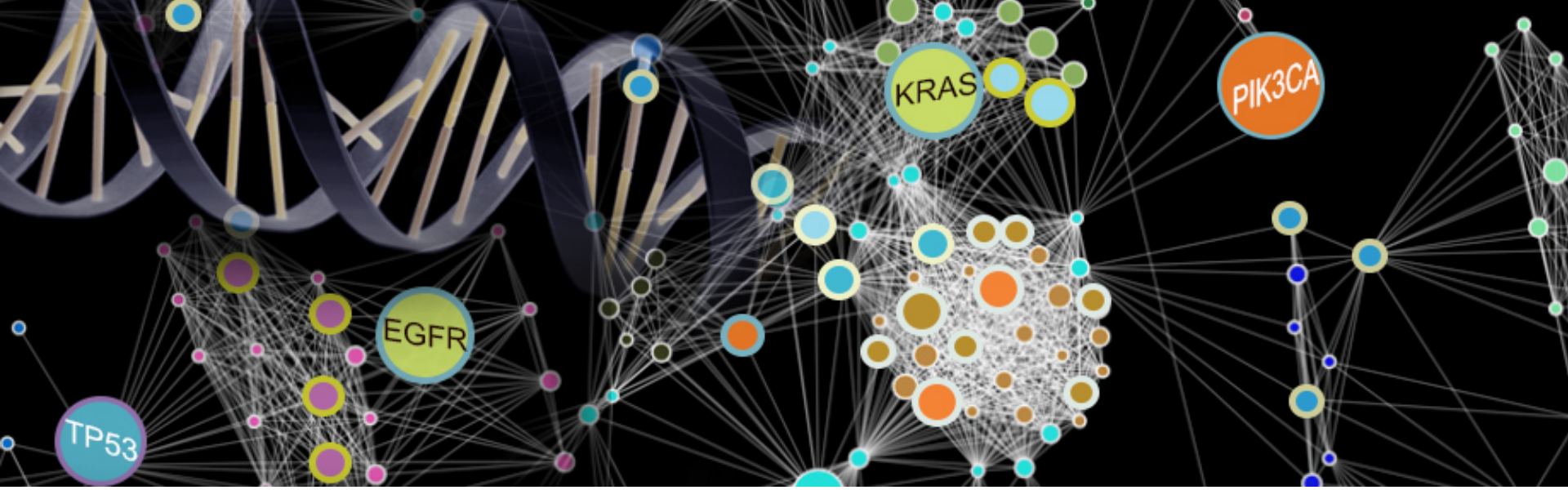
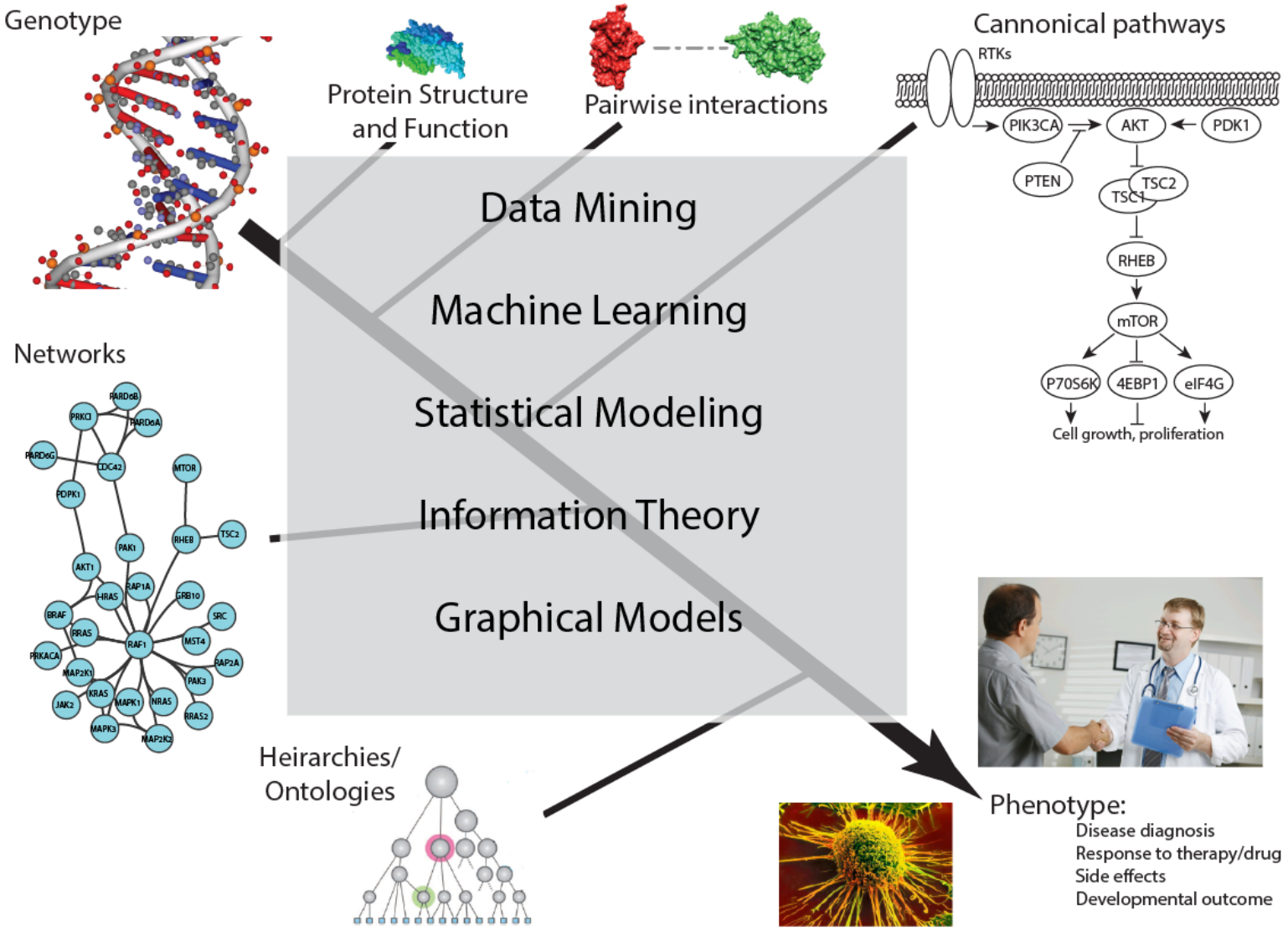


Mutational Burden of Tumors

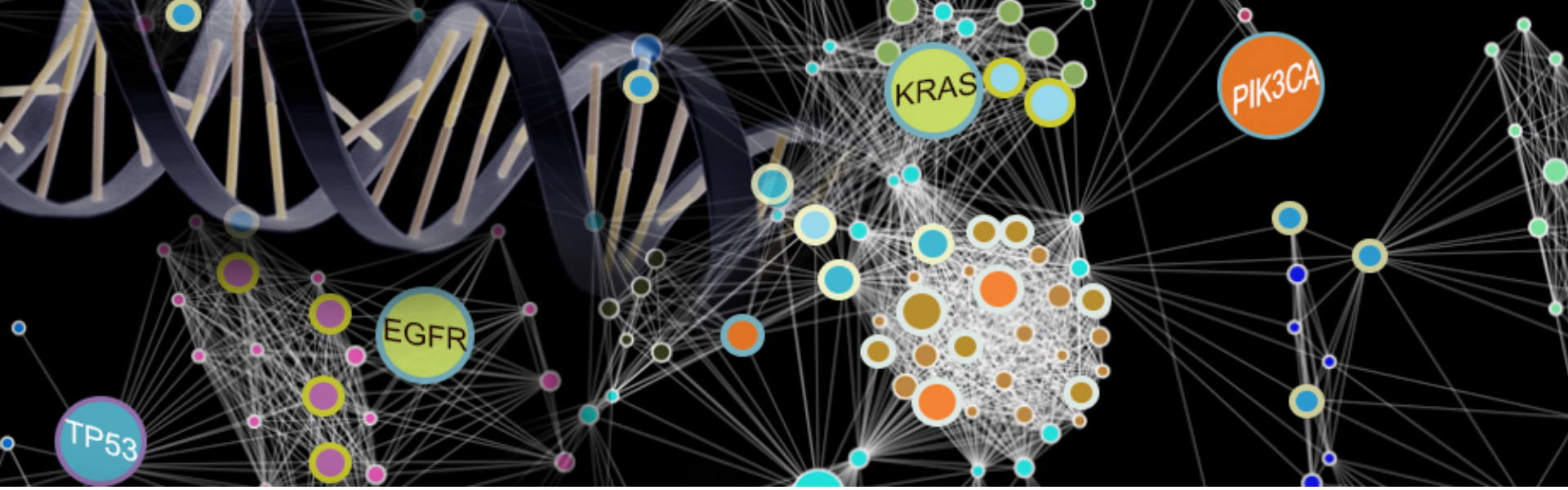
May 28, 2018
Hannah Carter, PhD
University of California, San Diego
Department of Medicine

Studying mutations at multiple scales



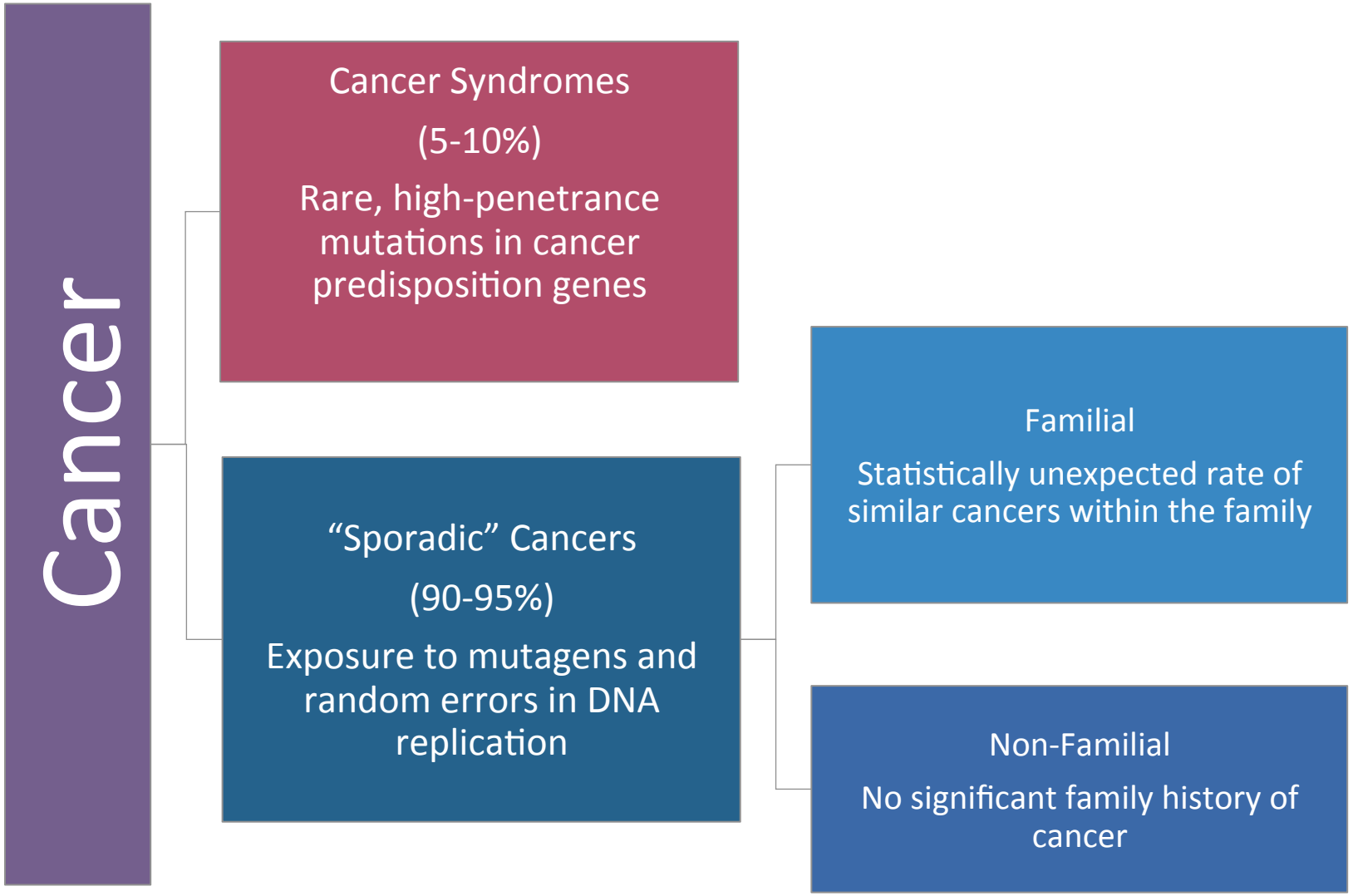
Outline

- Cancer as a genetic disease
- Mutation burden in tumors
- Identifying driver mutations in cancer
- Tumor heterogeneity
- Example: predicting driver mutations



Cancer as a genetic disease

Cancer is a genetic disease



Familial cancer risk

Updated estimates of cancer heritability from a study of > 80,000 twins

Heritability of cancer overall was 33% (95% CI, 30%-37%)

Cancer	Heritability	95% CI
Melanoma	58%	43-73%
Prostate	57%	51-63%
NonMelanoma Skin	43%	26-59%
Ovarian	39%	23-55%
Kidney	38%	21-55%
Breast	31%	11-51%
Corpus Uteri	27%	11-43%

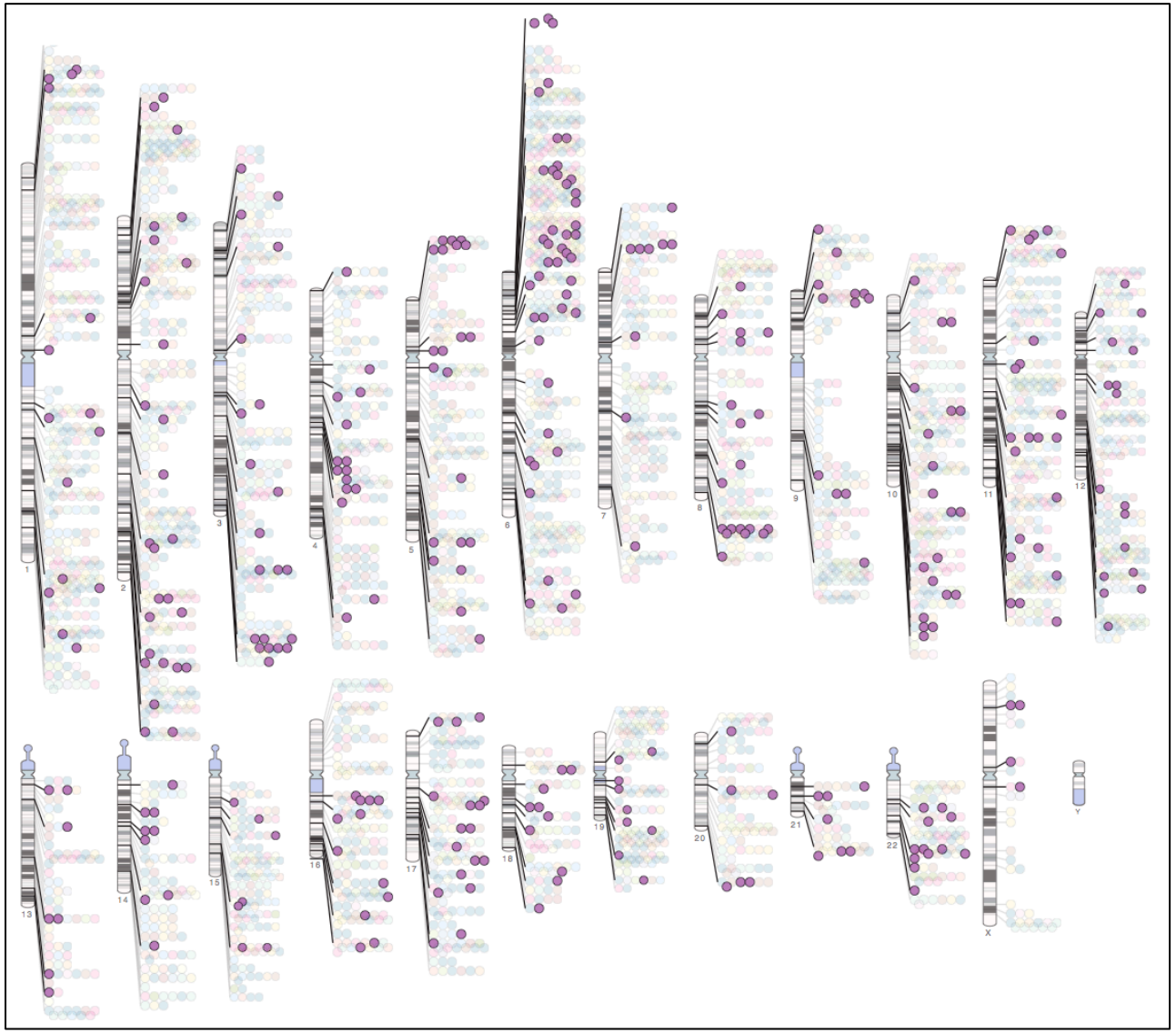
Heritability: proportion of variance in cancer risk due to inter-individual genetic differences

