



Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by



This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomeksi français français (CA) Galego മലാളി Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenščina srpski (latinica) Sotho svenska
中文 繁體 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full licence) available in the following languages:

[English](#) [French](#)

[Learn how to distribute your work using this licence](#)

Module 1

Introduction to Cancer Genomics

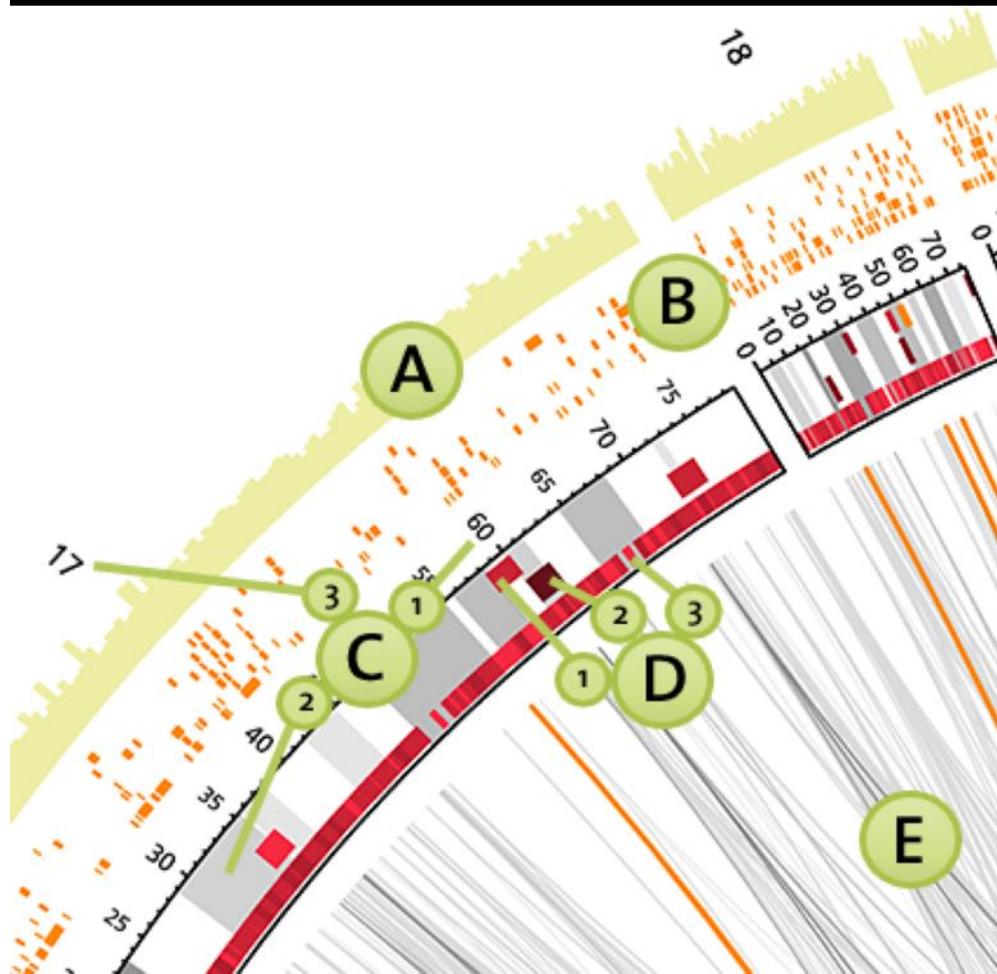


bioinformatics.ca

Trevor Pugh, PhD, FACMG

Bioinformatics for Cancer Genomics

June 3-7, 2019



Princess Margaret
Cancer Centre  UHN



UNIVERSITY OF
TORONTO

Learning Objectives of Module

- Describe how cancer genomes differ from normal genomes and various sources of cancer genome variation
- Understand different bioinformatic approaches to detecting cancer genome variation
- Learn strengths and weaknesses of different approaches to cancer genome analysis (whole genome, exome, RNA-seq, targeted sequencing, etc.)
- Inform how cancer genome analysis may be used to guide cancer patient management

We are all made of cells

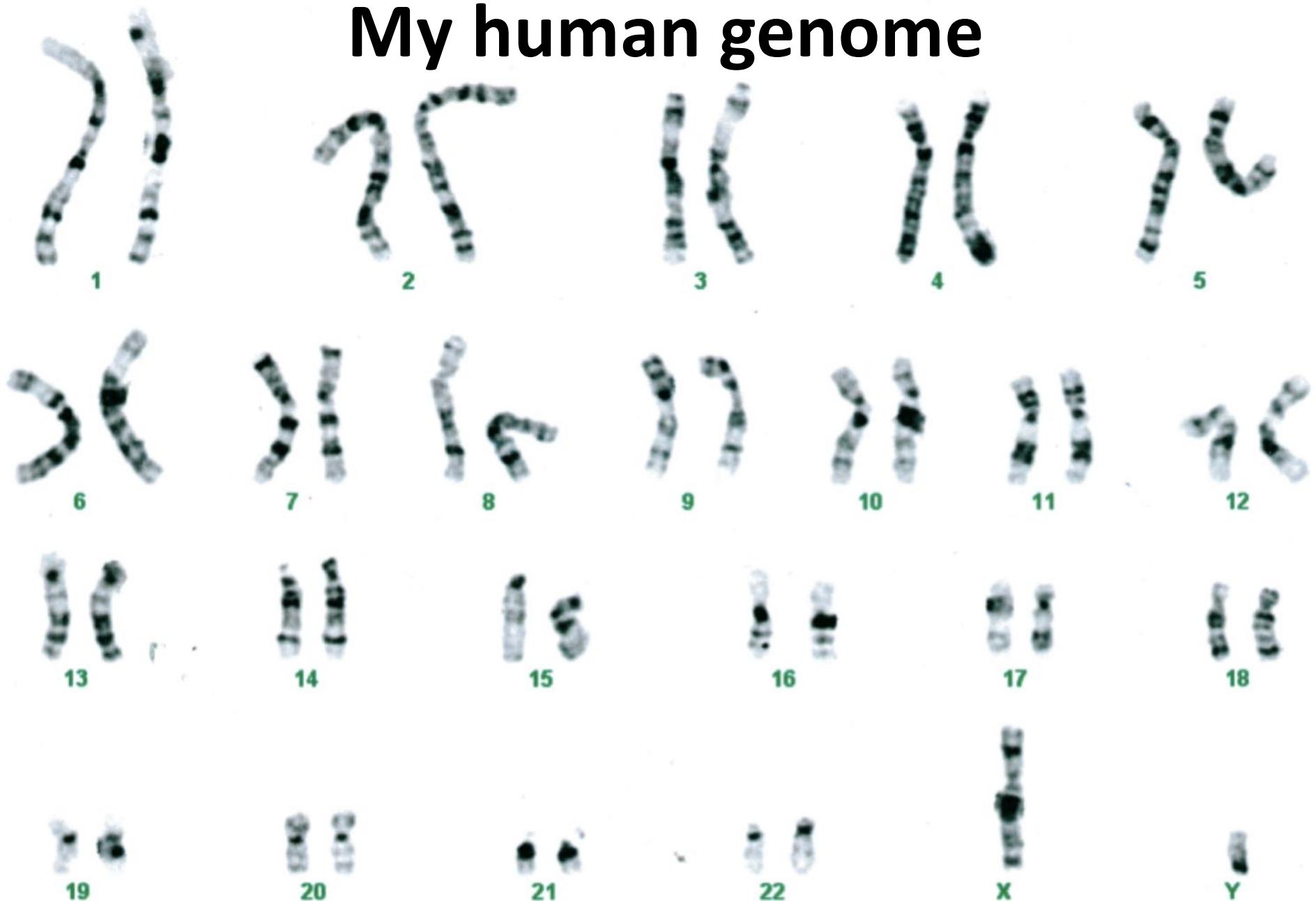


<https://genographic.nationalgeographic.com/science-behind/genetics-overview/>

My human genome



My human genome



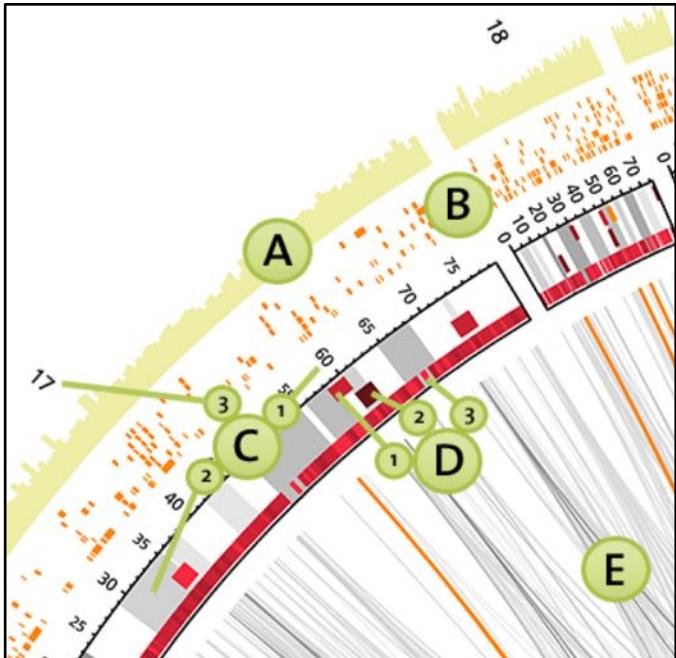
Cancer is a disease of the genome



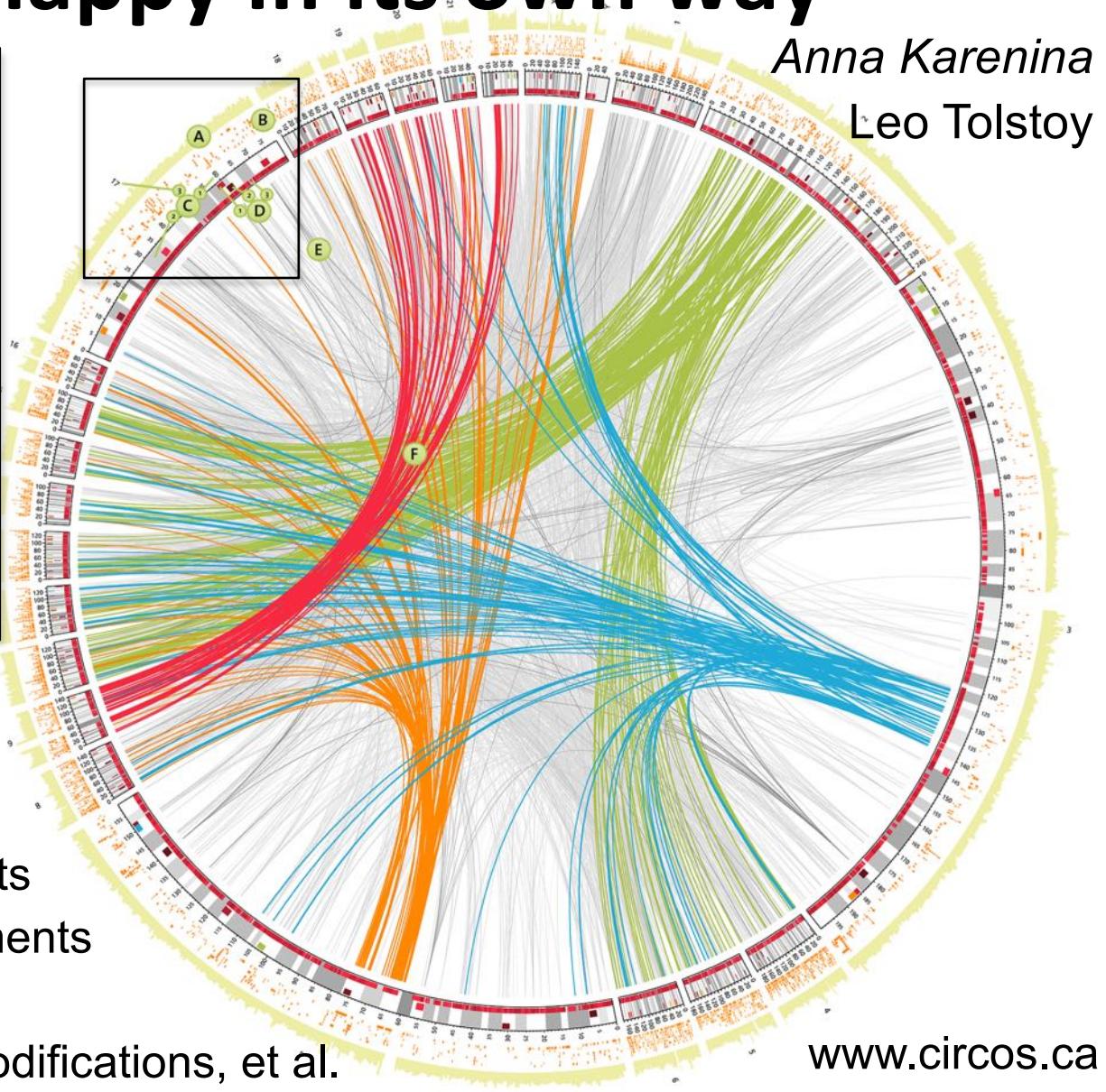
Glioma karyogram (GTG-banding). 78,<4n>,XXXX,-2,-5,-6,del(6)(q21q23)x2,+7,-8,del(8)(q22q24.1),del(9)(p10)x2,-10,-10,-11, -11,-12,-13,-13,-14,-14,-16,-19,del(19)(p10),-21,+22

www.cityofhope.org/research/support/cytogenetics/

"All happy families are alike; each unhappy family is unhappy in its own way"



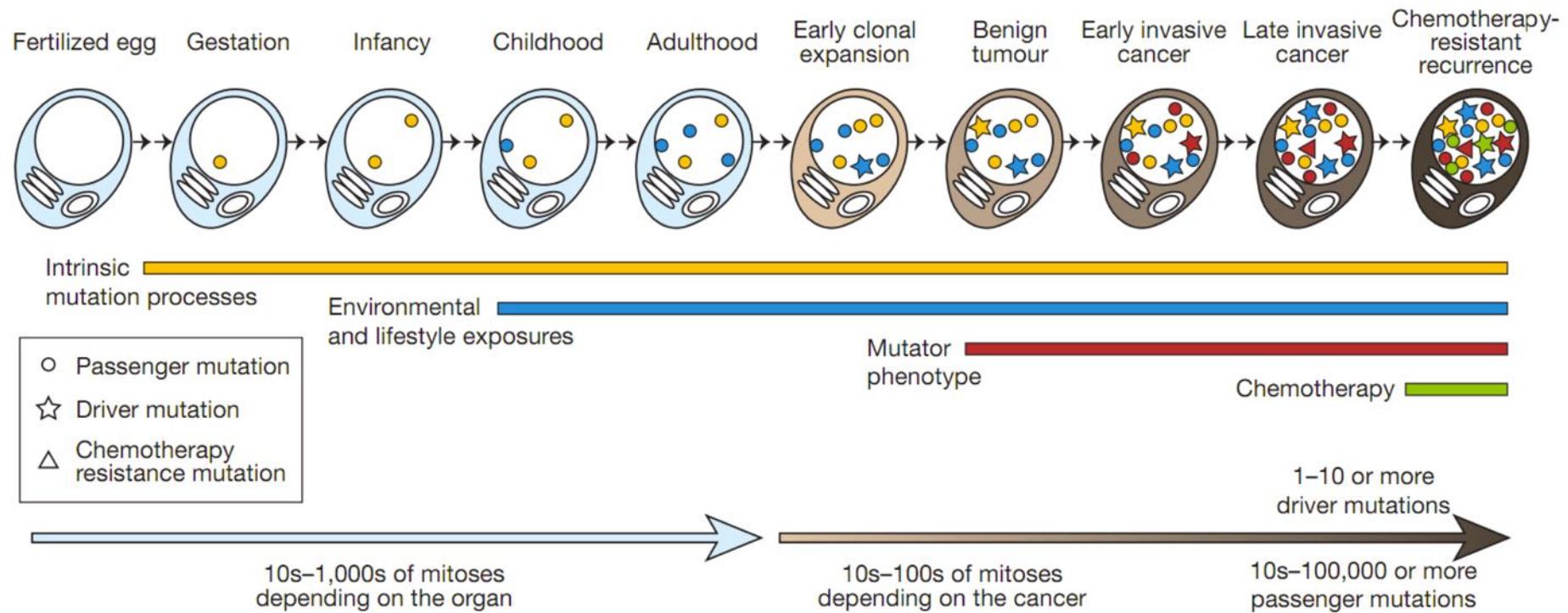
- A** Point mutations
- B** Copy number alterations
- C** Structural rearrangements
- D** Genes & regulatory elements
- E** Pathways
- F** Homologs, epigenetic modifications, et al.



Anna Karenina
Leo Tolstoy

www.circos.ca

Cancer cells accumulate somatic alterations over time

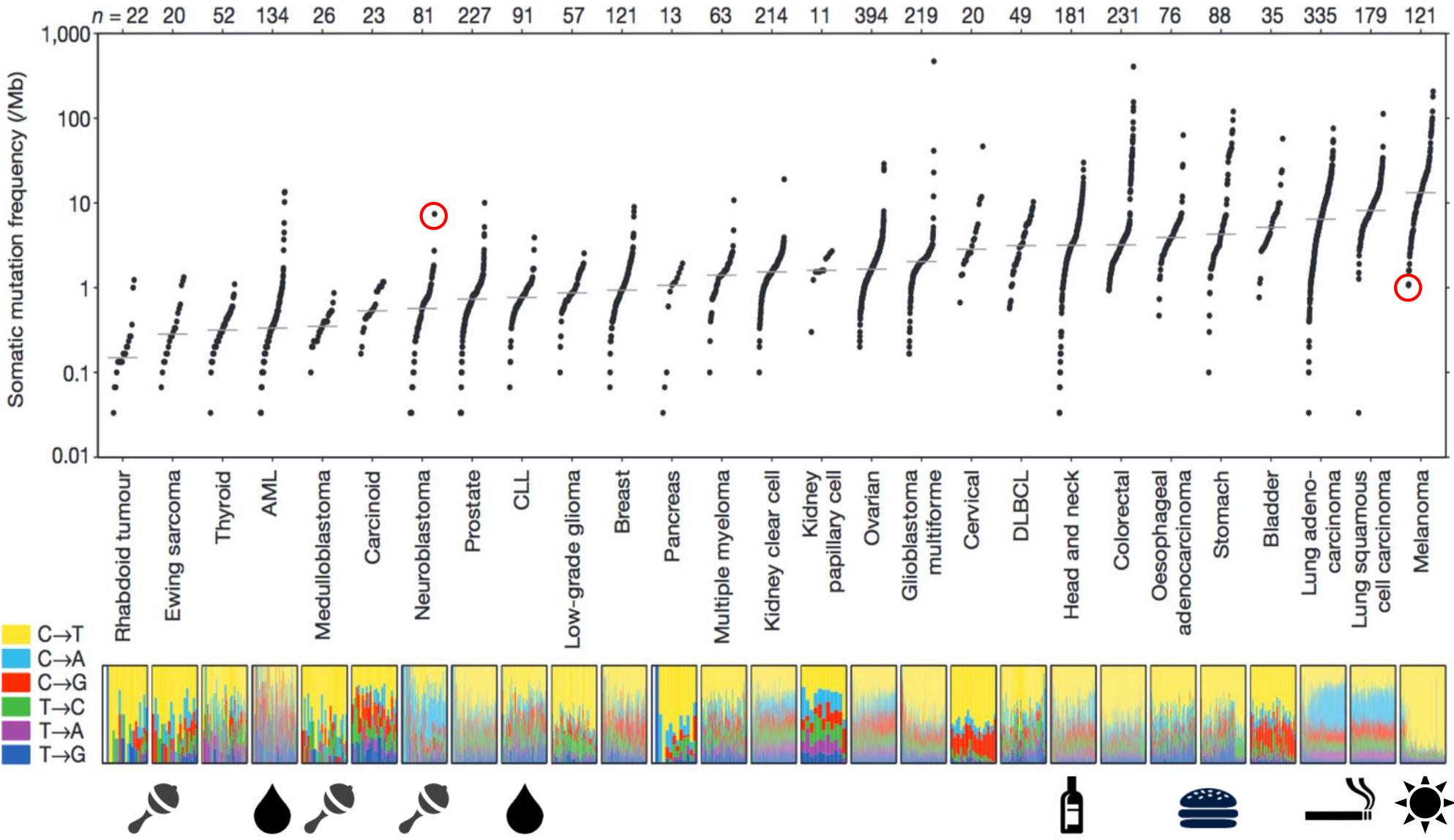


Mutation frequency depends on cancer type

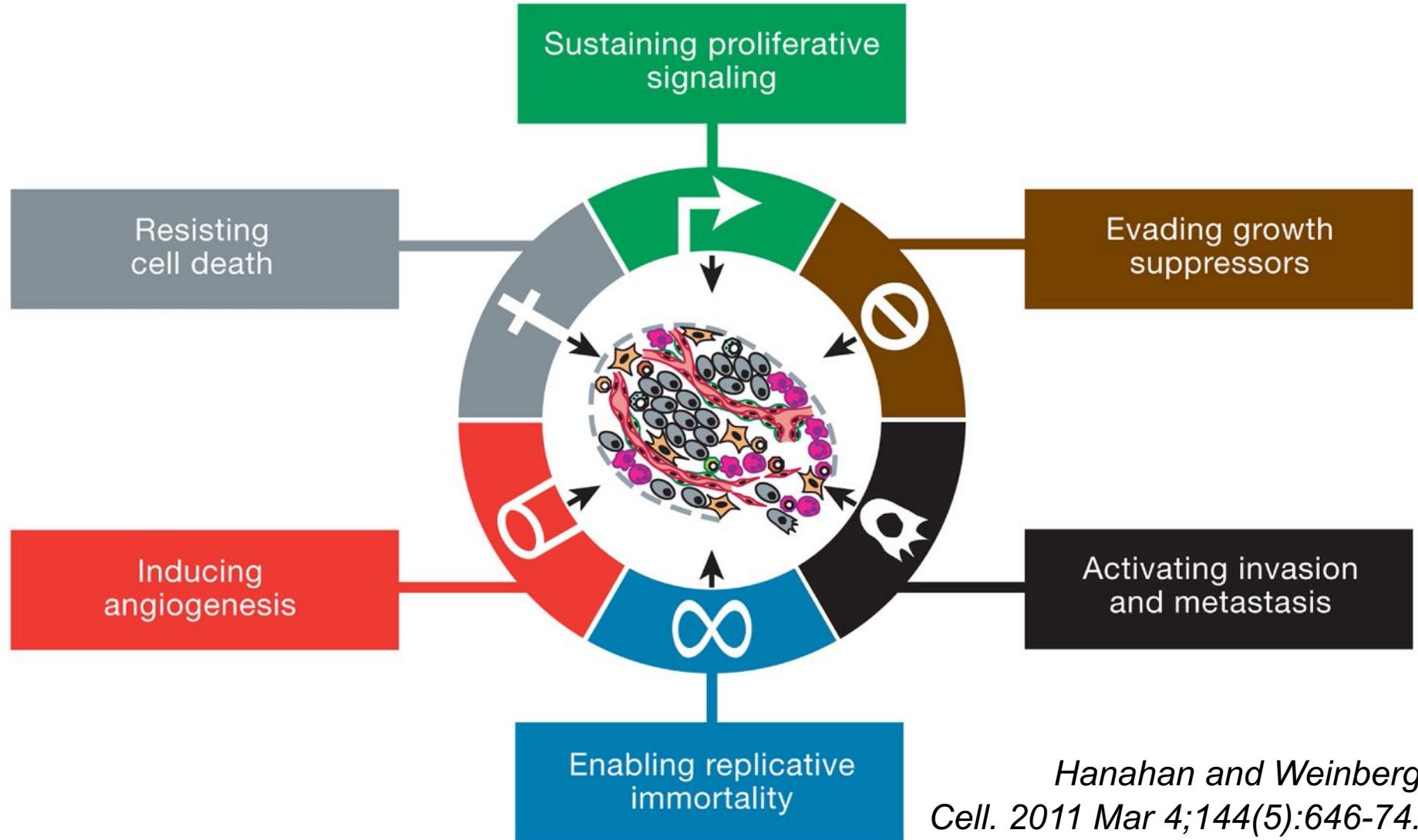
External forces (e.g. drug treatment) can select for specific clones (e.g. cells with resistance mutations)

Stratton MR, Campbell PJ, Futreal PA. *Nature*. 2009 Apr 9;458(7239):719-24. Review.

Mutation burden varies by cancer type, exposure, age of onset, & DNA repair ability

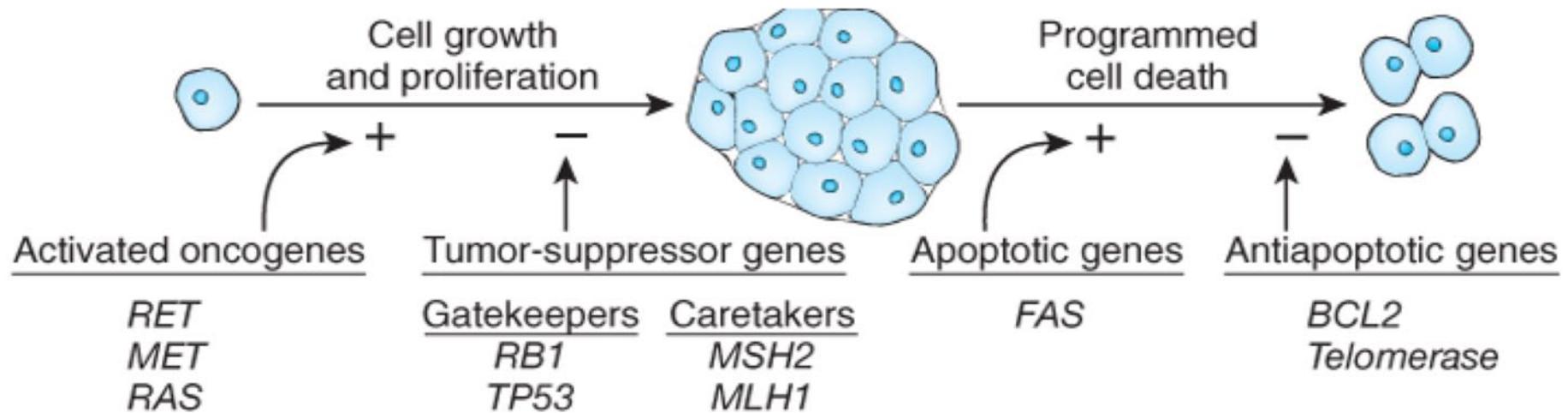


Tumour cells acquire abnormal abilities by co-opting normal cell behaviour



Hanahan and Weinberg
Cell. 2011 Mar 4;144(5):646-74.

Oncogenic somatic alterations target a core set of biological functions



© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

Differentiating “driver” from “passenger” mutations is a central challenge of cancer genome analysis

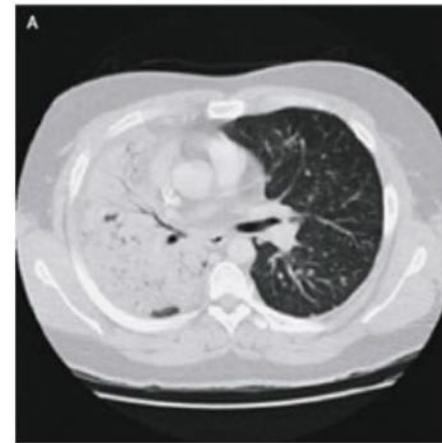
Targeted therapies exploit specific mutations

Lung adenocarcinoma with activating mutations of *EGFR*



Inhibiting EGFR shrinks tumors

Lynch et al. N Engl J Med. 2004 May 20;350(21):2129-39



Metastatic melanoma with activating mutations of *BRAF*



Inhibiting BRAF shrinks tumors

Bollag et al. Nat Rev Drug Discov. 2012 Nov;11(11):873-86

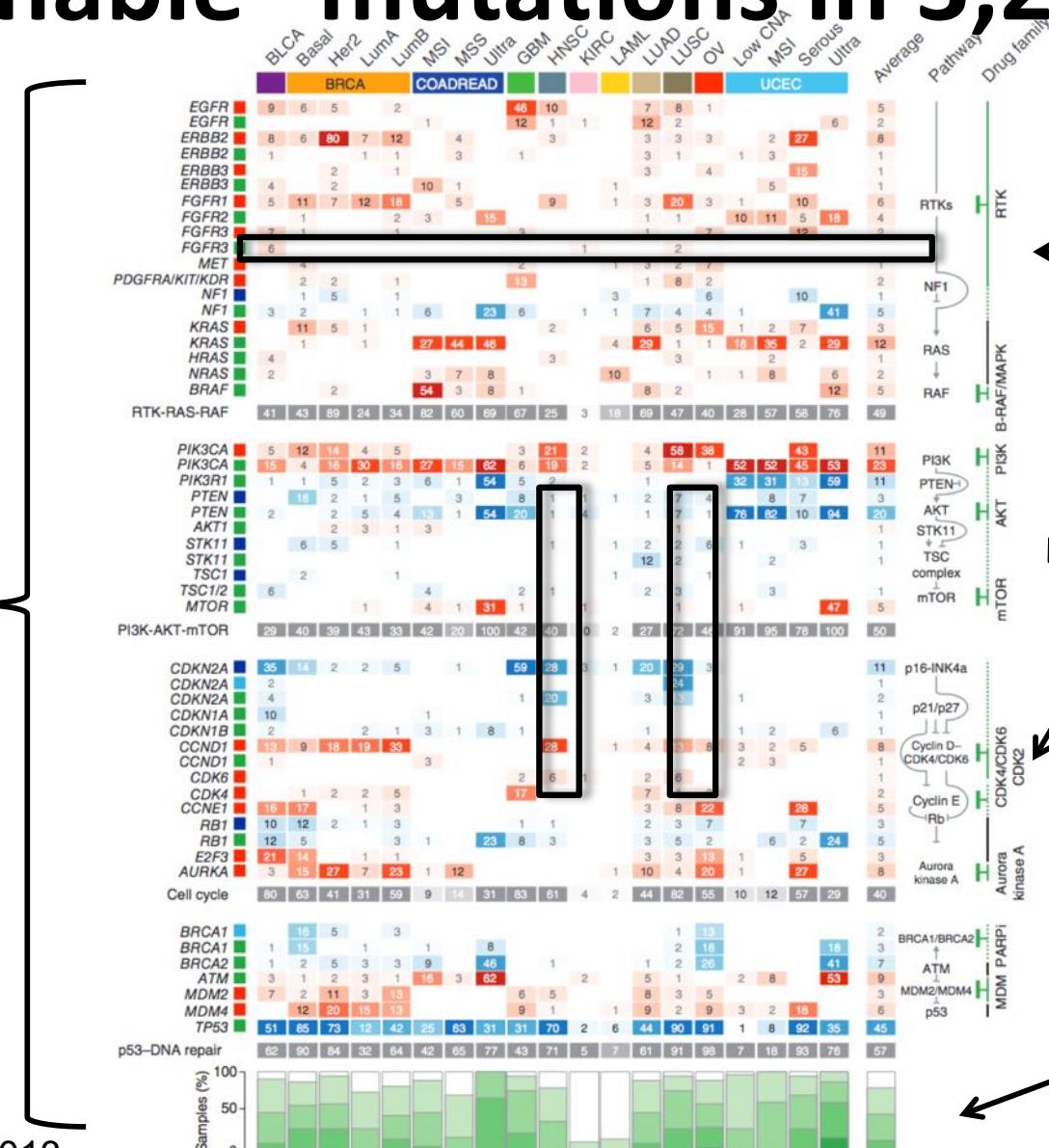


Resistance inevitably arises

“Actionable” mutations in 3,299 tumours

37 genes from
4 pathways
linked to
treatments

12 genes
disrupted by
multiple
mechanisms



Ciriello et al.
Nature Genetics 2013

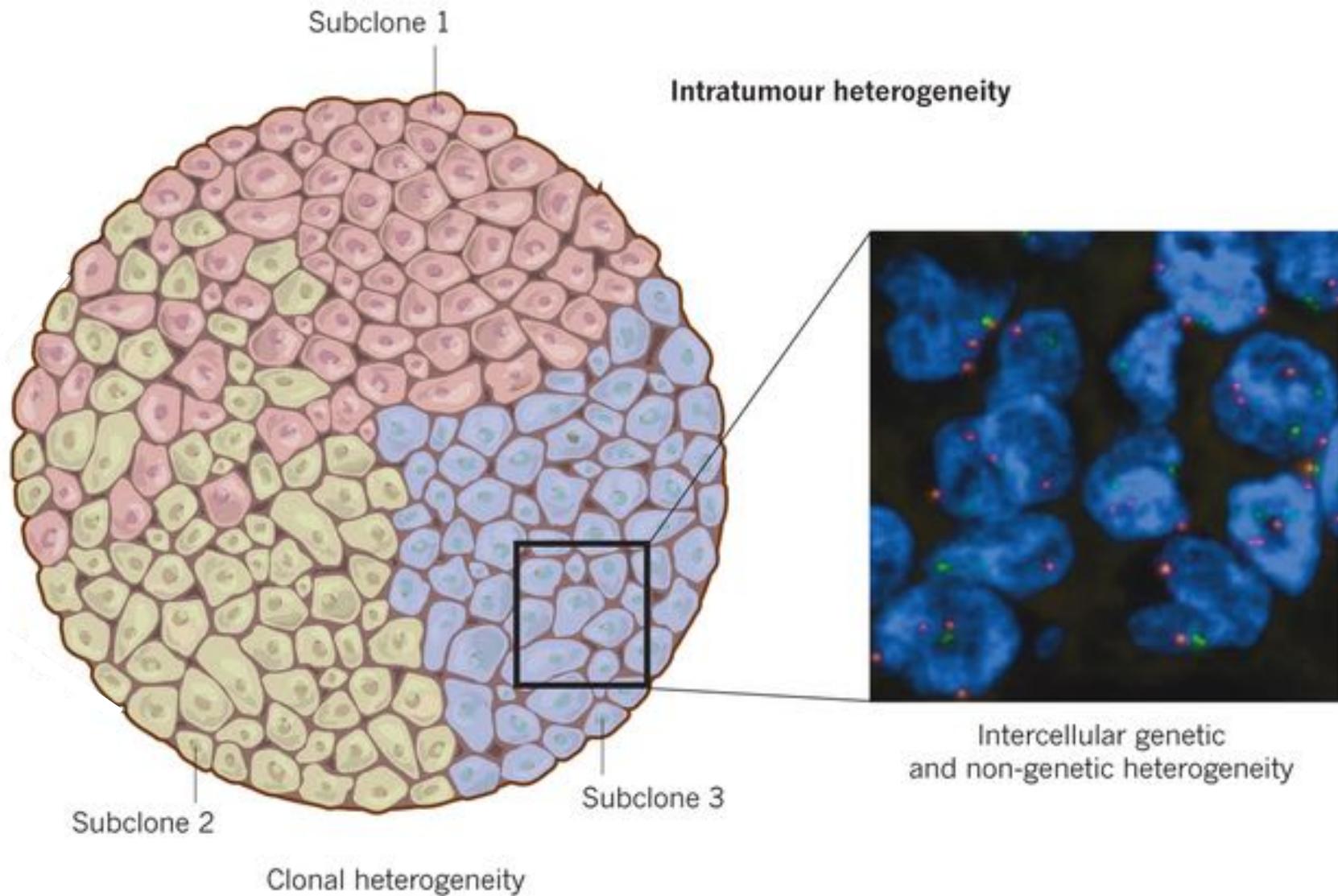
Legend: Homozygous deletion (dark blue), High-level amplification (red), Somatic mutations (green), DNA hypermethylation (light blue). Percent altered (8%). Inactivating (blue), Activating (orange). Number of pathways altered (0, 1, 2, 3, 4). Currently targetable (No, Direct, Indirect).

