set is like building lego ..



The long tail motivates broad genomic profiling



Sum(the long tail of actionability) – can be a large fraction of patients

- Metastatic prostate cancer
 - Increased neoantigens on the cell surface
 - Potential drug: Immunomodulator e.g. pembrolizumab
 - MSI+/TMB-H/Tandem duplication phenotype: ~10%
 - Homologous recombination deficiency
 - Potential drug: Parp inhibitor e.g. Niraparib
 - BRCA-complex genes only: ~10%

How to determine the baits?



- Clinical requirements
 - Cost-efficient sequencing of (all) clinically actionable regions
 - Deep sequencing to achieve high sensitivity if purity is low
 - Maintain performance in low-quality samples such as FFPE
- Research
 - Include other target regions/genes of relevance
 - Allows for exploration in large samples sets

How to find the optimal panel (mini genome)?

Sources to determine the panel

- **Literature, literature, literature**
 - Significantly mutated genes
 - TCGA pan cancer atlas
 - DNA repair genes, which are relevant for Parp/Carboplatin therapy?
 - How to design for gene fusions?
 - How many tiled probes are needed to achieve robust CNV calling?
- **Databases**
 - Ensembl
 - FDA approved list
 - Cancerhotspots.org
 - Cosmic
 - OncoKB
 - Other designs (via project Genie)
- **QC**
 - UCSC
 - Gene structure overview
 - cBioPortal
 - To investigate frequency of variants
- **From the sources above: 1069 genes in total**

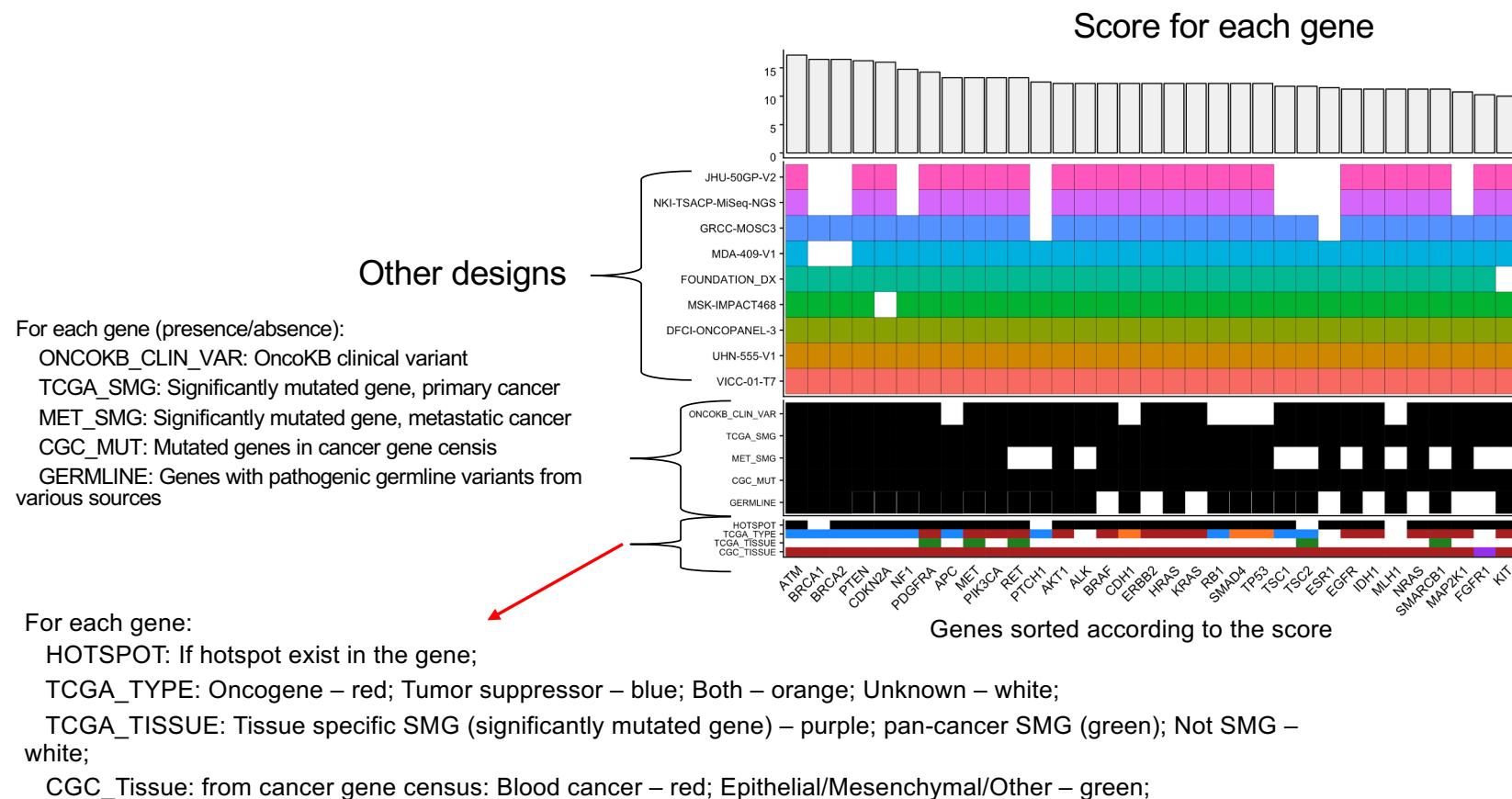
Other panels

- DFCI-ONCOPANEL-3 – 447 genes (capture, unclear ..)
- MSK-IMPACT468 – 468 genes (capture, Nimblegen)
- GRCC-MOSC3 – 78 hotspot (Ampliseq, PCR - Iontorrent)
- JHU-50GP-V2 – 50 hotspot (Ampliseq , PCR - Iontorrent)
- MDA-409-V1 – 409 genes (Ampliseq , PCR - Iontorrent)
- NKI-TSACP-MiSeq-NGS – 48 genes (Truseq - PCR - Illumina)
- UHN-555-V1 – 555 genes (Capture, Agilent)
- VICC-01-T7 – 429 genes (Capture, Foundation Medicine)
- Foundation Dx – 294 genes (Capture, Foundation Medicine)