Mucci *et al 2016 Jama*

Common cancer risk variants: Genome-Wide Association Studies

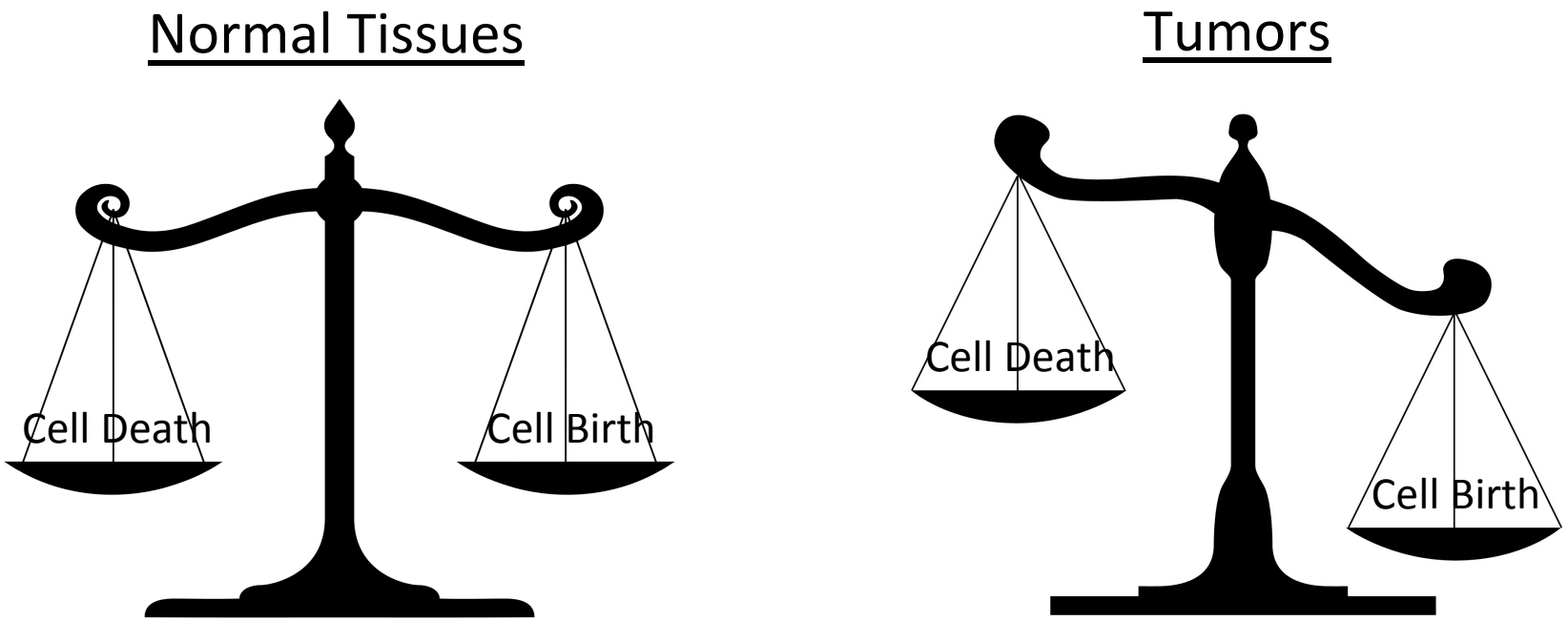
NHGRI GWAS Catalog:

- 165+ cancer studies
- Over 500 risk SNPs
- Population specific



Welter *et al*, *Nucleic Acids Research* 2014

Somatic genetic alterations in cancer

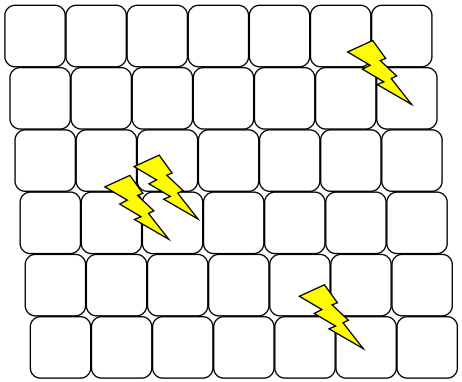


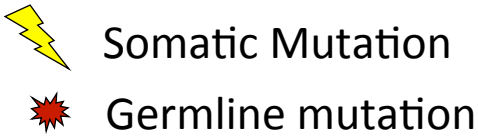
Accumulating genetic and epigenetic changes result in cancer by altering the balance between cell birth and cell death.

Concept: Germline-somatic interactions

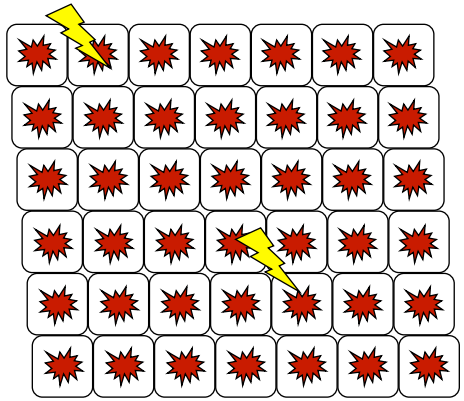
Cancer results from multiple mutations co-occurring in the same cell

- Sporadic cancers
  - Somatic mutations only

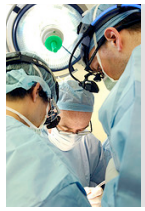




- Genetic background
  - Germline variants could modify the fitness effects of somatic mutations

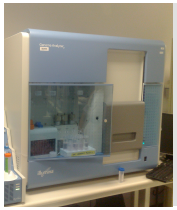


Cataloging mutations in tumors by DNA sequencing



Tumor DNA

Normal DNA



Sequencing reads

Sequencing reads

Alignment to reference

Alignment to reference

MuTect
eBayes
Strelka
…

Compare

Somatic Mutations

Types of cancer mutations

Small-Scale Mutations	
Point mutation	Change in a single nucleotide
Silent mutation	Amino acid sequence is not changed
Missense mutation	Amino acid sequence is changed
Nonsense mutation	Amino acid sequence is changed to a stop codon, therefore truncating the protein
Insertion	Addition of one or more extra nucleotides
Deletion	Removal of one or more nucleotides
Large-Scale Mutations	
Amplification	Multiple copies of chromosomal region
Deletion	Loss of chromosomal region
Translocation	Interchange of regions from different chromosomes
Inversion	Reversal of the orientation of a chromosomal region
Loss of heterozygosity	Loss of one allele

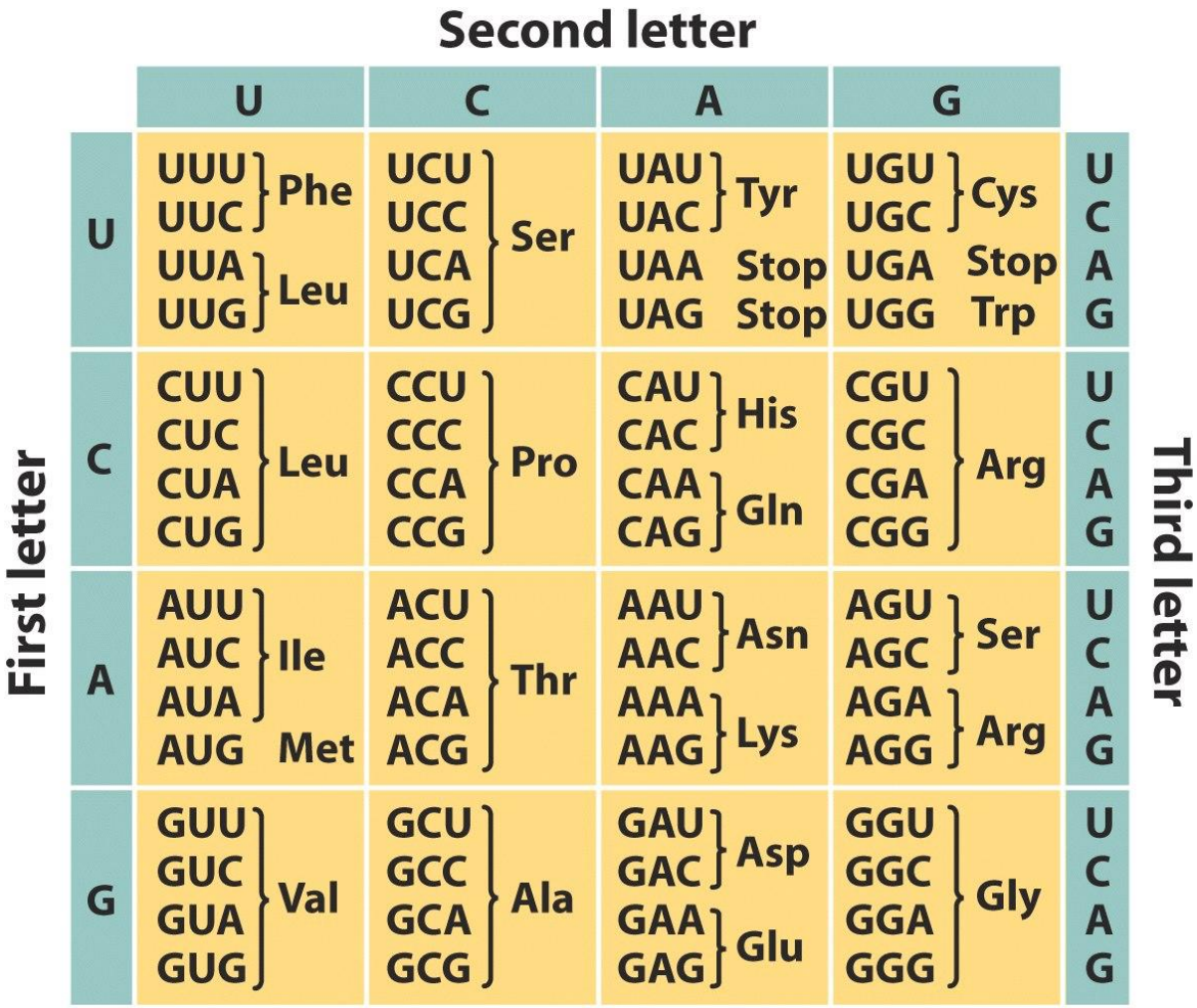
Most tumor sequencing data is whole exome

Most analyses still focus predominantly on point mutations

Mutation burden = number of mutations per sequenced MB

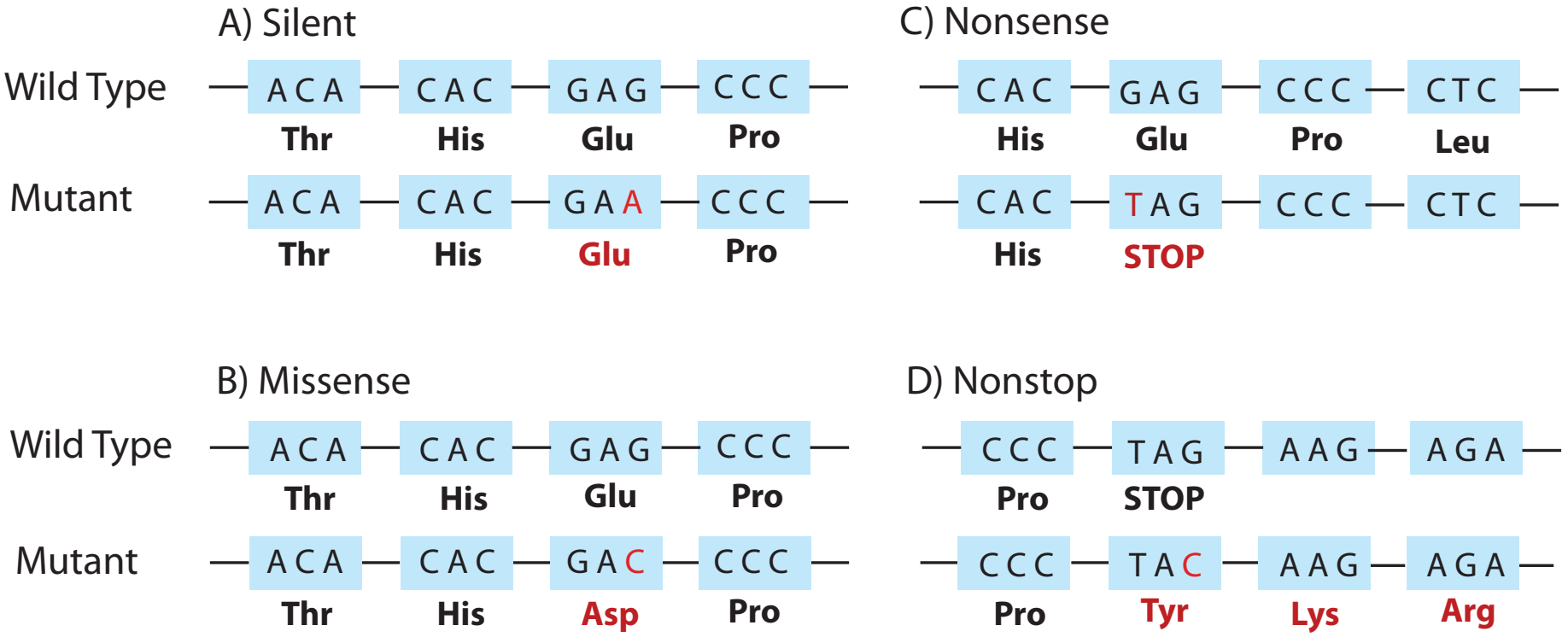
Coding Base Substitutions

- Genetic Code
  - 3 bases = 1AA
  - "Degeneracy"
    - 20 AAs, 64 possible 3 base combinations
    - Not every base substitution will result in an AA change
    - 3rd position is least likely to result in a change



Coding Base Substitutions

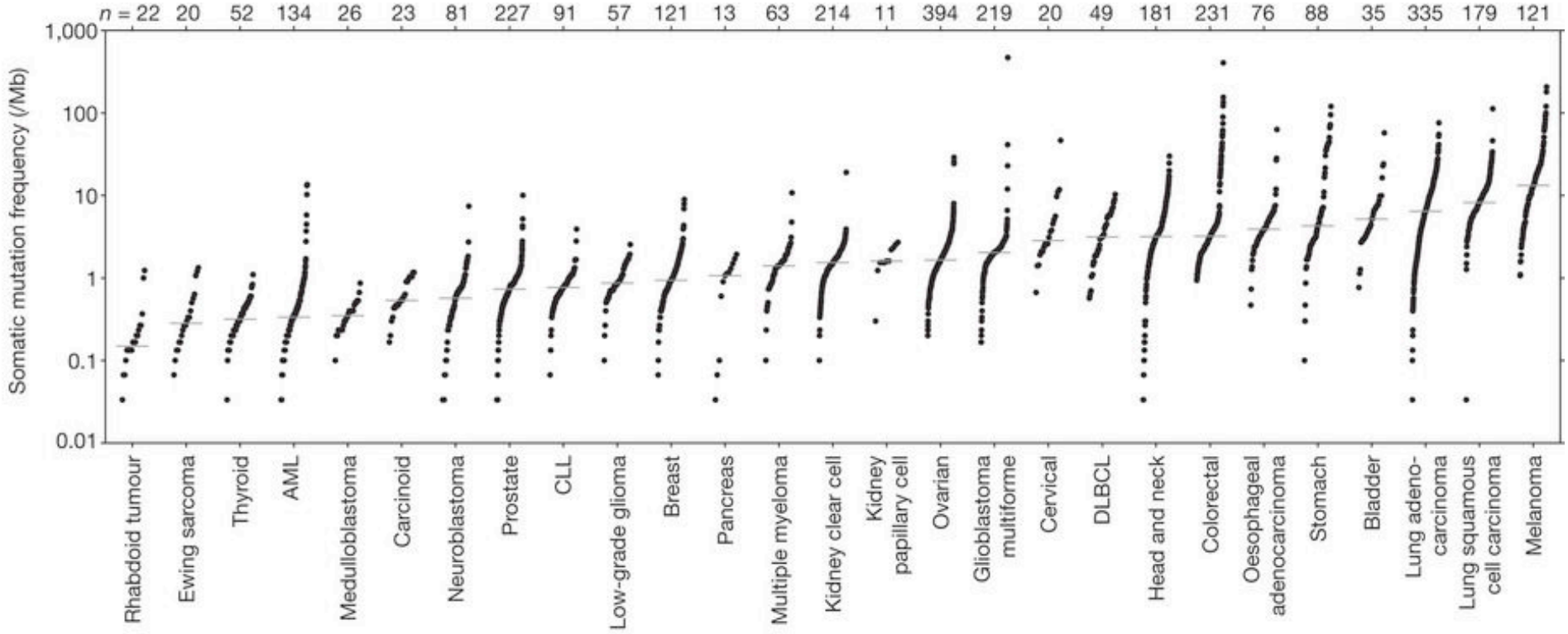
Synonymous (or silent) Vs Non-synonymous



Somatic mutations uncovered by large-scale tumor exome sequencing

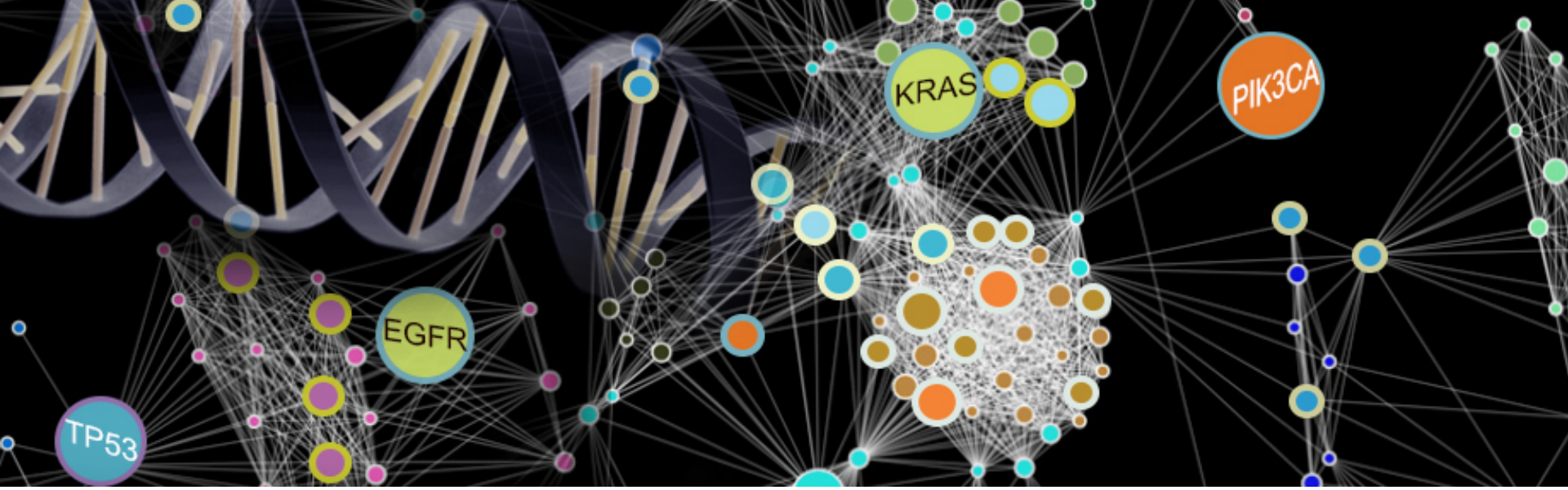
Tumor Type	Samples	Total Mutations	Missense	Silent	Nonsense	Nonstop	Splice	Frameshift	Muts/Samples
UCEC	542	492965	321267	100586	34237	403	10556	25916	909
SKCM	468	348935	212749	113524	16329	96	4525	1712	745
COAD	433	157925	95723	31742	8605	117	3049	18689	364
LUAD	567	174019	115051	38250	10162	165	4417	5974	306
STAD	439	133647	79156	28037	4767	108	2595	18984	304
LUSC	494	144624	95239	32962	7824	157	3581	4861	292
BLCA	412	107105	68343	26183	7785	138	1924	2732	259
READ	158	40260	27559	7134	3368	41	765	1393	254
CESC	305	64083	40664	15391	4003	93	1103	2829	210
GBM	396	63395	42546	13060	3449	60	1471	2809	160
HNSC	509	80761	51156	19274	4972	67	1686	3606	158
ESCA	184	25301	16519	5578	1155	27	505	1517	137
OV	443	60955	39682	11906	2962	62	1816	4527	137
PAAD	180	22487	14220	5522	1863	10	444	428	124
DLBC	48	5764	3742	1407	285	6	135	189	120
UCS	57	6800	4594	1238	505	9	137	317	119
LIHC	375	39321	25590	8804	1811	48	1136	1932	104
BRCA	1044	93317	58034	19406	6254	96	2360	7167	89
KIRP	288	23996	14740	5597	1047	30	521	2061	83
ACC	92	7200	4642	1568	422	7	170	391	78
SARC	255	17371	11369	4217	720	20	387	658	68
KIRC	339	21058	12161	4112	1290	22	577	2896	62
LGG	512	26084	16627	5886	1365	13	689	1504	50
LAML	145	5763	3363	1092	241	27	719	321	39

Mutation burden by tumor type



Lawrence et al 2013

Where do these mutations come from?