1) Druggable alterations cut across cancer types

2) Combination therapies may be effective in tumors with compound pathway alterations

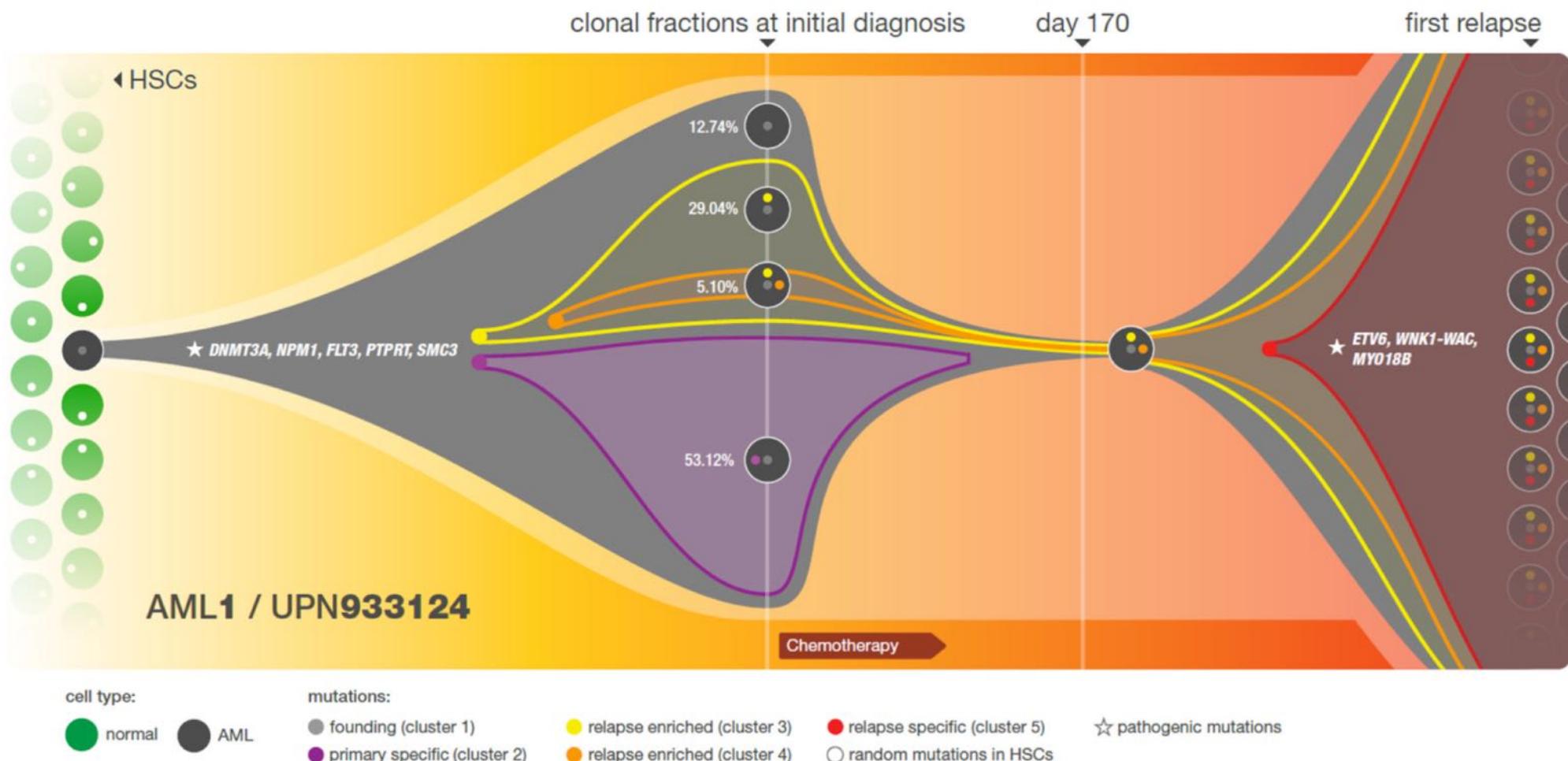
3) 50% of tumors have at least two disrupted, druggable pathways

Even within the same tumour mass, different cells may have different mutations,



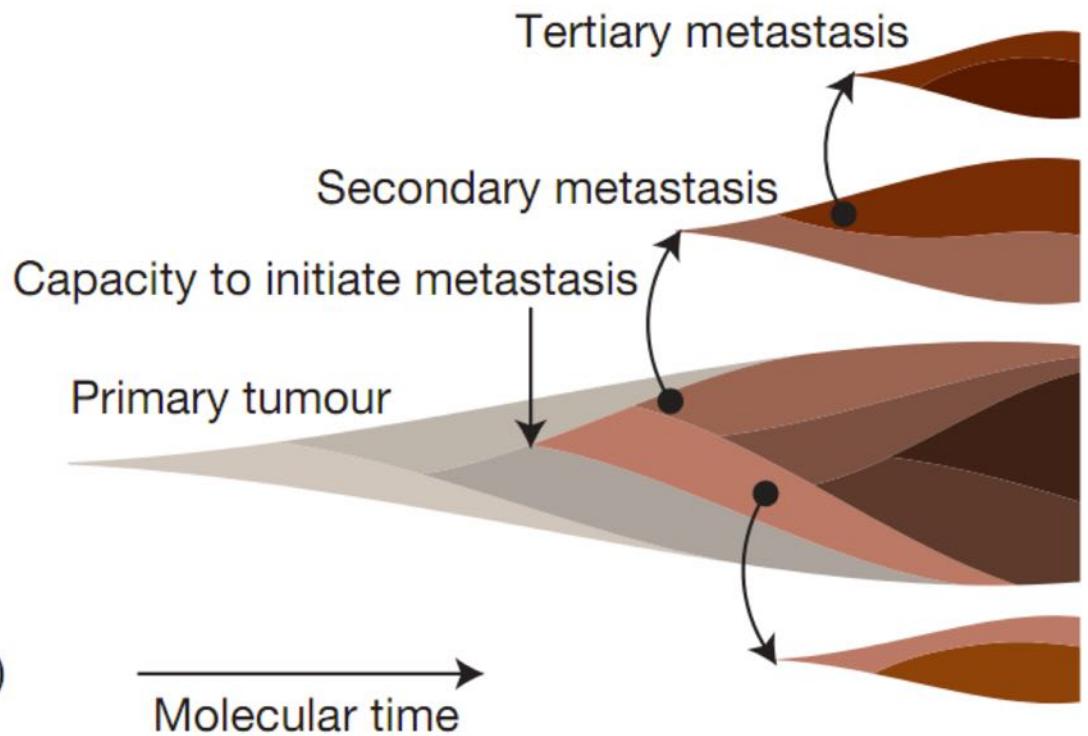
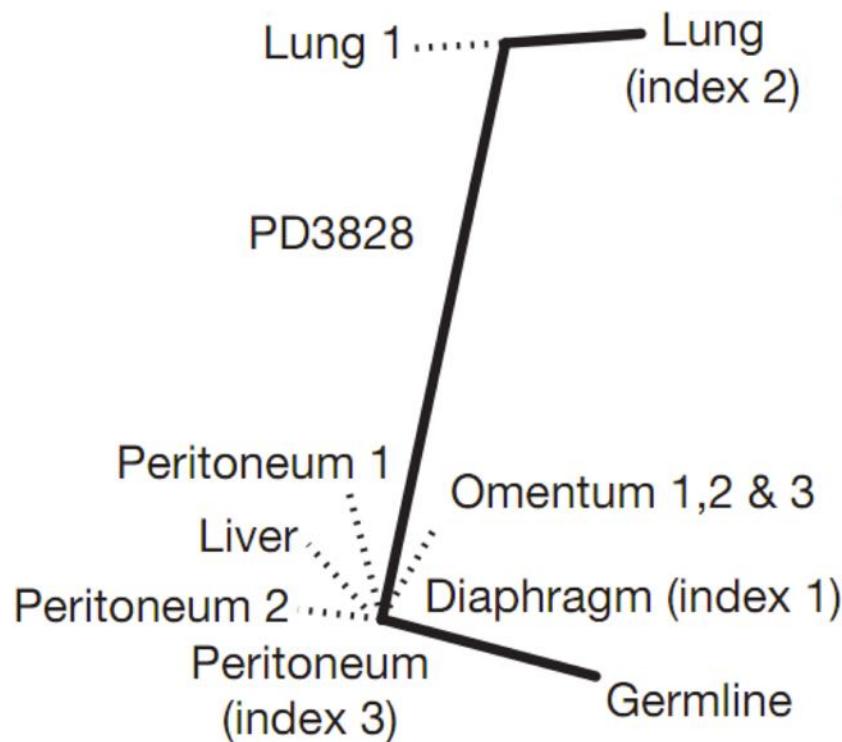
Burrell et al. Nature. 2013 Sep 19;501(7467):338-45.

Cancers are a mix of subclones that can respond differently to therapy



Ding et al. Nature. 2012 Jan 11;481(7382):506-10.

Subclones and metastases are genetically related, diverge to form subpopulations



Campbell et al. Nature. 2010 Oct 28;467(7319):1109-13.

Reasons for molecular testing for cancer

Treatment

treatment susceptibility, predict adverse side-effects, watch-and-wait versus aggressive treatment

Drug resistance and metabolism

pre-existing resistance mutations, drug dosage

Inherited cancer syndromes

additional primary tumors, at-risk family members

Prognosis

molecular subtyping, benign vs. malignant, determine unknown primary, predict metastatic potential

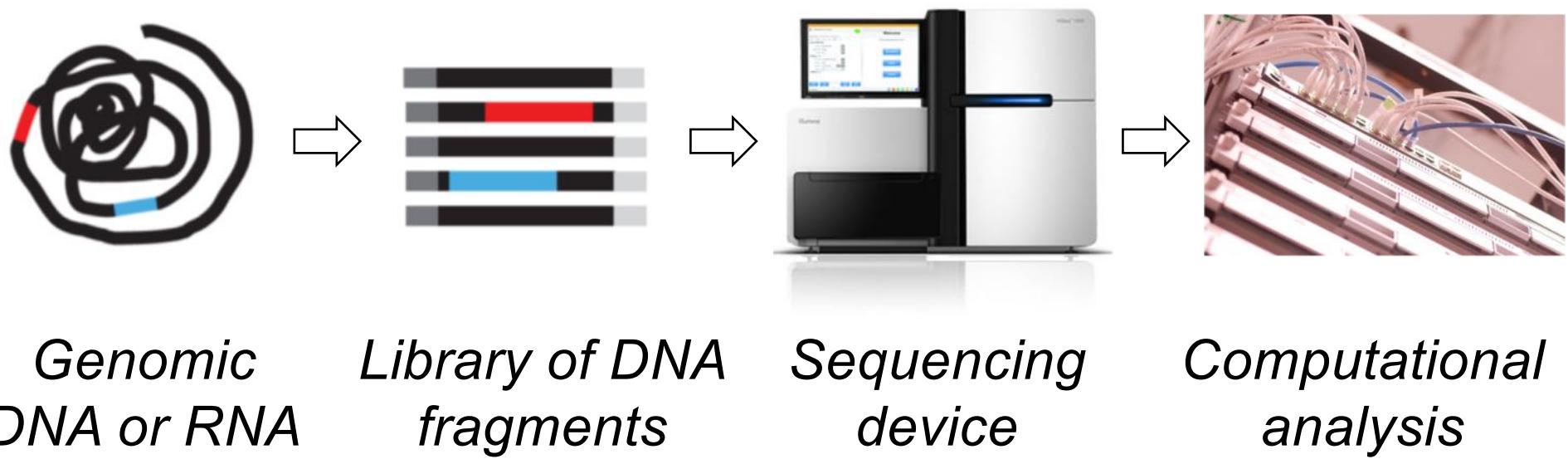
What are the targets in *my* cancer...

Sequence	mutations, polymorphisms
Structure	copy number variation, translocation, loss-of-heterozygosity
Function	expression, exon-usage
External	Viruses, bacteria

...and what can be done about them?

Applications of next-generation DNA sequencing technology to cancer

General DNA sequencing workflow



*Genomic
DNA or RNA*

*Library of DNA
fragments*

*Sequencing
device*

*Computational
analysis*

75 bp

← →
AAATGTCAAGGAGTCCTTCTGACTAGTCTGTGCCTATATGGTTGGATATTTTATGATTACCCAACCAACCACATC
AGCAGCATTGTCAGAAACTACTTGATGTTGCATTCAAGTCACAGAATTGAACACTCCCTTCACAGAGCAG
TGCTGATCCTATTGGTAAGTGCTCATACCTGCAGCATCTCACCTGTCTTCTATATTCACCTGTGACTAAA
GCCCTCTCCCTGATTCACCCCTCAGCAGCATCGTCTCCCTCACCATCTTCTACCTTCAAGCCTTCCCTCT
AGAAAACCTACCTGTCACTTTAACATTATCGATCCACTGGATCTCATGGACTCAAATGTCCATAAGCCG
GCTTATAATCTAGCCATGCTGGCAGCTGATTAGATTGTGCCACTCAGATTAAGAGTCTGCCTTCCCAGCCCAC
CTGGGGATGGGATTAGGAAAACCTCCCTAACGGGTTTGAAATGCAGTCCCTAACCTCTTAAGCATCTCC
TTGTTCTGAGGATAGGTTACATAGATGATTCTAATCCATTATTATTTACATTCCAGGCTACCTAAAAGAAG
GCACTTGGAGGCCAACGGTTGGATCATGAAGTCAGGAGTTGGAGACCACCTGACCAGCATGGTAAACCC
GTGTGTCACAAATCTTGTACAGAAAAGGCCGGAGATTCCACCTGTTAACCAAAGGGCTGTATCACTACA
ATGACCTCATCTAACATTGACATGCAATGGCCCTGTTCTAAATAGGTACACTCTGAGGTACTGGAGATT
GATTCCATTGATGTTGATTCCATTGAGTCCATTGATCATTACATTGATTCTATTGATGATTCCATT
GCAATACGGACTCAGTAGGAAAGCAACTGAAGGGTGTCAAAATTATGCTGCAGATTGCGAATGAGCTACTGT
TTGGTCTCCTCAAATTACTTTATTCTCACCTCAGGATTGATATGTTTACTTTATATAACTTAA
TAACTTTGATACAATGAAGTAAGTTCTATTTTTTTACCACTCTAACATATAAGACCTGGAAATAC
AGTGCCAGGAGCTACCAGGGCTGTGCCAACGGAGAAAATGTATCAACACCTTGGAGTTCCAGCGCCGCTGTC
AGGATAAAAGCAAGATGTAAGCTCAAATTAGTCCAACCTGCTACTTTAGTCACTTATTAAAGCATTGGGA
ATGCTATGTTATTGCGGCACTATTACAATAGCACAGACTGGAACCAACCCAAATGATAGACTGGACTAAGAA
AGCCCTGTCCCTCCTCACGGTTCTCAAGCATGTCCTGCCTGCACCTACCTCAGAGAGCCAACATTGACT
AAATTCAAAAAACAAACTCTGGTCCAGTTCAAGTATAATTCTAACAGGTACAGCCTCGGACTGTCCA
TTGTTGAGCTCCTATATATTCTGGAATTAAATCCCTGTCACCGAGTAGTTGCAAGTATTCTTCGCTTC
TGGGGACTAAATGTAATGCTGGAAAAATTCTTTGCAGAGGCATATCCACCAACCTTCTCCAGACCTAAC
AGAATTCCCTTGGTTCCCTTCAGGAAGTAGATTGAAGAAATGAAGAATCTGACTTCAGTGTATTGTC
ATGAACATTCCCTATCTATTCAAAATTACATGAATCACAATAAGAATAACACAATTCTGAACAATGTGTGAA
ATGAAAATAATAATATCTTTAAAGTCCCAGAGGTGCCAACAGGTAGCTGTCCCAGAACAGAAGGTGCCTGAAGCT

25 reads

On an Illumina HiSeq 2500, one lane of an eight-lane flow cell generates >600 million reads

“Pipeline”

Variants



www.hickerphoto.com

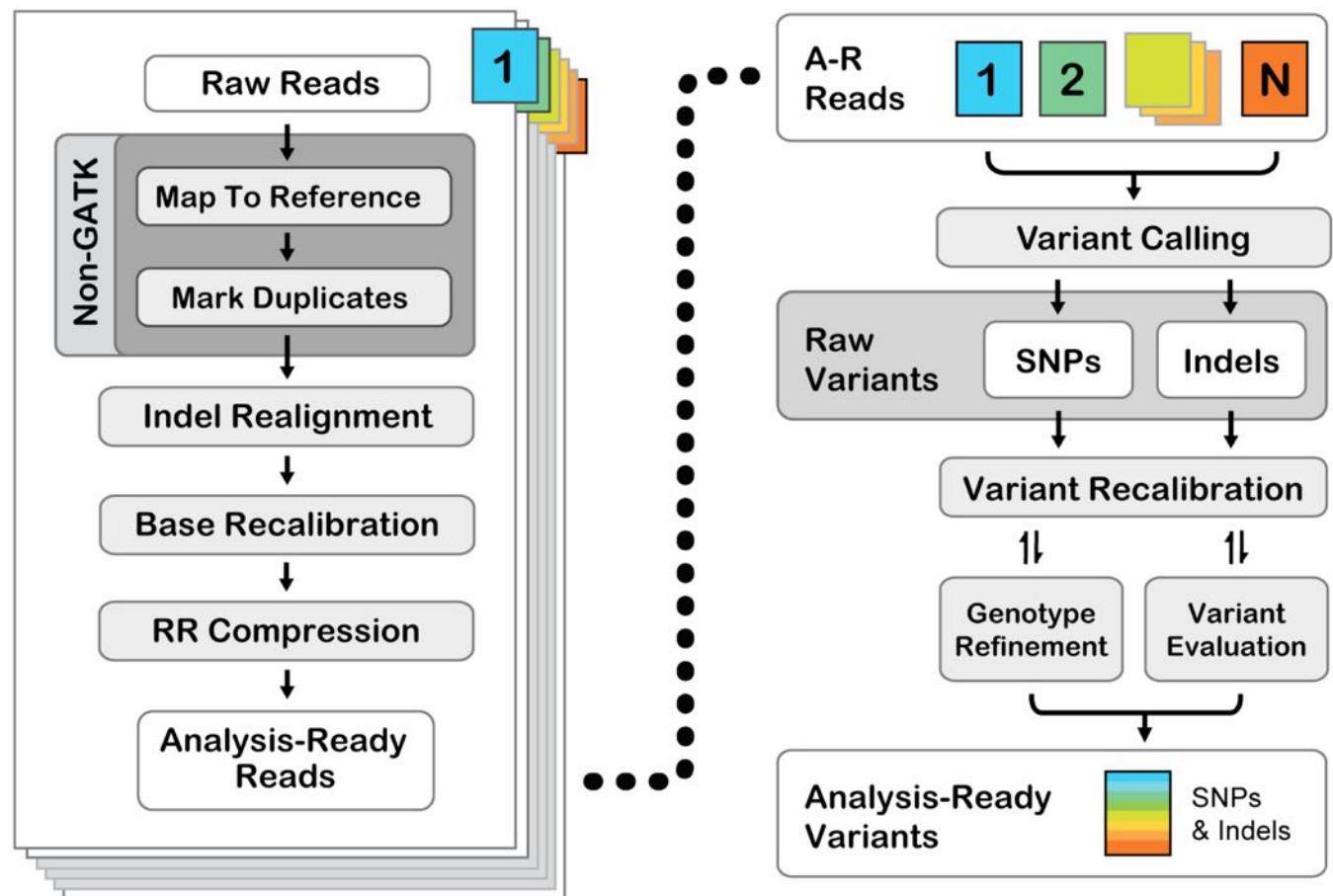
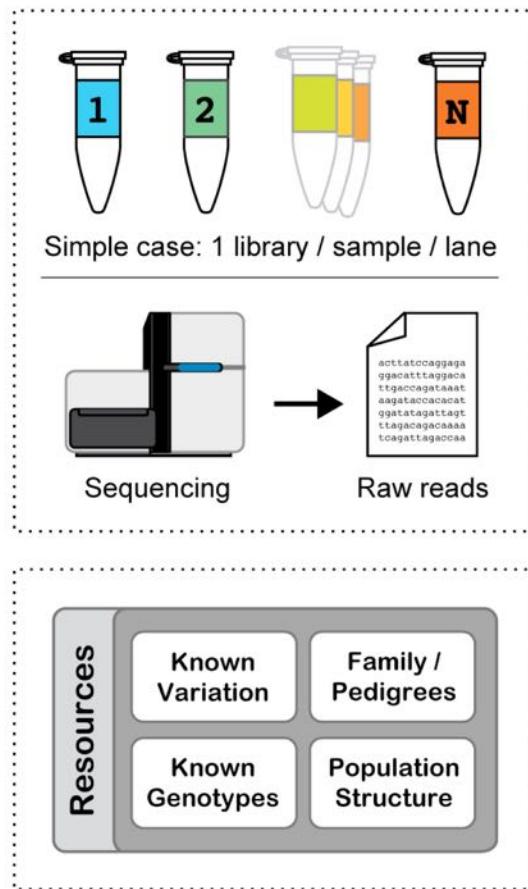
“Pipeline”

- 1) Alignment
- 2) Pre-processing
- 3) QC metrics
- 4) Variant calling
- 5) Interpretation
- 6) Clinical report



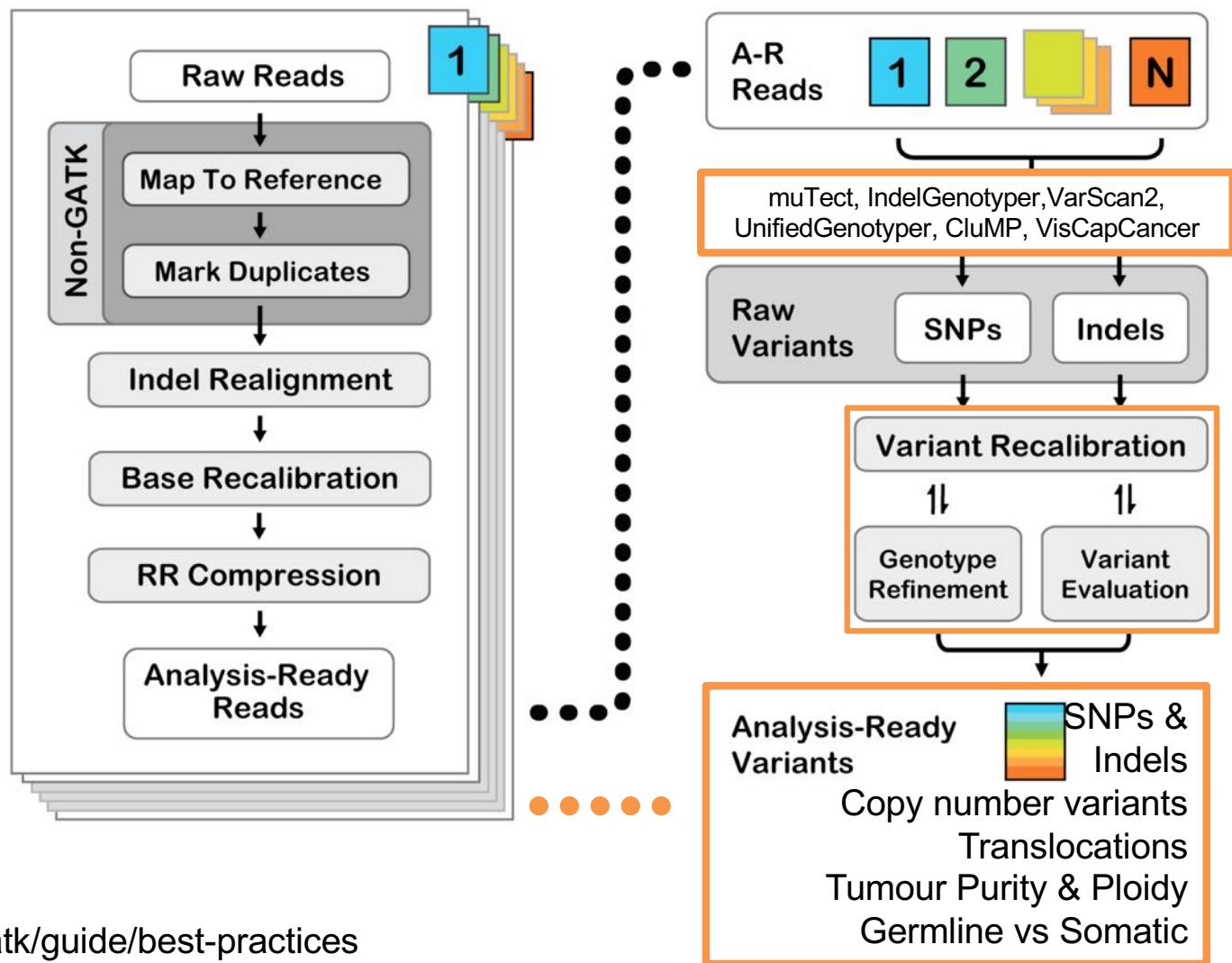
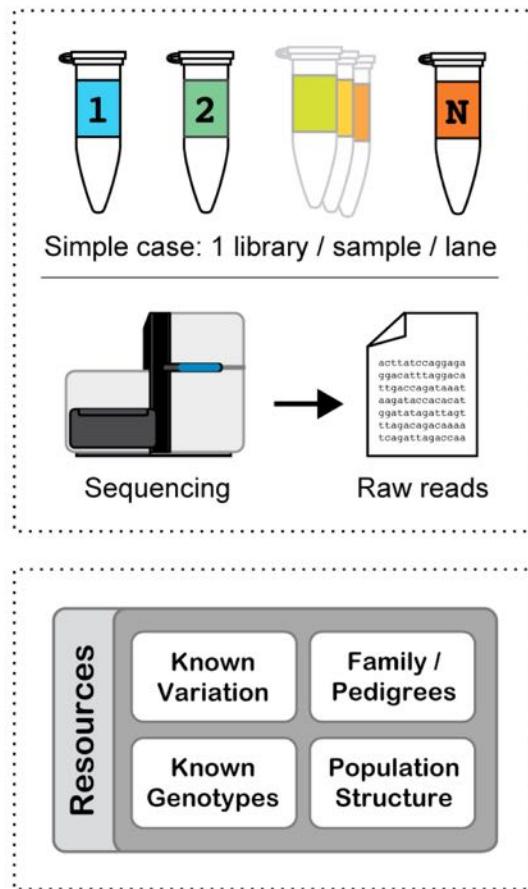
www.vnf.com

Adapting Genome Analysis Toolkit Best Practices for analysis of cancer genomics data



<http://www.broadinstitute.org/gatk/guide/best-practices>
<http://picard.sourceforge.net/>