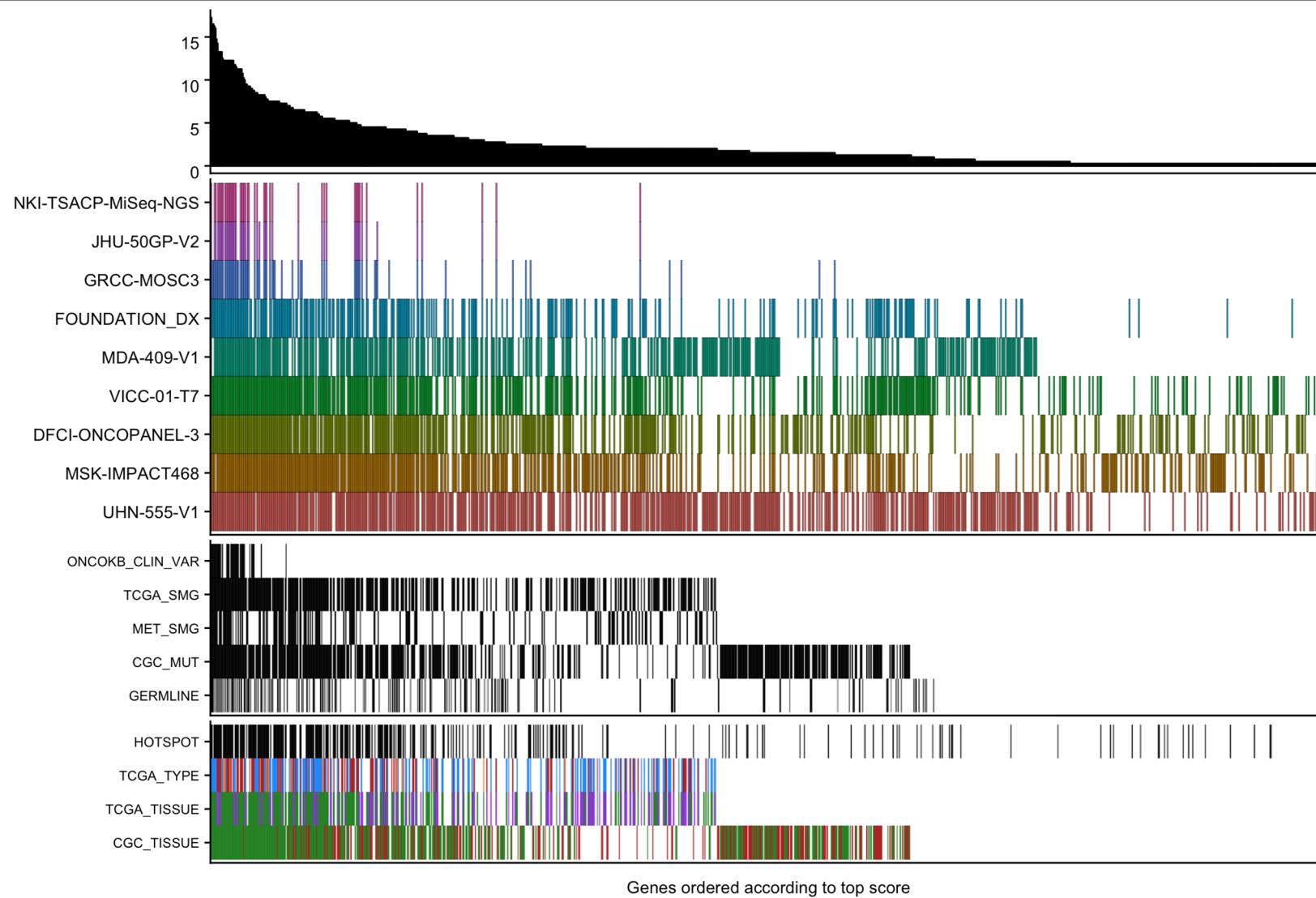
Scoring system to select exonic regions

- OncoKB: 4 p
- TCGA SMG: 2 p
- Advanced cancer SMG: 2 p
- Advanced “integrated” mutation list (less stringent, no new genes): 1 p
- CGC_mutation: 1 p
- BROCA pathogenic germline: 1 p
- TCGA pathogenic germline: 1 p
- CGC germline: 1 p
- ColorGenomics germline: 1 p
- Integrative germline (500 met cases): 1 p
- In panel from other source: 0.25 p

Overview heatmap, exonic regions – Zoom in



Overview heatmap, exonic regions



Evaluation using cBioPortal - 23576 samples in 33 studies

Adrenocortical Carcinoma (TCGA, PanCancer Atlas) 92 samples
Cholangiocarcinoma (TCGA, PanCancer Atlas) 36 samples
Bladder Urothelial Carcinoma (TCGA, PanCancer Atlas) 411 samples
Ewing Sarcoma (Institut Cuire, Cancer Discov 2014) 112 samples
Colon Adenocarcinoma (TCGA, PanCancer Atlas) 439 samples
Rectum Adenocarcinoma (TCGA, PanCancer Atlas) 155 samples
Breast Invasive Carcinoma (TCGA, PanCancer Atlas) 1084 samples
Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples
Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples
Cervical Squamous Cell Carcinoma (TCGA, PanCancer Atlas) 297 samples
TCGA data for Esophagus-Stomach Cancers (TCGA, Nature 2017) 559 samples
Stomach Adenocarcinoma (TCGA, PanCancer Atlas) 440 samples
Uveal Melanoma (TCGA, PanCancer Atlas) 80 samples
Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas) 523 samples
Kidney Renal Clear Cell Carcinoma (TCGA, PanCancer Atlas) 512 samples
Kidney Renal Papillary Cell Carcinoma (TCGA, PanCancer Atlas) 283 samples
Liver Hepatocellular Carcinoma (TCGA, PanCancer Atlas) 372 samples
Small Cell Lung Cancer (U Cologne, Nature 2015) 110 samples
Lung Adenocarcinoma (TCGA, PanCancer Atlas) 566 samples
Lung Squamous Cell Carcinoma (TCGA, PanCancer Atlas) 487 samples
MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017) 10945 samples
Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas) 585 samples
Pancreatic Adenocarcinoma (QCMG, Nature 2016) 456 samples
Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas) 184 samples
Pediatric Neuroblastoma (TARGET, 2018) 1089 samples
Metastatic Prostate Cancer, SU2C/PCF Dream Team (Robinson et al., Cell 2015) 150 samples
Prostate Adenocarcinoma (TCGA, Provisional) 499 samples
Skin Cutaneous Melanoma (TCGA, PanCancer Atlas) 448 samples
Sarcoma (TCGA, PanCancer Atlas) 255 samples
Testicular Germ Cell Tumors (TCGA, PanCancer Atlas) 149 samples
Thymoma (TCGA, PanCancer Atlas) 123 samples
Thyroid Carcinoma (TCGA, PanCancer Atlas) 500 samples
Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas) 529 samples