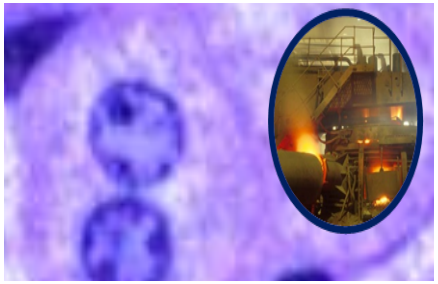


Mutation burden in tumors

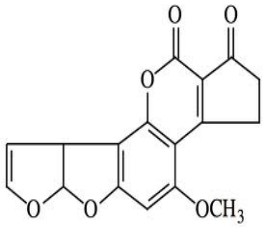
Why do tumor cells carry so many mutations?

Why is the burden different for different types of cancer?

Somatic mutations occur in all cells of the body throughout life



.....ATCGGGAATCGGACCCGATG.....



Classification of base substitution mutations

C>T
C>A
C>G
T>A
T>C
T>G

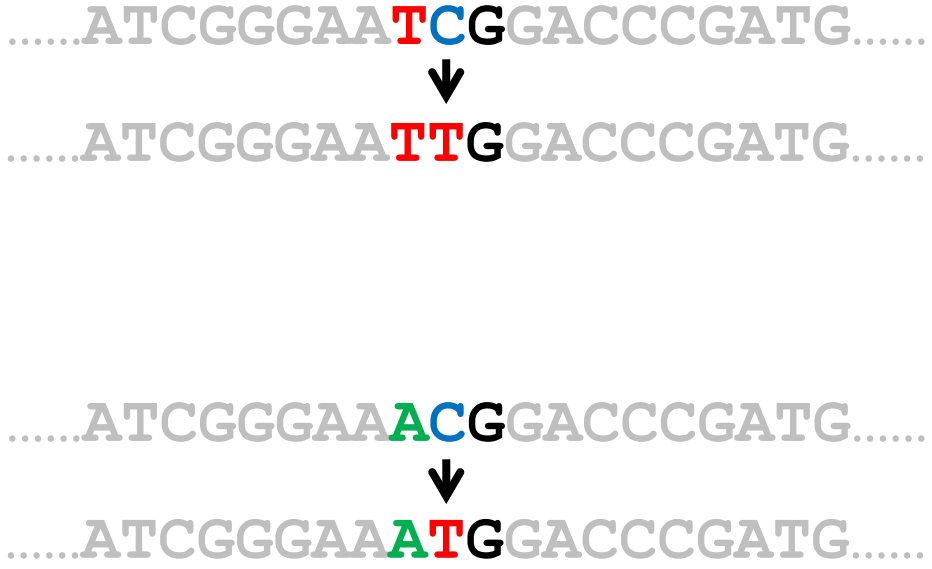
Since DNA bases are paired every possible substitution can be summarized with 6 categories

By convention, these are defined with respect to the pyrimidine (C or T)

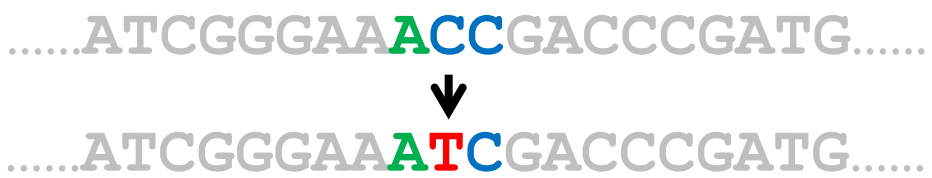
Classification of base substitution mutations

Classification of base substitution mutations

Classification of base substitution mutations



Classification of base substitution mutations

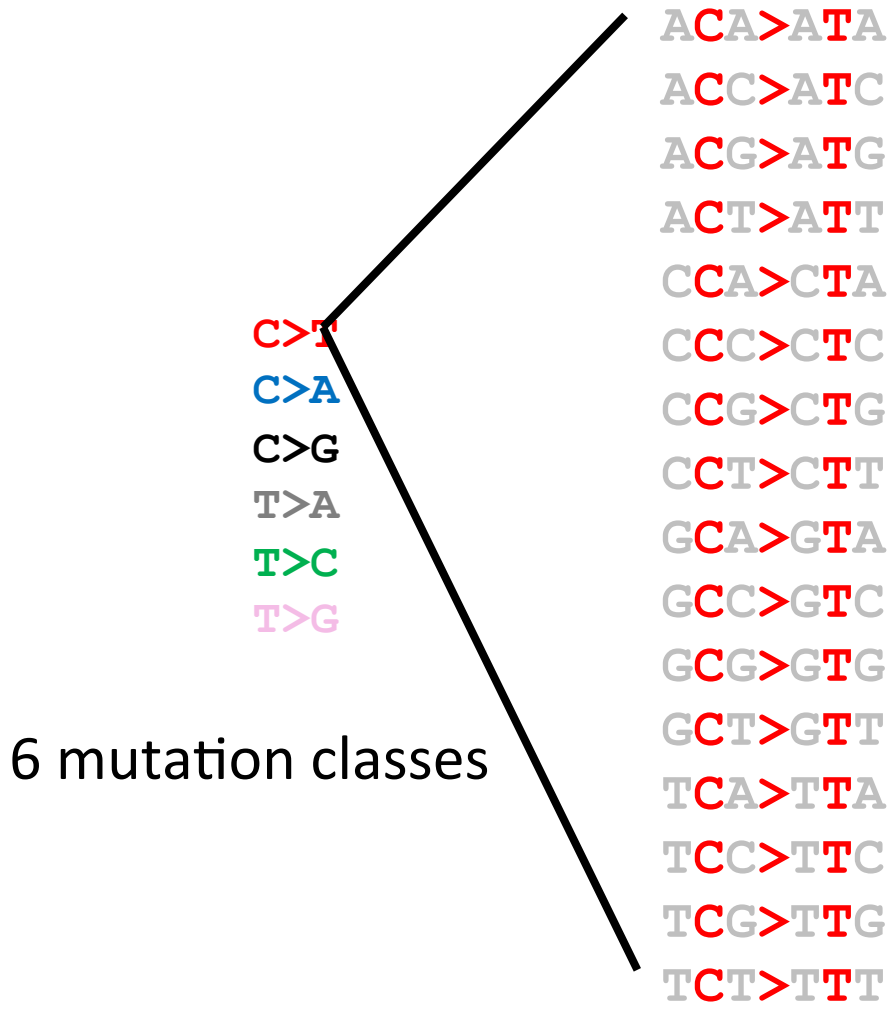


Classification of base substitution mutations

C>T
C>A
C>G
T>A
T>C
T>G

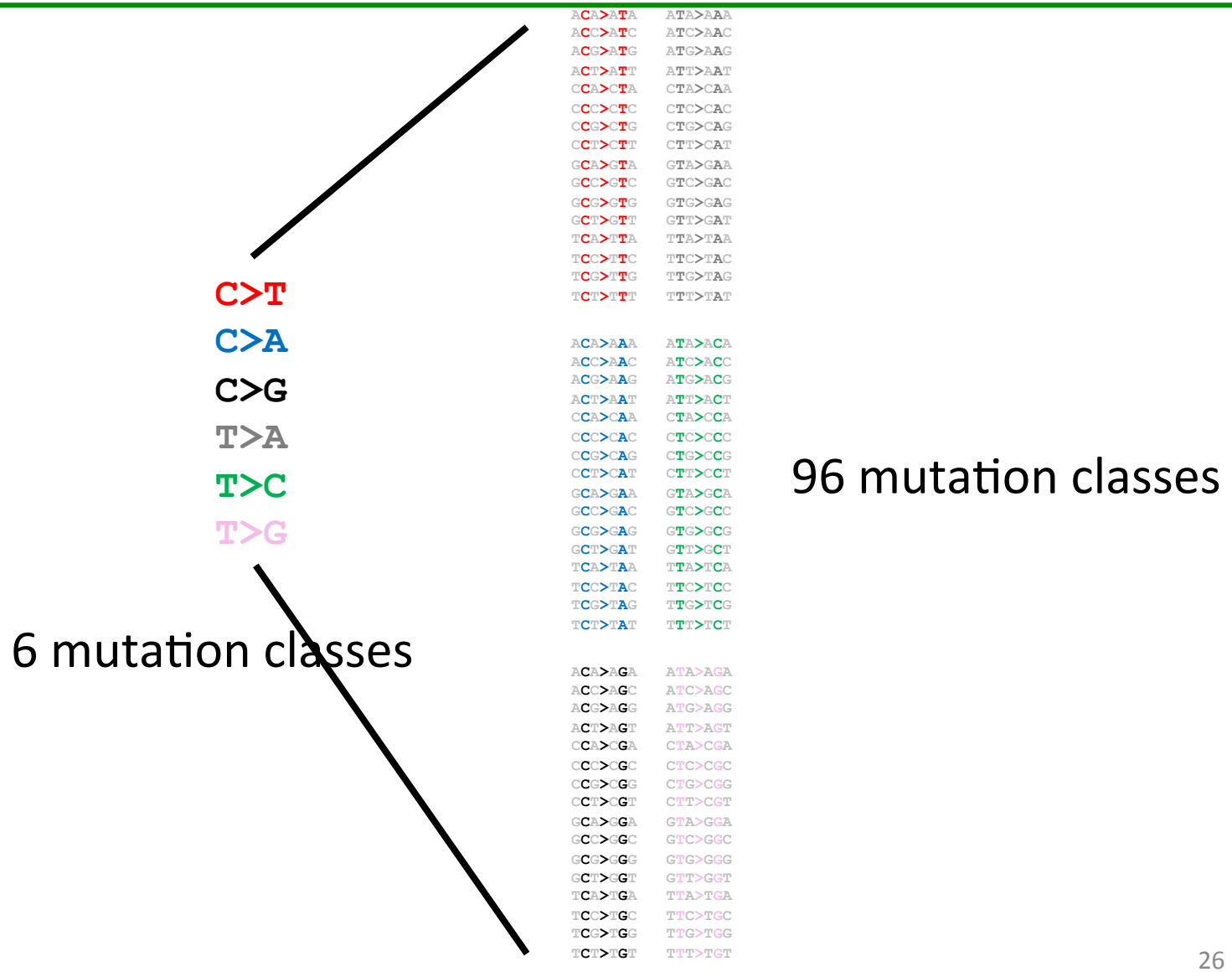
6 mutation classes

Classification of base substitution mutations

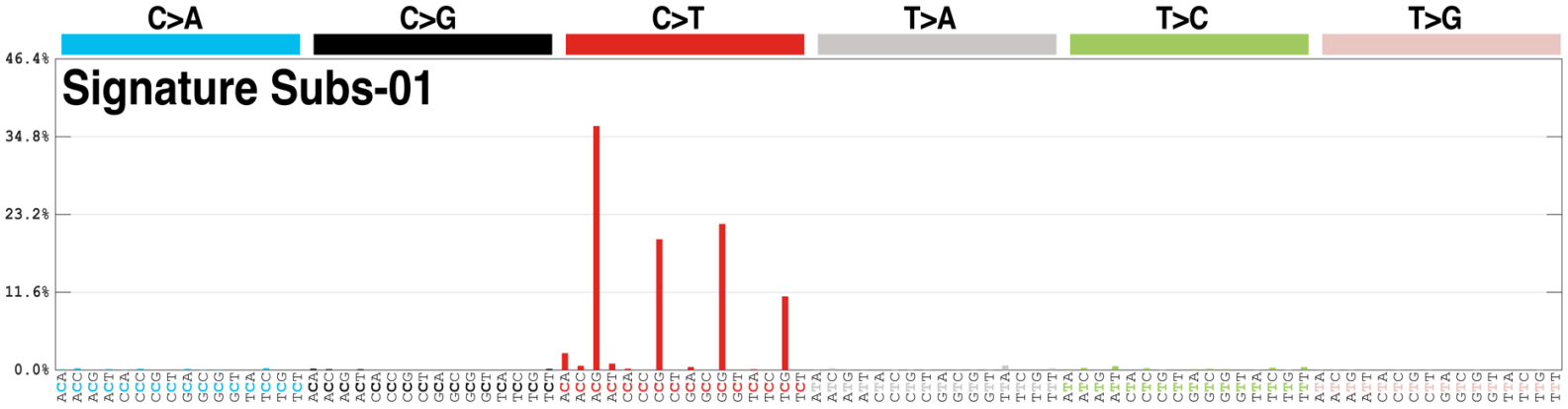


6 mutation classes

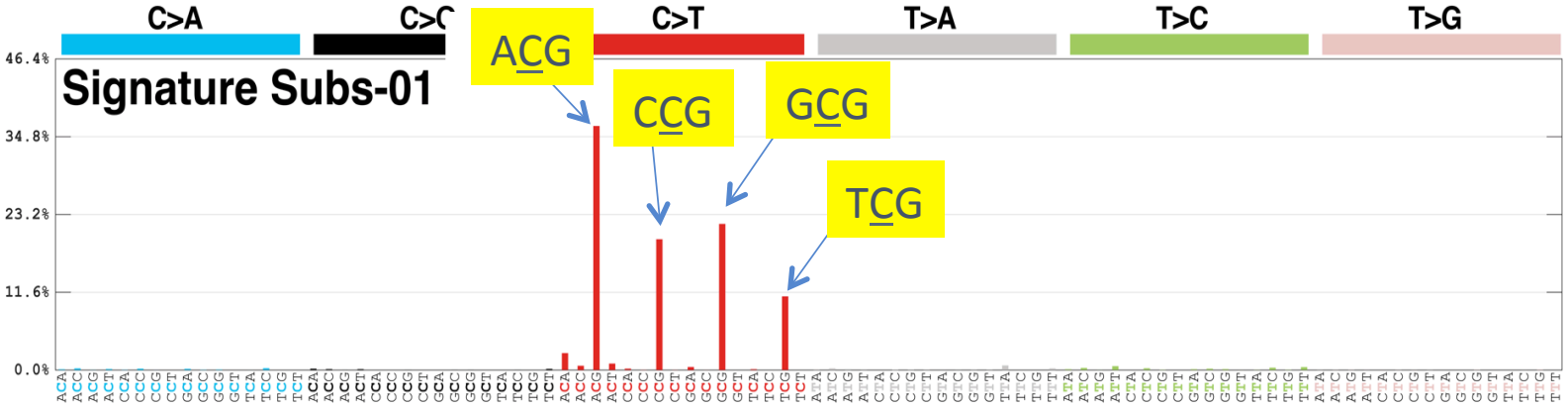
Classification of base substitution mutations