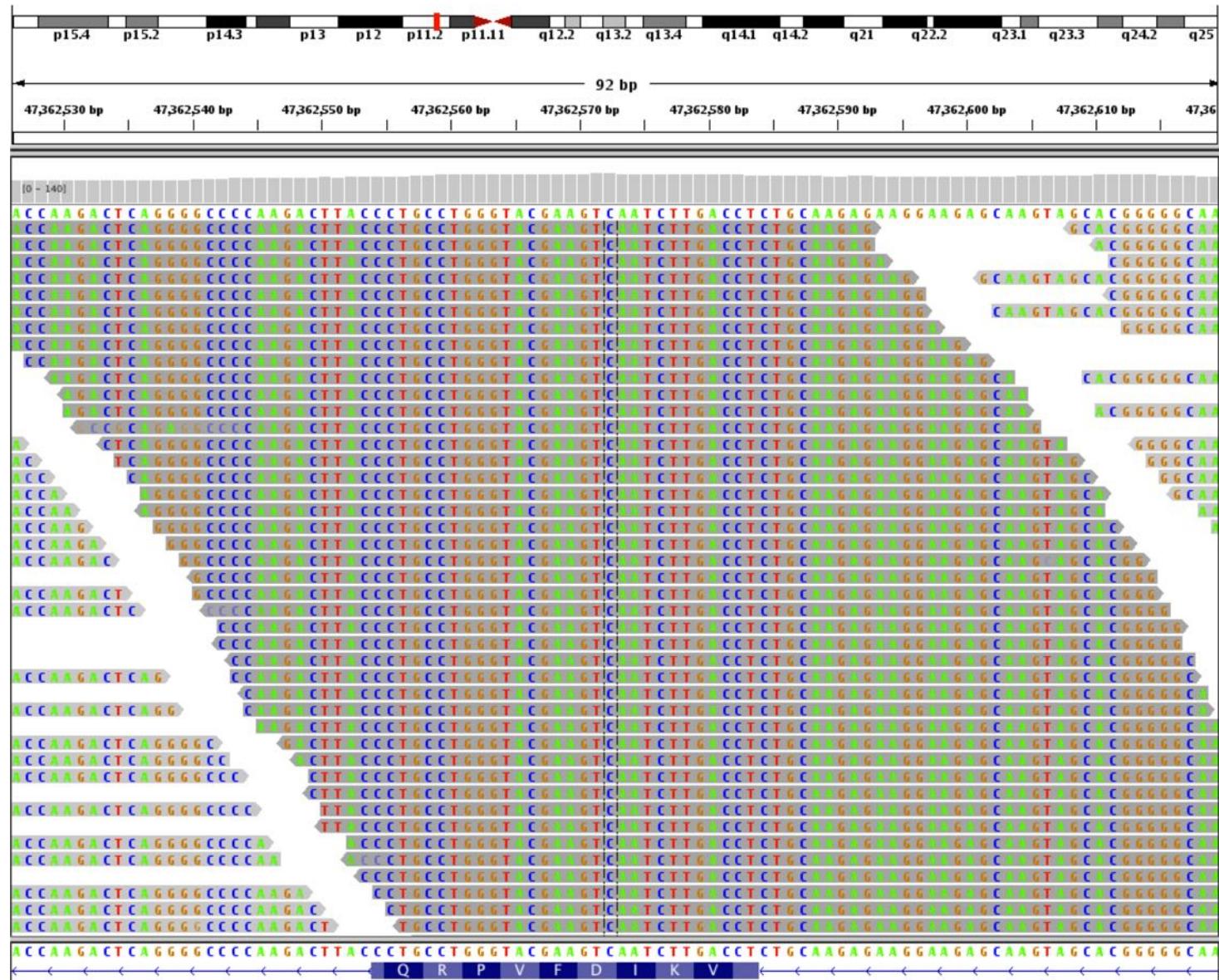
Adapting Genome Analysis Toolkit Best Practices for analysis of cancer genomics data



<http://www.broadinstitute.org/gatk/guide/best-practices>

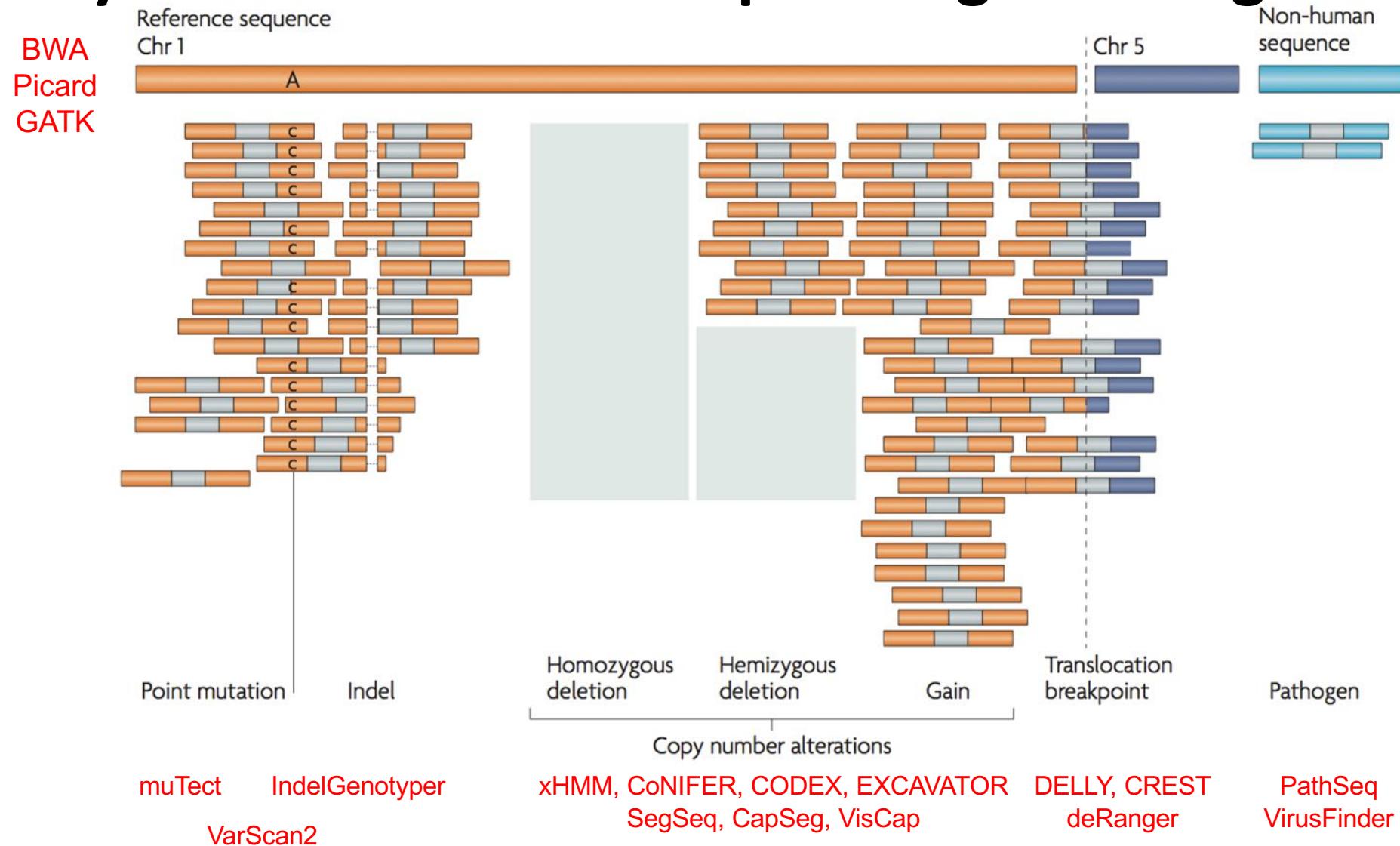
<http://picard.sourceforge.net/>

An alignment of reads to a human genome reference



Bwa, Novoalign, Stampy, SOAP, and >50 more listed on Wikipedia under “Short-Read Sequence Alignment”

Multiple types of cancer genome variation may be inferred from sequencing read alignments



Meyerson M, Gabriel S, Getz G. Nat Rev Genet. 2010 Oct;11(10):685-96. Review.

DNA sequencing approaches to cancer

Whole genome sequencing (WGS)

Whole exome sequencing (WES)

Targeted gene sequencing

Targeted variant genotyping

Epigenome modification (bisulphite)



DNA

Transcriptome sequencing (RNAseq)

miRNA sequencing



RNA

Protein/DNA interaction mapping

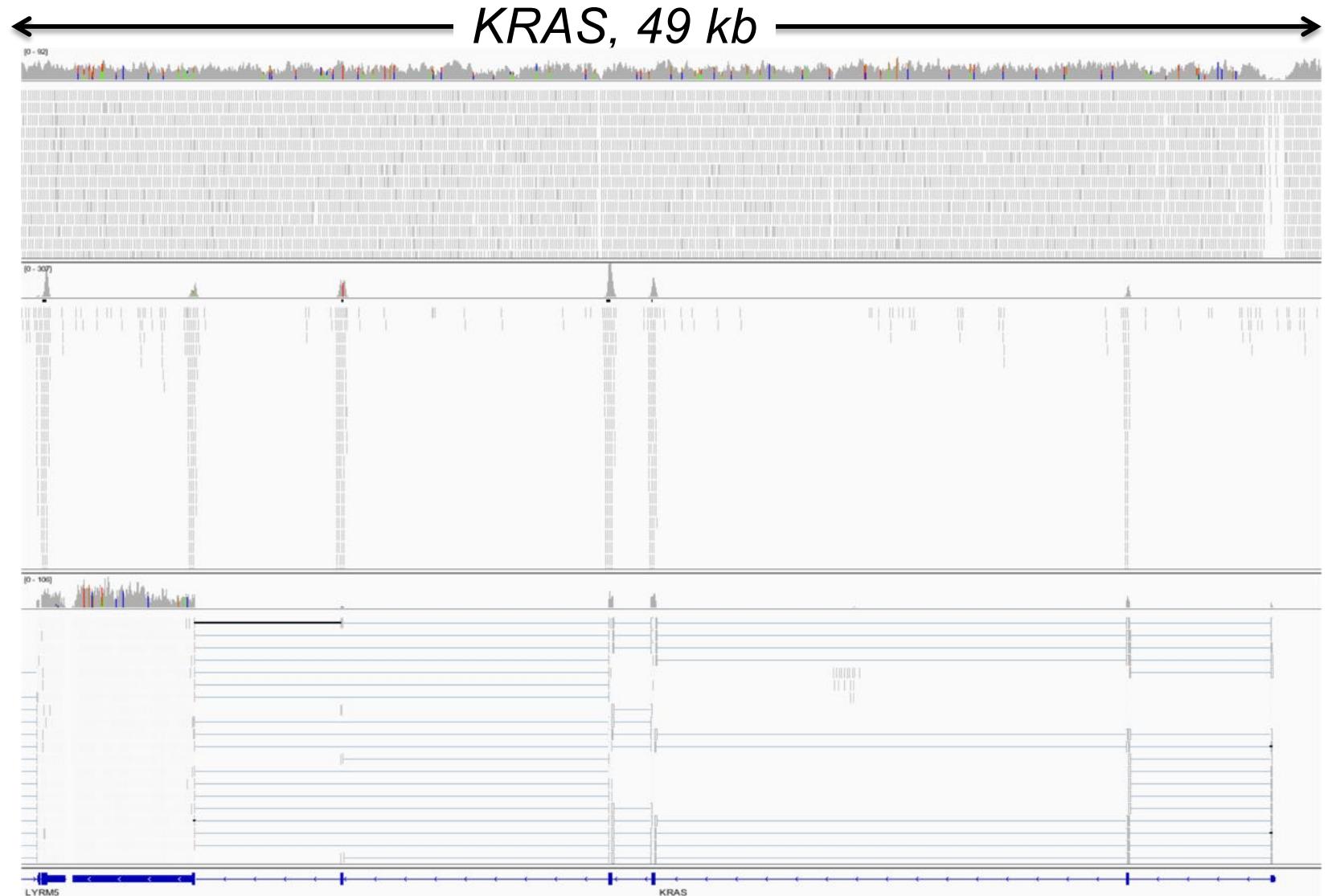
Epigenome mapping (histones)



Protein
(DNA footprint)

Coverage of genome, exome, & RNAseq data

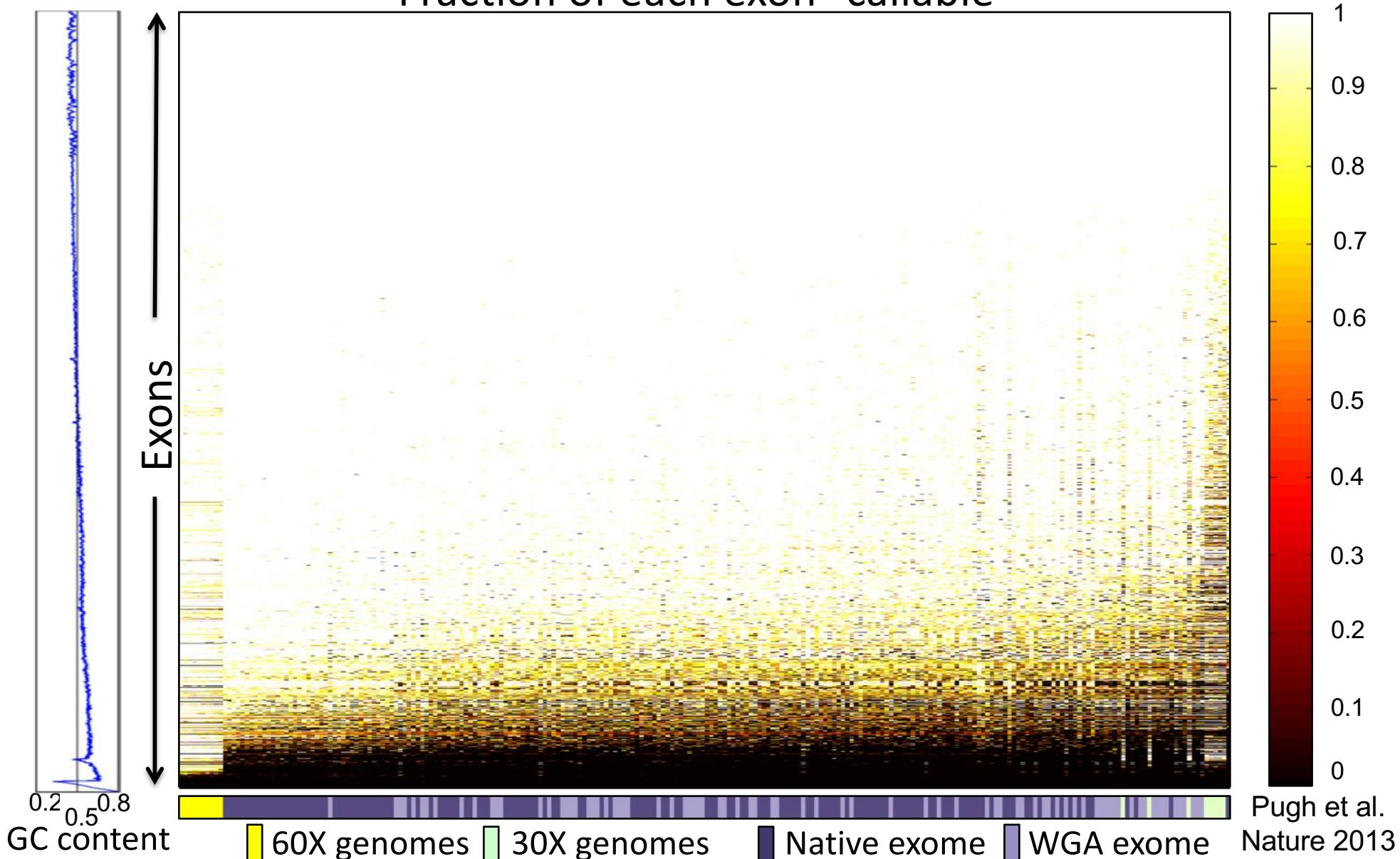
Genome
~40X
(0-92X)



www.broadinstitute.org/igv

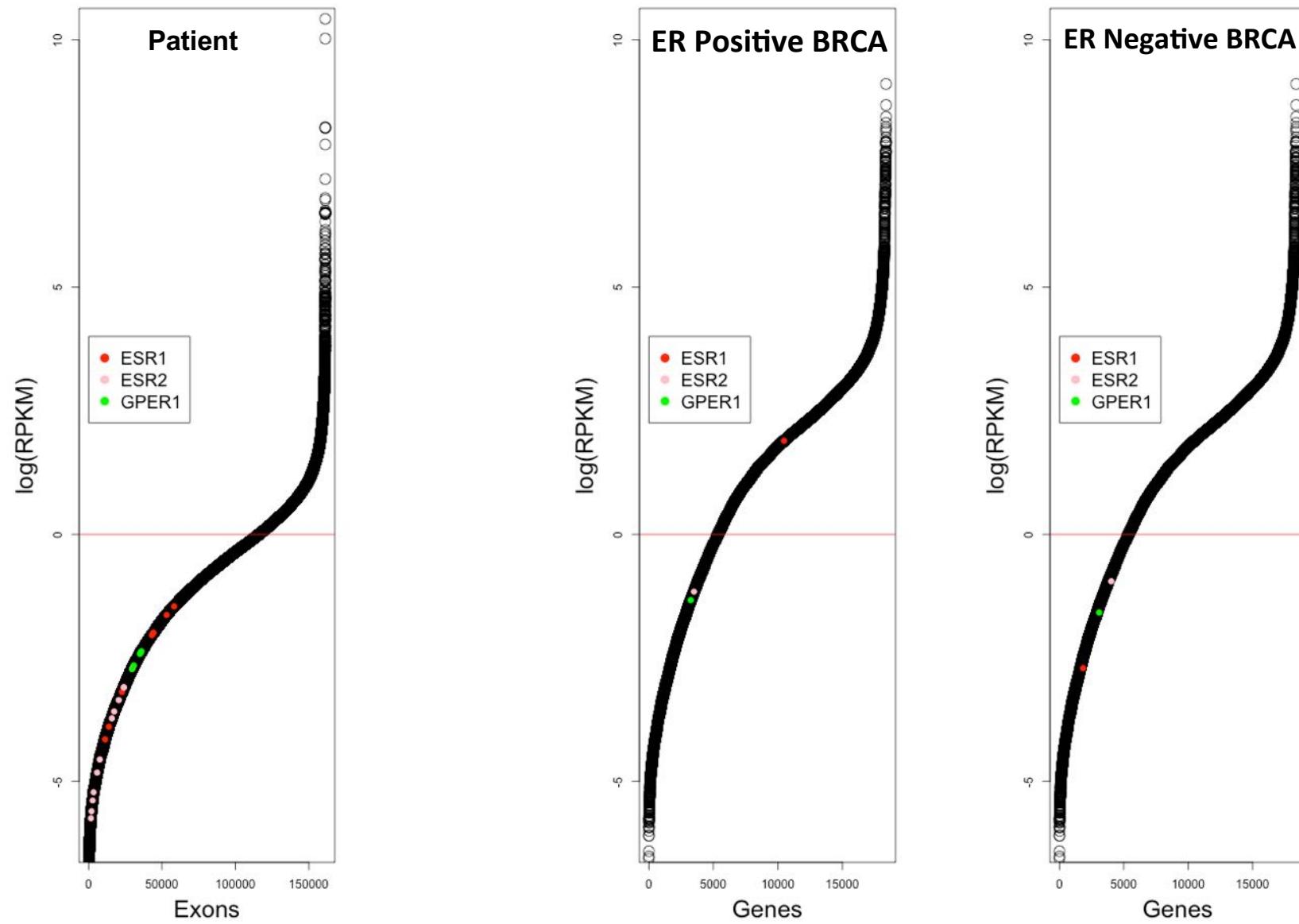
Exomes & Genomes don't measure everything

Fraction of each exon “callable”



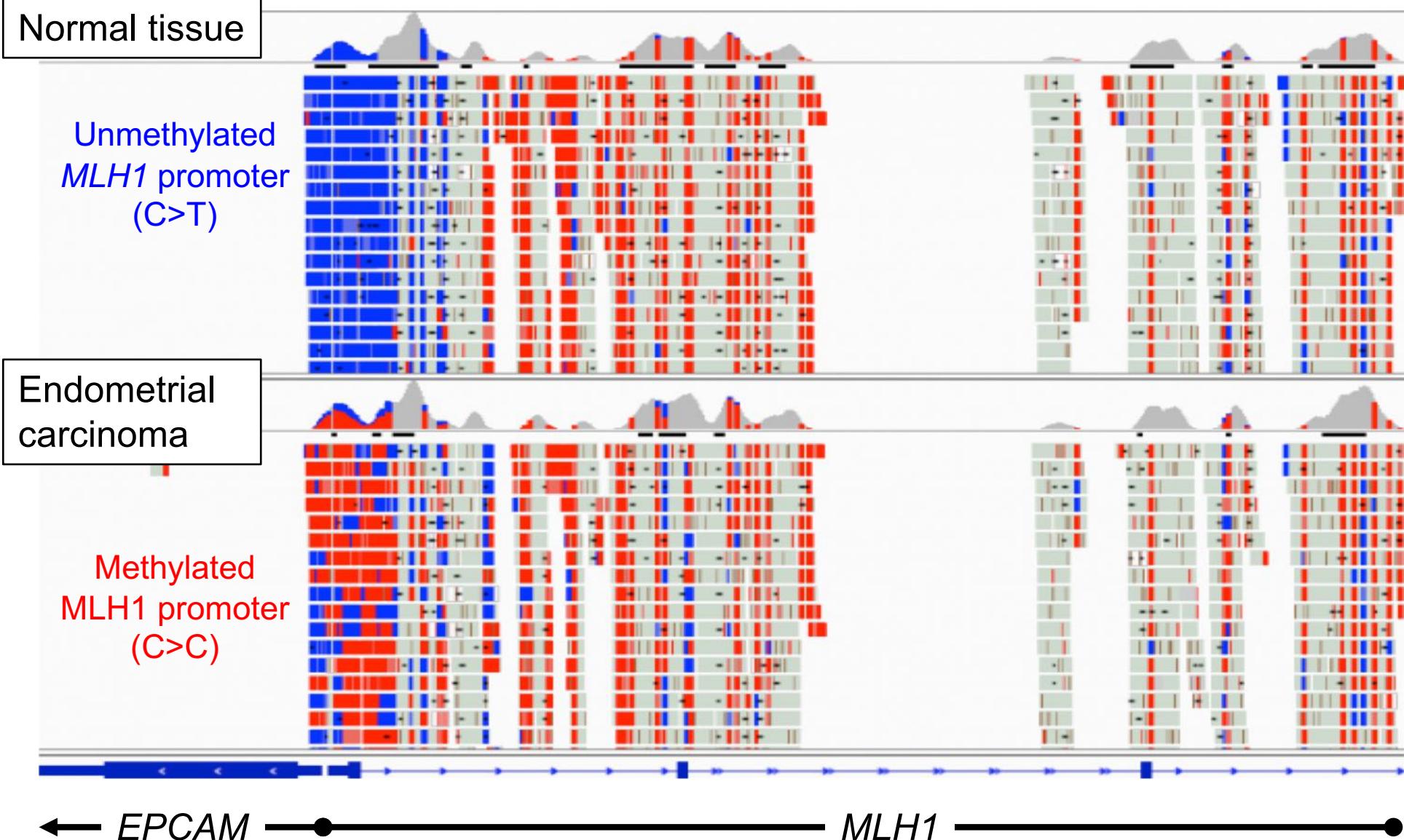
Transcriptome coverage depends on expression

Expression level



Genes ordered by rank expression

Epigenetic modifications that regulate gene expression detectable by bisulphite sequencing, even in admixed tissues



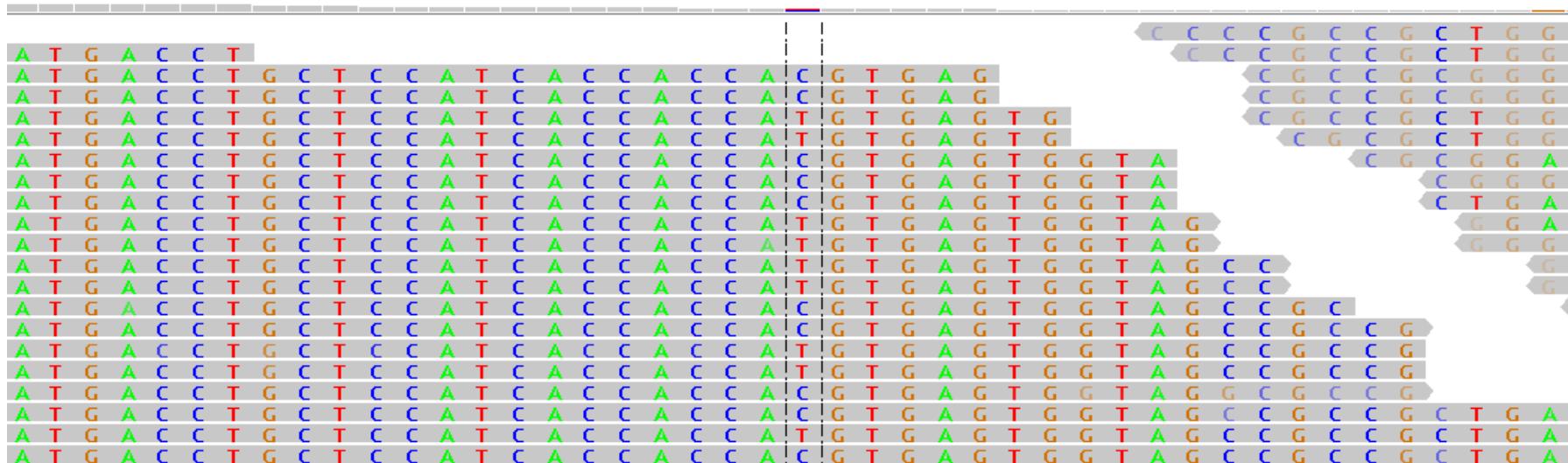
Data types are confirmatory and complementary

	Genome	Exome	RNAseq	MethylSeq
Sequence mutations				
Copy number				
Rearrangements				
Pathogens				
Purity/Ploidy				
Low purity or subclonal mutations				
Gene expression				
Allele- or exon- specific expression				

Somatic Mutations

Germline variants are detectable at low coverage

|D - 394|



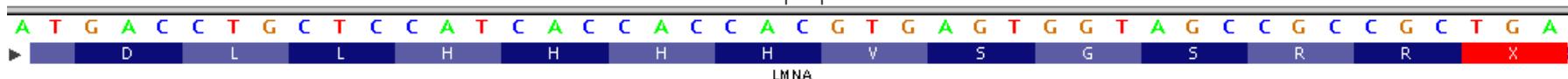
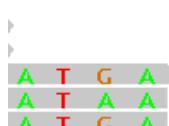
Heterozygous

Total coverage = 19

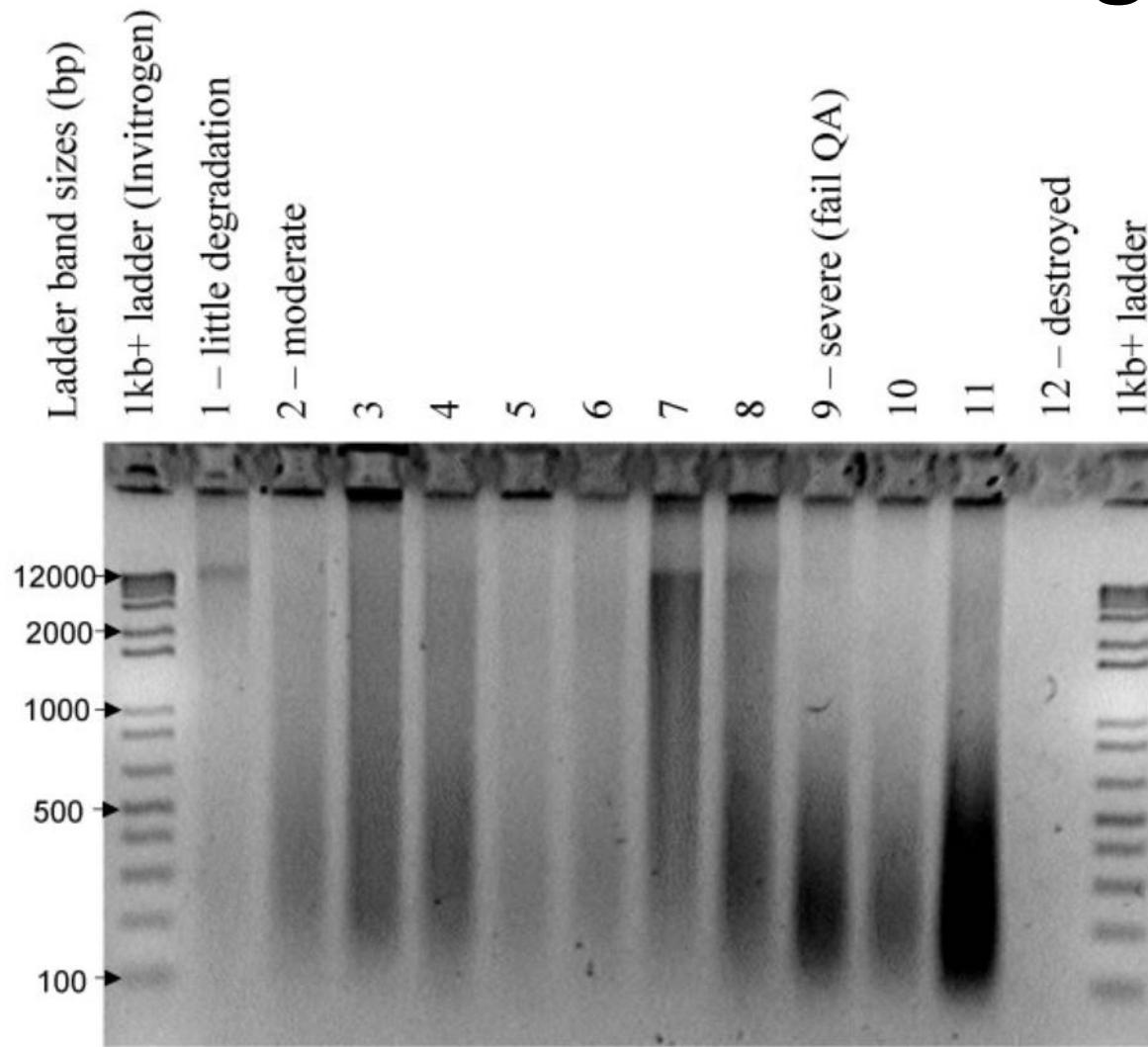
C = 10

T = 9

Allele balance = 0.47



Variable DNA quality & quantity from tissues used for routine diagnosis



Pugh *et al.* BMC Cancer. 2007 Jul 13;7:128.

Tumors are a mix of cancer & normal cells ("purity" or "tumour content")



Pugh *et al.* BMC Cancer. 2007 Jul 13;7:128.