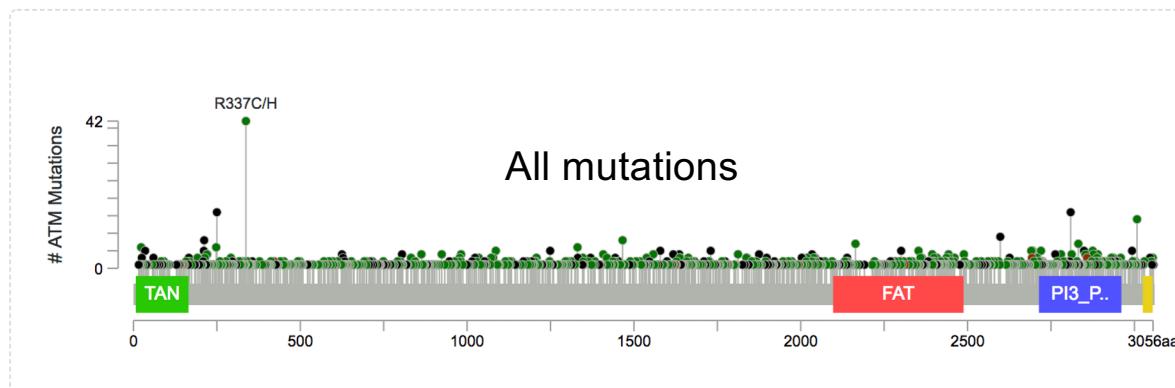
Selecting genes for mutations/germline alterations



Karolinska
Institutet

Combined Study (21381 samples)
Querying 20762 patients / 21381 samples in 33 studies

ATM IDH1



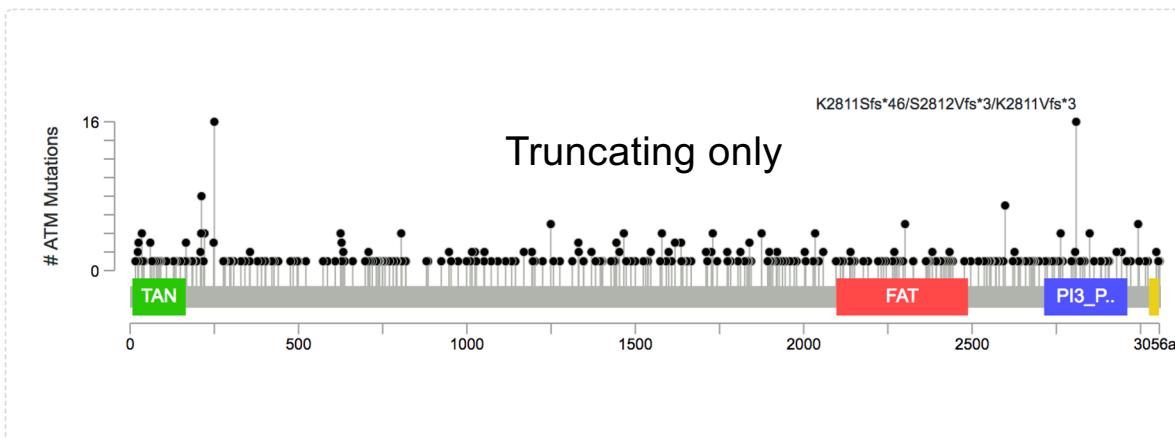
ATM

RefSeq: NM_000051
Ensembl: ENST00000278616
CCDS: CCDS31669
UniProt: ATM_HUMAN

Somatic Mutation Frequency: 5.1% ⓘ

951 Missense 436 Truncating
19 Inframe 11 Other

[View 3D Structure](#)



ATM

RefSeq: NM_000051
Ensembl: ENST00000278616
CCDS: CCDS31669
UniProt: ATM_HUMAN

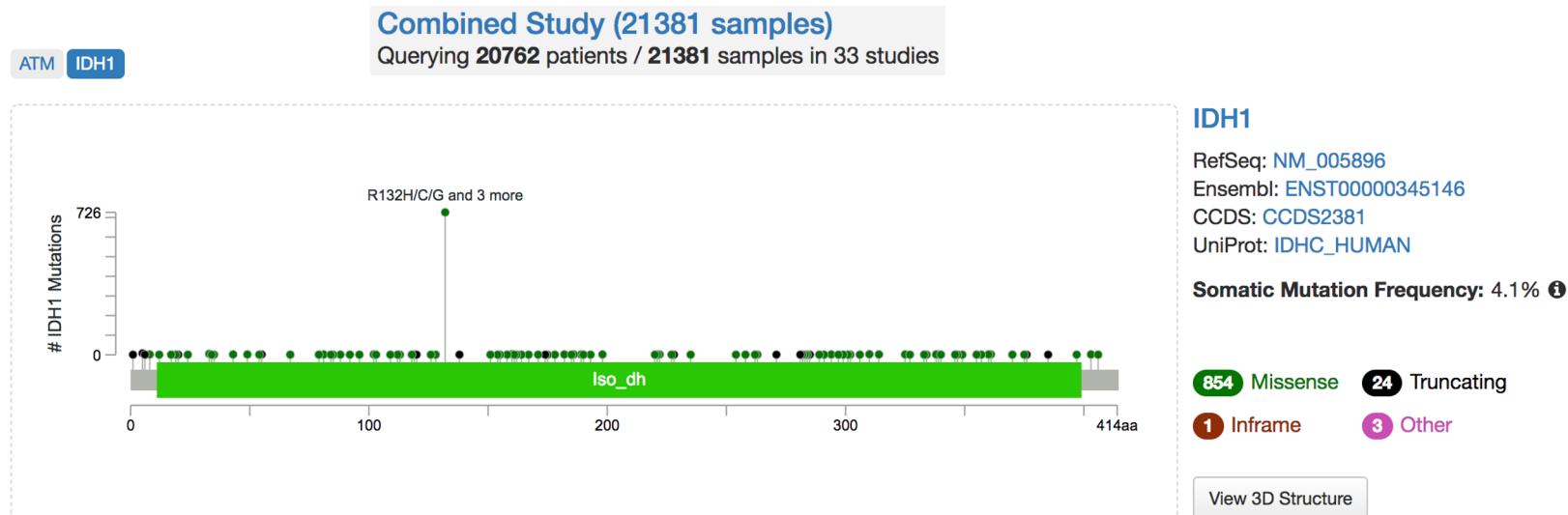
Somatic Mutation Frequency: 5.1% ⓘ

0 Missense 436 Truncating
0 Inframe 0 Other

[View 3D Structure](#)