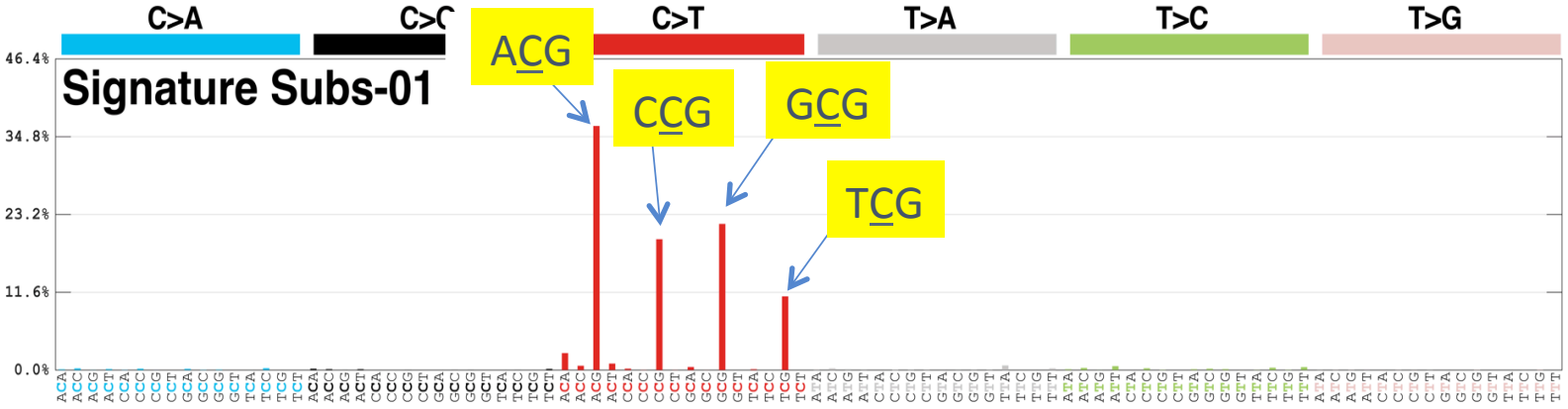
Mutational signatures in human cancer

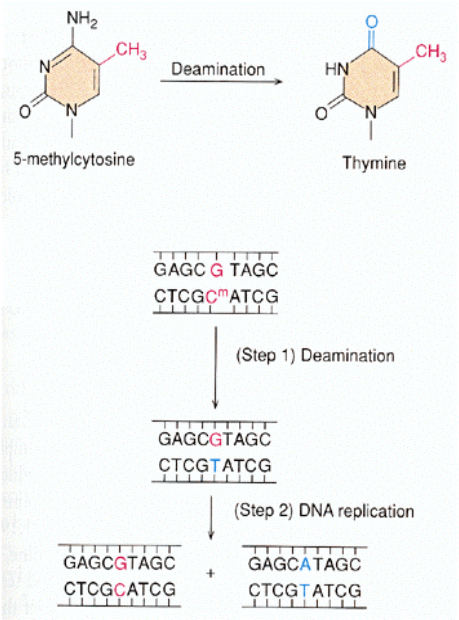


Mutational signatures in human cancer

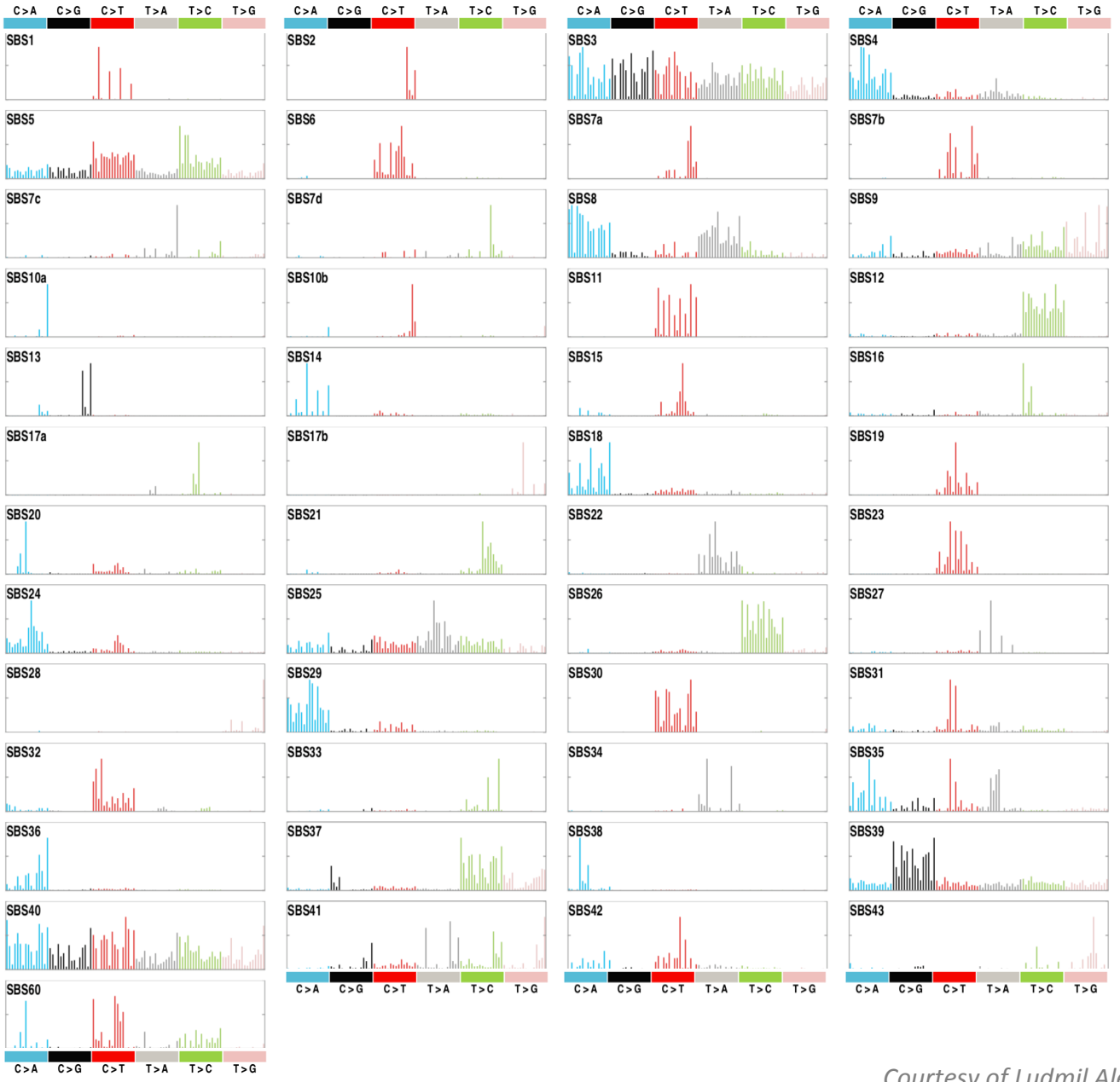


Mutational signatures in human cancer





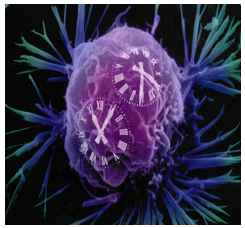
Aetiologies of substitution mutational signatures



Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Clock-like processes



Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Tobacco chewing

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Ultraviolet light

Courtesy of Ludmil Alexandrov

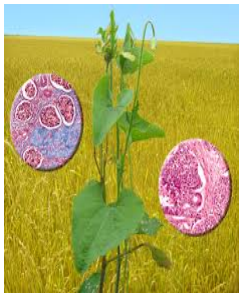
Aetiologies of substitution mutational signatures

Aflatoxin

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Aristolochic acid



Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Temozolomide

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Platinum therapy

Courtesy of Ludmil Alexandrov

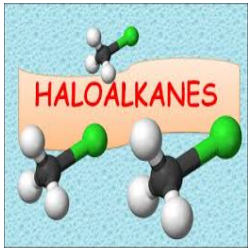
Aetiologies of substitution mutational signatures

Azathioprine

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Haloalkanes



Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Defective BRCA1, BRCA2, homologous recombination repair

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Defective base excision repair

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Defective polymerase epsilon activity

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Infidelity of polymerase eta activity

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

APOBEC cytosine deamination

Courtesy of Ludmil Alexandrov

Aetiologies of substitution mutational signatures

Unknown aetiologies 19/49

Courtesy of Ludmil Alexandrov

Origins of mutations observed in tumors

- DNA damaging environmental exposures
- Hereditary factors that impair the DNA damage response
- Replication error
  - Tissues with higher cell turnover rate will accumulate more mutations with age (Tomasetti and Vogelstein 2015)
  - The “bad luck” paper

  ⇒Mutations due to replication errors may account for more risk than hereditary and exposure related mutations
- Half of the mutations may pre-exist in cells prior to tumor initiation (Tomasetti et al 2012)

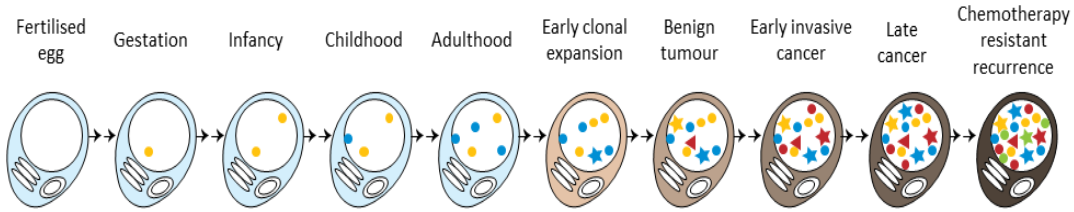
How do these many mutation sources contribute to the mutation burden of a tumor?

From fertilised egg to cancer cell

Chemotherapy resistant recurrence

Courtesy of Ludmil Alexandrov

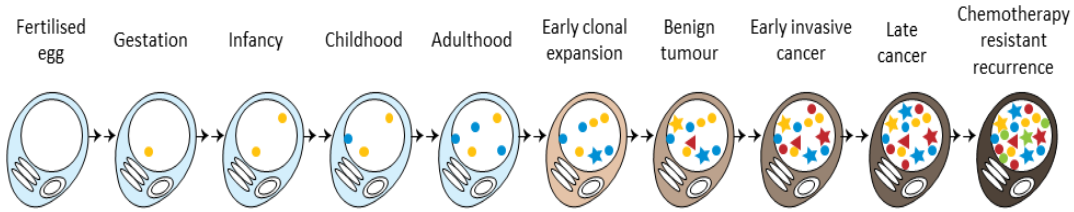
From fertilised egg to cancer cell



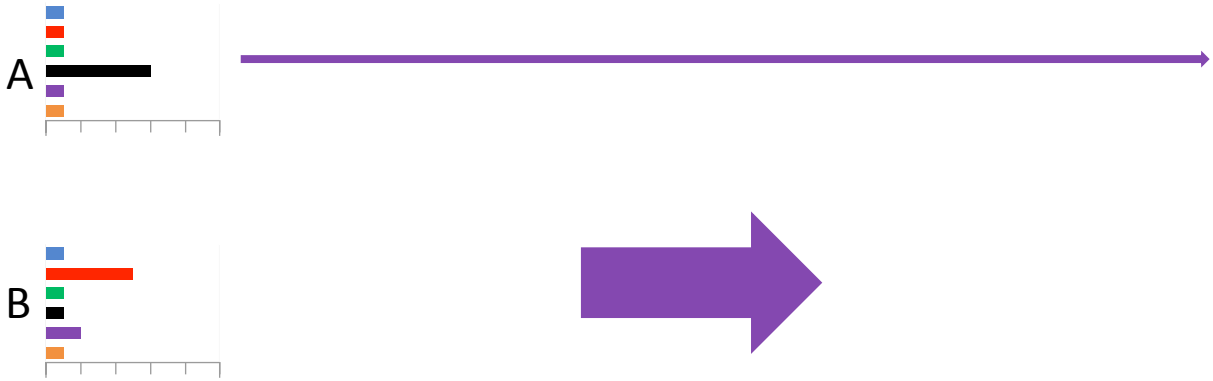
Mutational processes

Courtesy of Ludmil Alexandrov

From fertilised egg to cancer cell

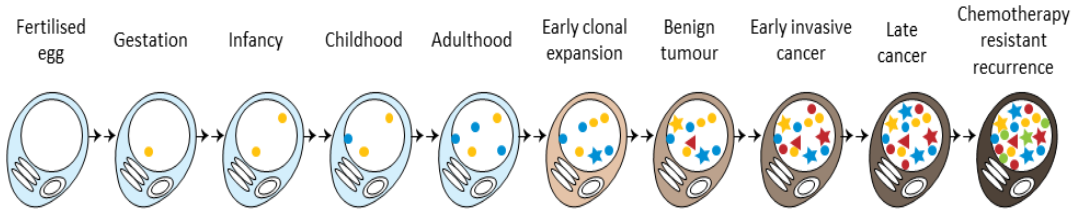


Mutational processes

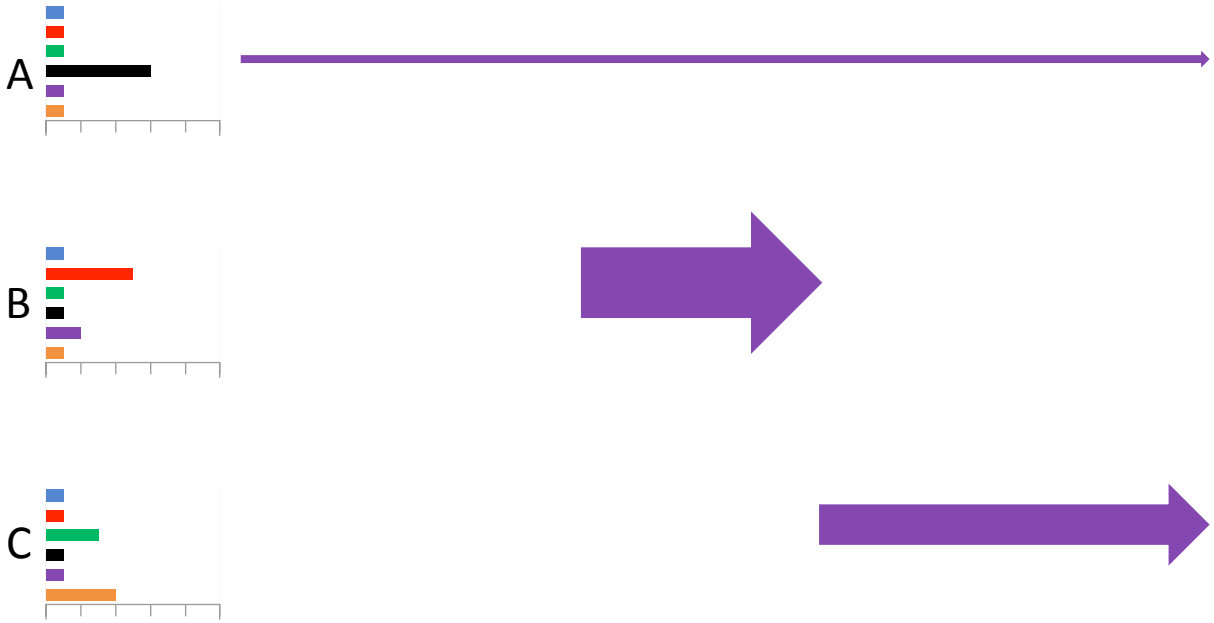


Courtesy of Ludmil Alexandrov

From fertilised egg to cancer cell

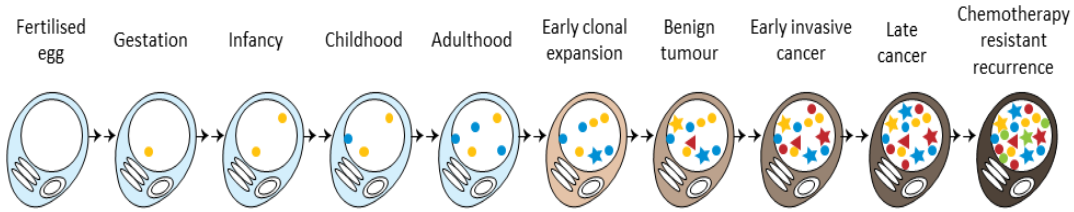


Mutational processes

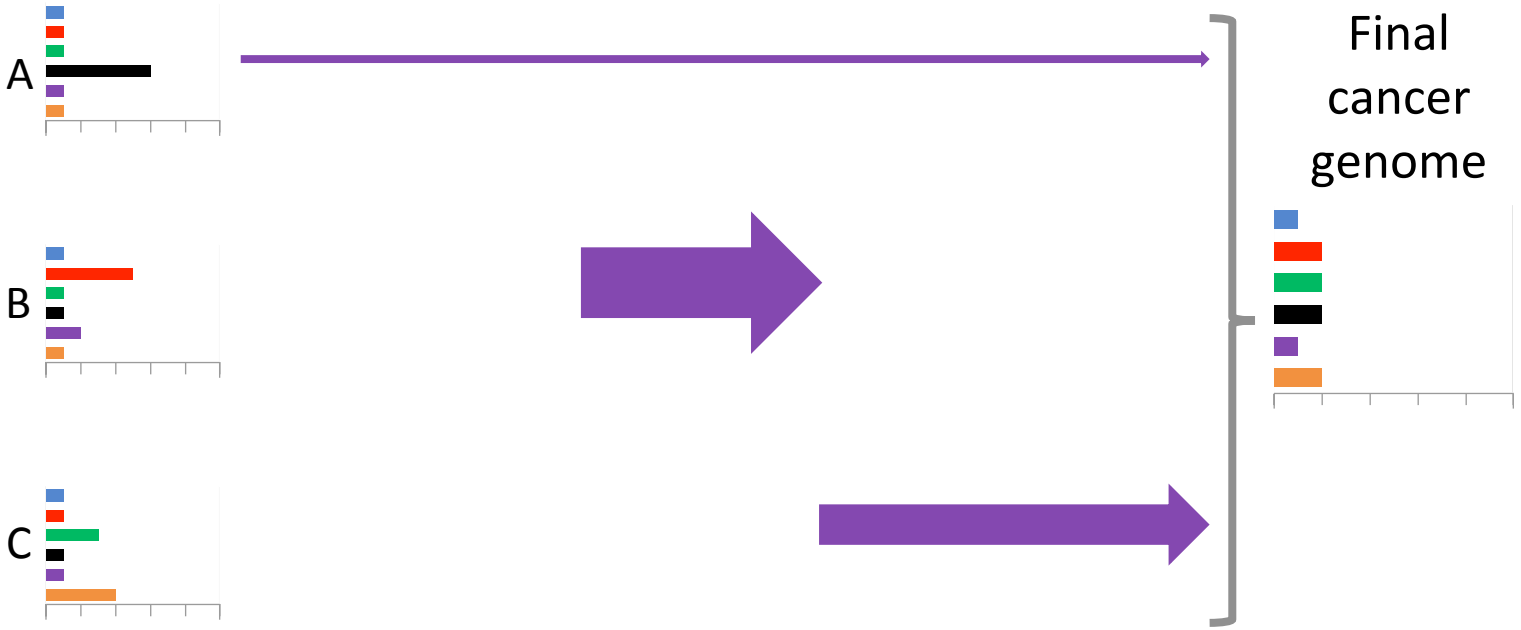


Courtesy of Ludmil Alexandrov

From fertilised egg to cancer cell



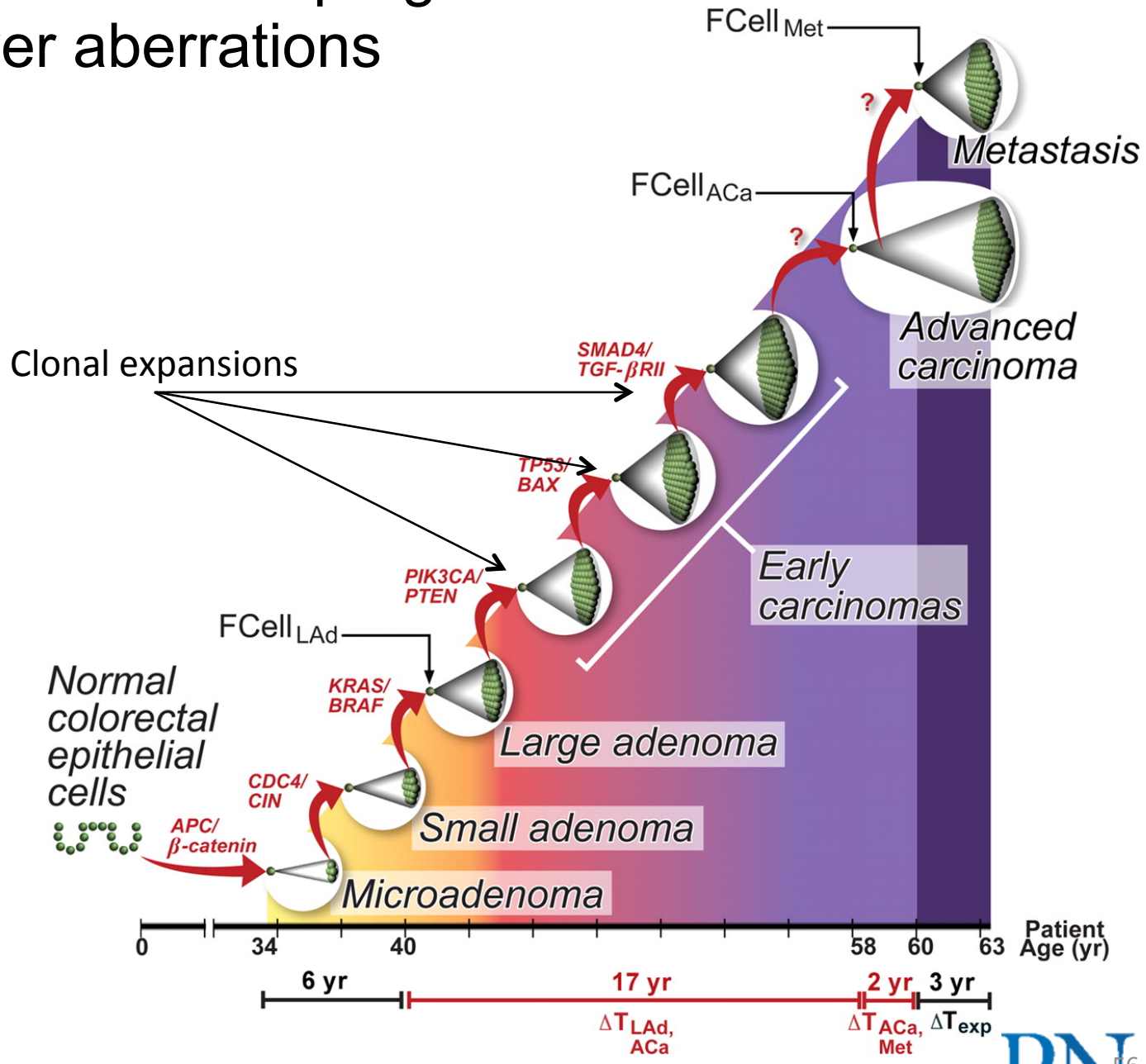
Mutational processes



Courtesy of Ludmil Alexandrov

How does mutation burden relate to cancer risk and tumor development?