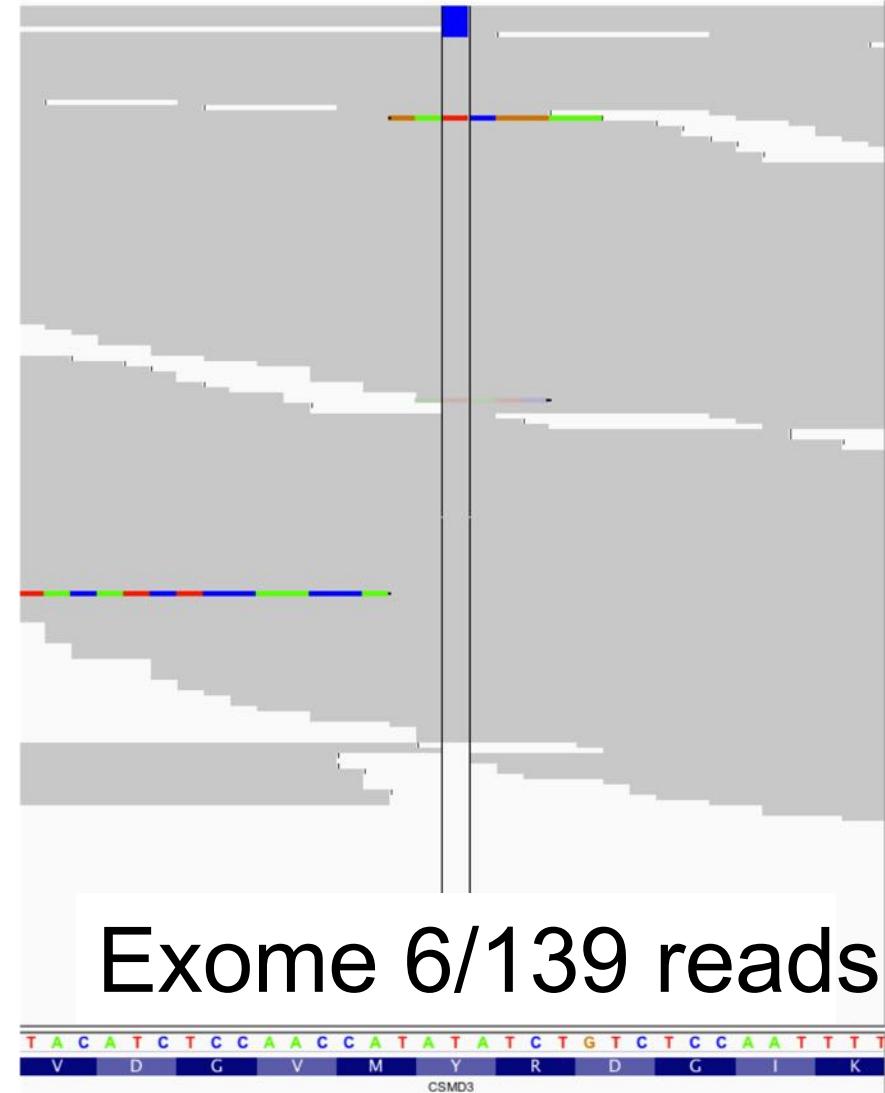
Tumors can have multiple genome copies (“ploidy”)



Glioma karyogram (GTG-banding). 78,<4n>,XXXX,-2,-5,-6,del(6)(q21q23)x2,+7,-8,del(8)(q22q24.1),del(9)(p10)x2,-10,-10,-11, -11,-12,-13,-13,-14,-14,-16,-19,del(19)(p10),-21,+22

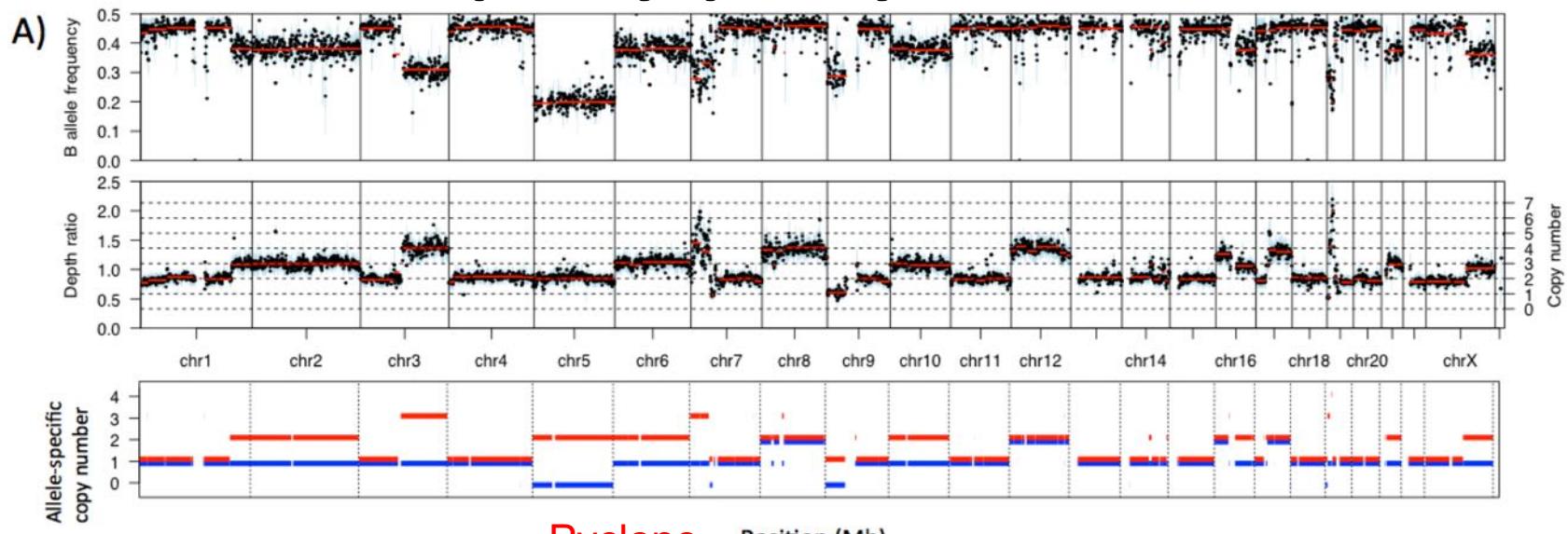
www.cityofhope.org/research/support/cytogenetics/

Deep coverage is necessary to detect mutations in low purity or high ploidy tumors

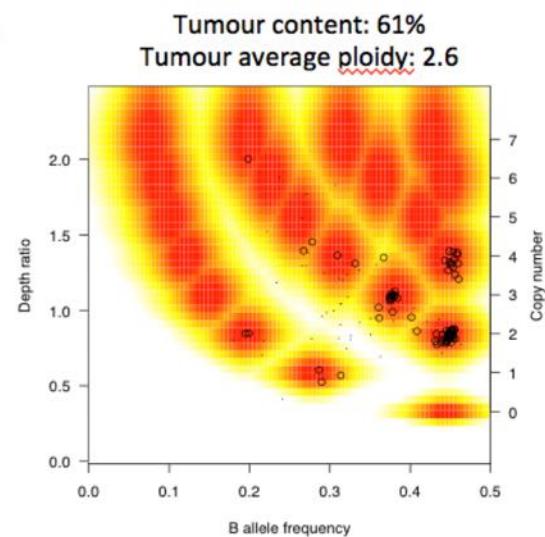


Deep (e.g. 250X) DNA sequencing coverage enables inference of tumour purity, ploidy, & subclonal structure

Sequenza



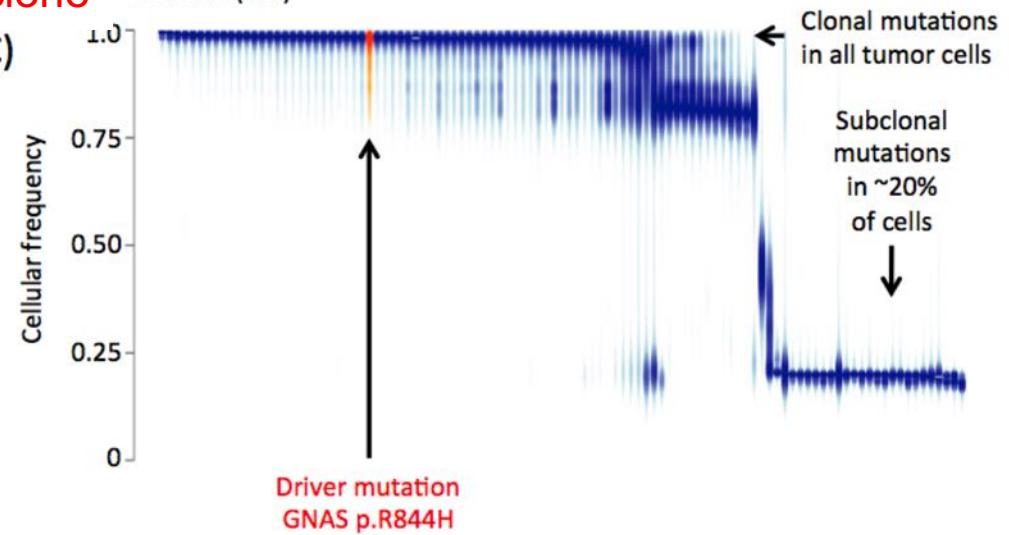
B)



Pyclone

Position (Mb)

C)



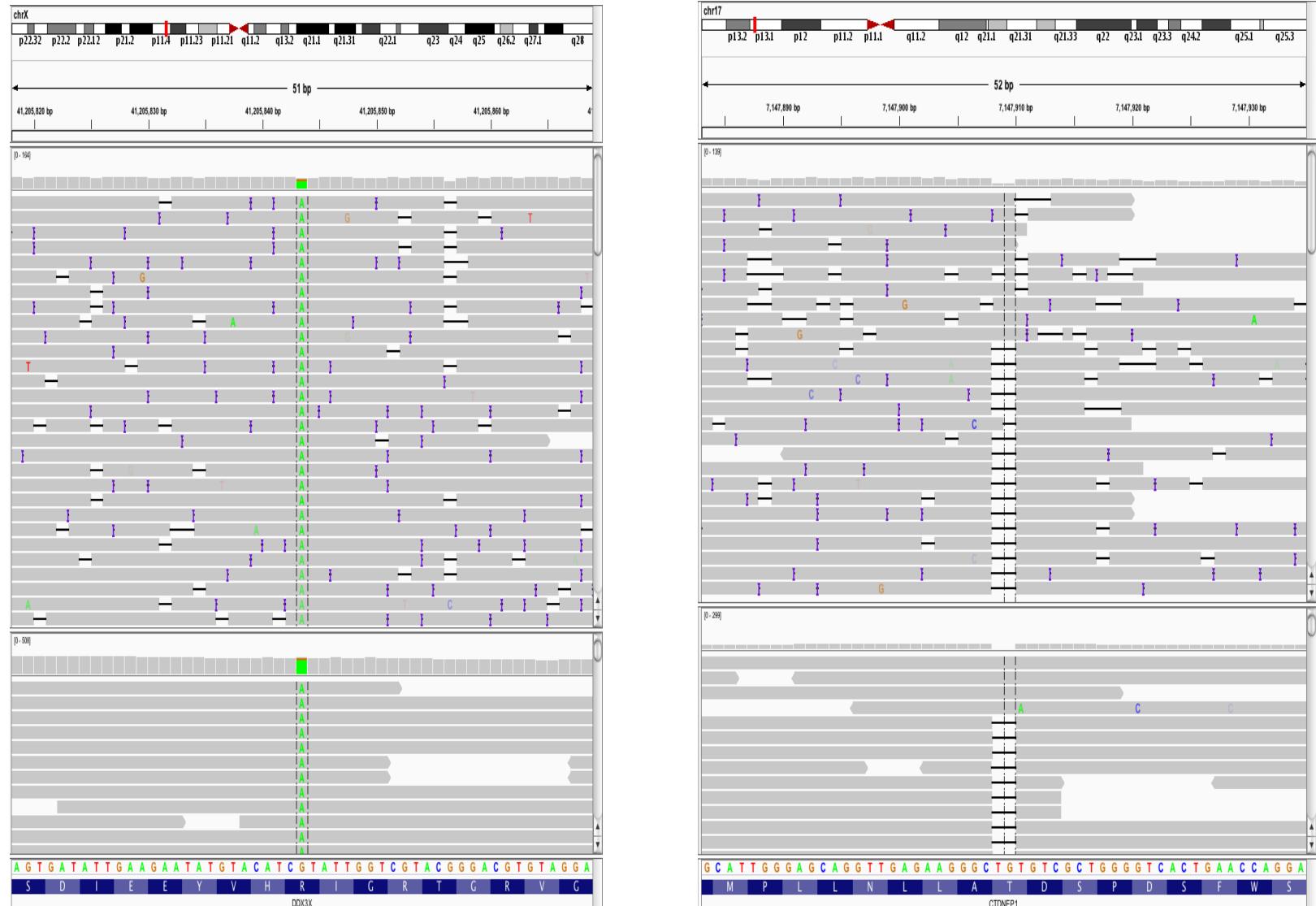
Favero et al. Ann Oncol. 2015 Jan;26(1):64-70.

Roth et al. Nat Methods. 2014 Apr;11(4):396-8.

Multiple DNA sequencing technologies enable mutation confirmation & cross-validation

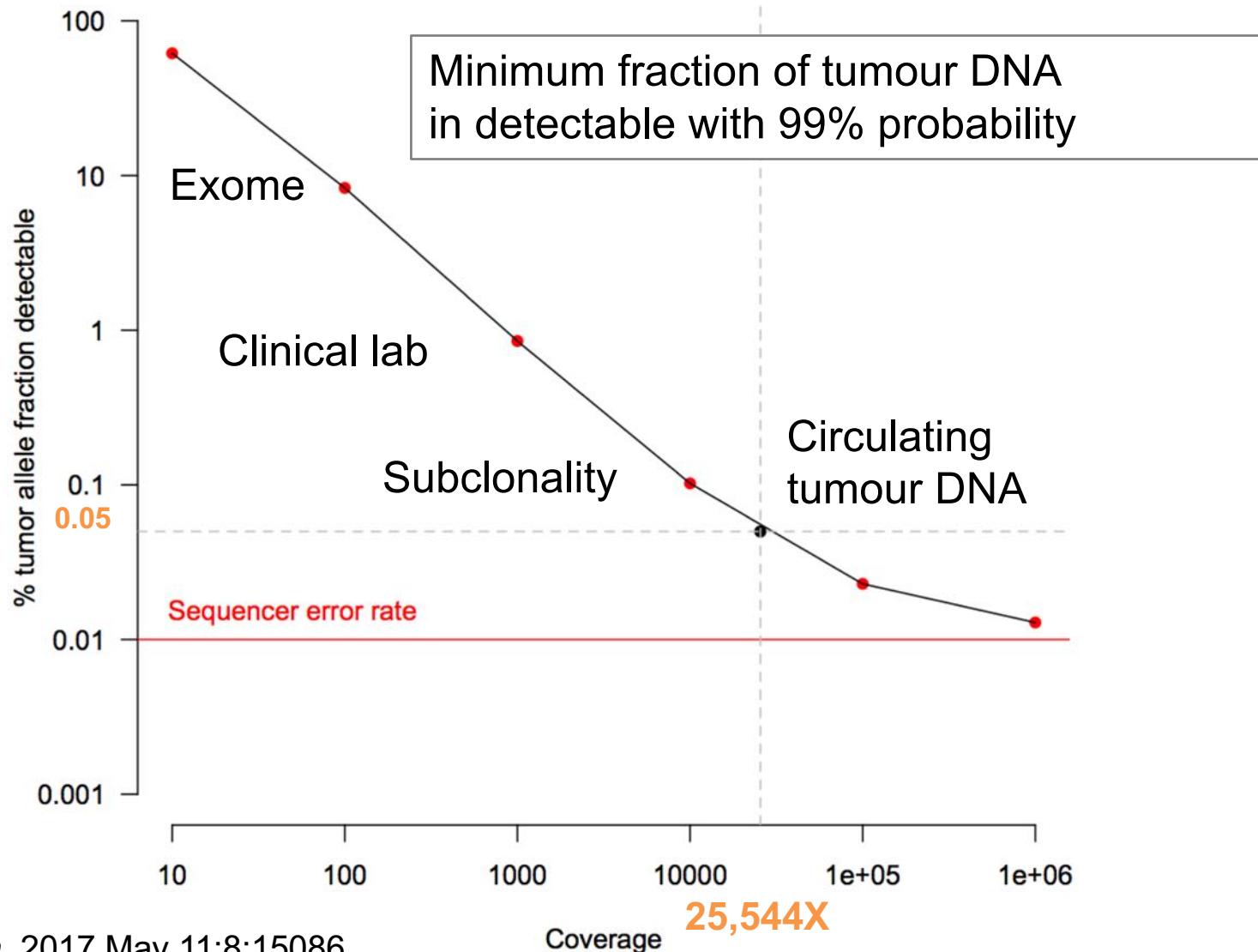
PCR/
PacBio

Hybrid
capture/
Illumina



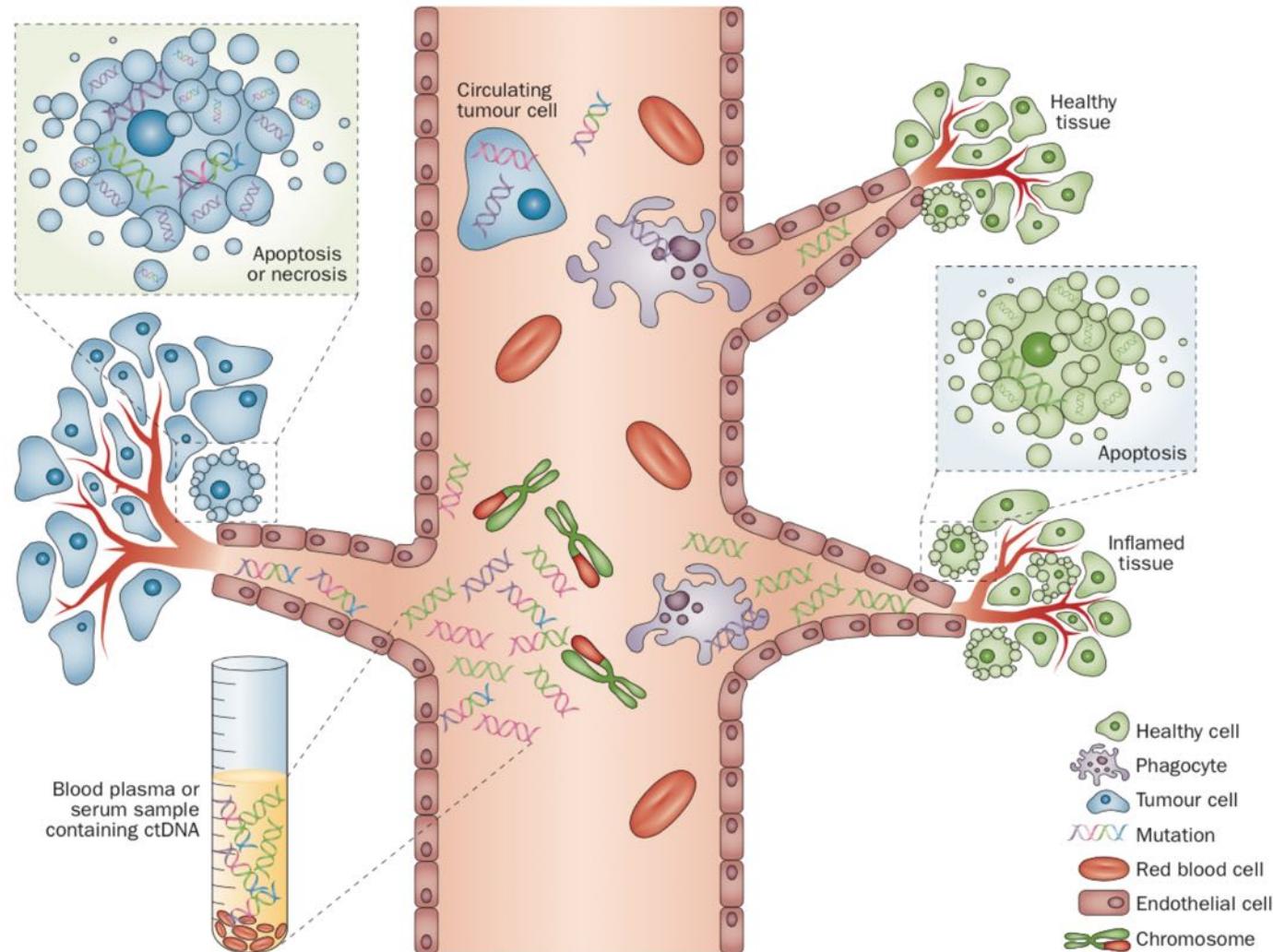
Pugh et al. Nature 2012.

Mutation detection sensitivity is dictated by sequencing depth and is limited by sequencer & polymerase error



Kis et al. *Nat Commun.* 2017 May 11;8:15086.

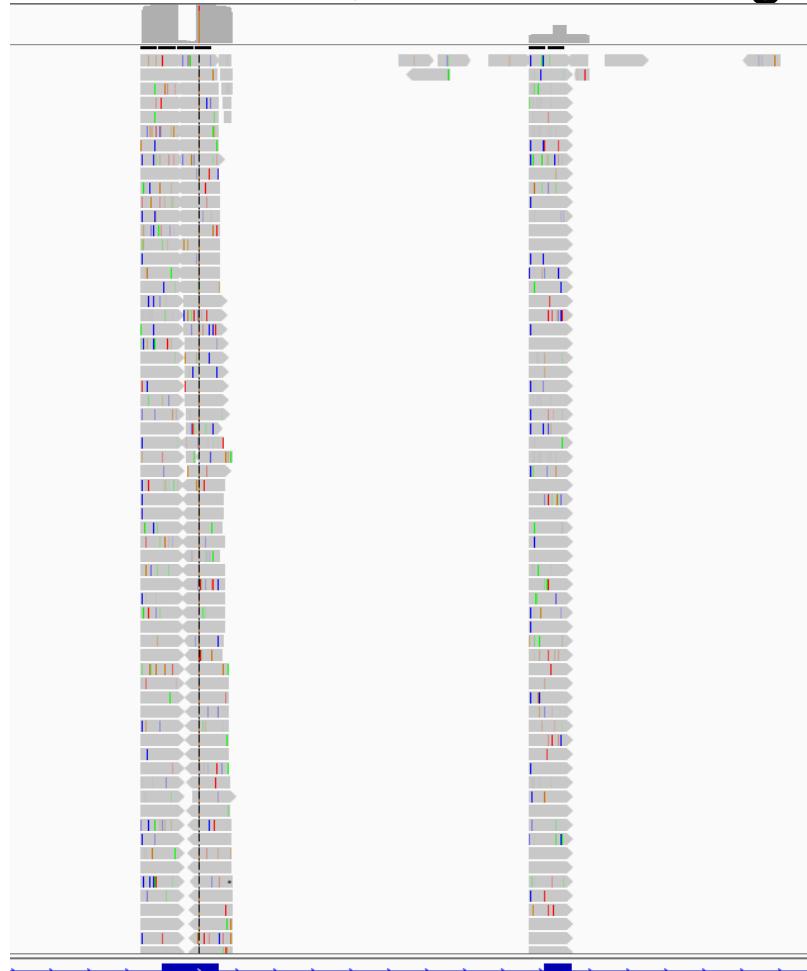
The ultimate complex mixture: Cell-free DNA dissolved in blood is derived from many normal cells and a few tumour cells



Crowley et al., *Nature Reviews Clinical Oncology*, 2013

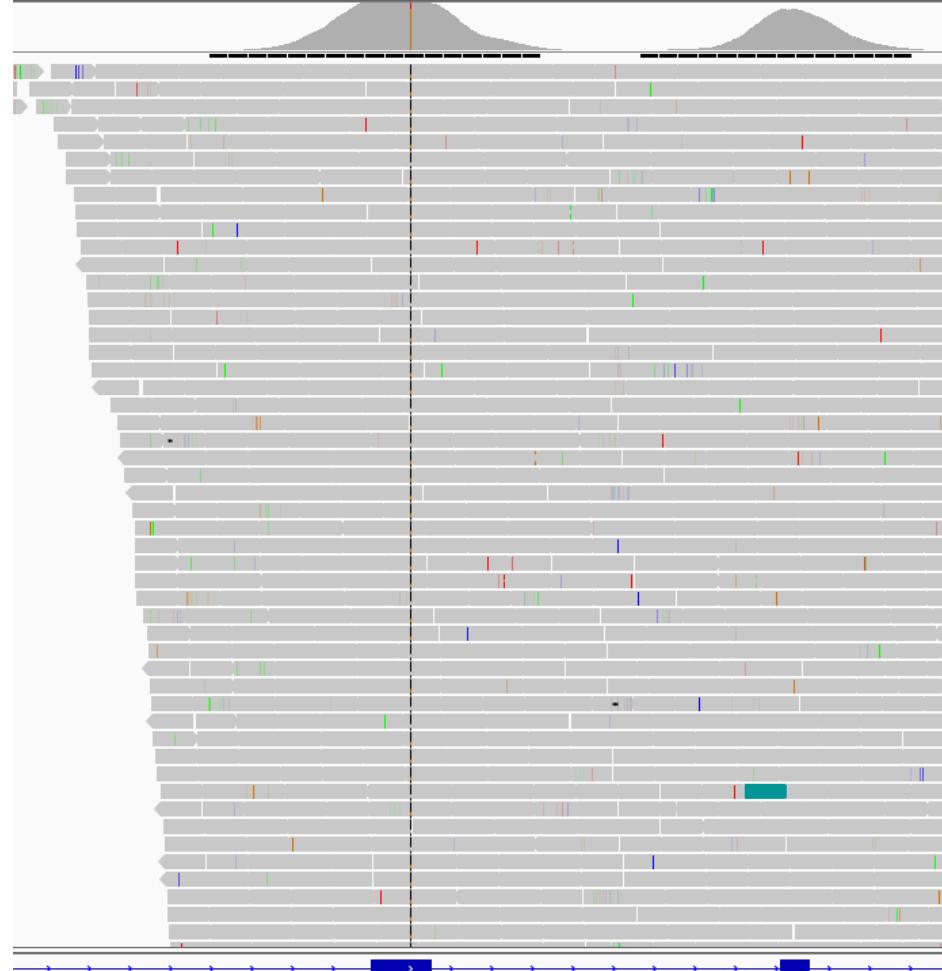
Coverage dependent on laboratory method used to target regions of interest

PCR >100,000X coverage



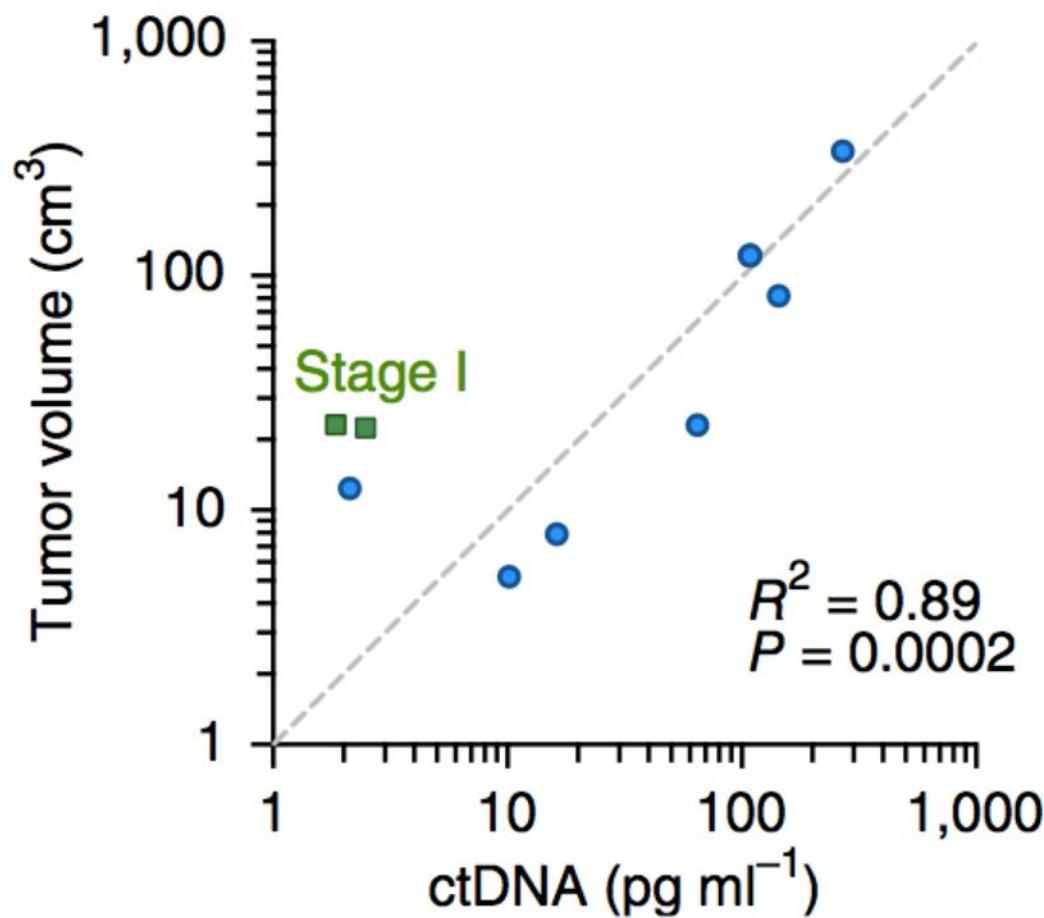
Tam-Seq: Forshew et al.
Sci Transl Med. 2012 May 30;4(136):136ra68.

Hybrid capture ~50,000X



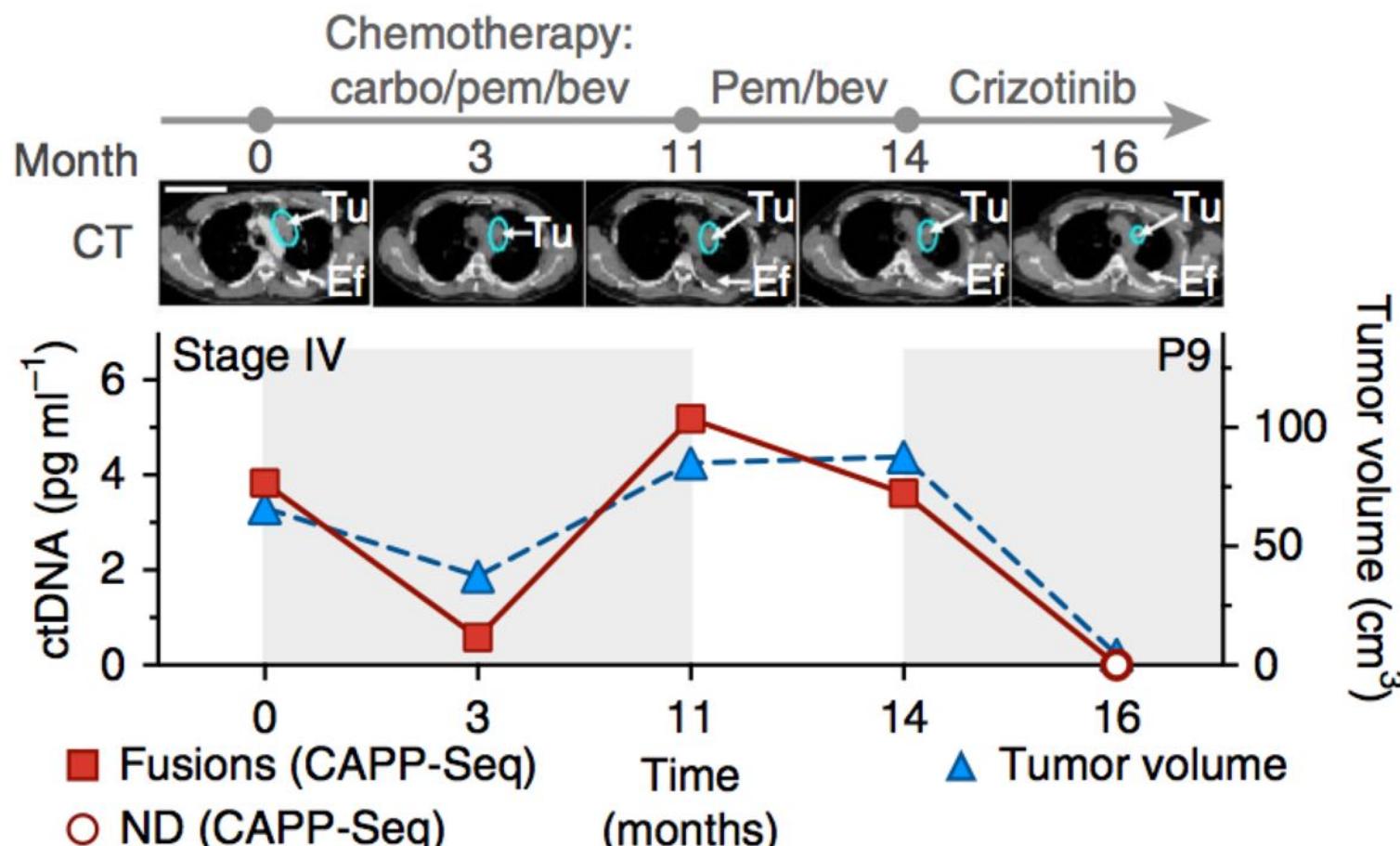
CAPP-seq: Newman, Bratman et al.
Nature Medicine. 2014 May;20(5):548-54.