Selecting genes for mutations/germline alterations

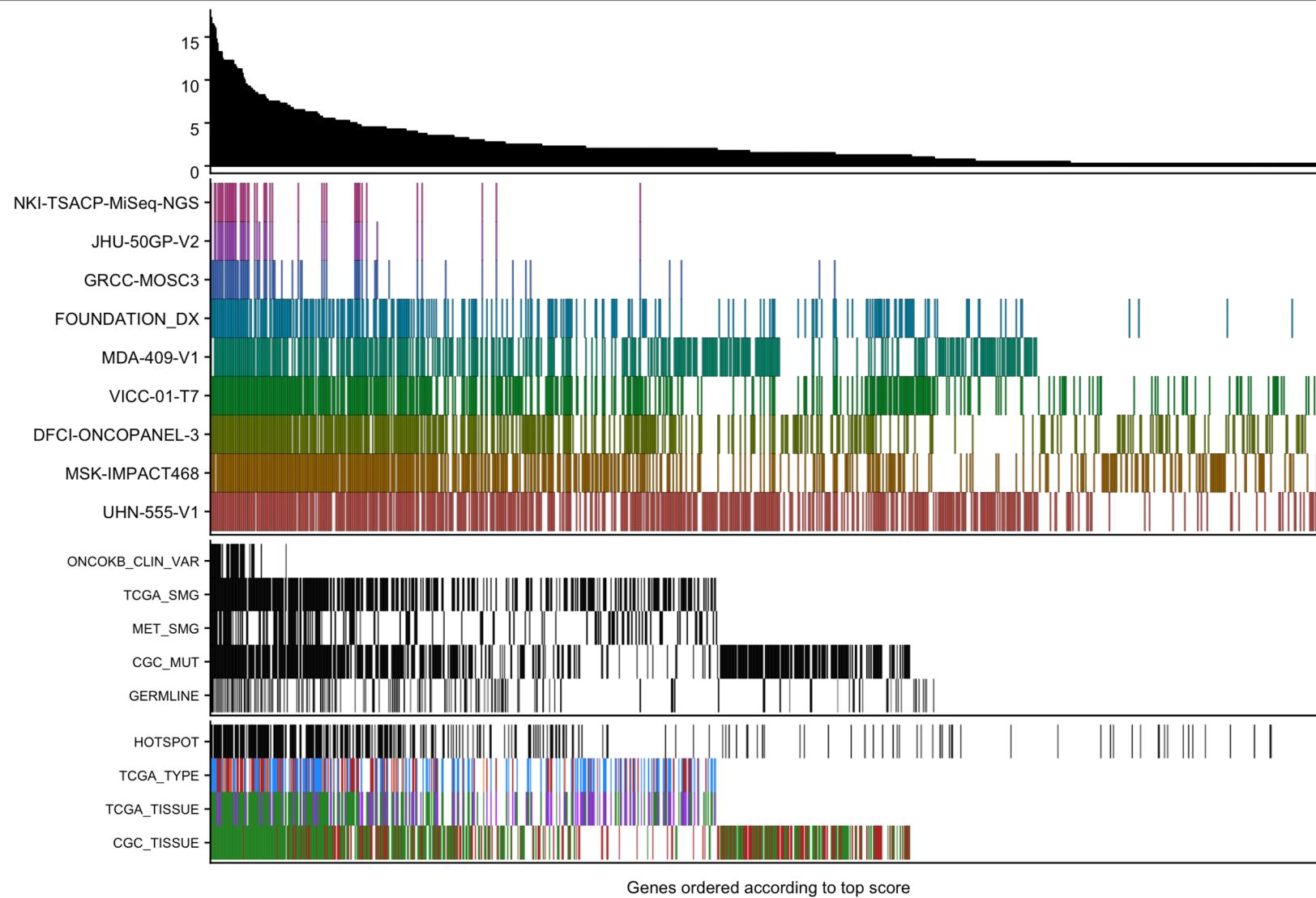
All mutations, only sequence hotspot



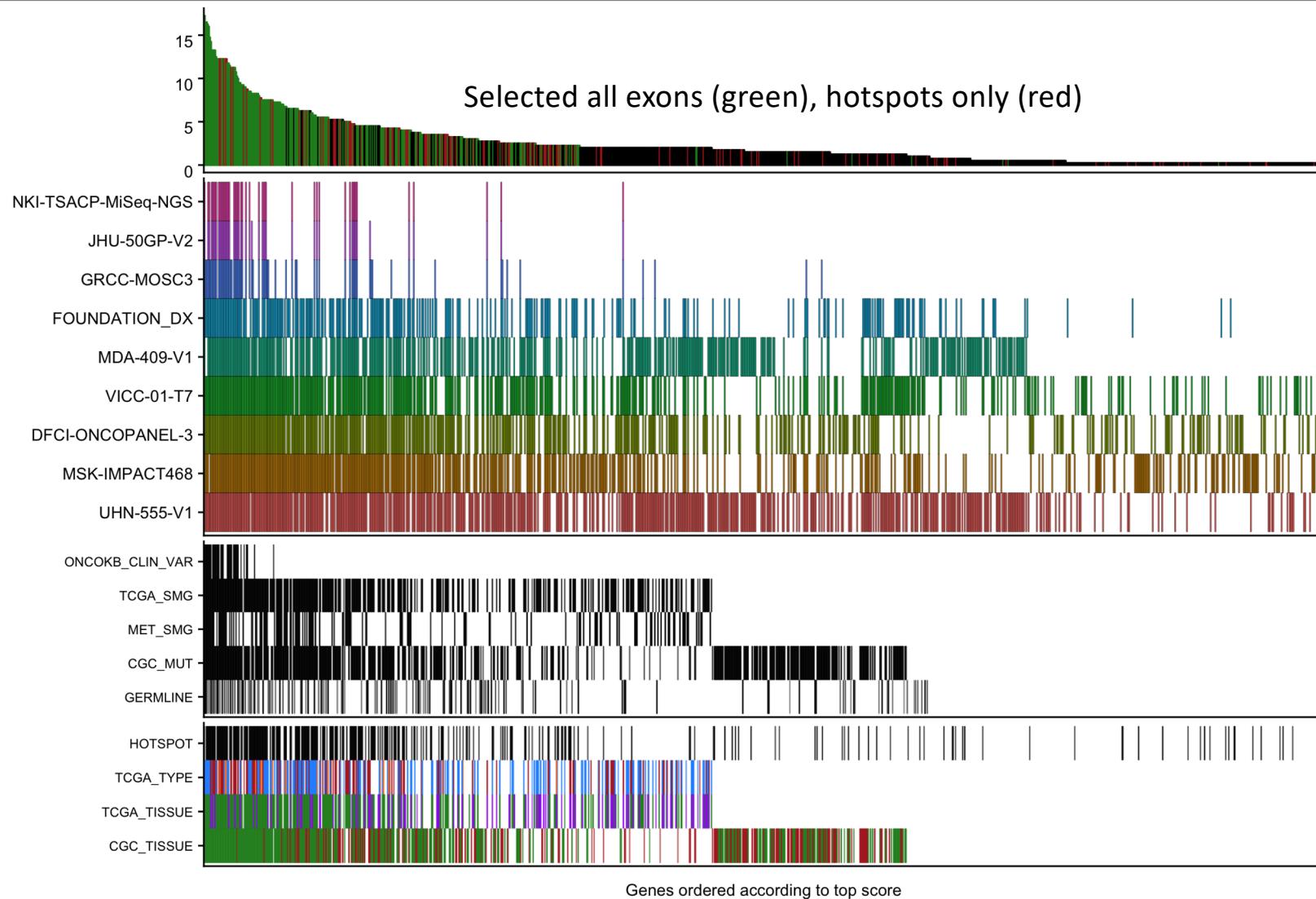
Selecting genes for mutations/germline alterations