Cancer is a disease of progressive genetic driver aberrations



Siân Jones et al. PNAS 2008;105:11:4283-4288

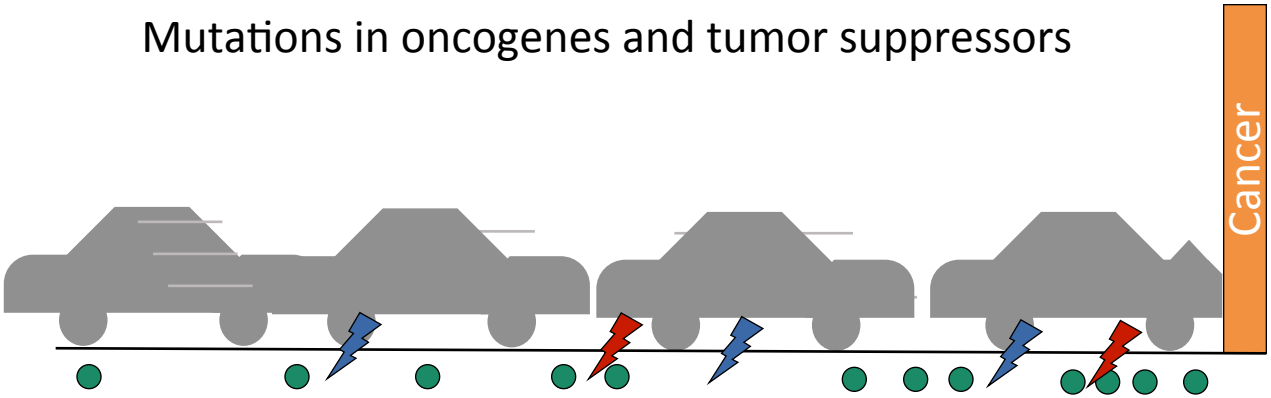
Driver versus Passenger mutations

Driver Mutations

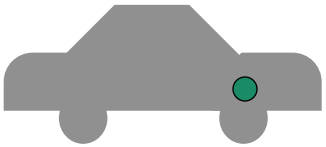
Oncogene = Gas

Tumor Suppressor = Brakes

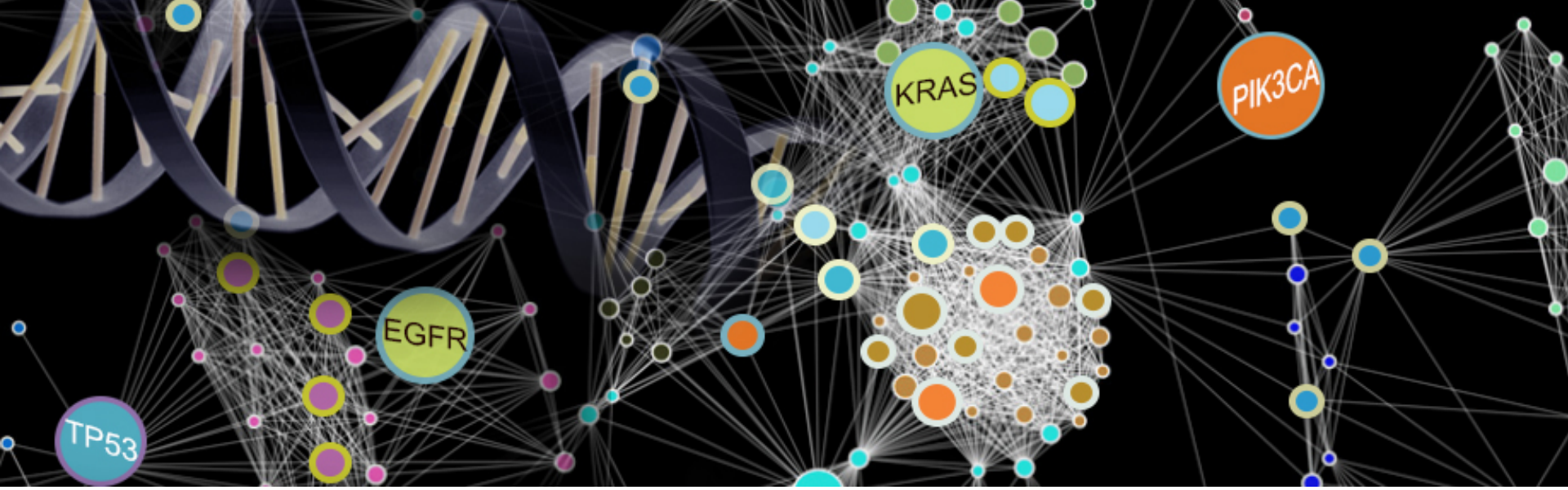
Mutations in oncogenes and tumor suppressors



Passenger Mutations

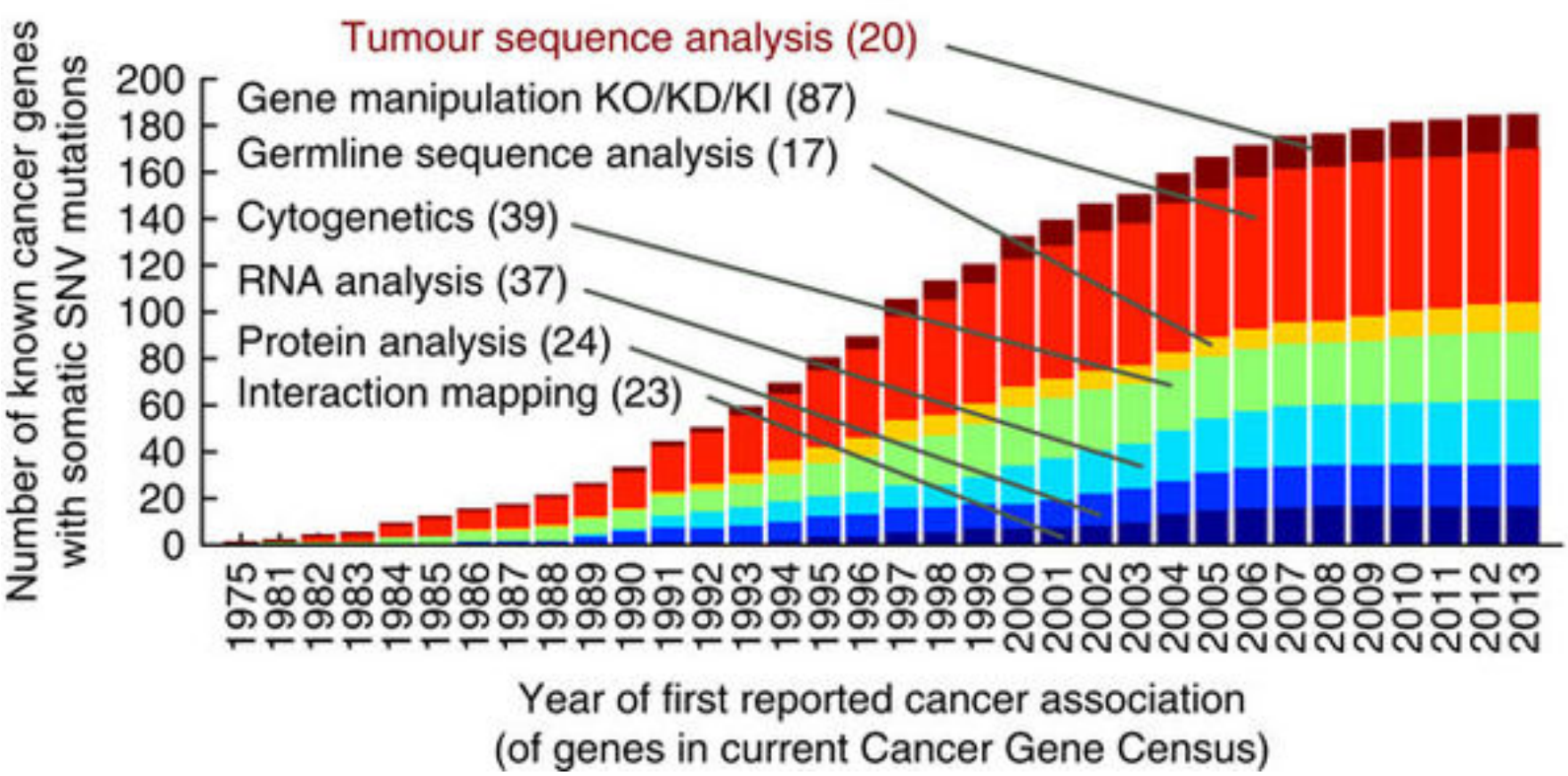


- Driver mutations provide a growth / survival advantage
- Passenger mutations can be preexist in a cell that becomes tumorigenic
- Passengers continue to accumulate after cancer initiation



Identifying driver mutations in cancer

Most known driver genes were found before NGS

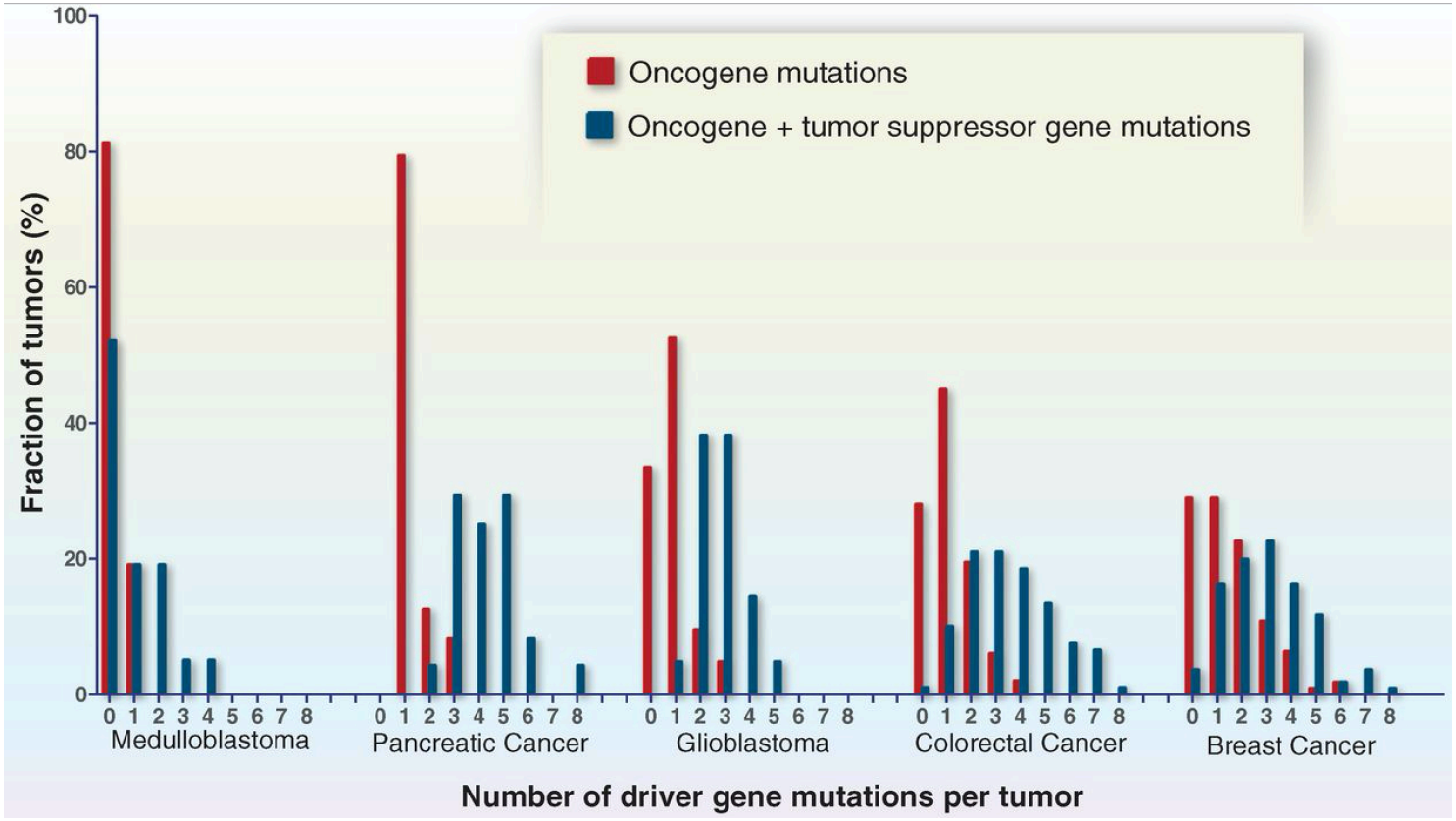


Hofree et al, Nat Comm 2016

Most mutations in tumors are “Passengers”

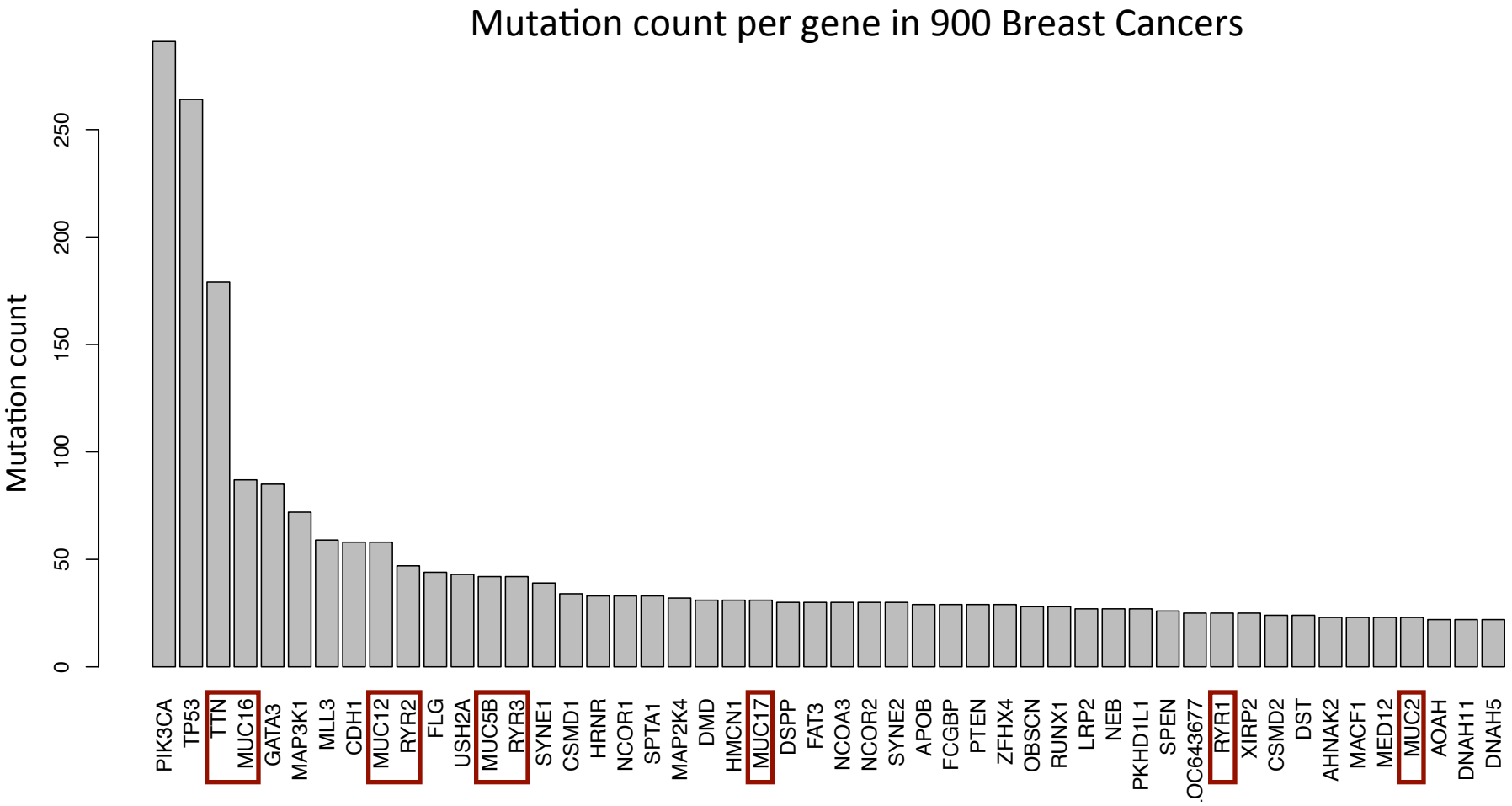
Early epidemiological studies suggested ~2-8 drivers per tumor

NGS studies



Bert Vogelstein et al. Science 2013;339:1546-1558

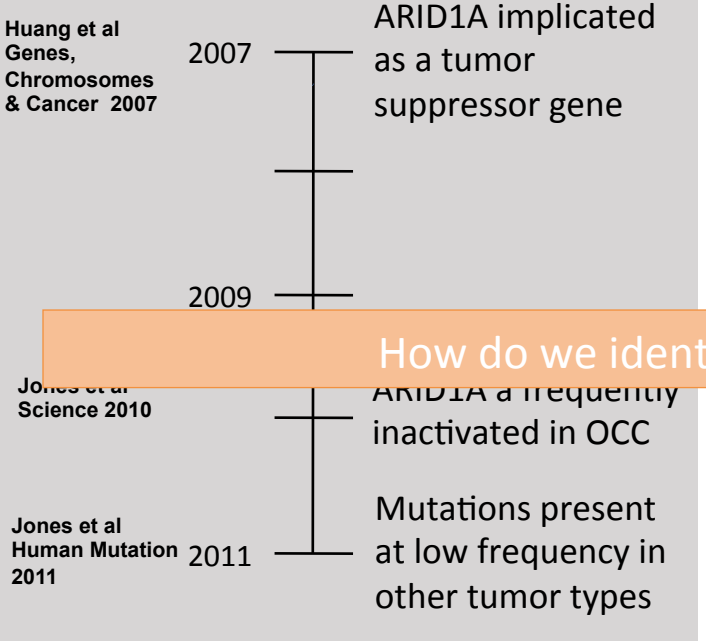
Frequently mutated genes



Frequency not a perfect indicator of a cancer gene

Evidence suggests *infrequently* mutated genes may harbor drivers

- ARID1A



Tumor Type	N samples	Samples with ARID1A inactivating mutations	Percentage
Breast	114	4	4%
Colon	119	12	10%
Gastric	100	10	10%
Medulloblastoma	125	3	2%
Pancreas	119	10	8%
Prostate	23	2	9%

Jones et al. Human Mutation 2011

How do we identify infrequently mutated cancer genes?