Quantity of circulating tumour DNA reflects cancer burden



Newman, Bratman et al. Nature Medicine. 2014 May;20(5):548-54.

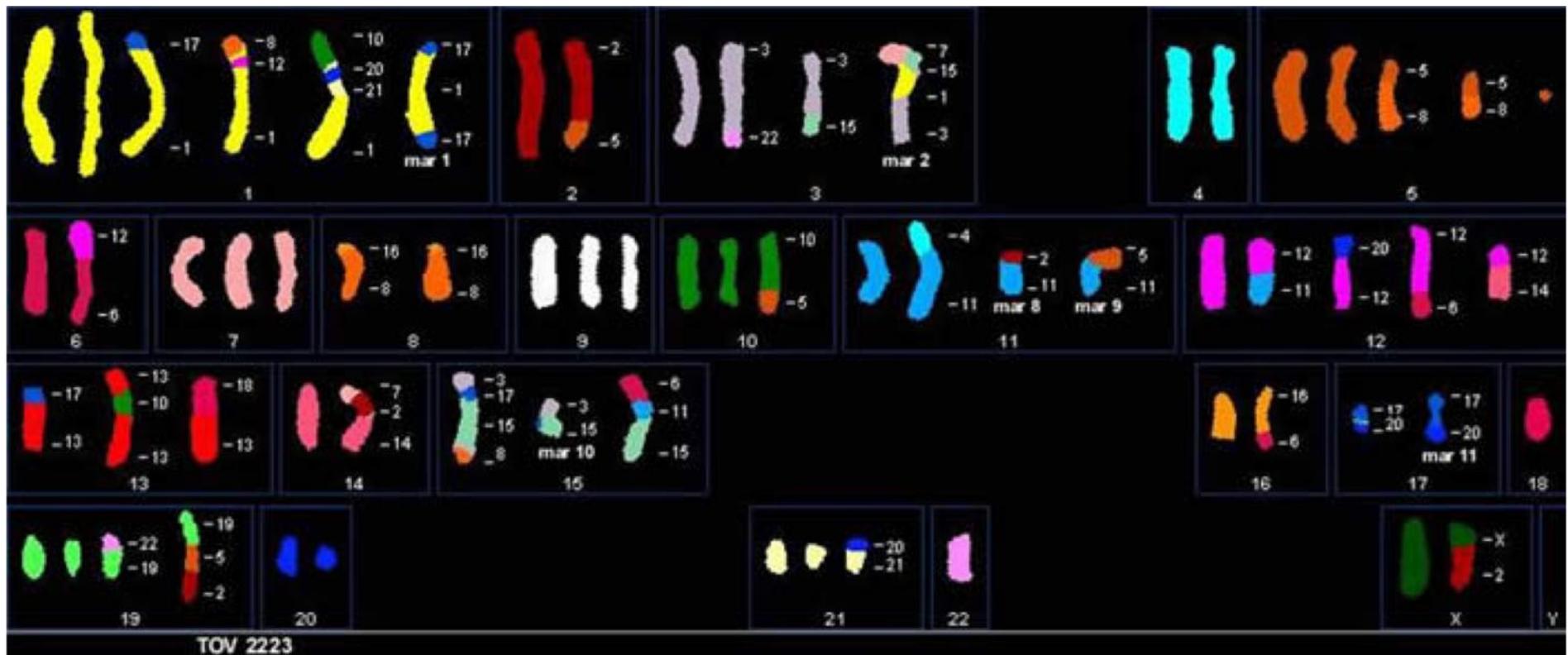
Quantity of circulating tumour DNA reflects clinical course



Newman, Bratman et al. Nature Medicine. 2014 May;20(5):548-54.

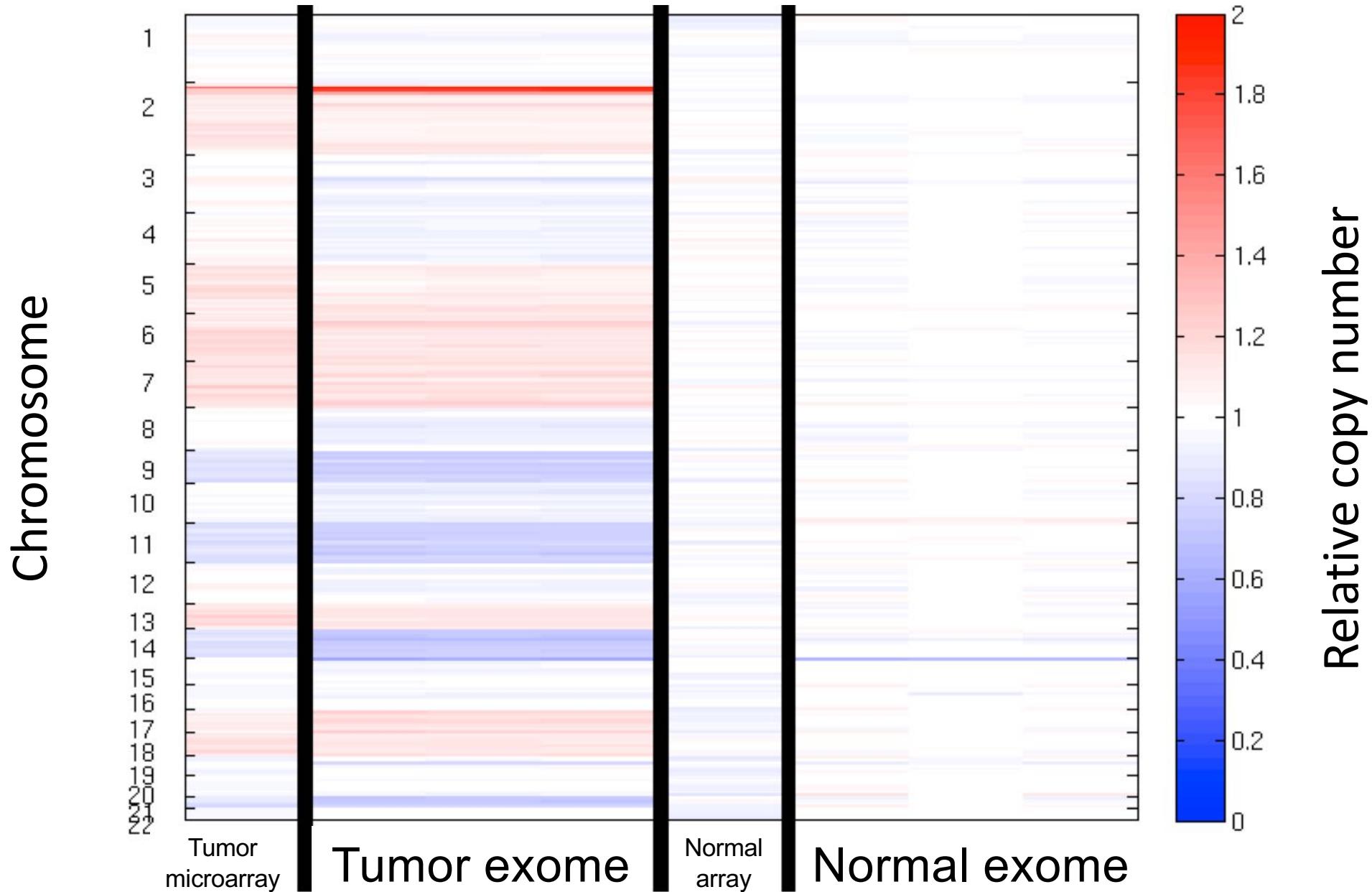
Somatic copy number alterations and rearrangements (structural alterations)

Tumors can have complex structural alterations

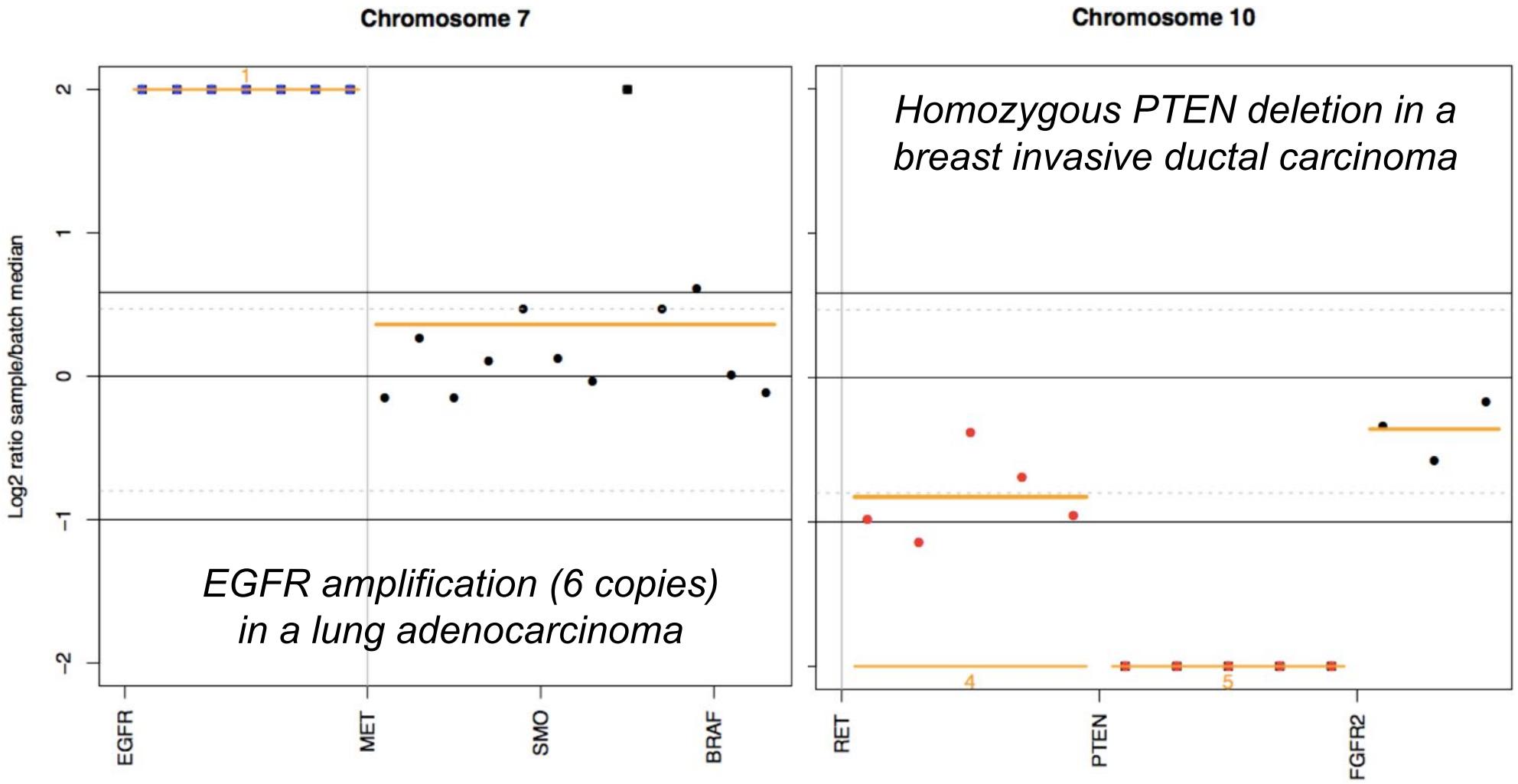


<http://www.bclq.org/en/index.html>

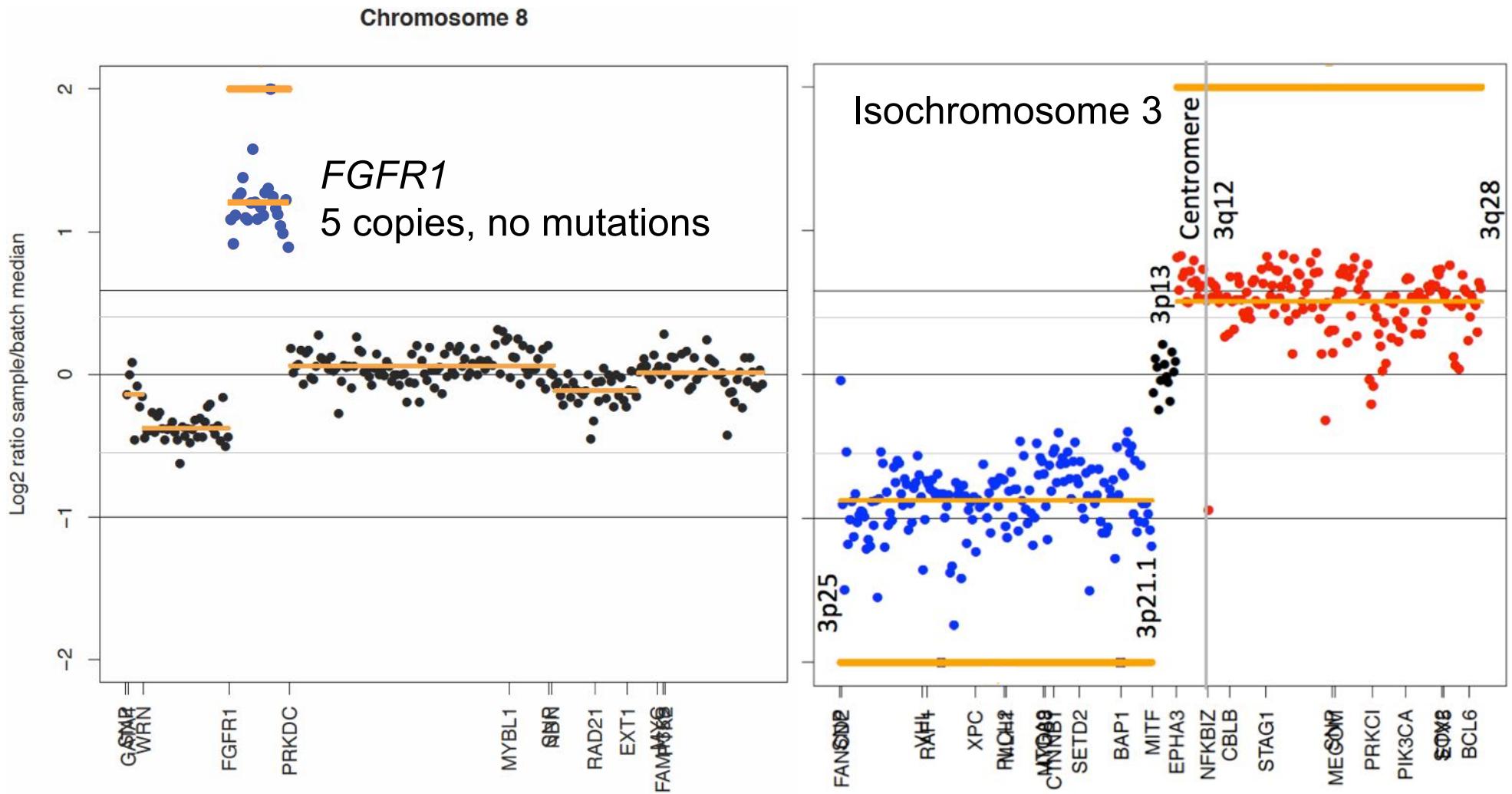
Gains and losses evident from sequence data



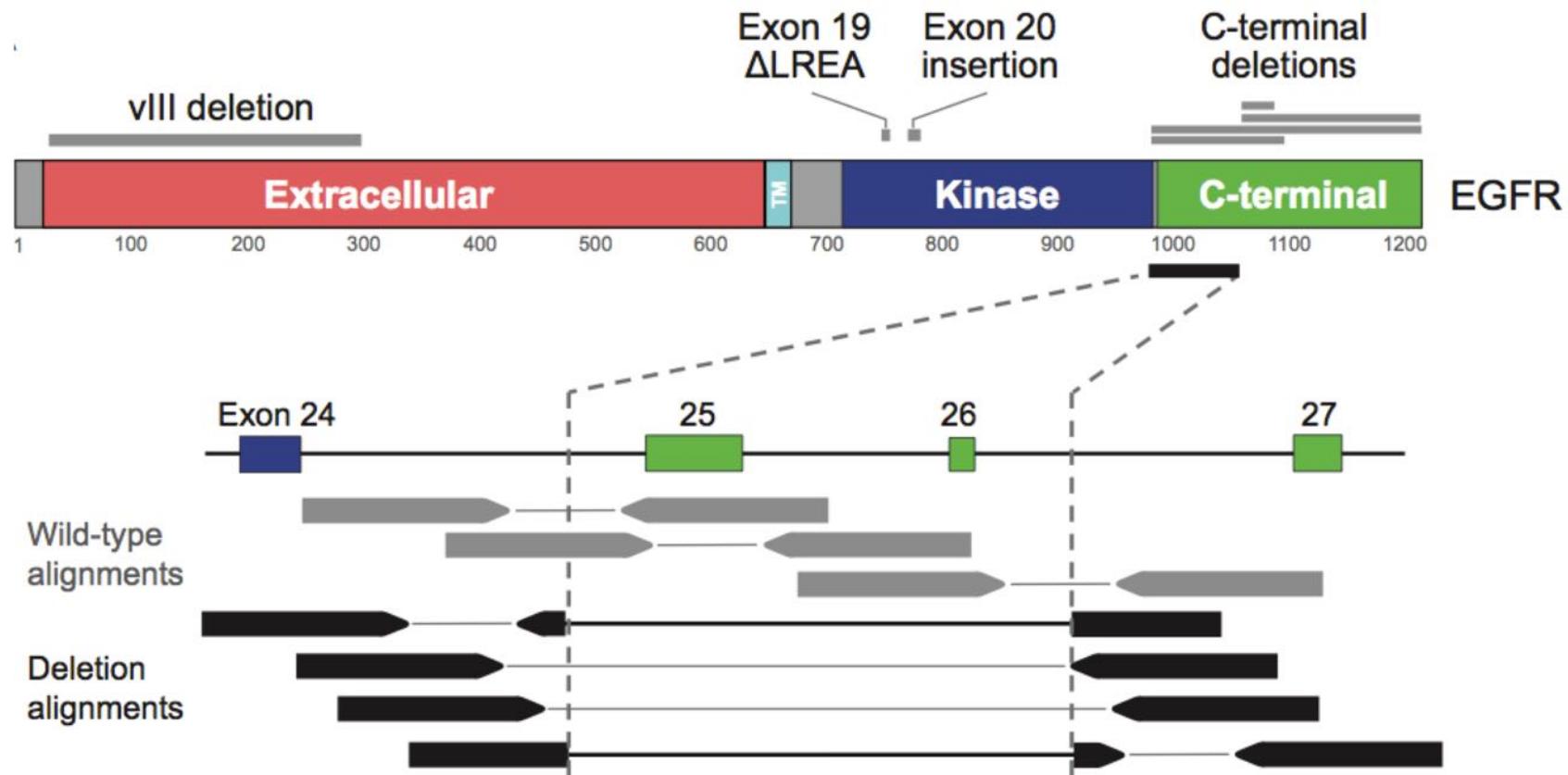
Re-analysis of existing clinical sequence data provides additional diagnostic value



Additional probes provide richer copy number profiles (183 gene panel, all exons plus select introns)



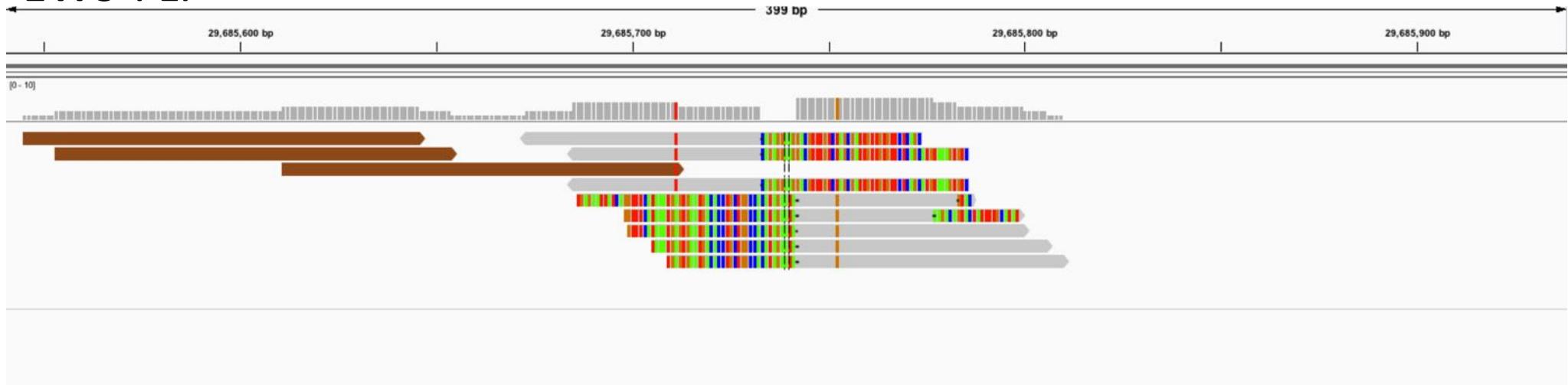
Exact breakpoints can be detected by examining reads that span the region



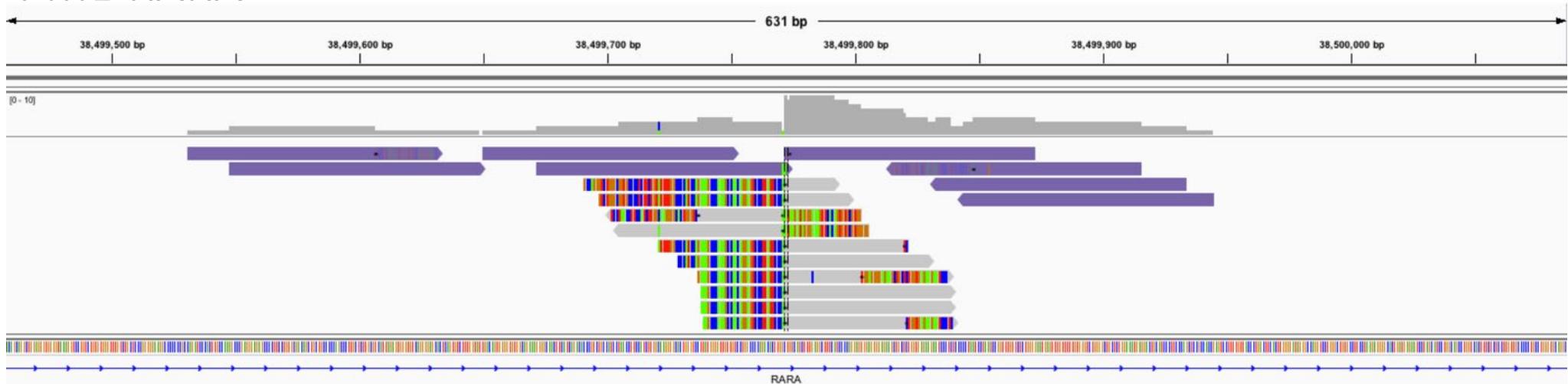
Imielinski et al. Cell. 2012 Sep 14;150(6):1107-20.

Translocation detection from soft-clipped bases plus large-insert or intrachromosomal read-pairs

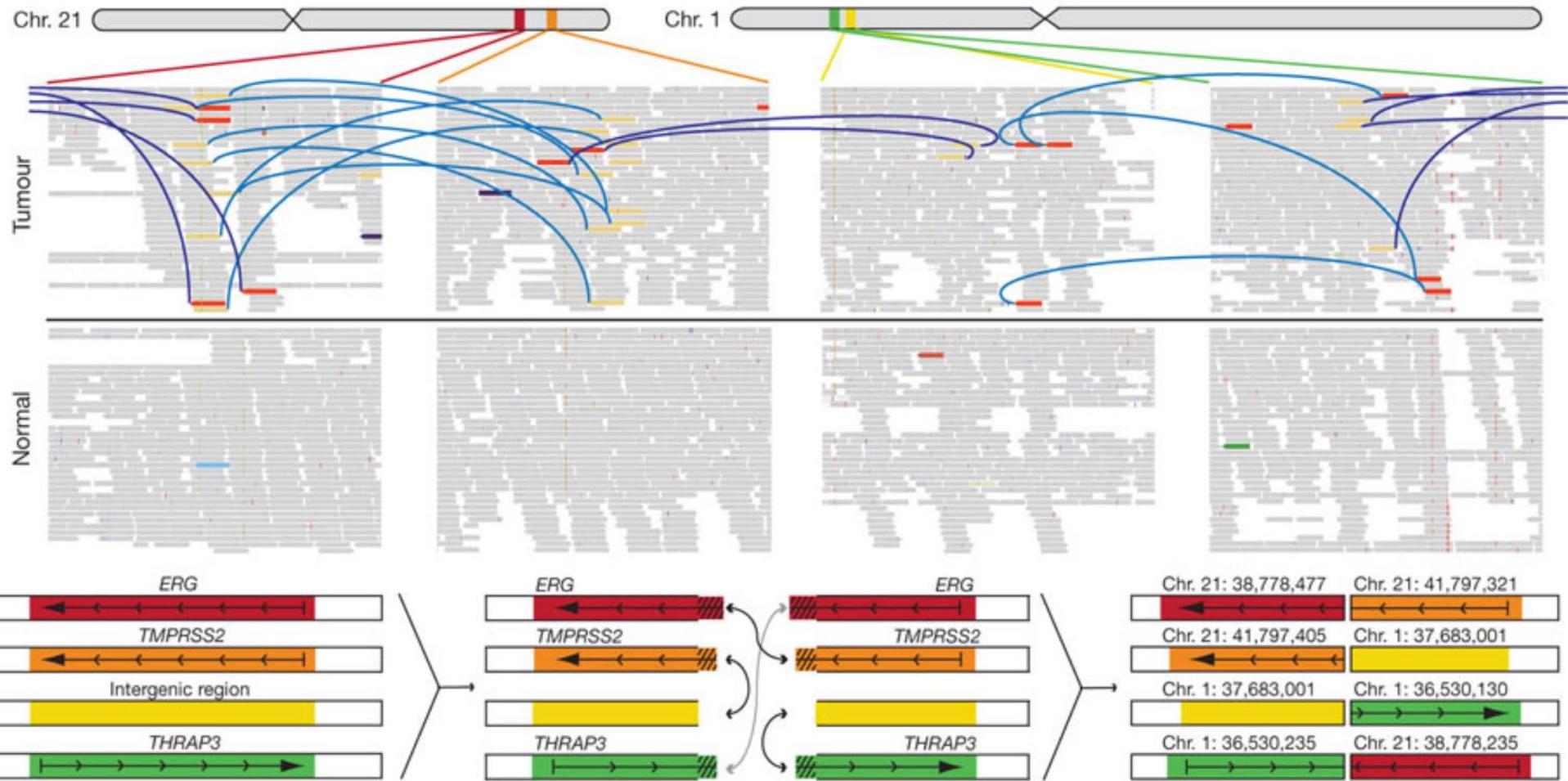
EWS-FLI



PML-RARA



Rearrangements can be highly complex and detectable at base-pair resolution

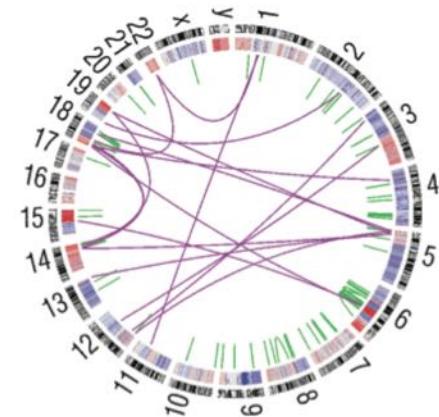


Berger *et al.* Nature. 2011 Feb 10;470(7333):214-20.