- Genes binned according to score
 - ≥4p
 - 198 genes
 - >2p
 - 35 significantly mutated genes
 - 74 genes with hotspots
 - Clonal hematopoiesis hotspots
 - 19 genes
 - Genes with low score but included in for e.g. germline sequencing purposes or other anecdotal reasons
 - 22 genes
- 348 genes selected for exon sequencing
 - Hotspot only: 132 genes
 - All exons: 198 genes
 - Clonal hematopoiesis: 18 (hotspots) + 1 (all exons)

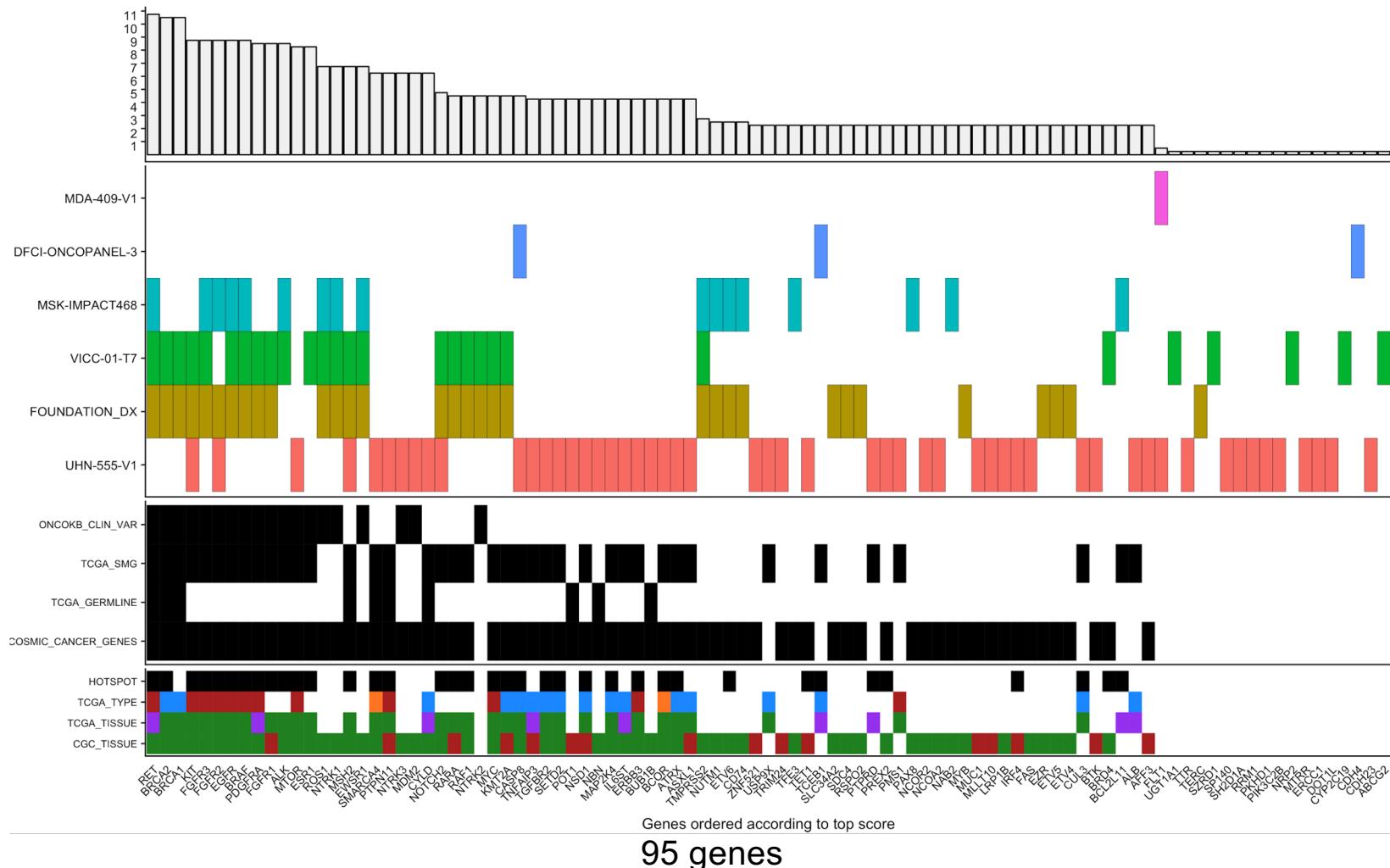
Overview heatmap, exonic regions

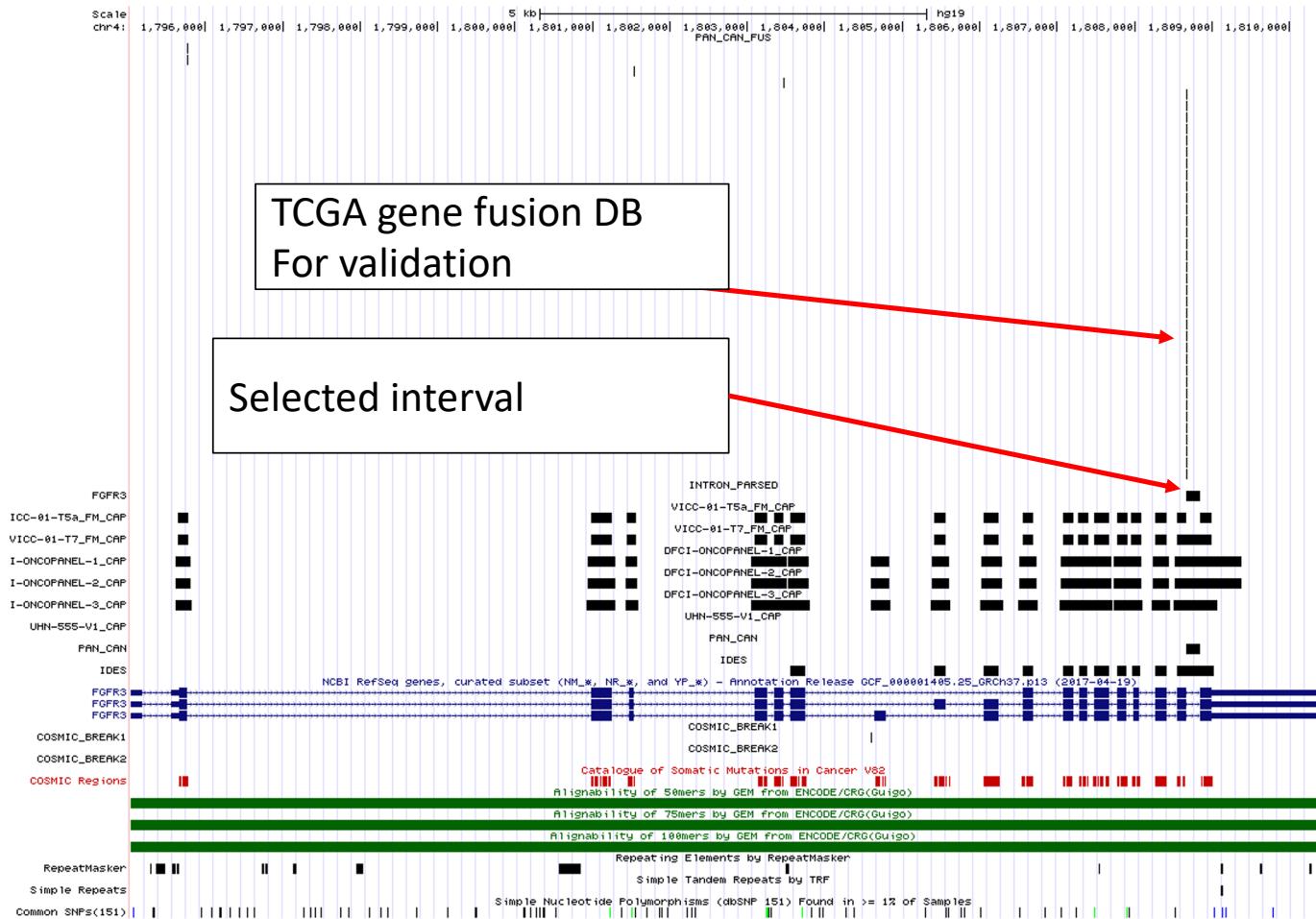


Overview heatmap, exonic regions

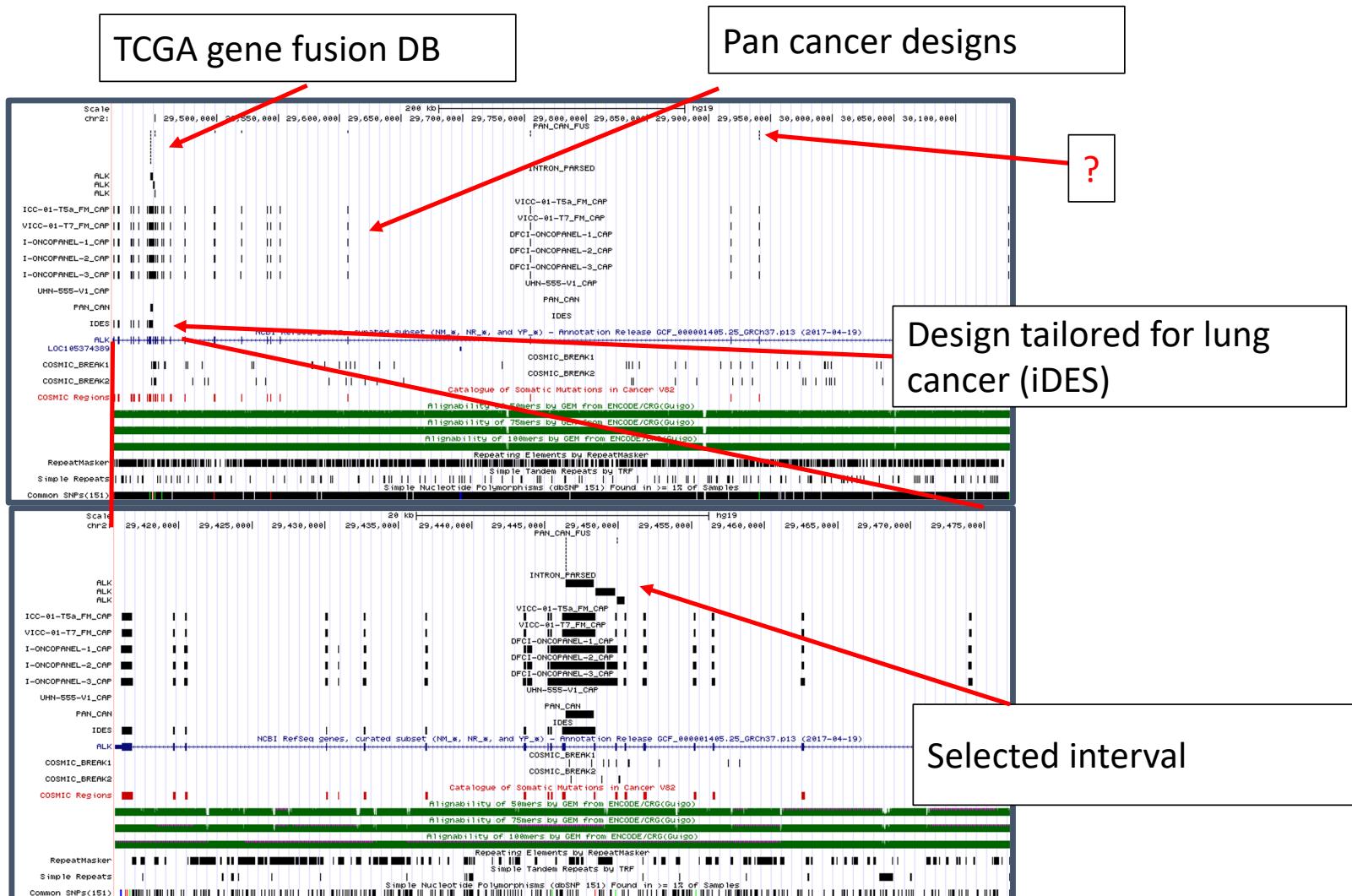