If genes mutated at low frequency include known cancer genes, they may also include unknown cancer genes.

Existing Tools For Variant Prioritization

Name	Type	Information	URL	Refs
MAPP	Constraint-based predictor	Evolutionary and biochemical	http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html	27
SIFT	Constraint-based predictor	Evolutionary and biochemical (indirect)	http://sift.bii.a-star.edu.sg/	39
PANTHER	Constraint-based predictor	Evolutionary and biochemical (indirect)	http://www.pantherdb.org/	41
MutationTaster*	Trained classifier	Evolutionary, biochemical and structural	http://www.mutationtaster.org/	40
nsSNP Analyzer	Trained classifier	Evolutionary, biochemical and structural	http://snpanalyzer.uthsc.edu/	44
PMUT	Trained classifier	Evolutionary, biochemical and structural	http://mmb2.pcb.ub.es:8080/PMut/	38
polyPhen	Trained classifier	Evolutionary, biochemical and structural	http://genetics.bwh.harvard.edu/pph2/	35
SAPRED	Trained classifier	Evolutionary, biochemical and structural	http://sapred.cbi.pku.edu.cn/	42
SNAP	Trained classifier	Evolutionary, biochemical and structural	http://www.rostlab.org/services/SNAP/	36
SNPs3D	Trained classifier	Evolutionary, biochemical and structural	http://www.snps3d.org/	51
PhD-SNP	Trained classifier	Evolutionary and biochemical (indirect)	http://gpcr2.biocomp.unibo.it/~emidio/PhD-SNP/PhD-SNP_Help.html	37

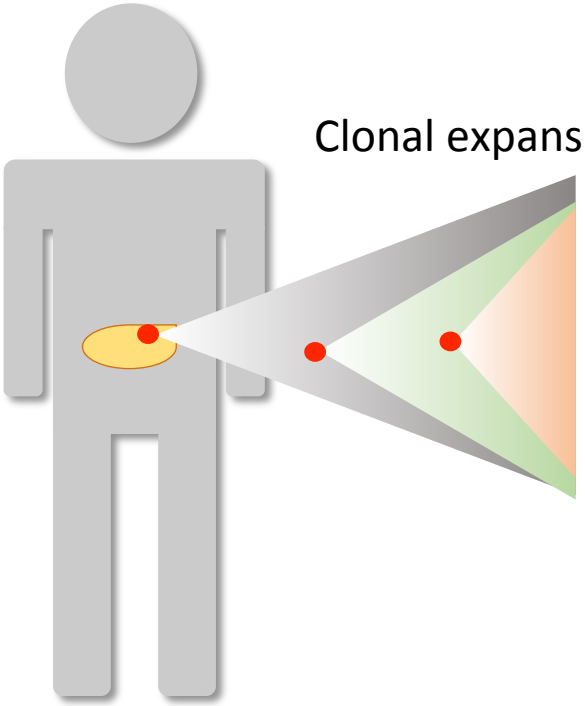
*Also makes predictions for synonymous and non-coding variant effects: for example, splicing. MAPP, Multivariate Analysis of Protein Polymorphism; polyPhen, polymorphism phenotyping.

And many, many more …

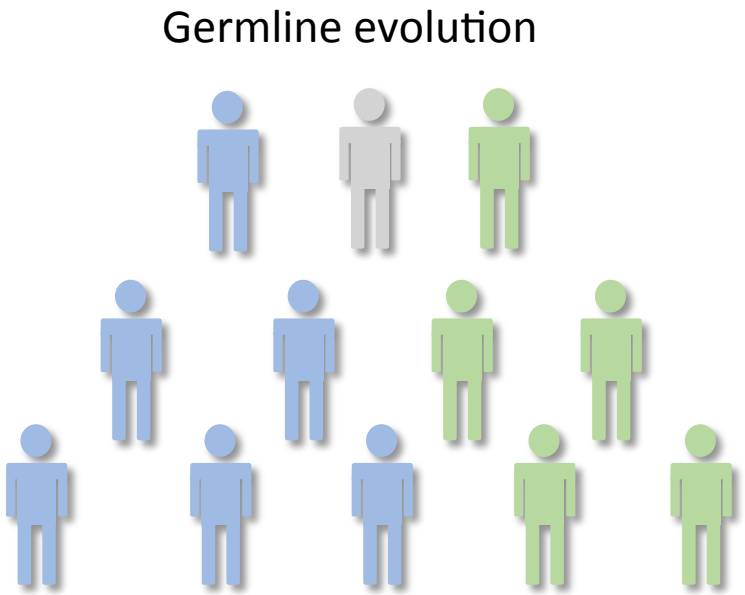
Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* **12,** 628–640 (2011).

Why do we need a new tool for cancer?

Accumulation of Mutations



Somatic mutation

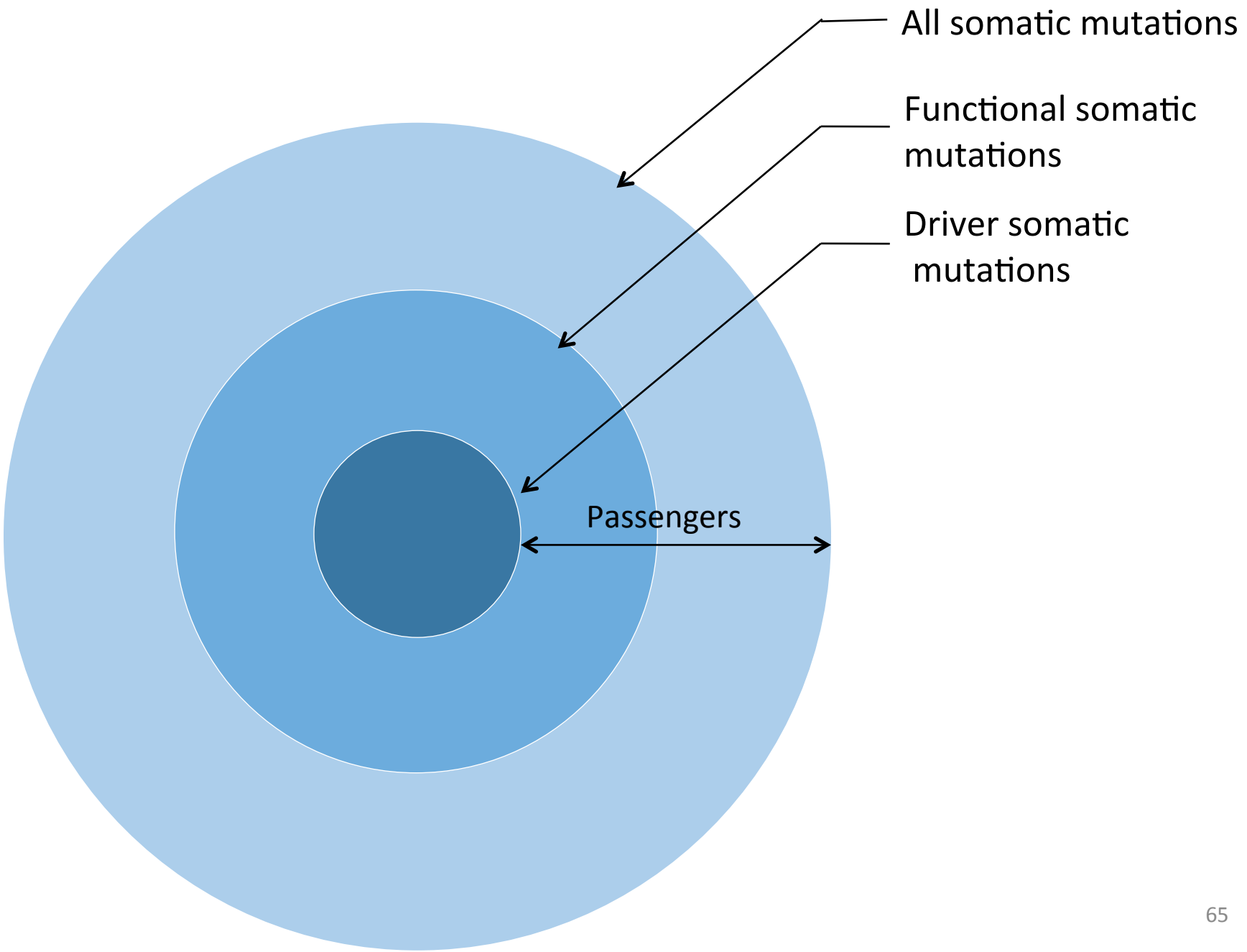


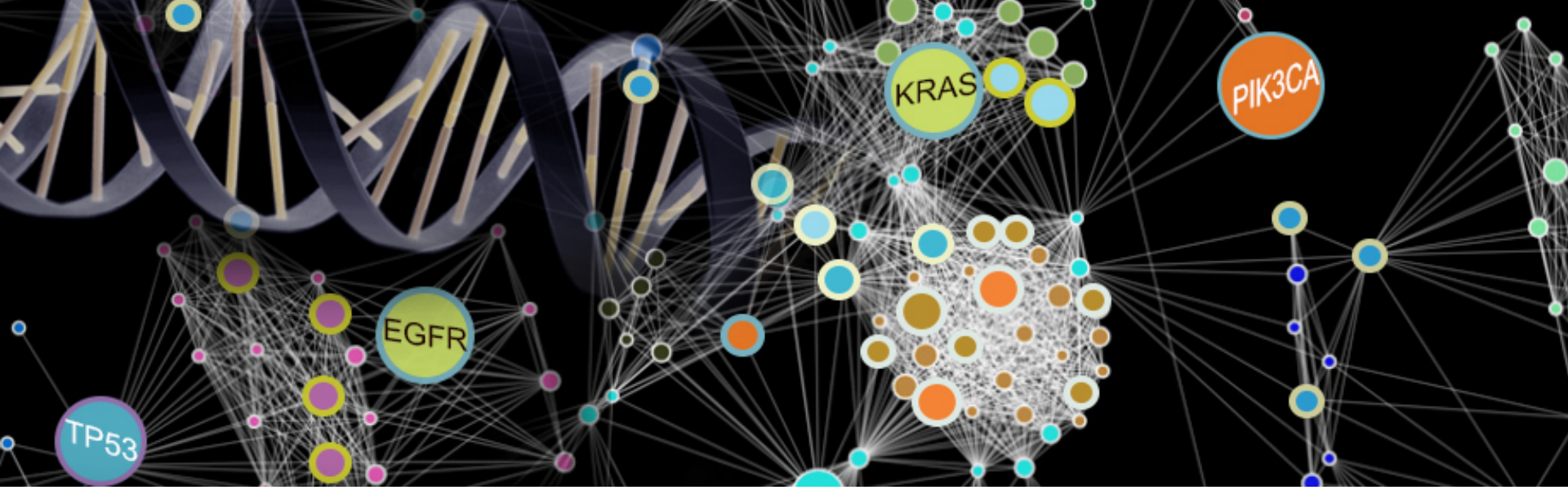
Rare inherited disease mutation

Common inherited disease mutation

Common neutral polymorphism

Cancer evolution is different from germline evolution

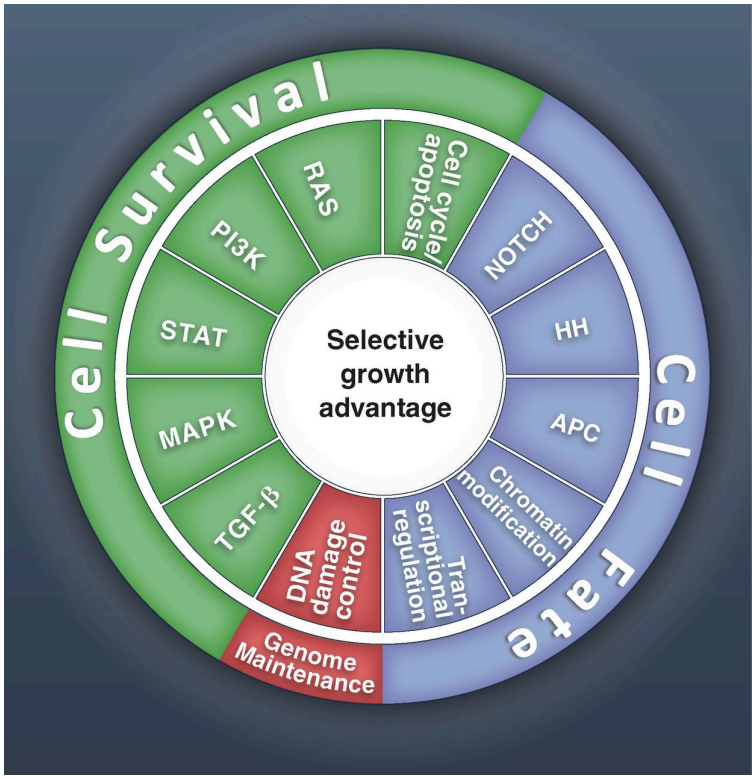

All somatic mutations
Functional somatic mutations
Driver somatic mutations
Passengers



Tumor heterogeneity

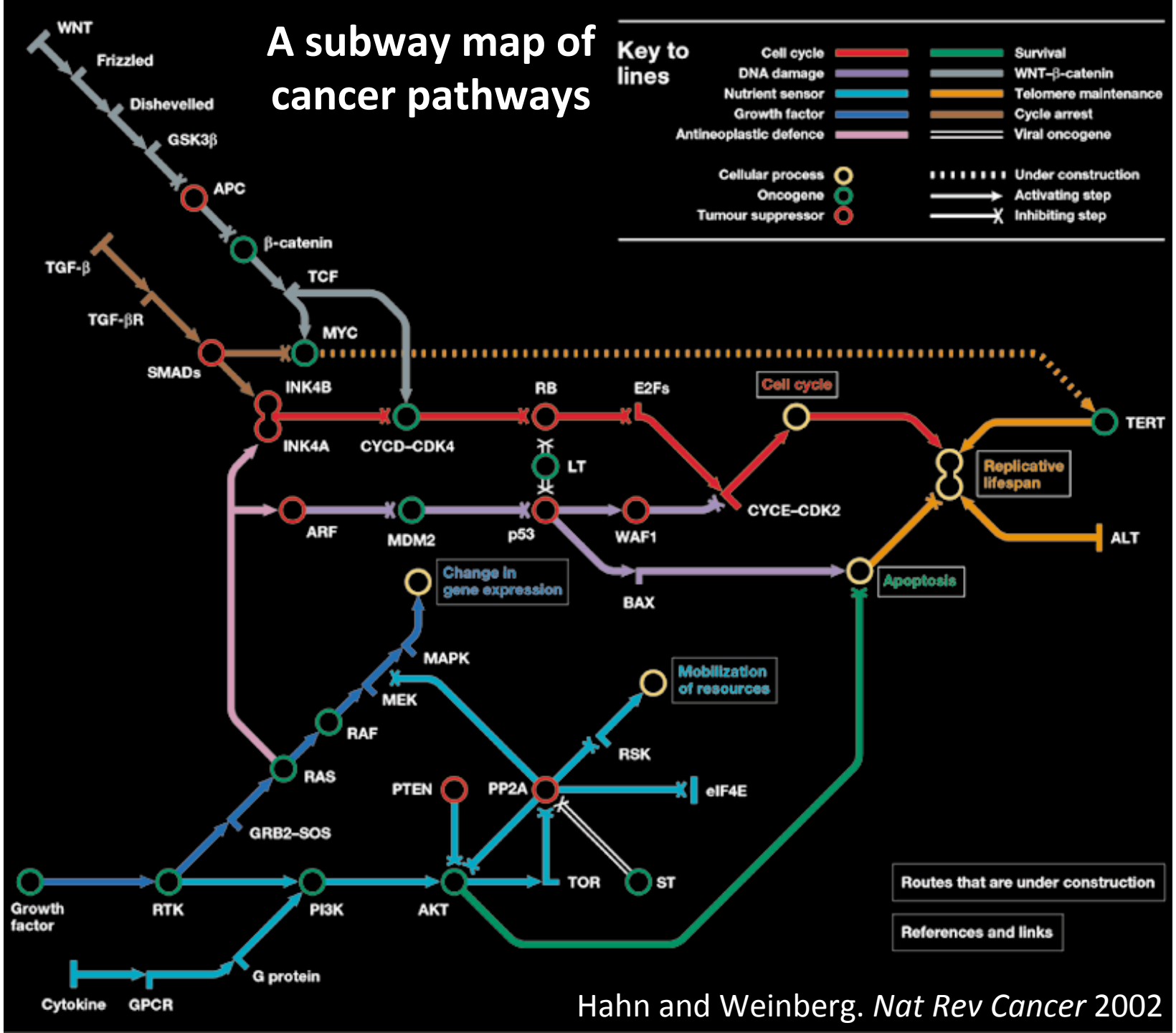
Cancer: Different mutations, same phenotypic characteristics