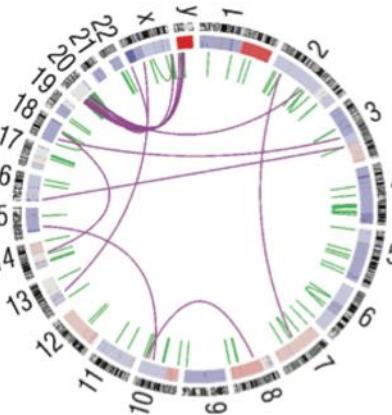
Some tumor genomes are highly rearranged, some are quiet, some have no rearrangements

Imielinski et al. 2012
Lung adenocarcinoma

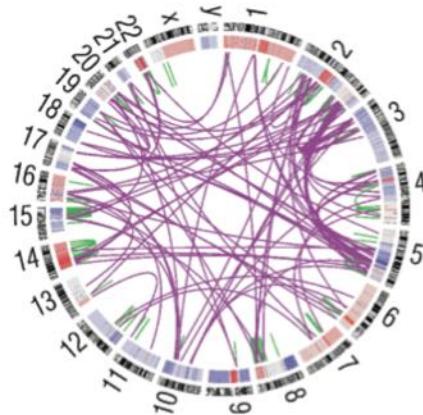
C9orf53-CDKN2A
(antisense fusion)
ROCK1
(exon 10-27 duplication)



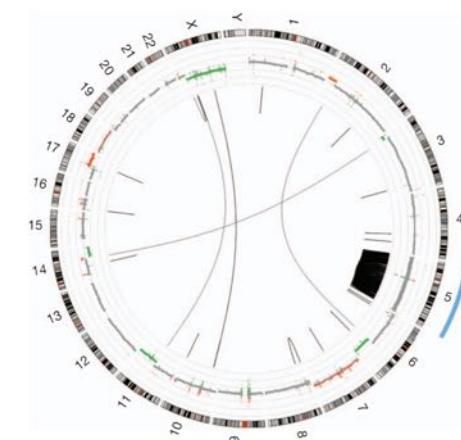
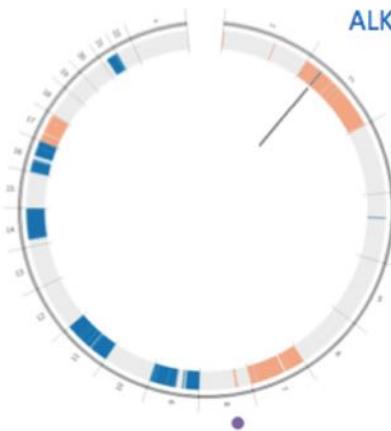
STK11 (deletion of translational start site)



EGFR
(exon 25-26 deletion)



Pugh et al. 2012
Neuroblastoma

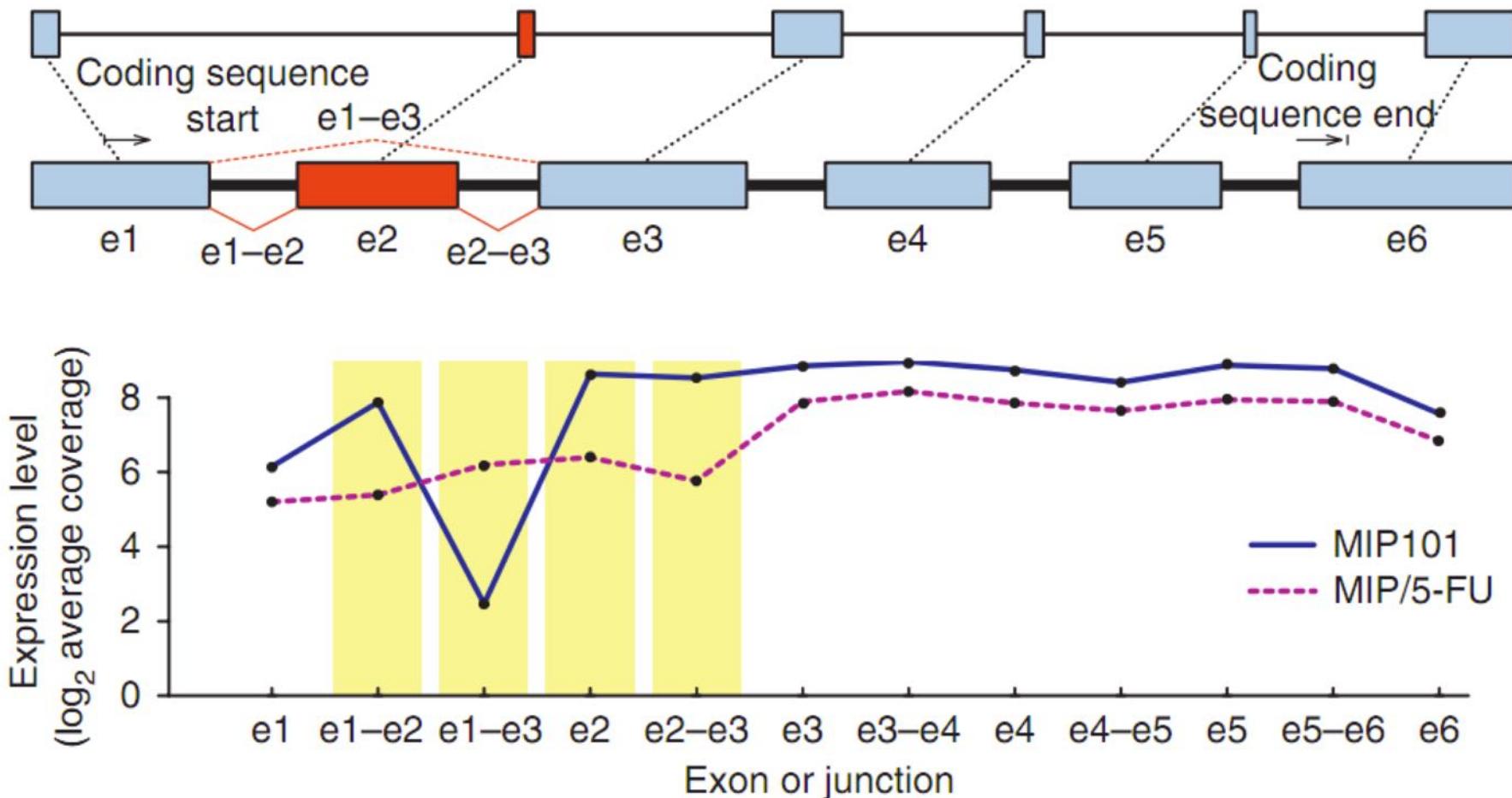


Molenaar et al. 2012.

Transcriptome sequencing

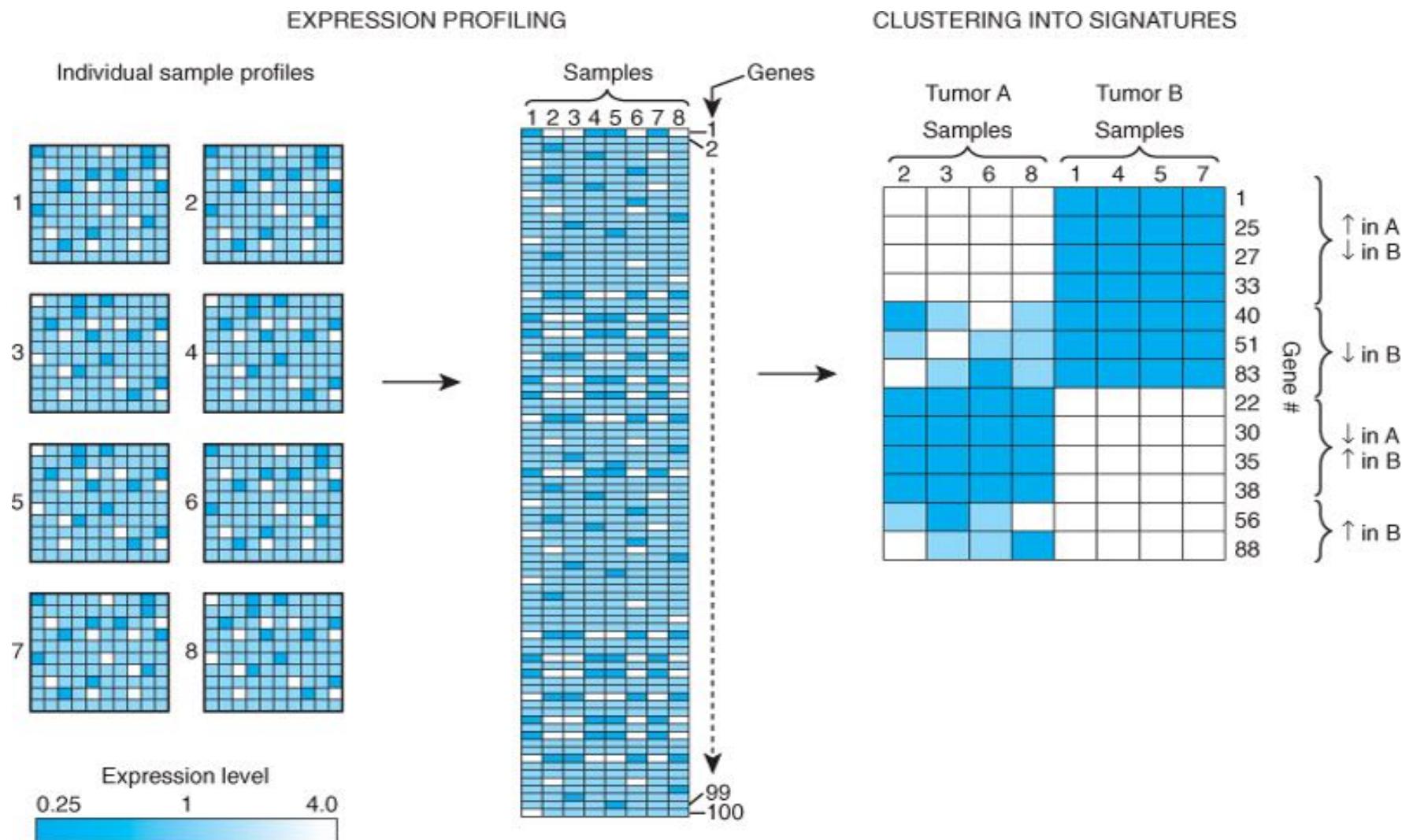
RNAseq can detect differences in exon usage that underlie resistance to treatment

UMPS gene model: 6 exons, 14,416 bases, 2,246 exon bases



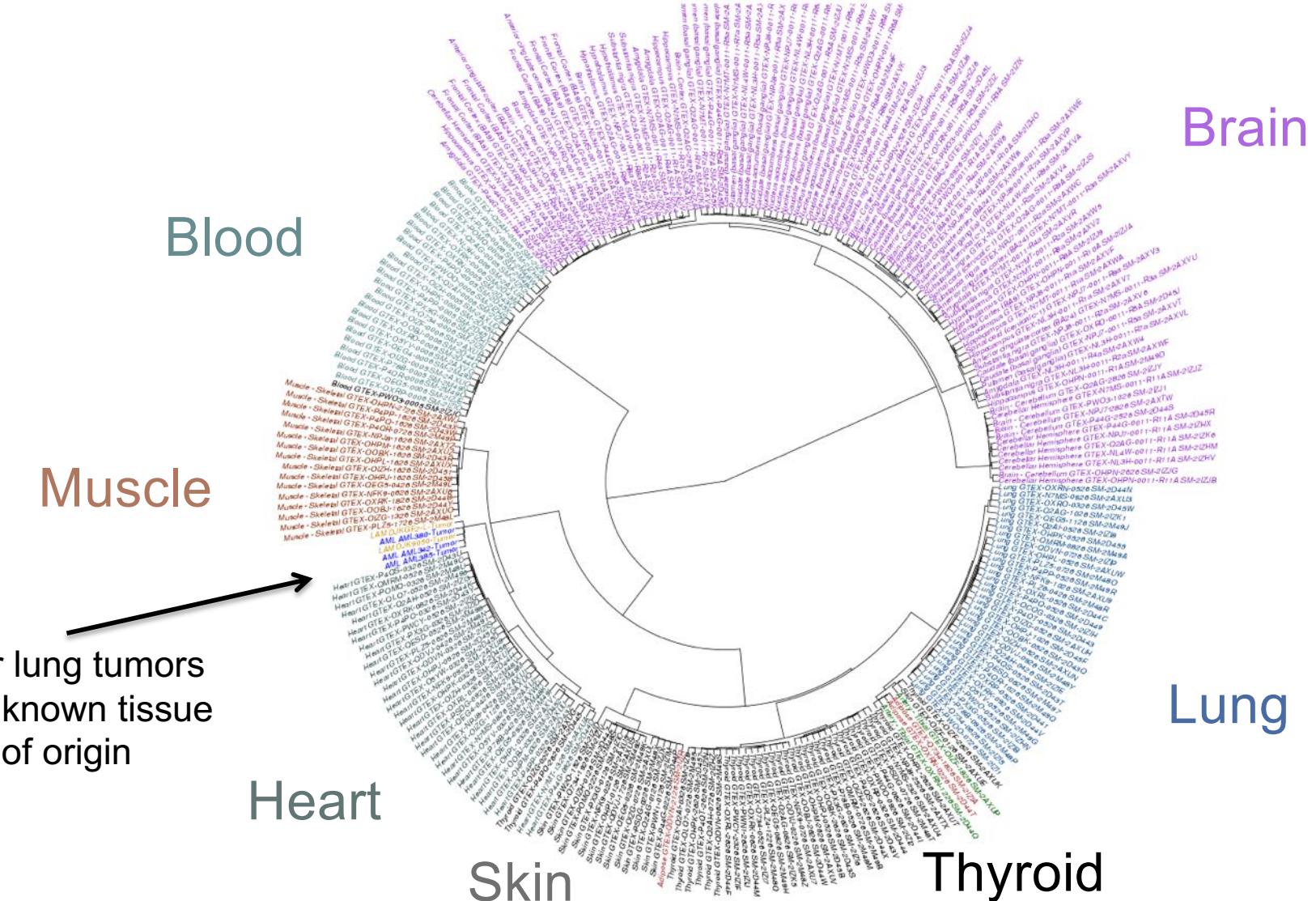
Griffith et al. Nat Methods. 2010 Oct;7(10):843-7.

RNAseq enables expression profiling & subtyping

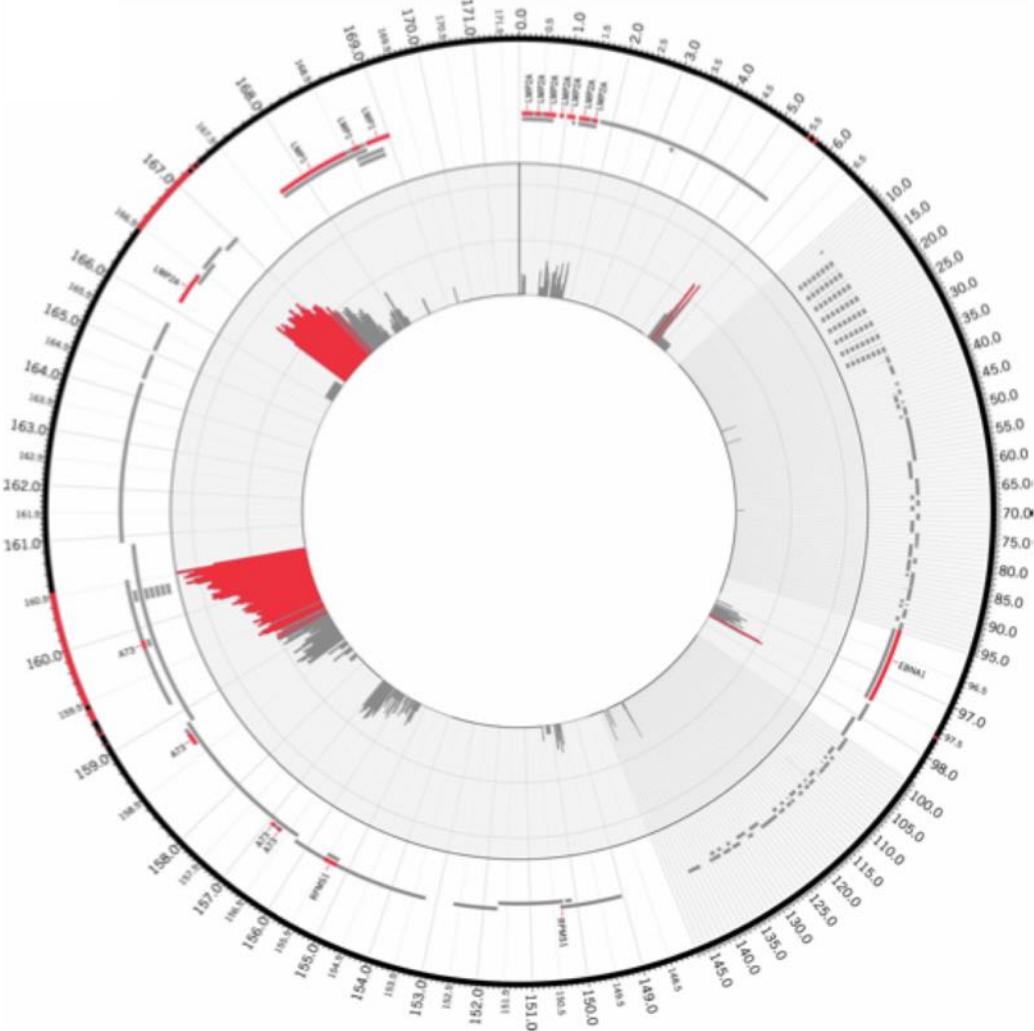


© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

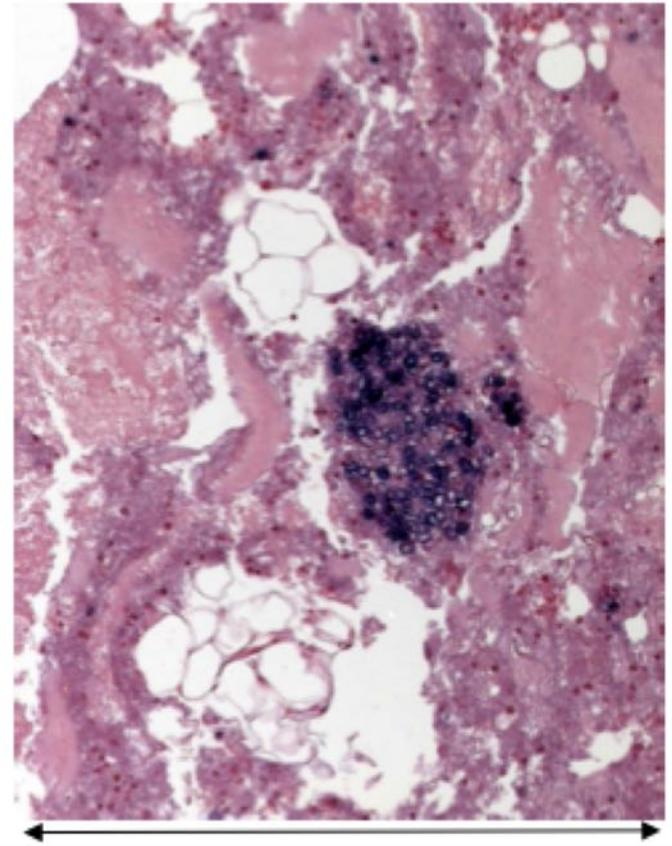
Primary tissue of origin evident from expression (if normal samples are available for comparison)



Pathogens seen in Genomes, RNAseq, sometimes Exomes

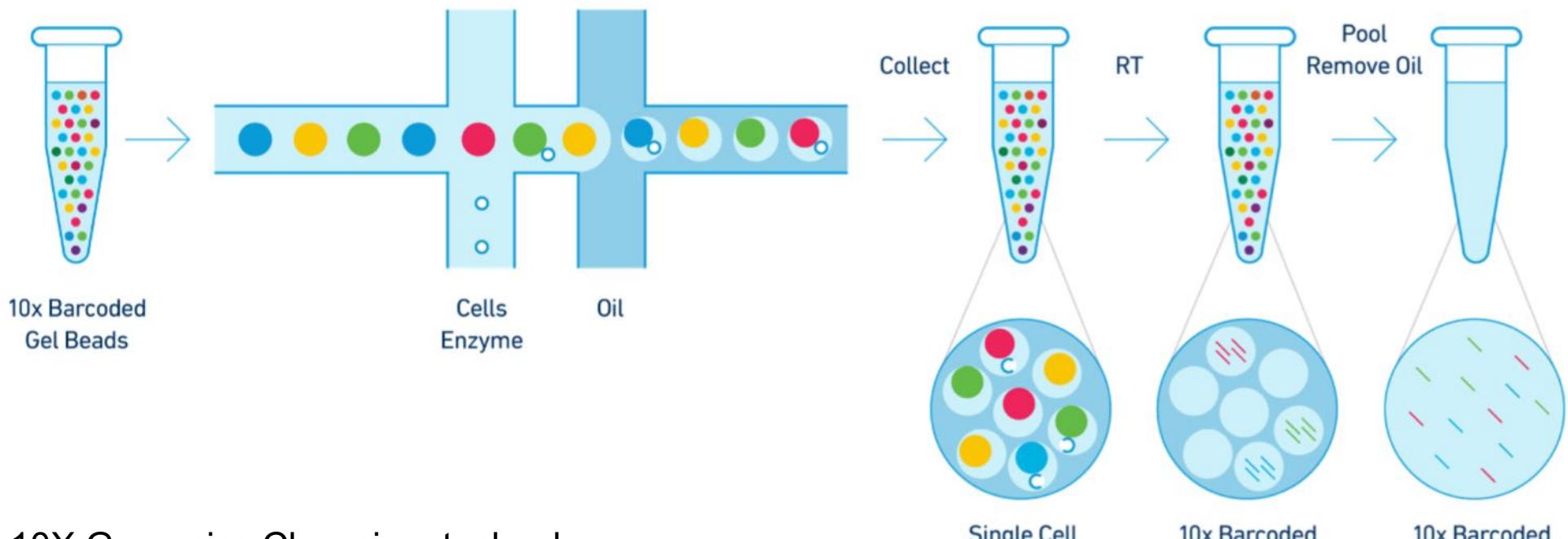


Unusual mapping of RNAseq reads to Epstein-Barr virus in lung adenocarcinoma



0.8 mm
ISH shows EBV confined
to tumor cells, tumor is actually a
lymphoepithelioma-like carcinoma

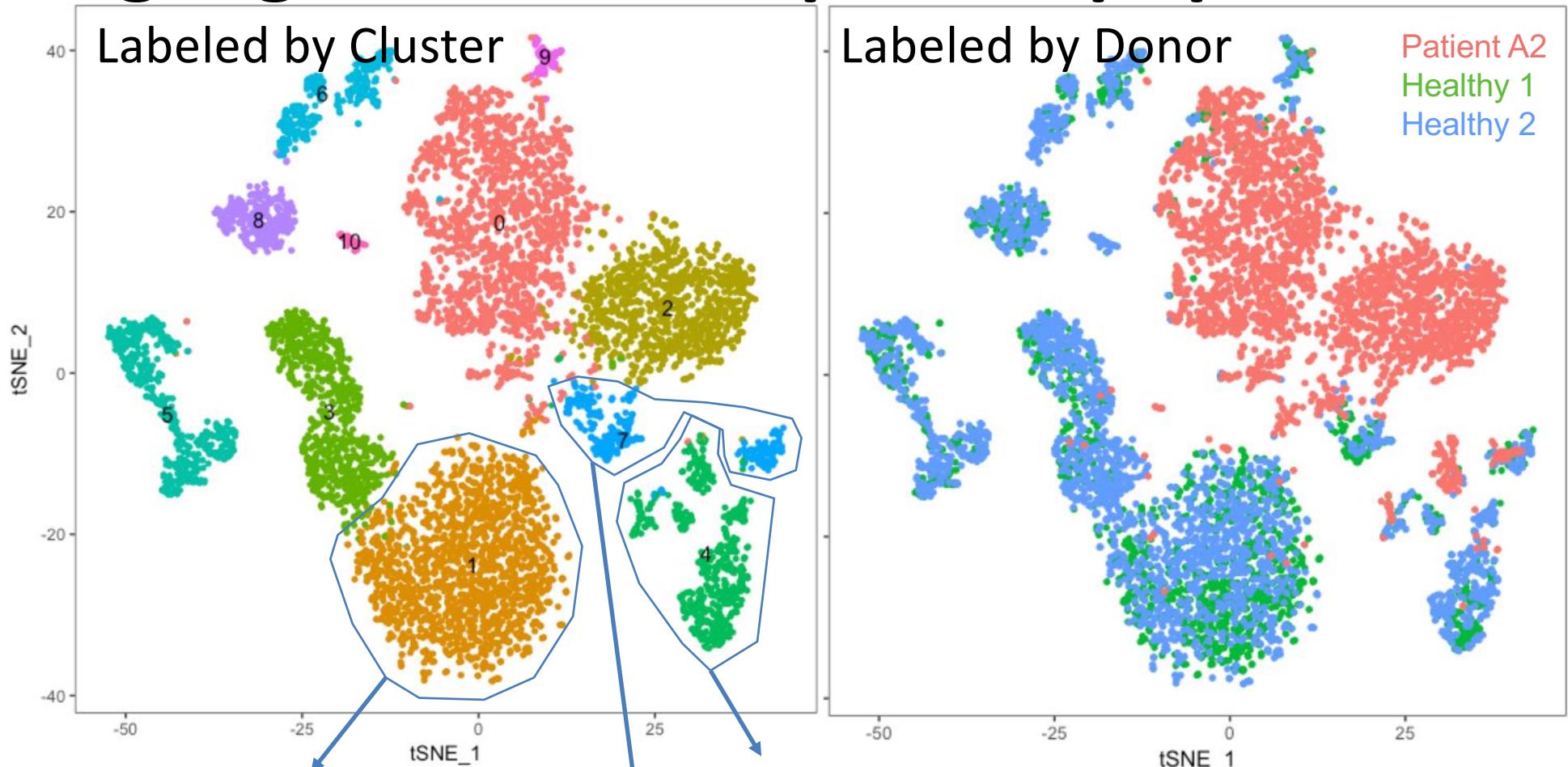
Transcriptomes from single cells can now be amplified and sequenced *en masse*



10X Genomics Chromium technology
www.10xgenomics.com/solutions/single-cell/

Princess Margaret Genomics Centre (a service provider)
www.pmgenomics.ca/pmgenomics/services/single_cell_genomics.html

Combining cancer & healthy cell data highlights shared & private populations



LTB: induces inflammatory response

CD3D: part of *TCR/CD3 complex*

TMEM66 (SARAF): cytoplasmic calcium signalling

JUNB: transcription factor, growth factor response

CD3E: part of *TCR/CD3 complex*

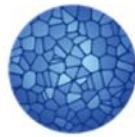
LYZ: present in cytoplasmic granules of *macrophages*

FCN1: pattern recognition molecule secreted by monocytes/*macrophages*

PRTN3: serine protease enzyme, expressed mainly *by neutrophil granulocytes*

MPO: peroxidase enzyme, abundantly expressed in *neutrophil granulocytes*

HumanCellAtlas.org: Scientific and Funding Opportunities



HUMAN
CELL
ATLAS

[Home](#) [Human Cell Atlas](#) [Areas of Impact](#) [Contact](#)

MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

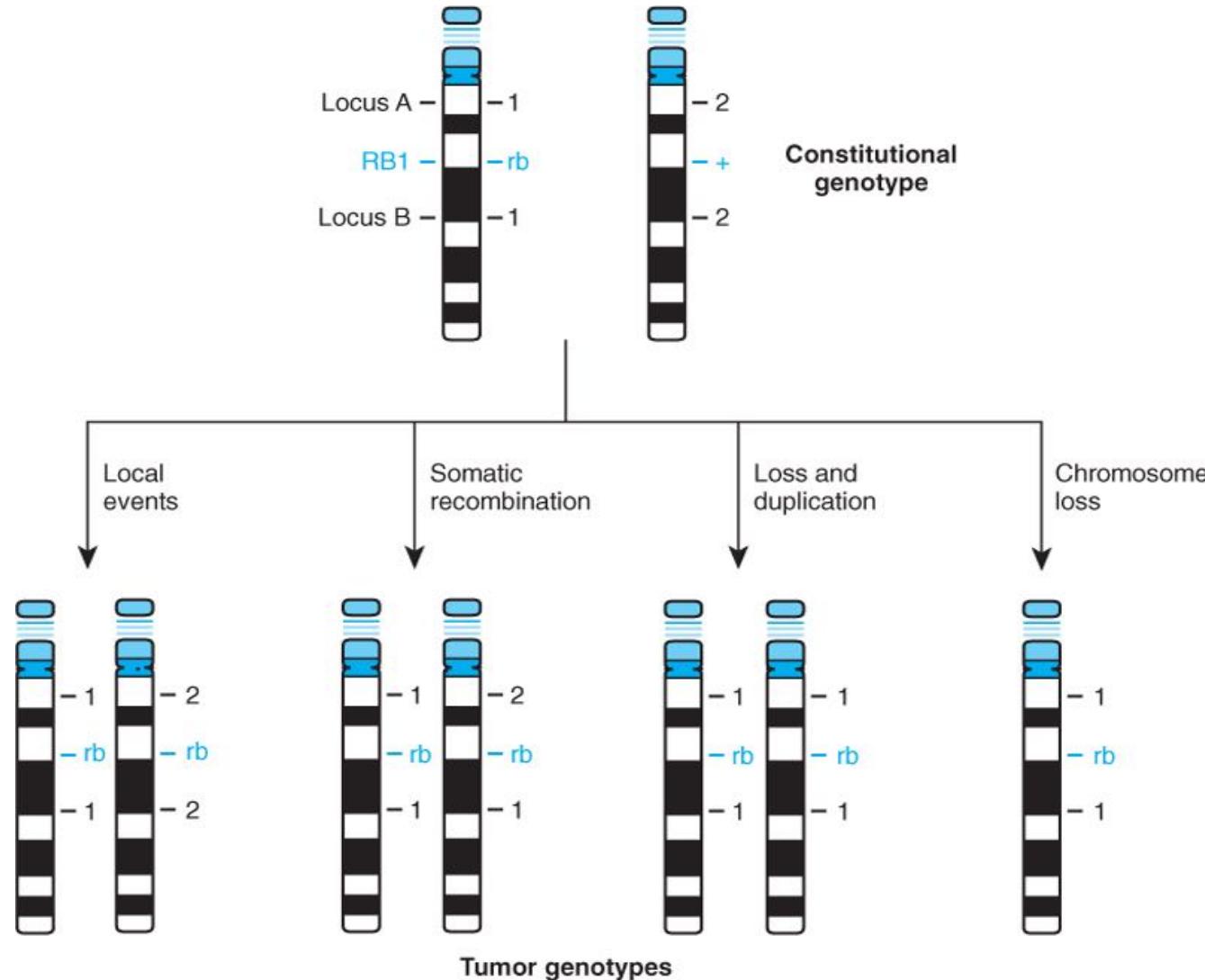


Chan Zuckerberg Initiative funding pilot projects in areas of technology development, tissue sourcing, and computational methods

chanzuckerberg.com/human-cell-atlas

Germline cancer genetics (Heredity Cancer Syndromes)

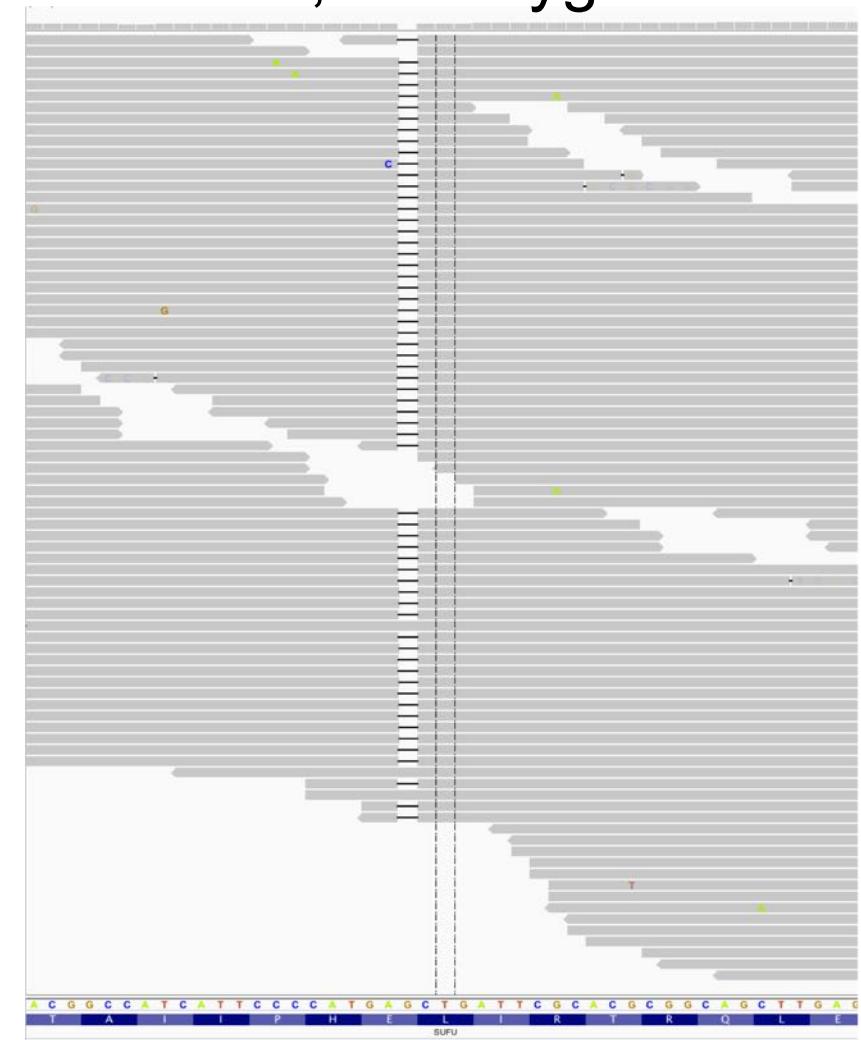
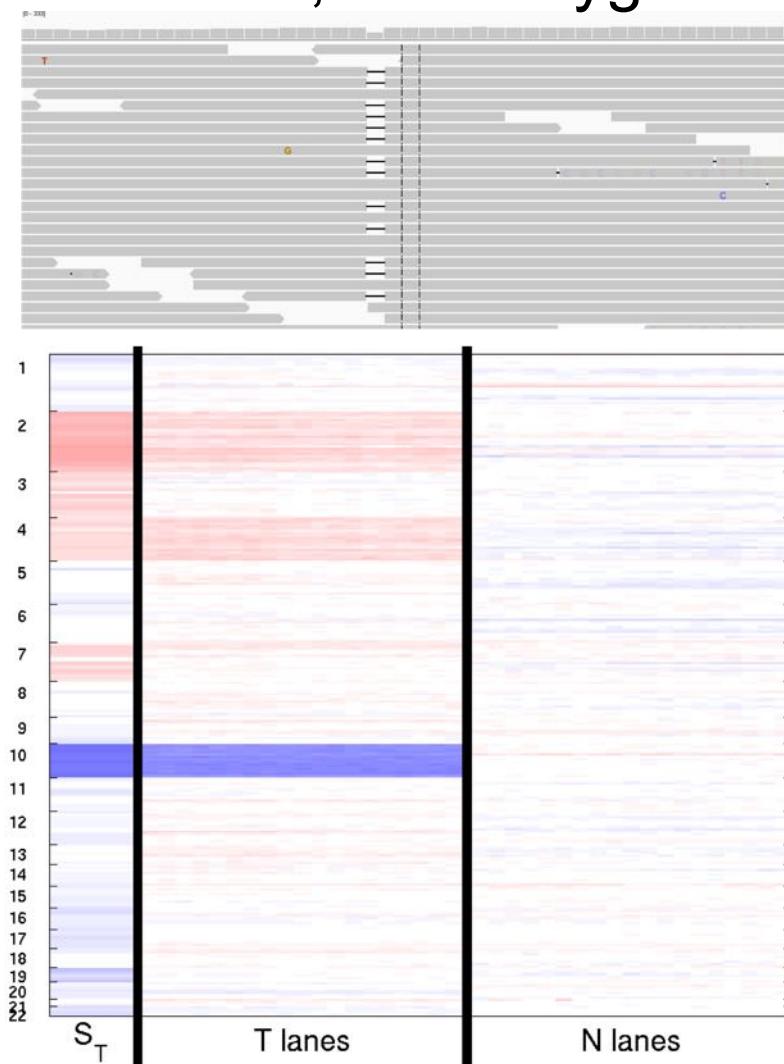
“Second hit” can unmask a pathogenic germline allele