Overview heatmap, intronic and promoter regions

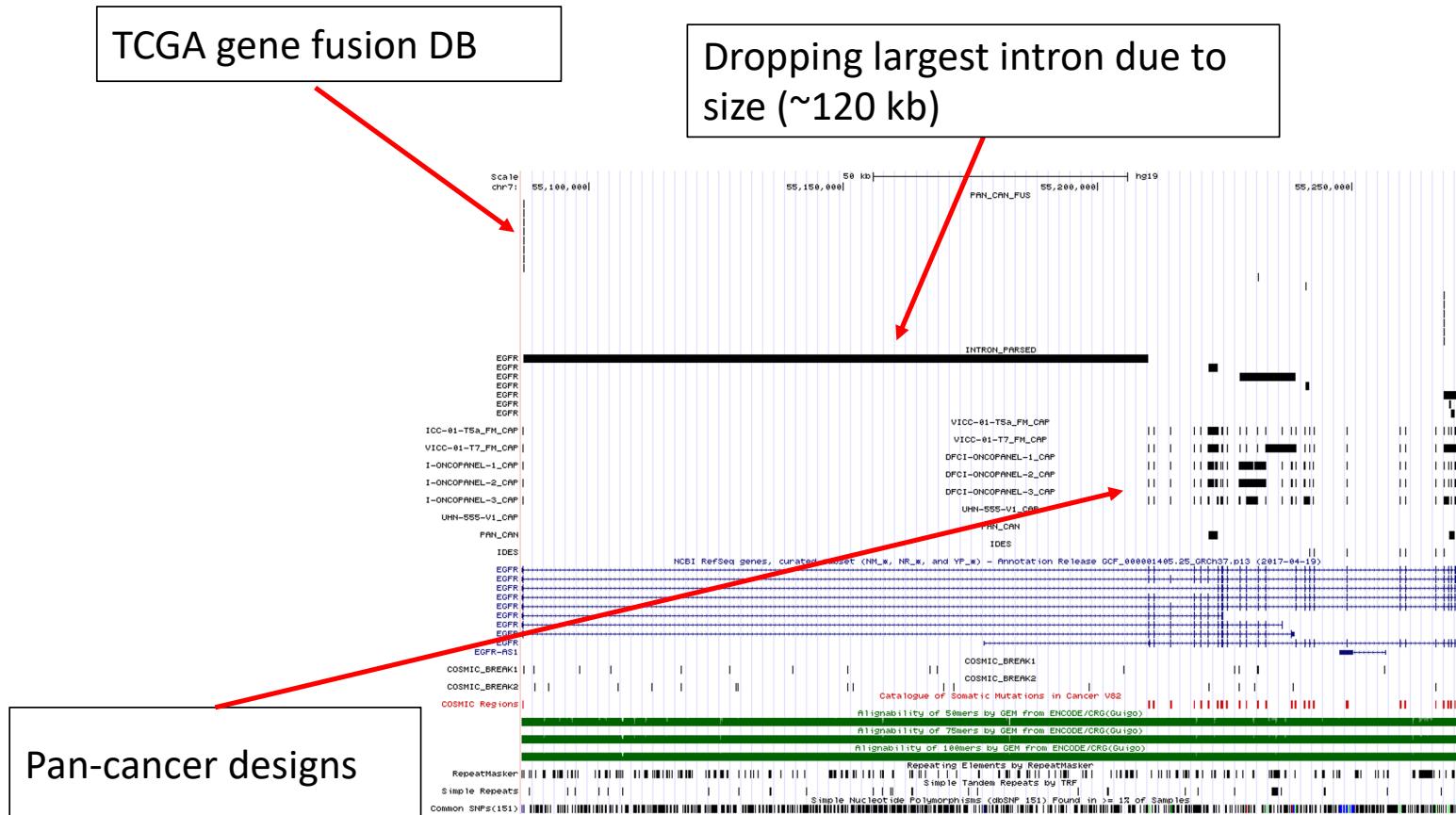




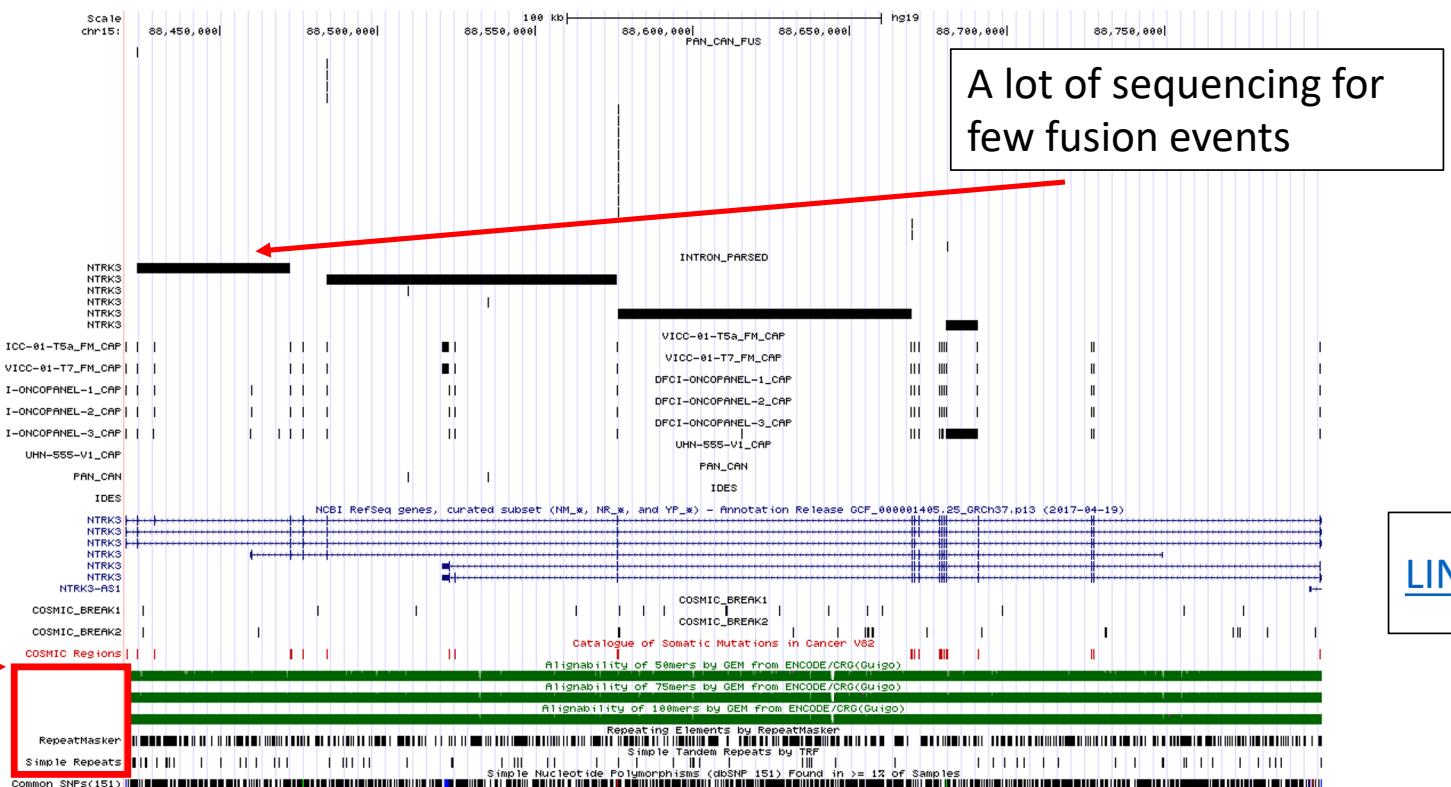
Manual examination of candidates - ALK



Evaluation - EGFR



NTRK3



Repeats/low mappability – impossible to catch all variants via WGS/targeted sequencing