Hanahan, Weinberg **Hallmarks of Cancer: The Next Generation** Cell Volume 144, Issue 5, 2011, 646 – 674



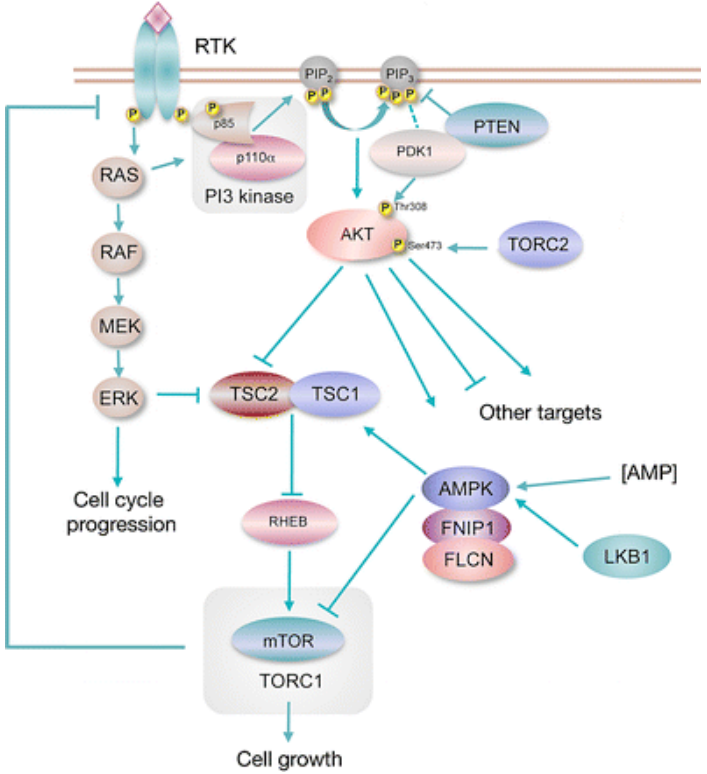
B Vogelstein et al. **Cancer Genome Landscapes** Science 2013, 339, 1546-1558

A subway map of cancer pathways



Hahn and Weinberg. *Nat Rev Cancer* 2002

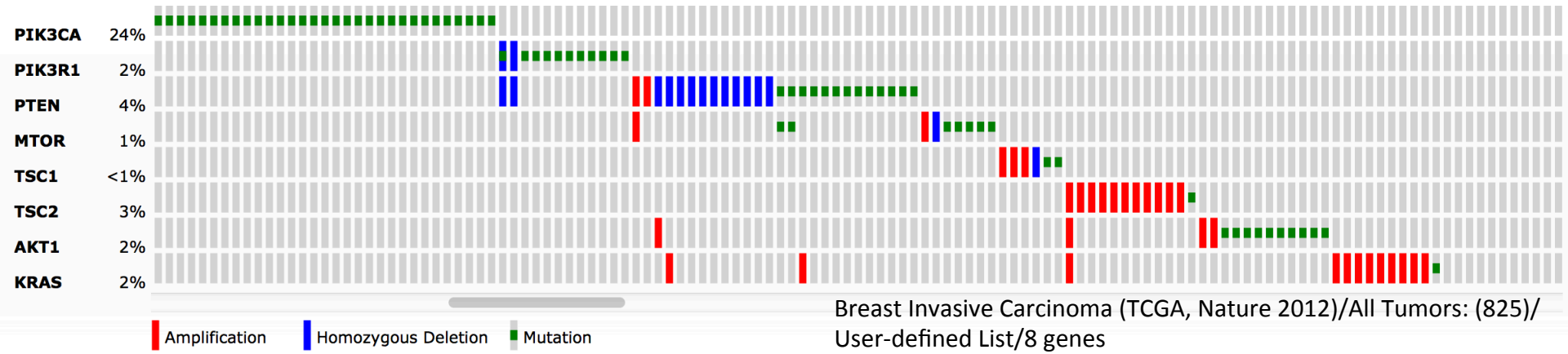
Mutual exclusivity within pathways



Knowles *et al* **Phosphatidylinositol 3-kinase (PI3K) pathway activation in bladder cancer** Cancer and Metastasis Reviews 2009

Memorial Sloan-Kettering Cancer Center

cBioPortal for Cancer Genomics

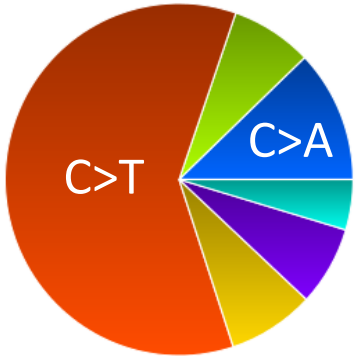


Breast Invasive Carcinoma (TCGA, Nature 2012)/All Tumors: (825)/ User-defined List/8 genes

Tissue-specific mutation signatures can cause tissue-specific heterogeneity in drivers

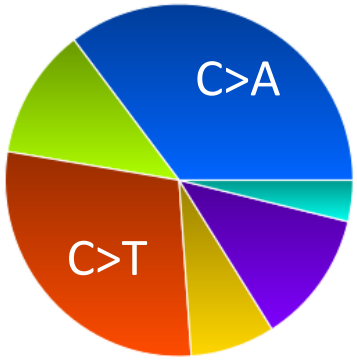
TP53: Most frequently mutated driver gene in cancer

686 skin cancers



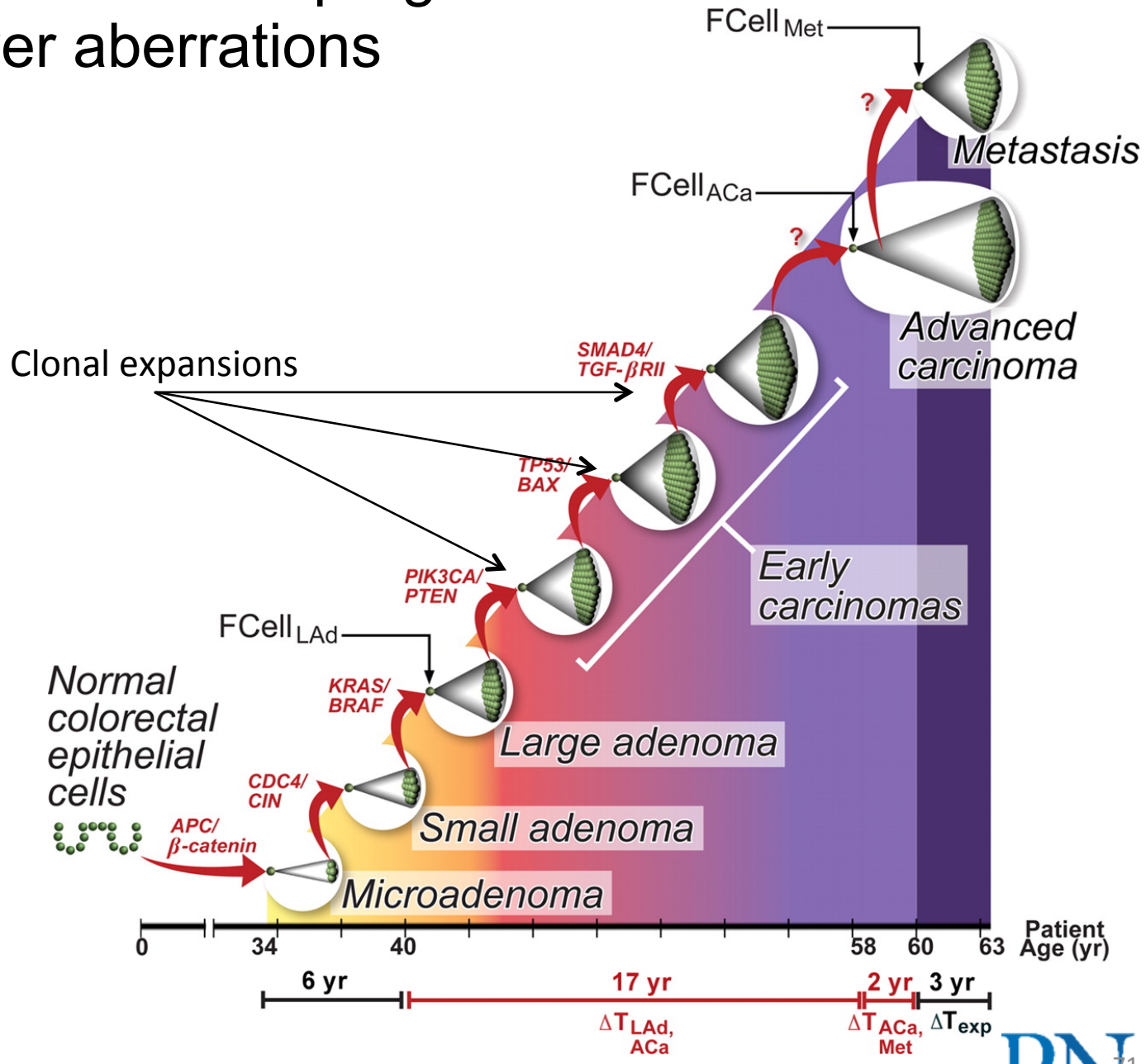
Ultraviolet light causes C>T mutations

1647 lung cancers



Tobacco carcinogens cause C>A mutations

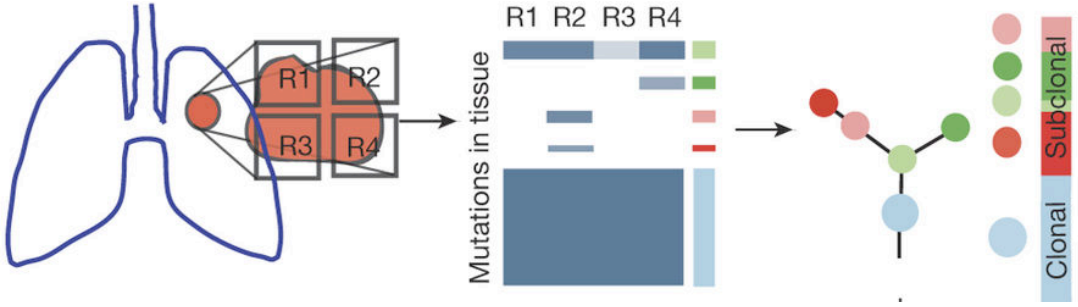
Cancer is a disease of progressive genetic driver aberrations



Siân Jones et al. PNAS 2008;105:11:4283-4288

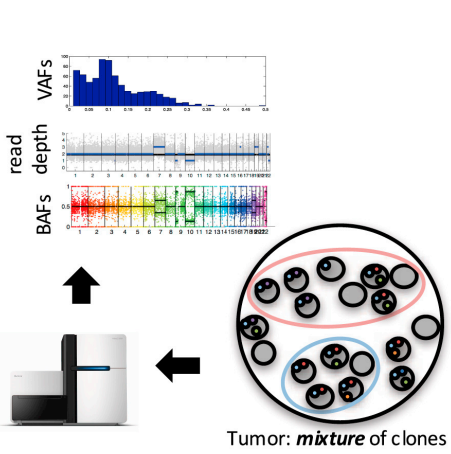
Clonal expansion creates spatial heterogeneity of mutations

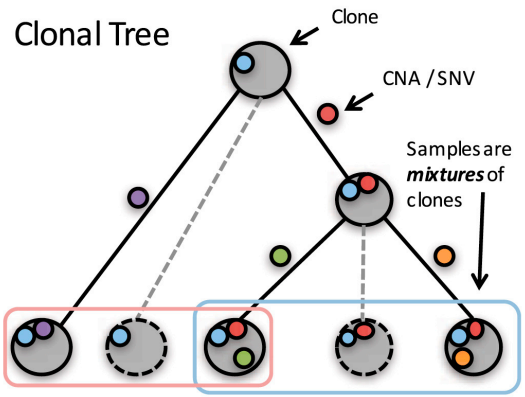
- Sequencing different locations reveals different mutations

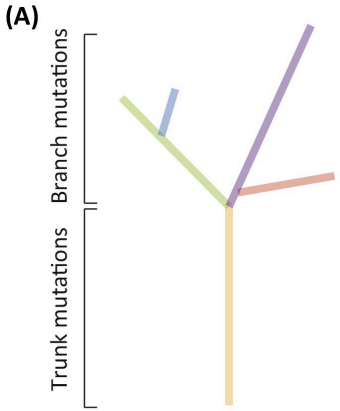


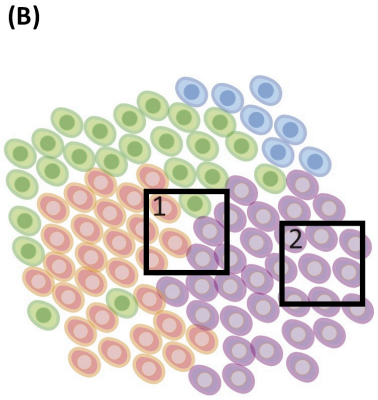
Abbosh et al 2017

- Sequencing coverage of mutations can be used to create hypothetical models of clonal composition but is challenged by sequencing bias









El-Kebir et al,. 2016

Burrel and Swanton, 2016

Mutation burden in tumors

More than just a count of mutations

Contains the history of DNA damage experienced by the tumor cell lineage

Not all mutations co-exist in the same tumor cells

- Subclonal heterogeneity a major barrier for successful treatment
- Need to eliminate all tumor cells but therapies may only target a subset

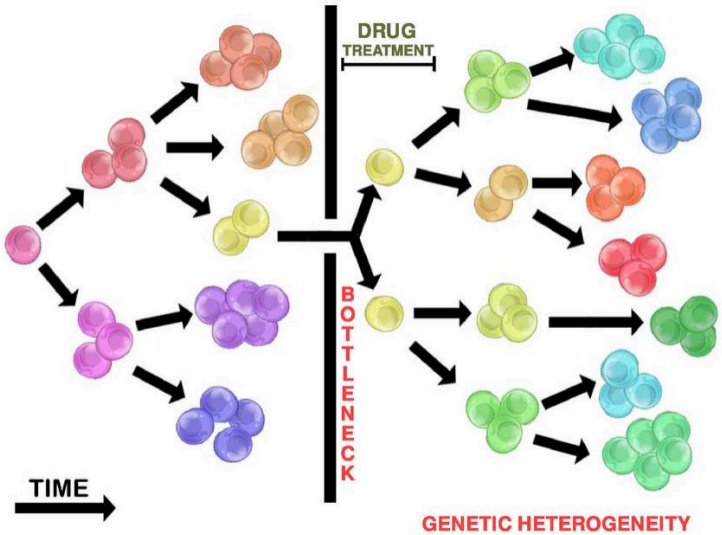
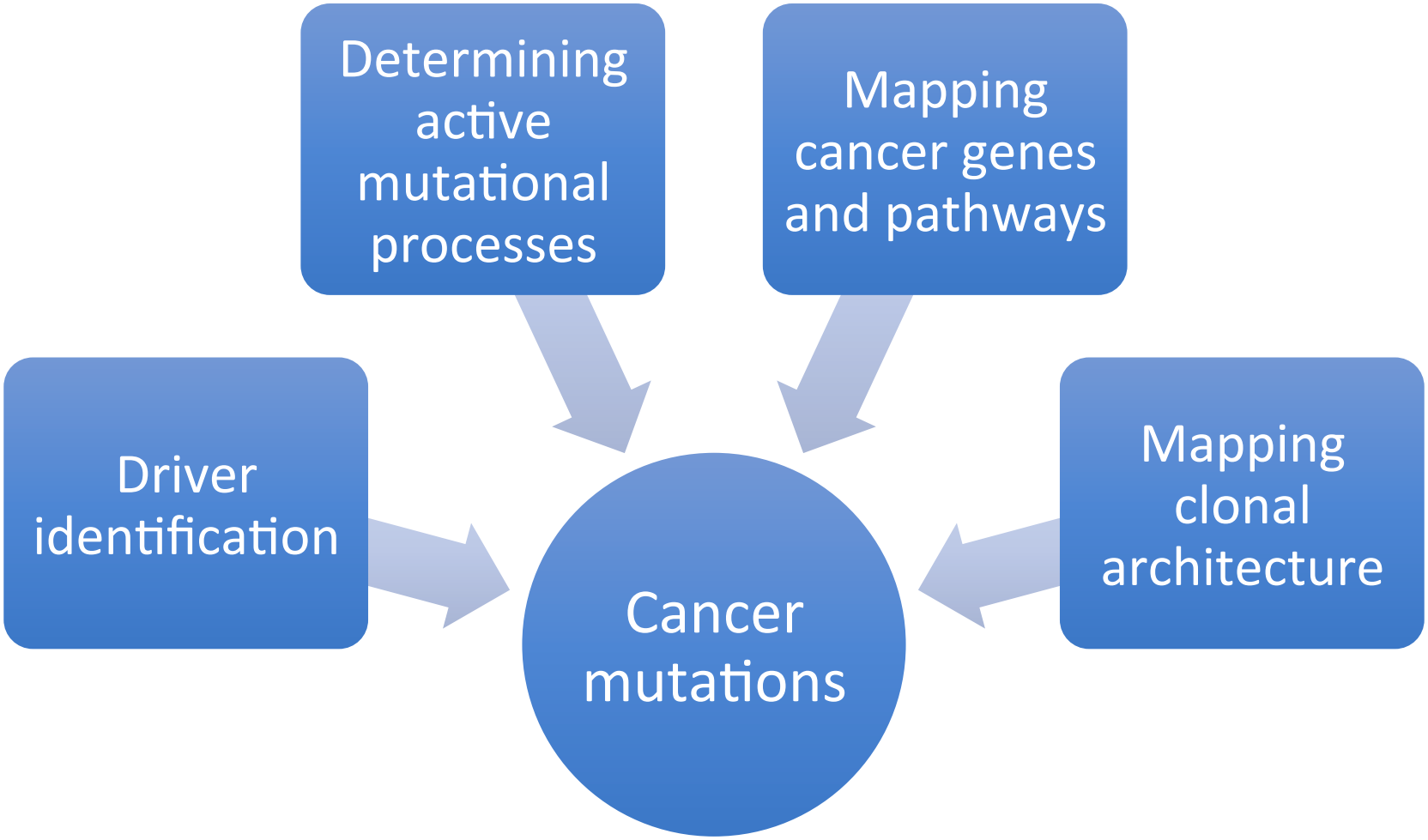
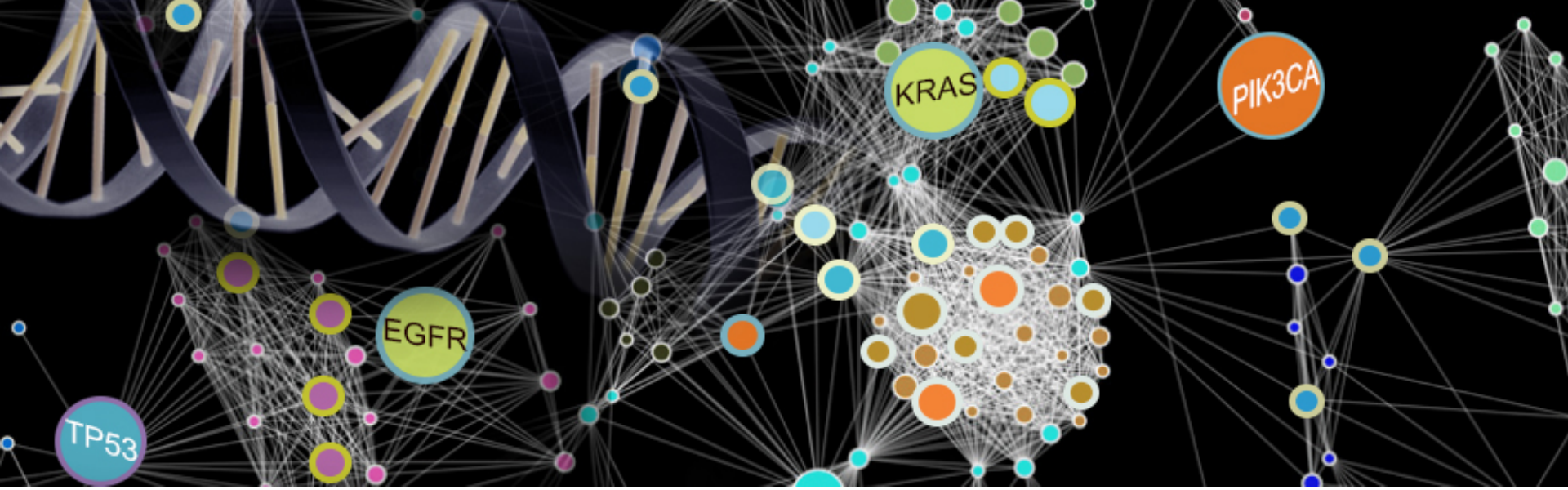