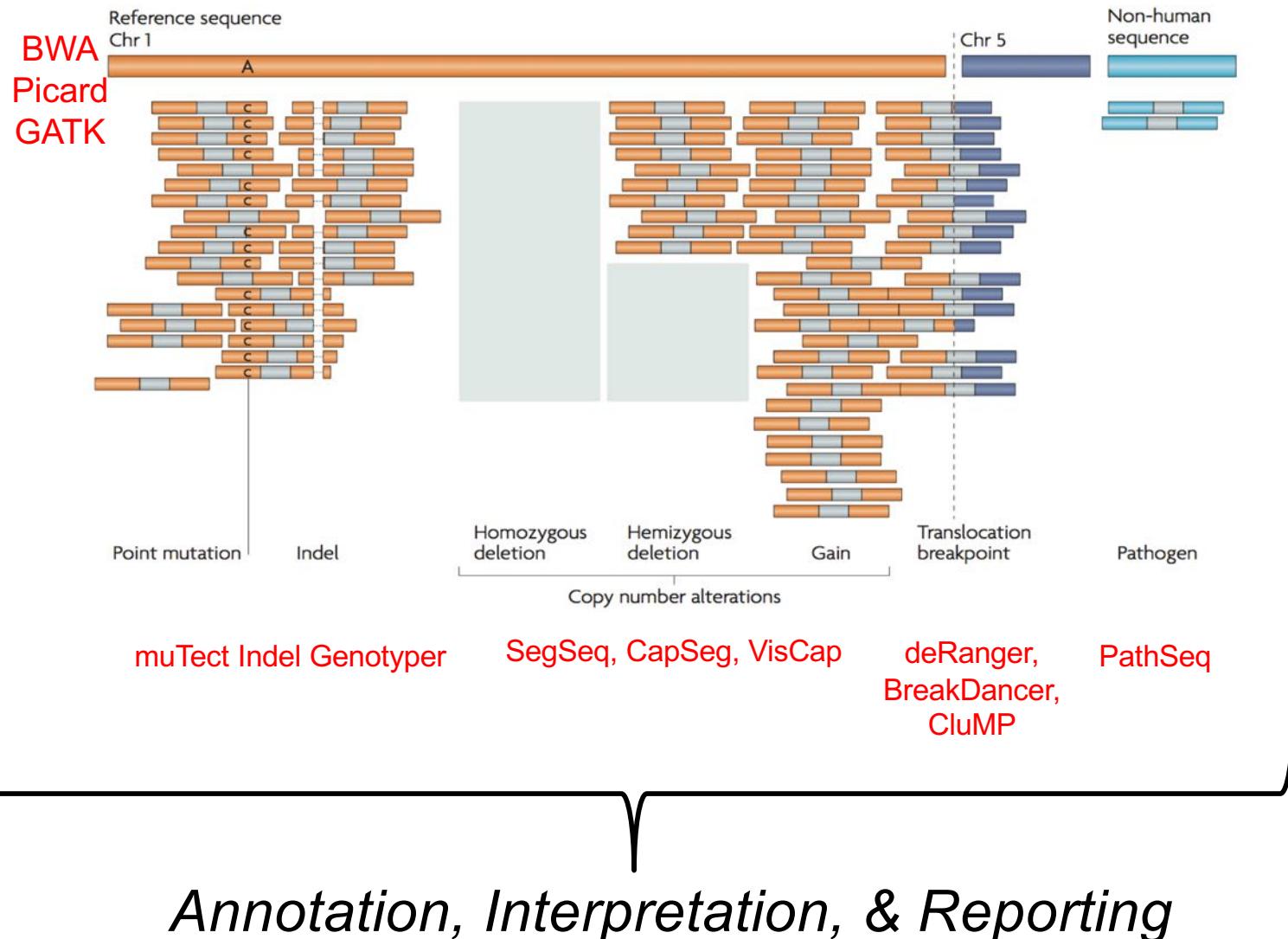
© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

Carrier for a pathogenic *SUFU* mutation developed medulloblastoma (10-)

Normal, heterozygous Tumor, hemizygous



Once found, variants (of all types) require annotation to guide interpretation



How do we interpret variants en masse?

Clinical indication & history (family members, de novo)

Locus-specific and internal variant databases

Population frequency

Public databases (disease-specific & general)

Publications

Amino acid & biochemistry conservation

Substitution prediction tools

(Polyphen2, SIFT, AlignGVGD, SarcomerePolyphen)

Splicing prediction tools

(SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splice Finder)

Druggable targets and pathways

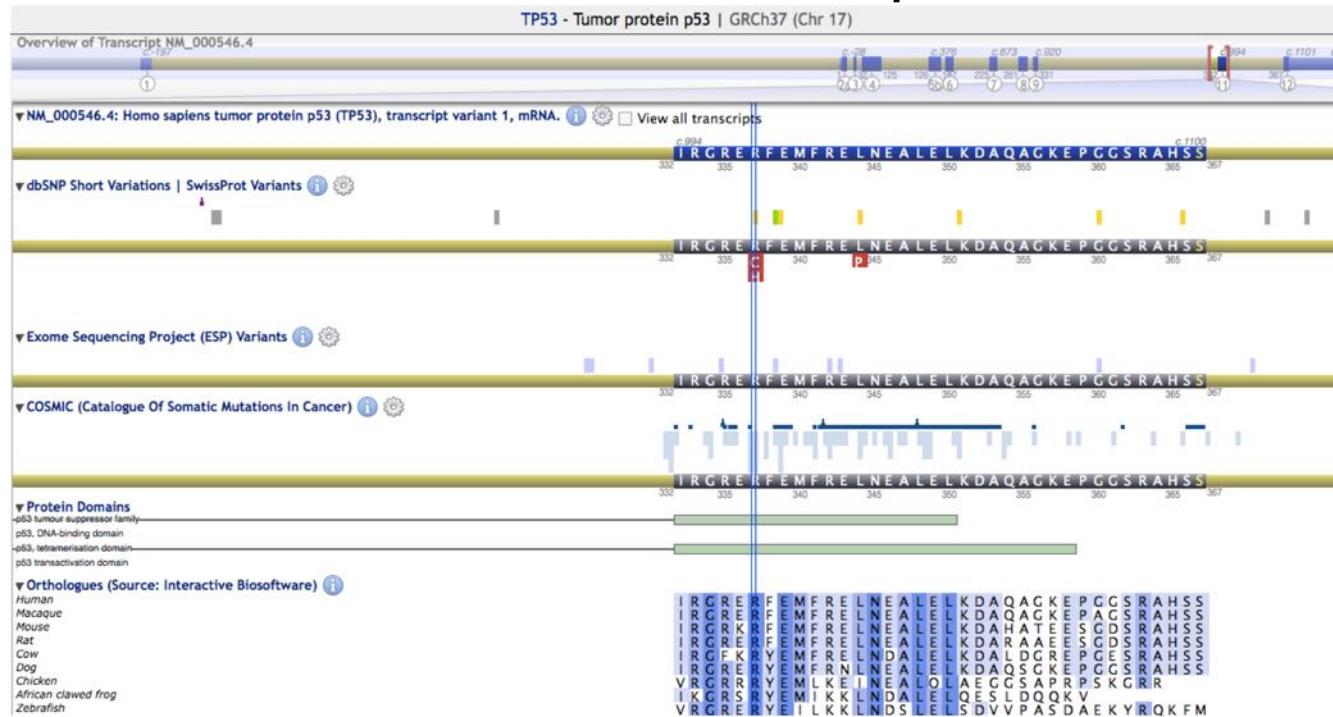
Drug Gene Interaction Database (DGIdb), Drug Bank

Command-line versus Manual Annotation

Oncotator – plain text table with >150 columns

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	## muTector v1.0.47986																																	
2	contig	position	context	ref_allele	alt_allele	tumor_name	normal_name	score	dbSNP_site	covered	power	tumor_pwe	normal_pwe	total_pairs	improper_pa_map_Q0	rest_id	fstar	tumor_f	contaminant	contaminant_t_ref_count	t_alt_count	t_ref_sum	t_alt_sum	t_ref_max	t_alt_max	t_ins_count	t_del_count	normal_best	init_n_low	n_ref_count	n_alt_count	n_ref_sum	n_alt_sum	judgement
3	chr3	41266097	CTGxACT	G	T	PMH13-142	PMH13-142	0	DBSNP+COSI	COVERED	0.52183	0.522161	0.999366	72	2	0	13.749453	0.666667	0.02	0.210543	2	4	69	135	60	60	0	0 GG	5.116689	17	0	592	0 KEEP	
4	chr4	1807414	GGCGTG	G	A	PMH13-142	PMH13-142	0	NOVEL	UNCOVERED	0.036511	0.036765	0.993072	32	0	0	6.392785	0.4	0.02	-0.052637	3	2	101	69	60	60	0	0 GG	3.310702	11	0	375	0 KEEP	
5	chr11	534171	GGCGCA	C	T	PMH13-142	PMH13-142	0	NOVEL	COVERED	0.229028	0.229173	0.999366	55	1	0	8.832265	0.272727	0.02	-0.131577	8	3	269	102	60	60	0	0 CC	5.115816	17	0	558	0 KEEP	
6	chr12	121435468	CAAGTG	G	A	PMH13-142	PMH13-142	0	COSMIC	COVERED	0.80019	0.800245	0.999931	115	1	0	6.549984	0.5	0.02	-0.298279	2	2	70	68	60	60	0	0 GG	7.524151	25	0	862	0 KEEP	
7	chr1	1168428	TGAAAC	C	T	PMH13-142	PMH13-142	0	DBSNP	UNCOVERED	0	0.353947	0	57	1	0	0	0	0.02	-0.166685	2	0	70	0	60	0	0	0 CT	-8.320054	8	3	280	102 REJECT	
8	chr1	1168420	ACCTCT	G	A	PMH13-142	PMH13-142	0	DBSNP	UNCOVERED	0	0.387407	0	56	0	0	0	0	0.02	-0.175452	6	0	206	0	60	0	0	0 GG	2.708785	9	0	310	0 REJECT	
9	chr1	1187893	CAACAA	T	C	PMH13-142	PMH13-142	0	DBSNP	UNCOVERED	0	0.024074	0	13	1	0	15.907936	1	0.02	-0.035092	0	4	0	140	0	60	0	0 CC	-31.819773	0	8	0	280 REJECT	
10	chr1	1199486	CTTAGA	C	A	PMH13-142	PMH13-142	0	NOVEL	UNCOVERED	0.098977	0.099721	0.99254	32	0	0	0	0	0.02	-0.0877	2	0	71	0	60	0	0	0 AC	-3.944378	7	2	250	57 REJECT	
11	chr1	1199541	AGGAACA	G	A	PMH13-142	PMH13-142	0	DBSNP	UNCOVERED	0	0.641347	0	68	2	0	0	0	0.02	-0.245612	7	0	229	0	60	0	0	0 AG	-2.339804	12	2	403	56 REJECT	
12	chr1	1120558	CATGCA	C	T	PMH13-142	PMH13-142	0	DBSNP	COVERED	0.799979	0.800245	0.999667	121	3	0	29.915045	1	0.02	-0.307042	0	8	0	261	0	60	0	0 CT	-52.200904	15	16	518	539 REJECT	
13	chr1	11206970	ATAAAC	A	G	PMH13-142	PMH13-142	0	DBSNP	UNCOVERED	0	0.553887	0	72	1	0	0	0	0.02	-0.193901	5	0	158	0	60	0	0	0 AG	-11.391674	13	4	430	146 REJECT	

Alamut – Interactive desktop software



Variant Effect Predictor, ANNOVAR, SNPeff, VAT, & many others

Novel Variant Assessment is still a manual enterprise, aided greatly by sharing interpretations

ACMG STANDARDS AND GUIDELINES						
	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			

Richards et al. *Genetics in Medicine*. (2015) 17, 405–423

How do we report and share these results?

Overall result: Somatic variants identified:

BRAF (NM_004333.4) Heterozygous, c.1406G>T (p.G469V)

RB1 (NM_000321.2) Heterozygous, c.607+1G>A (splice)

TP53 (NM_000546.4) Heterozygous, c.488A>G (p.Y163C)

Class 4A mutation: BRAF p.Gly469Val (58% of reads)

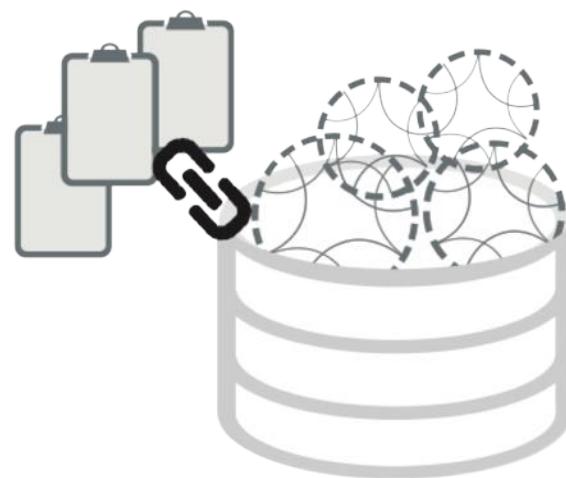
BRAF p.G469V is a rare mutation across cancer (76 of >193,000 cases in the Catalogue of Somatic Mutations in Cancer), and has not been previously reported in over 4,300 ovarian cancers. BRAF is recurrently mutated in ovarian cancers, but is rarely mutated in high grade serous carcinoma (mycancergenome.org). The most common variant found in ovarian tumours is p.V600E (mycancergenome.org). Currently, the impact of this mutation on patient prognosis or treatment in this tumour site is unknown.



American Association
for Cancer Research

FINDING CURES TOGETHERSM

AACR Project GENIE is an international, multiphase, multiyear project that will provide the “critical mass” of genomic and clinical data necessary to improve clinical decision making and catalyze new clinical and translational research.



www.aacr.org/genie

PROJECTGENIE

Genomics Evidence Neoplasia Information Exchange



GENIE will aggregate existing and ongoing genotyping efforts from the **seven phase 1 → project participants** into a single registry and link these data to select clinical outcomes, ultimately making these data publicly available.

- The Center for Personalized Cancer Treatment, The Netherlands
- Dana-Farber Cancer Institute
- Institut Gustave Roussy, France
- Johns Hopkins University's Sidney Kimmel Comprehensive Cancer Center
- Memorial Sloan Kettering Cancer Center
- Princess Margaret Cancer Centre, Canada
- Vanderbilt-Ingram Cancer Center

Molecular report of the future?

Concise, systematic (crowd-sourced?) genome interpretation alongside clinical annotations

 **cBioPortal**
for Cancer Genomics

Visualize, analyze, discover.

Memorial Sloan Kettering Cancer Center.
You are logged in as schultzn@mskcc.org. [Sign out.](#)

HOME DATA SETS TUTORIALS FAQ NEWS TOOLS ABOUT [VISUALIZE YOUR DATA](#)

P-0001104 75 years old, Colorectal Cancer (Colon Adenocarcinoma), LIVING, Recurred/Progressed (27 months)
↳ ① P-0001104-T03-IM3 Primary (Colon), ② P-0001104-T02-IM3 Metastasis (Liver), ③ P-0001104-T01-IM3 Metastasis (Liver), ④ P-0001104-T05-IM3 Metastasis (Liver)

Clinical Events
Months since diagnosis: 0, 6, 12, 18, 24
Specimen: ① ②
Treatment:

Genomic Overview
Chromosome plot showing mutation counts across chromosomes 1-22, X, and Y. A scatter plot shows mutation count vs fraction of copy number. A histogram shows variant allele frequency.

18 mutations

Tumors	Gene	Protein Change	Type	Allele Freq	Cohort	cBioPortal	COSMIC	Mutation Assessor	Drugs
① ② ③ ④	BRAF	V600E	Missense	---	5.1%	1446	23294	Low	
① ② ③ ④	TP53	R273H	Missense	---	42.2%	849	1312	Medium	
① ② ③ ④	EPHAS	P841S	Missense	---	3.2%	9	4	Medium	

www.cbiportal.org

Case study

Jones et al. *Genome Biology* 2010, **11**:R82
<http://genomebiology.com/2010/11/8/R82>



RESEARCH

Open Access

Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors

Steven JM Jones^{1*}, Janessa Laskin², Yvonne Y Li¹, Obi L Griffith¹, Jianghong An¹, Mikhail Bilenky¹, Yaron S Butterfield¹, Timothee Cezard¹, Eric Chuah¹, Richard Corbett¹, Anthony P Fejes¹, Malachi Griffith¹, John Yee³, Montgomery Martin², Michael Mayo¹, Nataliya Melnyk⁴, Ryan D Morin¹, Trevor J Pugh¹, Tesa Severson¹, Sohrab P Shah^{4,5}, Margaret Sutcliffe², Angela Tam¹, Jefferson Terry⁴, Nina Thiessen¹, Thomas Thomson², Richard Varhol¹, Thomas Zeng¹, Yongjun Zhao¹, Richard A Moore¹, David G Huntsman³, Inanc Birol¹, Martin Hirst¹, Robert A Holt¹, Marco A Marra¹

¹Genome Sciences Centre, British Columbia Cancer Agency

²British Columbia Cancer Agency

³Vancouver General Hospital

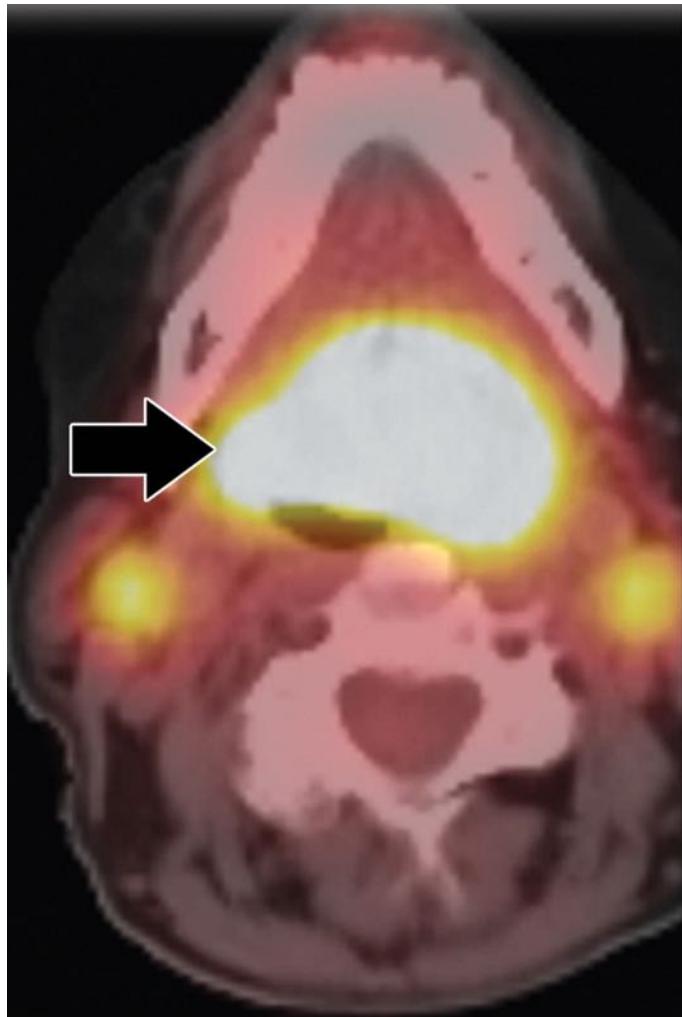
⁴Centre for Translational and Applied Genomics of British Columbia Cancer Agency and the Provincial Health Services Authority Laboratories

⁵Molecular Oncology, BC Cancer Research Centre

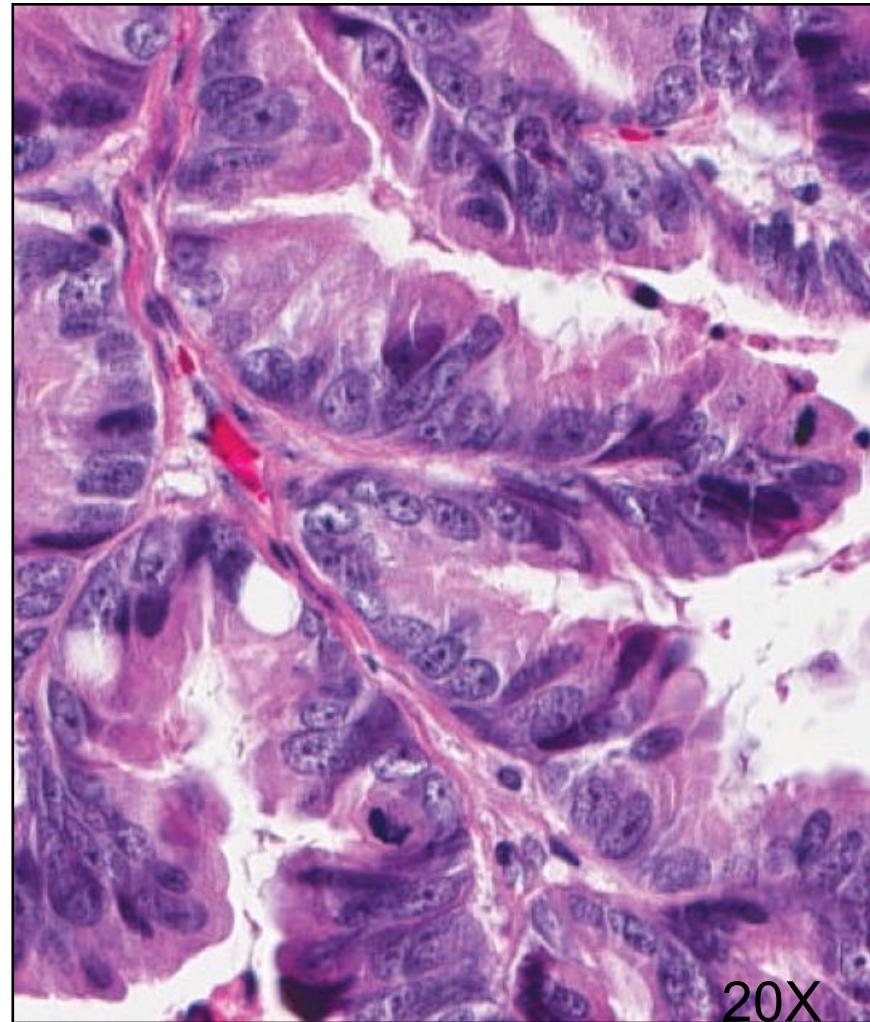
Initial presentation

- 78 year-old Caucasian man
- Fit and active
- Presented in August 2007 with throat discomfort
- Examination found a 2 cm mass at left base of tongue
- Minimal comorbidities
- No obvious risk factors for oropharyngeal malignancy

PET-CT scan & subsequent biopsy



Fukui et al. Radiographics.
2005 Jul-Aug;25(4):913-30.



Primary tongue mass,
H&E stain, 20X objective

Surgery & further pathology

Laser resection of tumor & lymph nodes

Primary: 1.5 cm poorly differentiated adenocarcinoma with micropapillary & mucinous features

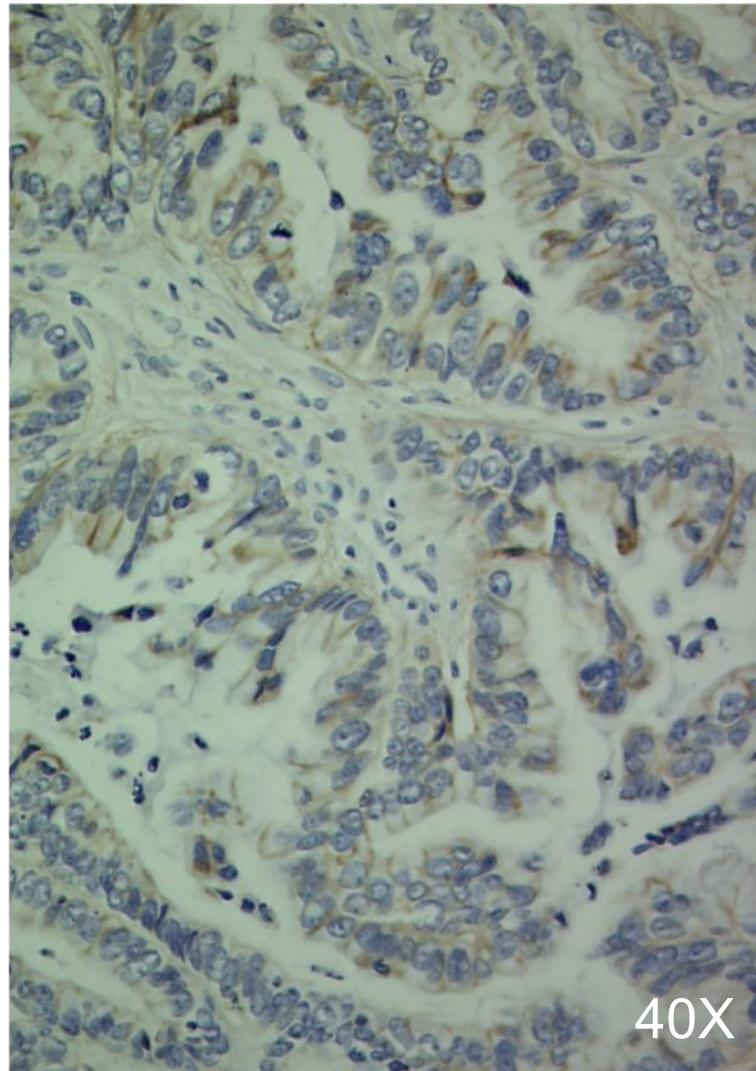
Lymph nodes: 3 of 21 neck nodes contain metastatic adenocarcinoma

60 Gy of adjuvant radiation therapy completed in Feb.

Good quality of life, returned to work for four months

Then...numerous small (< 1.2 cm) bilateral pulmonary metastases

EGFR IHC positive expression



+2 EGFR expression by Zymed
immunohistochemistry
protein expression assay

Treatments?

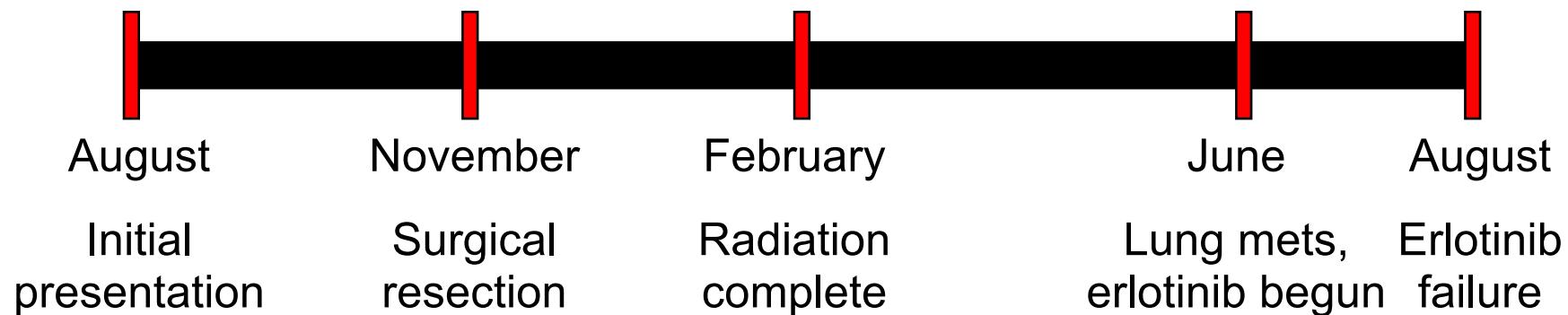
6 week trial of erlotinib

All pulmonary nodules grew while on erlotinib

Largest lesion grew from 1.5 to 2.1 cm

Erlotinib discontinued in August

Palliative care? What next?