Intronic sequencing

- Selected introns (19) genes
 - ALK
 - BRAF
 - EGFR
 - ERG
 - ETV6
 - EWSR1
 - FGFR1
 - FGFR2
 - FGFR3
 - GPR126
 - KIT
 - NTRK1
 - NTRK3
 - PDGFRA
 - PDGFRB
 - RET
 - ROS1
 - TMPRSS2
- Promoter
 - TERT

Covers 11 out of 12 gene fusions for solid tumor in oncoKB

Should be seen as opportunistic and not diagnostic, needs to be complemented by an RNA assay

Intronic sequencing

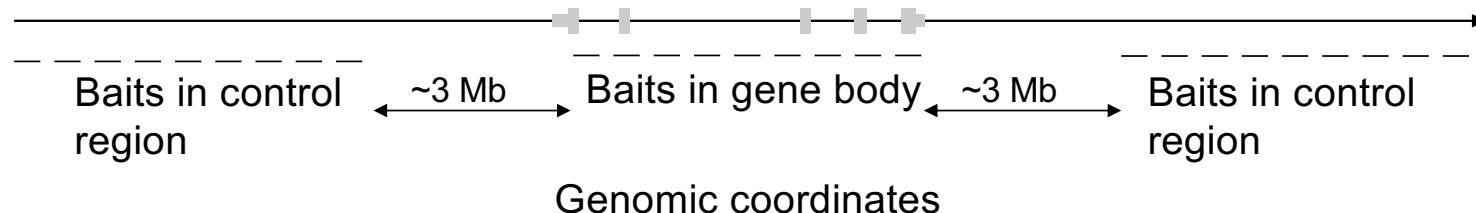
- Selected introns (19) genes
 - ALK
 - BRAF
 - EGFR
 - ERG
 - ETV6
 - EWSR1
 - FGFR1
 - FGFR2
 - FGFR3
 - GPR126
 - KIT
 - NTRK1
 - NTRK3
 - PDGFRA
 - PDGFRB
 - RET
 - ROS1
 - TMPRSS2
- Promoter
 - TERT

Covers 11 out of 12 gene fusions for solid tumor in oncoKB

Should be seen as opportunistic and not diagnostic, needs to be complemented by an RNA assay

Baits for copy-number profiling

- 2814 common SNPs (HapMap SNPs, pan-population)
 - Genome-wide CNA calling
 - B-allele frequencies to support CNA and enable detection of copy-neutral events
 - 64 “ID” SNPs from lifetech
- 86 CNA affected genes/regions with 20 tiled CNV baits and control regions to increase resolution (overlap with common SWE-SNPs)
- No gene-body tiles genes if all exons are anyways sequenced



GMCK panel summary – soon to be updated



Karolinska
Institutet

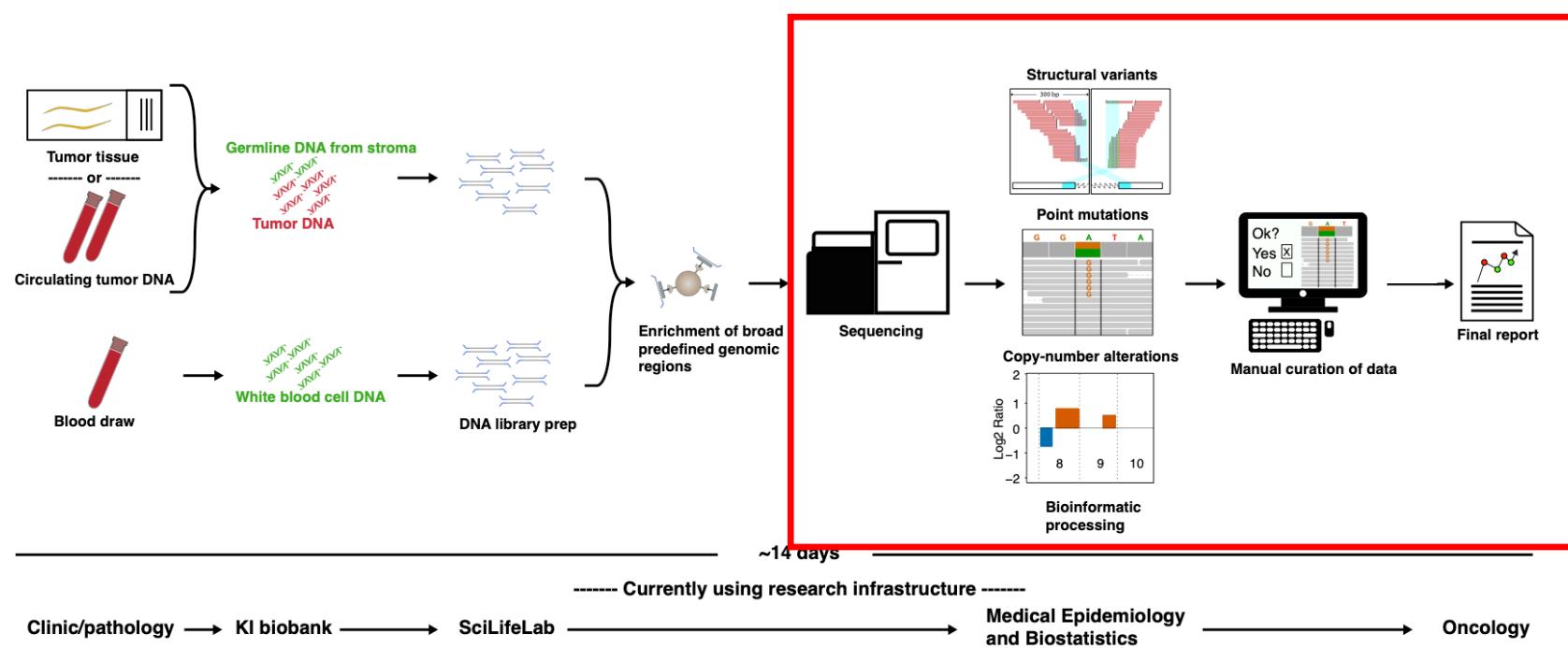
Mutations		OncoKB coverage	
All coding exons	198 genes	26/31	
Hotspots	132 genes	5/31	
Pharmacogenetic variants			
SNPs			9 genes
Copy-number alterations			
Tiled SNP for genome-wide CNV	2814 SNPs		
Directed analysis to increase sensitivity	86 genes	6/6	
Structural variation			
Genefusions by intronic sequencing	19 genes	9/10	
Gene-body sequencing (e.g. BRCA1/2)	9 genes		
Microsatellite instability & Hypermutation			
Microsatellites	63 in total		
Hypermutation, entire design footprint	yes		
Associated genes	7		
Total size (Mb)			2.4

If time allows (non-obligatory work) you will get to design your own targeted assay during the course.

The size of the panel was determined considering the sequencing price at Clinical Genomics @ SciLife

Practical considerations for performing cancer genomics

Covered in other lectures ...



Questions?