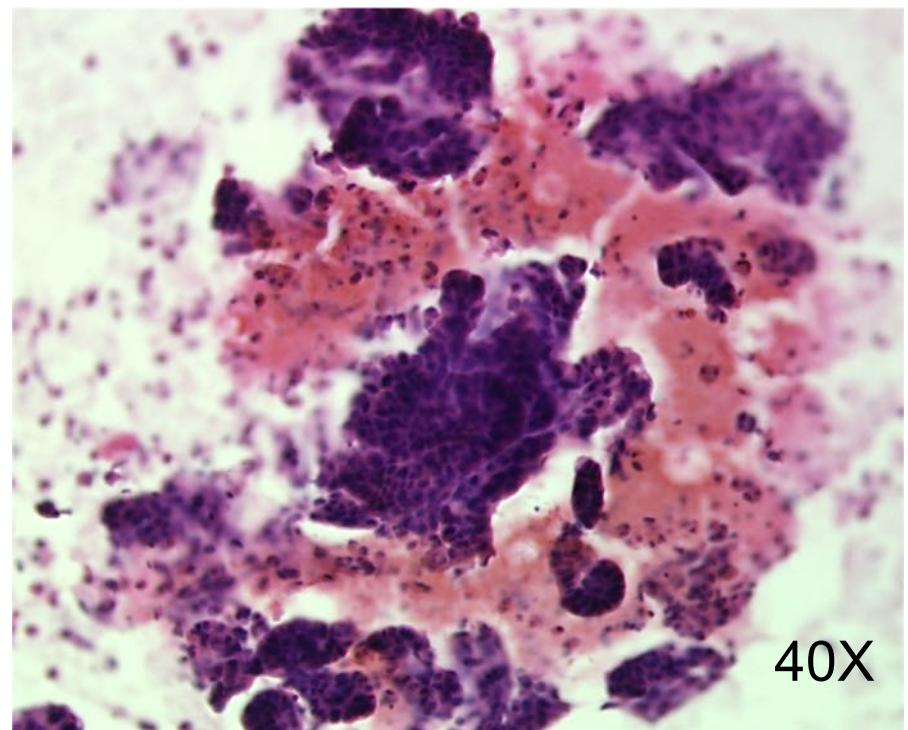
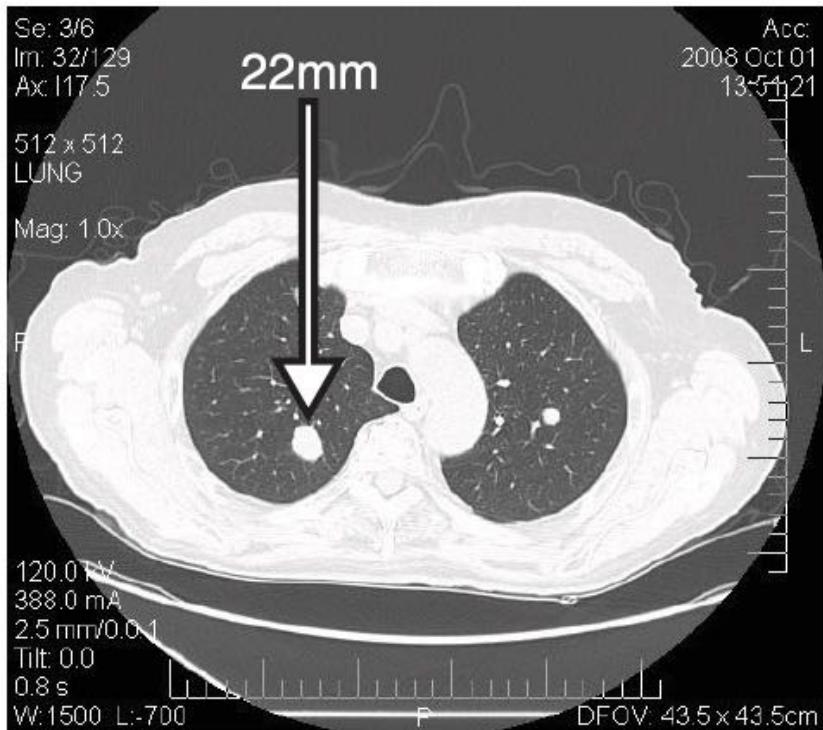
What are the targets in *our* case?

Having exhausted standard of care, BC Cancer Agency oncologist turned to the Genome Sciences Centre (a large research group) for new leads

Special meeting of the BCCA Research Ethics Board approved a single-case analysis for this patient

Patient consented to full genomic sequencing and analysis with the understanding that novel treatment options may be *suggested*

Fresh frozen biopsy taken for RNA-seq



Fine-needle aspirates taken from large lung lesion
Pathologist reviewed, used samples with highest tumor
content (~80%) for RNA sequencing

FFPE DNA from surgical resection used for DNA

Somatic Mutations

<i>TP53</i>	p.D259Y	Located in region of LOH Known somatic mutation in 7 cancer types (16 entries + 4 alternate mutations across 28 tumors in the Catalogue of Somatic Mutations in Cancer)
<i>RB1</i>	p.L234*	Novel truncating mutation Results in loss of 75% of protein sequence Loss of <i>RB1</i> and <i>PTEN</i> result in gefitinib resistance (<i>Albitar et al. Gynecol Oncol. 2007 Jul;106(1):94-104.</i>)
<i>ZNF187</i>	p.G63C	Novel mutation of unknown significance Zinc-finger protein
<i>ZFPM2</i>	p.K785E	Novel mutation of unknown significance Zinc-finger protein

Confirmed by secondary sequencing method (Sanger)

Copy number alterations

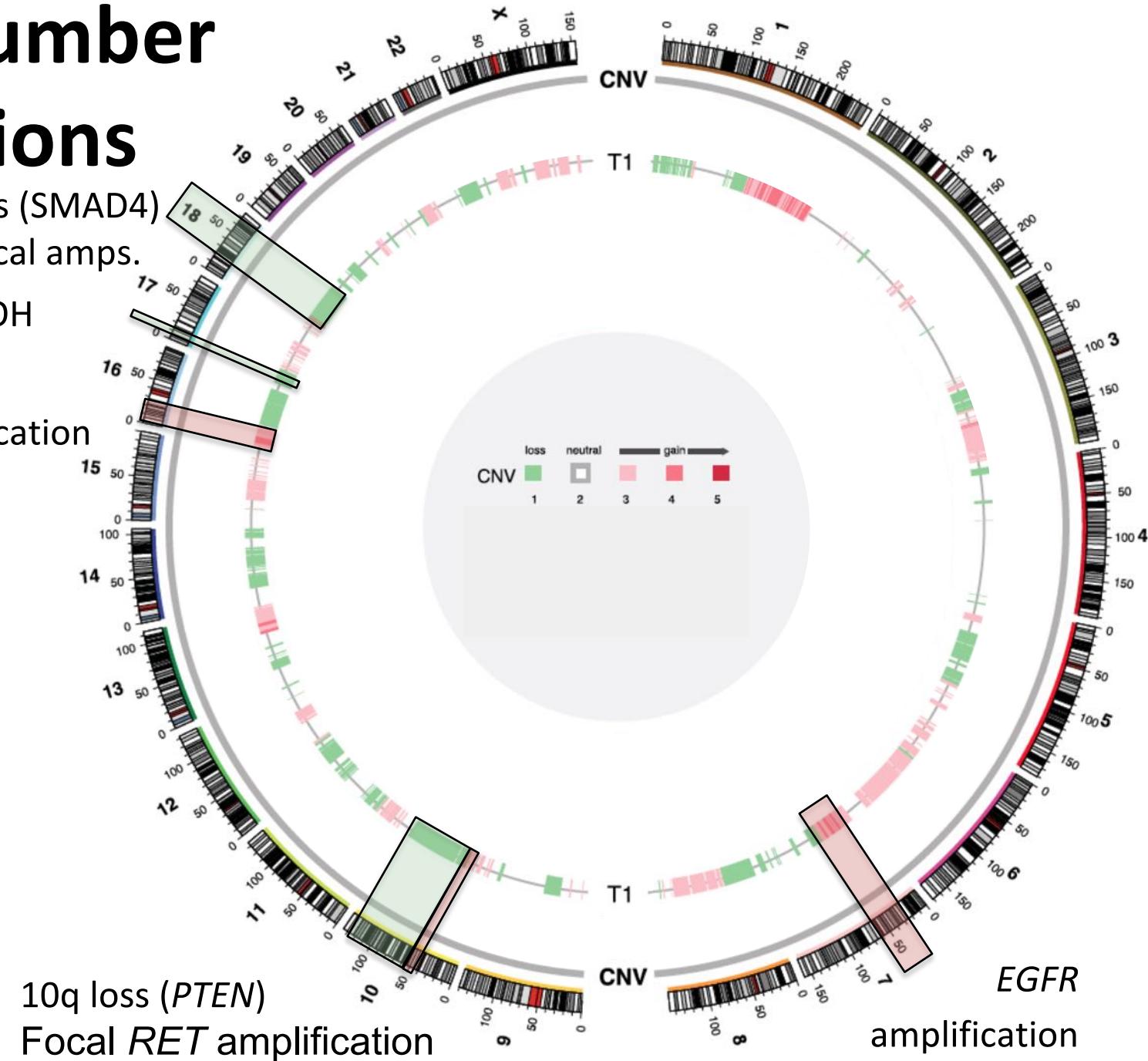
18q loss (*SMAD4*)

with focal amps.

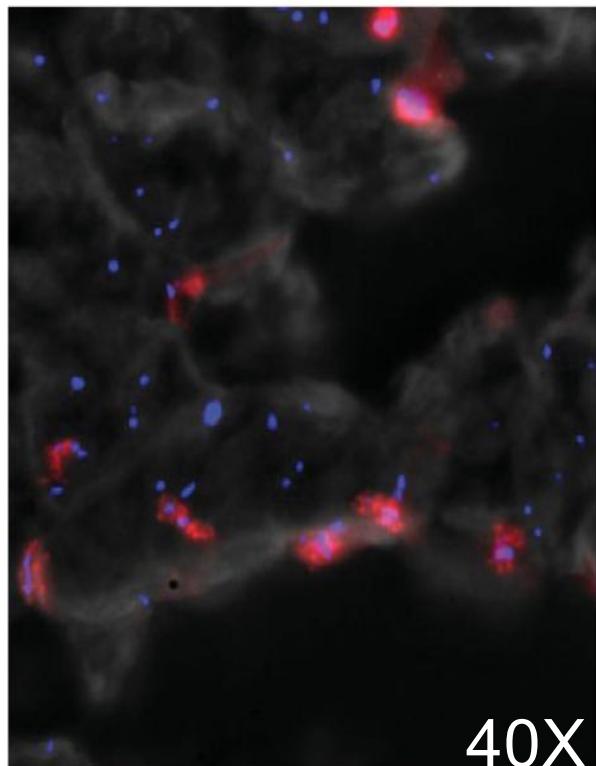
TP53 LOH

MAPK3

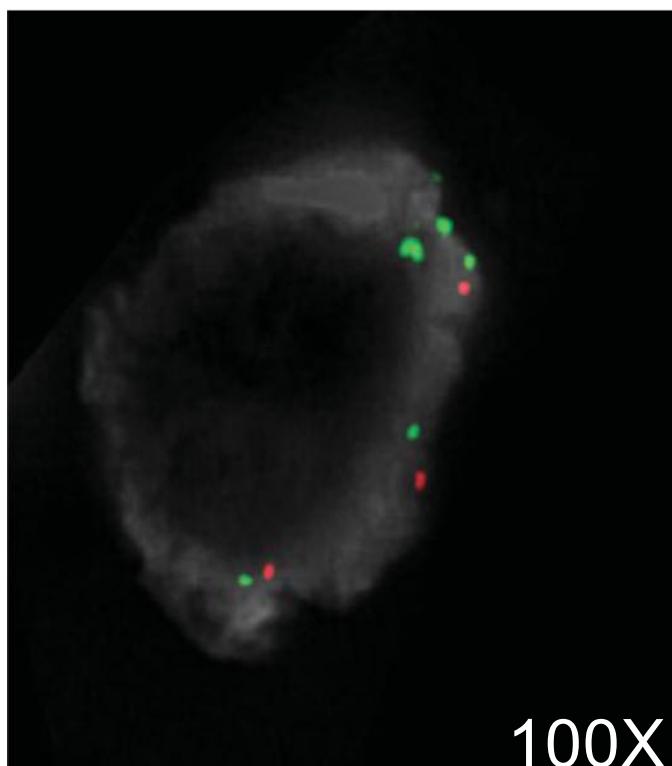
amplification



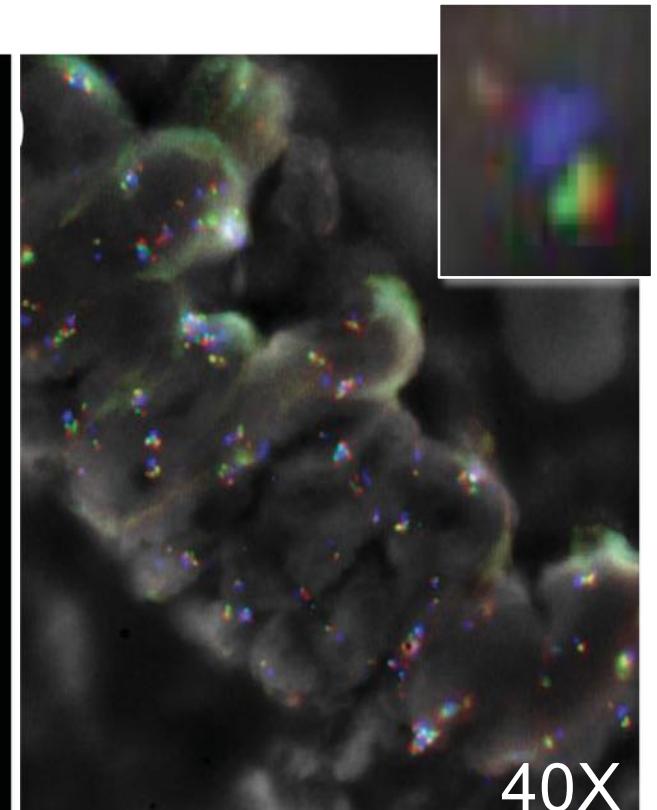
FISH confirmation of copy number alterations



RBBP8
amplification



PTEN single
copy loss



Focal *RET*
amplification

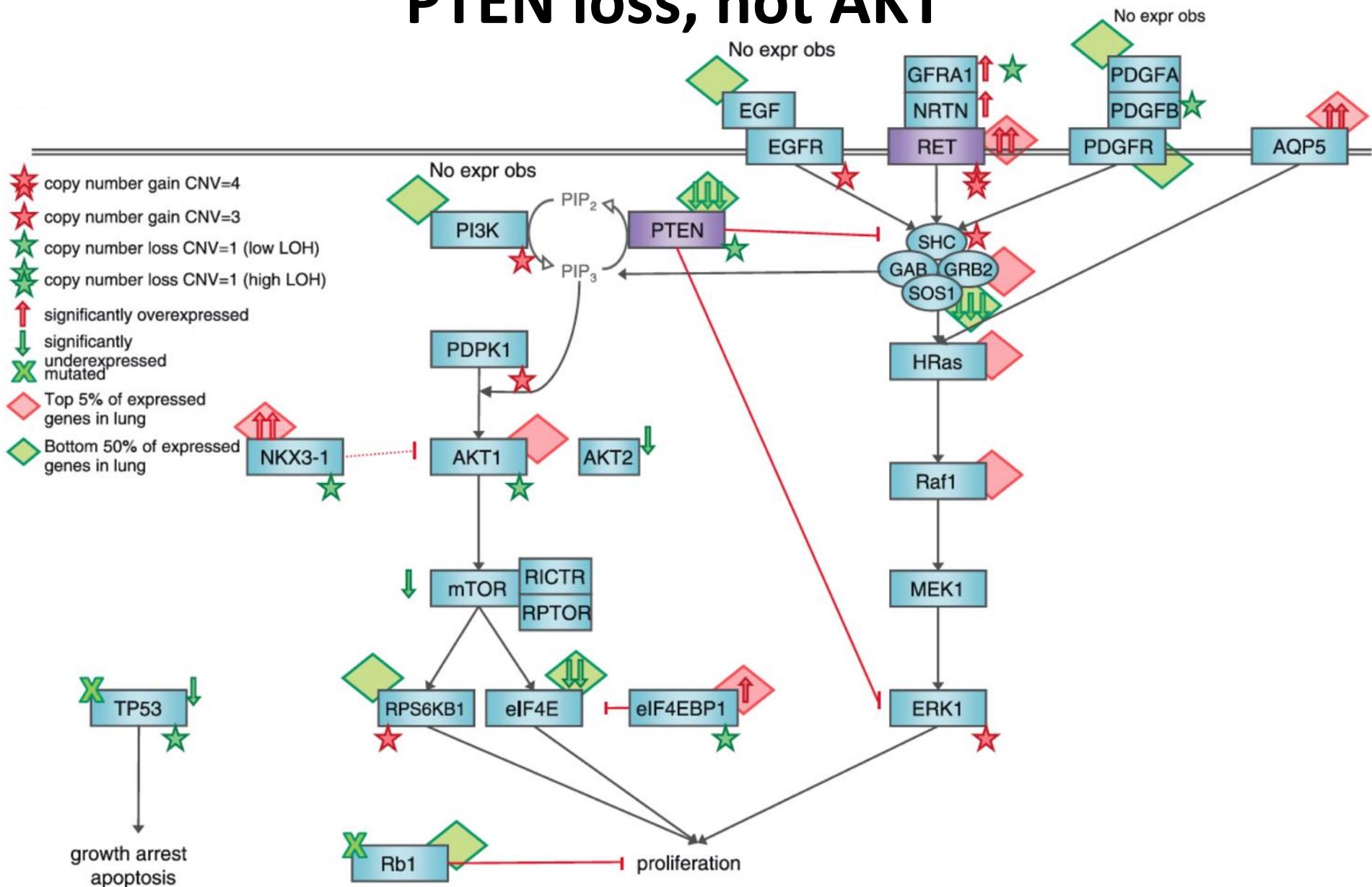
Tumour gene expression levels compared against reference set of 50 tumors & matched blood sample

SMAD4 expression 43X lower vs compendium
deleted & down-regulated, associated with metastasis of colorectal cancer

RET in top 5% of expressed genes, 34X vs compendium
most highly expressed oncogene

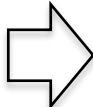
PTEN in bottom 5% of expressed genes,
significantly under-expressed

Tumor driven by RET up-regulation, PTEN loss, not AKT



Short list presented to oncologist

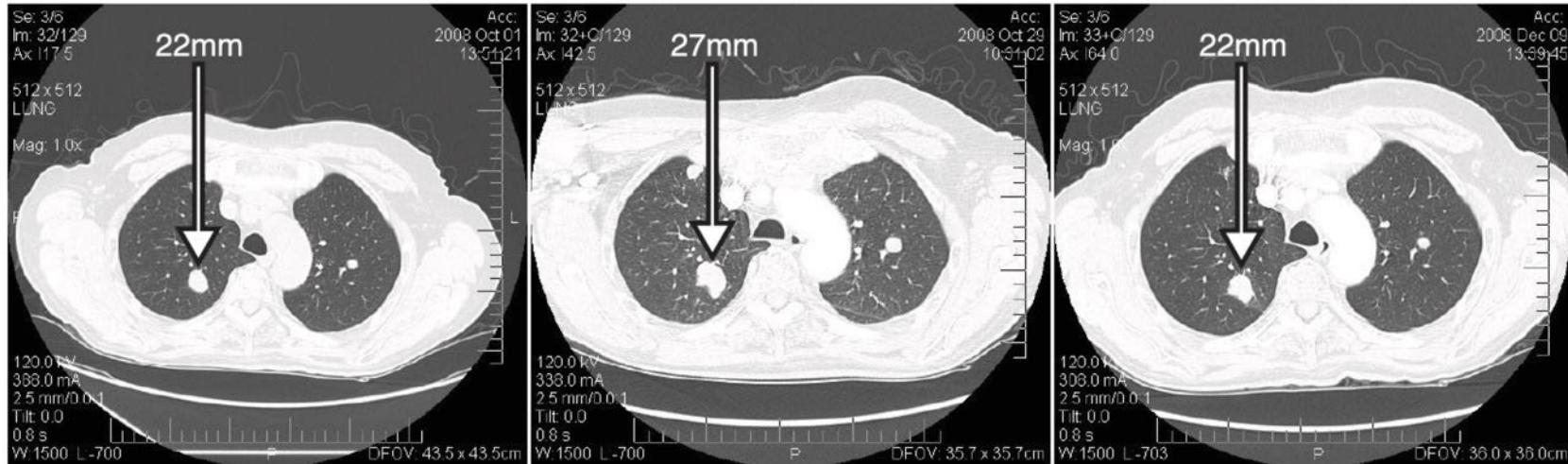
Targeted aberrations
• Up-regulation the MAPK pathway increases cell proliferation.
• RET, a validated thyroid cancer target, and its growth factors are amplified and overexpressed
• AQP5 a known activator of this pathway is overexpressed
• MAPK3 (ERK1) is amplified.
• PTEN, a suppressor of this pathway, is highly down-regulated.



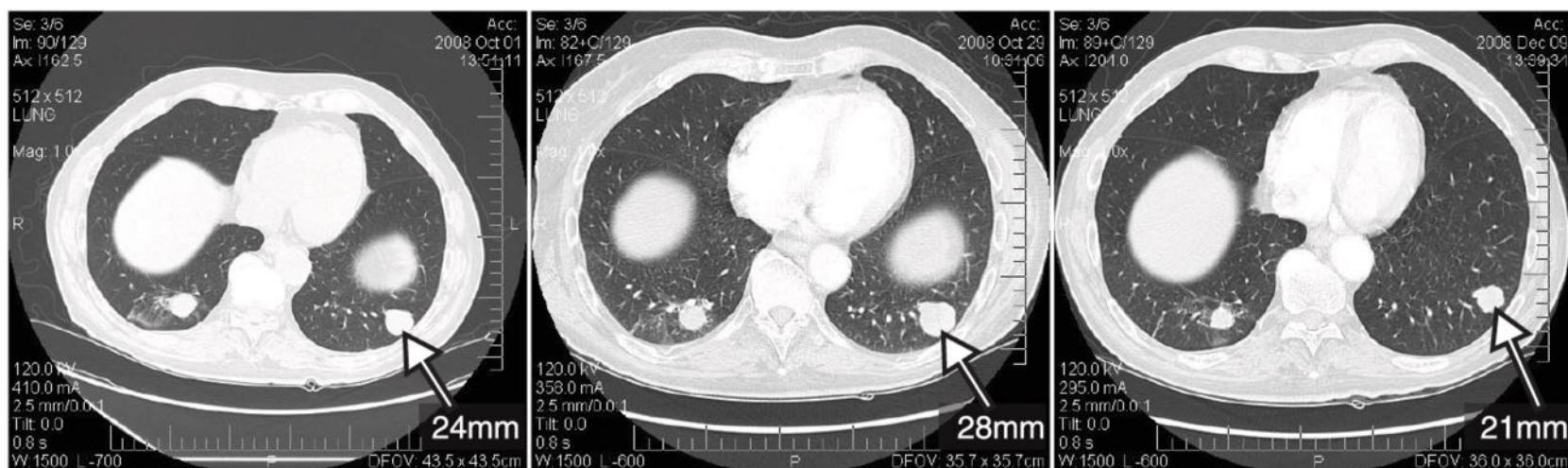
Drug	Known mechanism & indications
Sunitinib	Targets PDGFRs, VEGFRs, RET, KIT, CSF1R, FLT3. Approved for GIST and RCC. In trials for thyroid cancer.
Motesanib	Targets VEGFRs, PDGFRs, KIT, RET. In trials for thyroid cancer, GIST, NSCLC.
Sorafenib	Targets BRAF, RAF1, RET, VEGFRs, PDGFRB, KIT, FLT3. Approved for RCC and HCC. In trials for thyroid cancer.
Sulindac	An NSAID COX inhibitor for inflammation but also inhibits MAPK3 (ERK1).

22% decrease after 4 weeks on sunitinib (16% growth prior to sunitinib)

Nodule 1



Nodule 2



Oct 1st 2008

Biopsy taken

Oct 29th 2008

Sunitinib begun

Dec 9th 2008

4 wks 50mg daily Sunitinib

Stabilization for 7 months

Sunitinib dose reduced due to known side-effects, otherwise excellent quality of life

Repeated scans showed no new nodules and disease stabilization and 4 months

...then, existing lung mets began to grow

Switched to sorafenib & sulindac

Disease again stabilized within 4 weeks and continued for 3 months

Recurrent disease after 7 months

Recurrent disease at primary site on tongue

New neck skin nodule

Progressive & new metastases in lung

Deteriorating quality of life

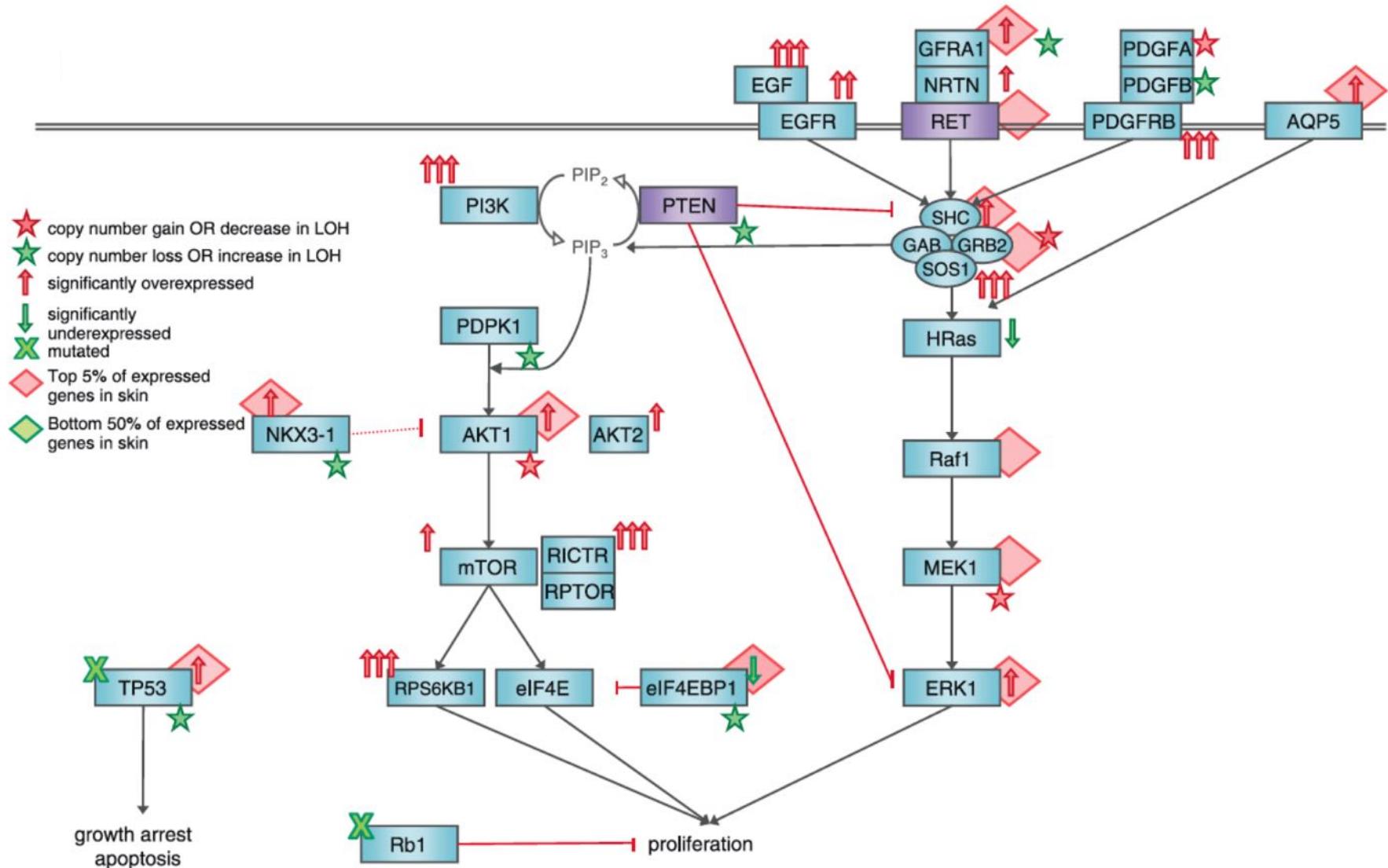
What changed?

Skin metastasis contains new mutations

ZMYM4	p.Q317H	Zinc-finger
DNAH7	p.V2590L	Force generating protein of respiratory cilia
CXCL13	p.R56H	B-cell attracting chemokine
HSD17B8	p.A141T	hydroxysteroid (17-beta) dehydrogenase 8
PCLO	p.T2759A	presynaptic cytomatrix protein
GRIA4	p.R872C	glutamate receptor
OR4K2	p.L197F	olfactory receptor
SYNE2	p.A302G	tethers the nucleus to the cytoskeleton
PTPRM	p.A929T	cell-cell adhesion, possibly signal transduction & growth

No evidence of these in the pre-treatment biopsy,
even at low frequency

Resistance due to intensified RET and new AKT signaling?



Overcoming mechanisms of resistance

Cocktail of targeted drugs against multiple members of activated and parallel pathways?

- RET, EGFR, mTOR, Akt, etc.
- untested, risk of adverse side-effects

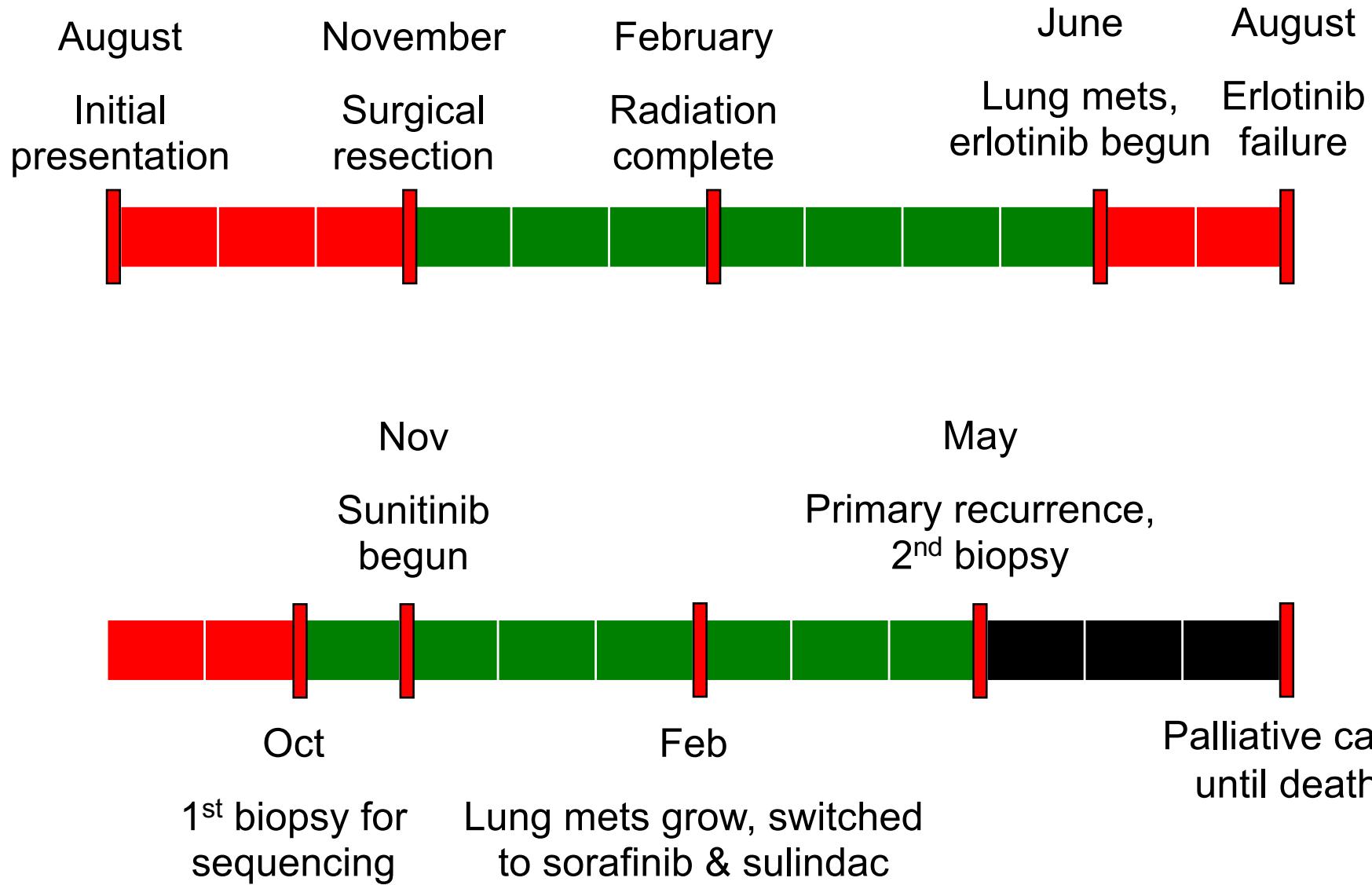
Detection of resistance mechanisms pre-treatment

→ none of the new mutations were evident pre-treatment

Can resistance be modeled & monitored?

- serial biopsies? blood test?

Timeline



We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics

MiCM McGill initiative in
Computational Medicine