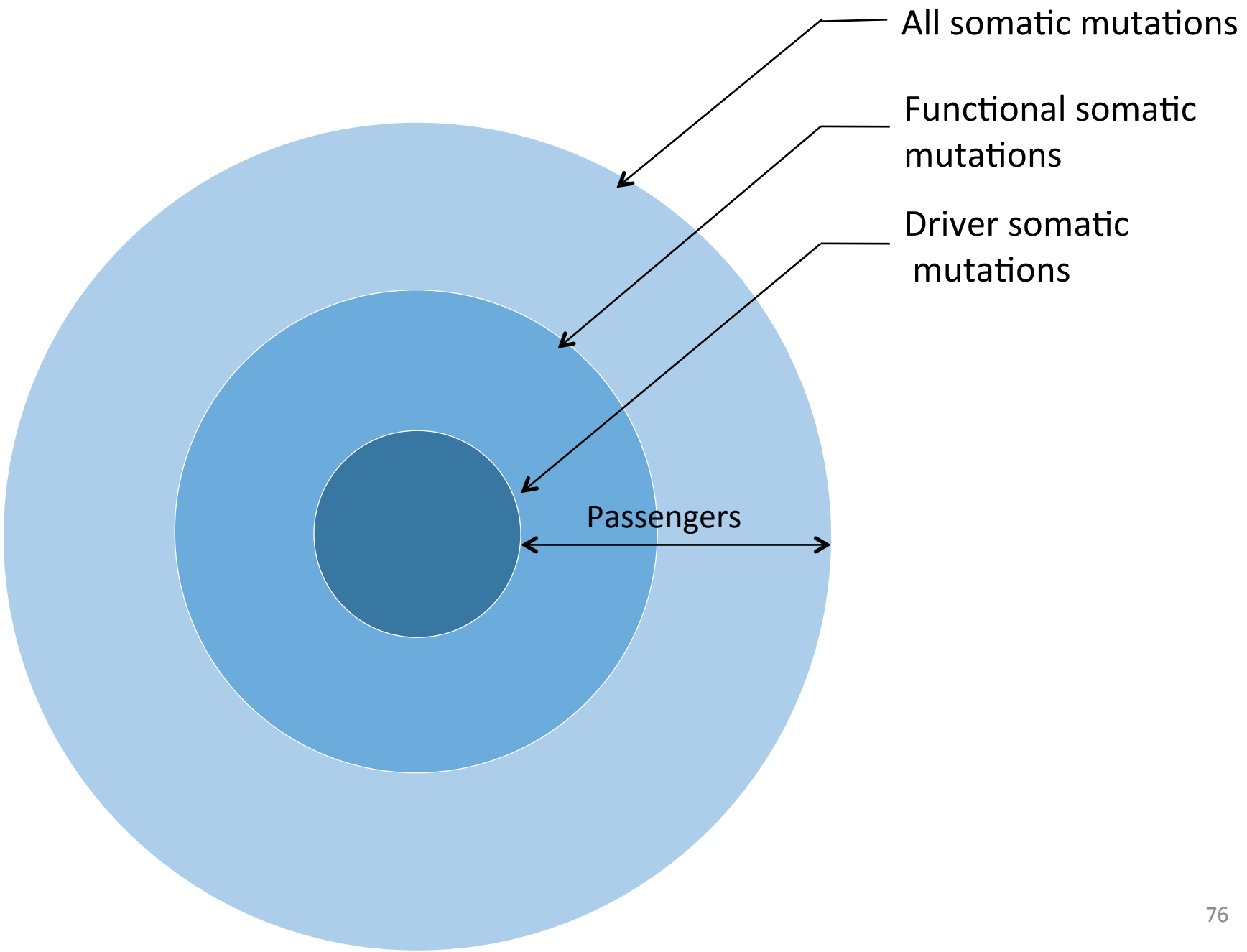
Image source: Wikipedia page on tumor heterogeneity

Major bioinformatic challenges for studying cancer mutations





Practical Example: Machine learning for driver mutation identification


All somatic mutations
Functional somatic mutations
Driver somatic mutations
Passengers

Machine Learning: Supervised Prioritization

- **Training Set**
  - **Positive examples (drivers)**
  - **Negative examples (passengers)**

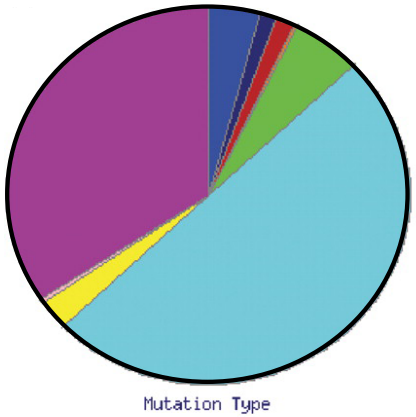
- Feature set
  - Informative characteristics of amino acid substitutions

Positive class: Driver Mutations

COSMIC

Curated data type	Curated data count
Experiments	2,760,220
Tumours	541,928
Mutations	136,326
References	10,383
Genes	18,490
Fusions	4946
Structural variants	2307
Whole cancer genomes	29

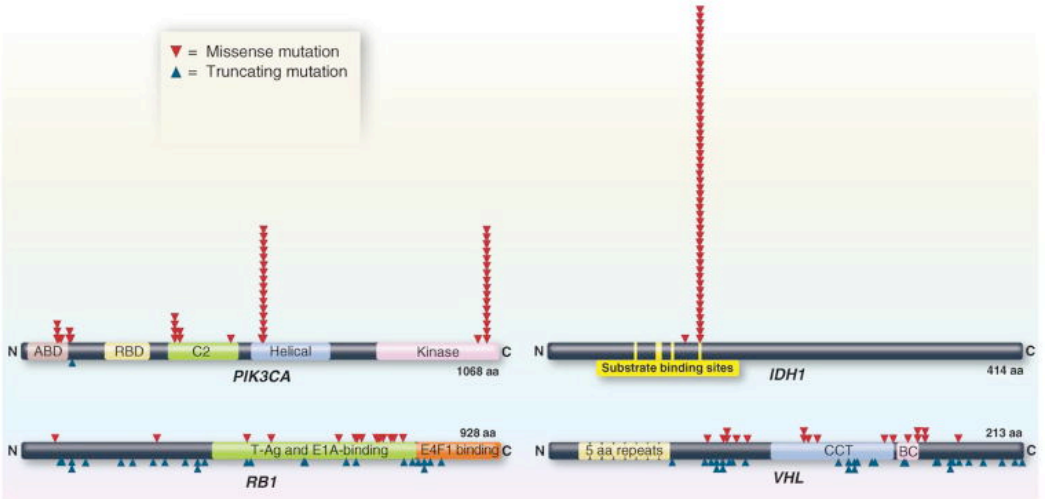
Forbes S A et al. Nucl. Acids Res. 2011;39:D945-D950



Details For Mutation Type Chart

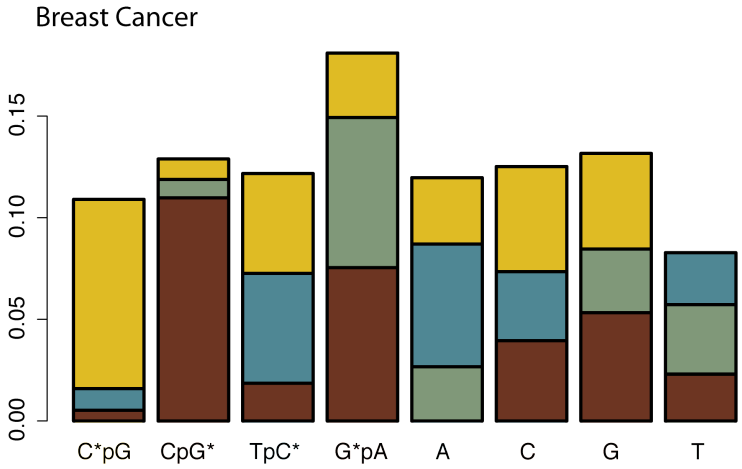
Mutation Type	Number	Percentage	Mutation Data
Deletion Frameshift	855	4.2 %	More Details
Deletion Inframe	275	1.3 %	More Details
Insertion Frameshift	317	1.5 %	More Details
Insertion Inframe	50	0.2 %	More Details
Substitution Nonsense	1139	5.5 %	More Details
Substitution Missense	10292	50.0 %	More Details
Substitution Synonymous	518	2.5 %	More Details
Complex	65	0.3 %	More Details
Other	7056	34.3 %	More Details
Total	20567	100%	More Details

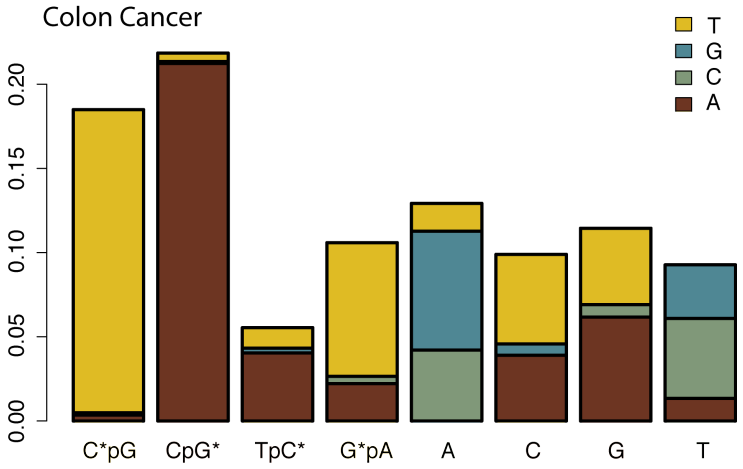
Selection Criteria:

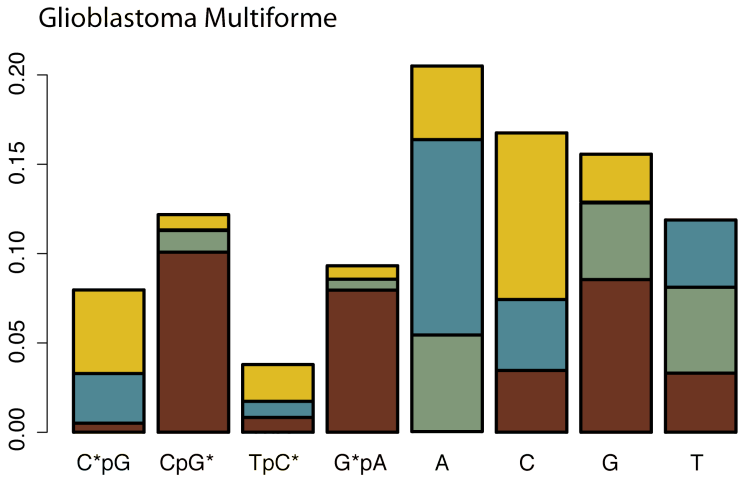


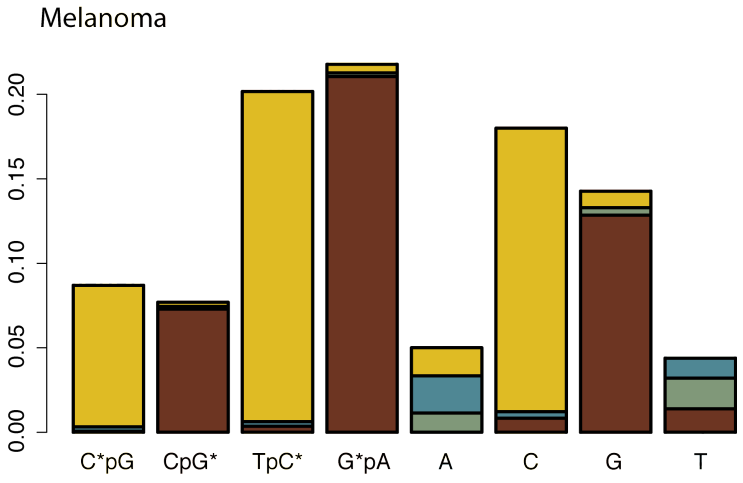
What do passenger mutations look like?

Assumption: The majority of mutations detected in tumor sequencing are passengers









Infamous Cancer Genes

TP53

CDKN2A

KRAS

PIK3CA

PTEN

RB

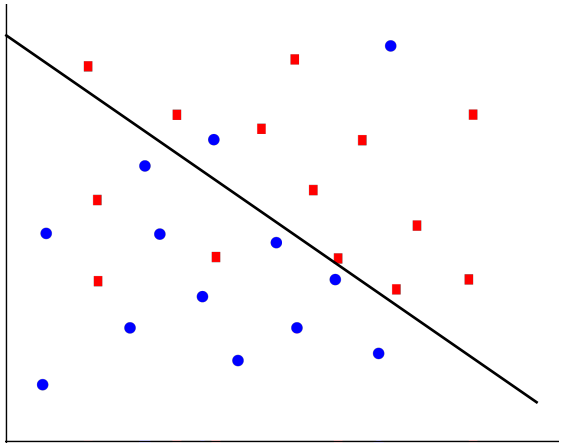
NF1

SMAD4

Machine Learning: Supervised Prioritization

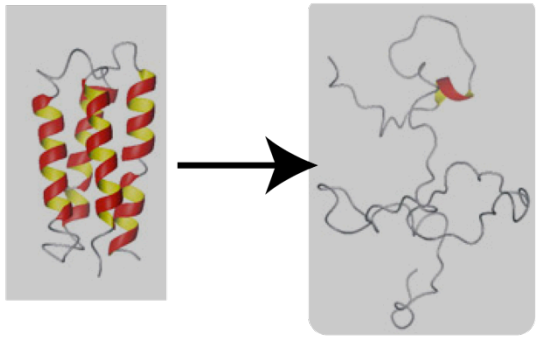
- Training Set
  - Positive examples (drivers)
  - Negative examples (passengers)

- **Feature set**
  - **Informative characteristics of amino acid substitutions**



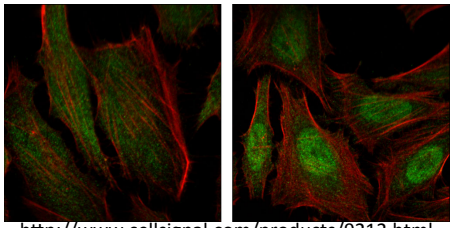
Functional approach: How do amino acid substitutions impact protein activity?

Destabilize the protein



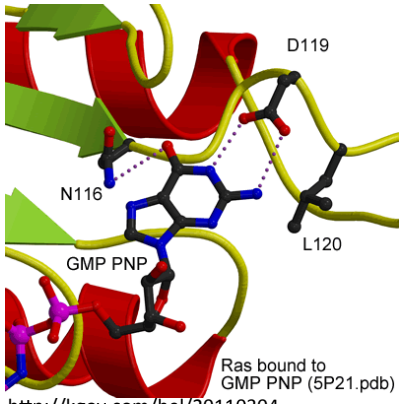
Adam Liwo

Cause improper localization



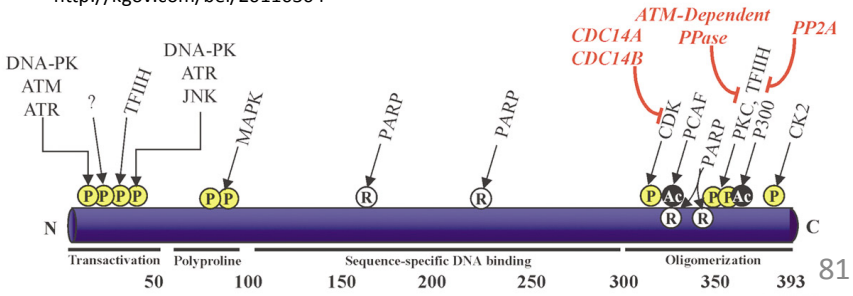
http://www.cellsignal.com/products/9212.html

Disrupt binding or active sites



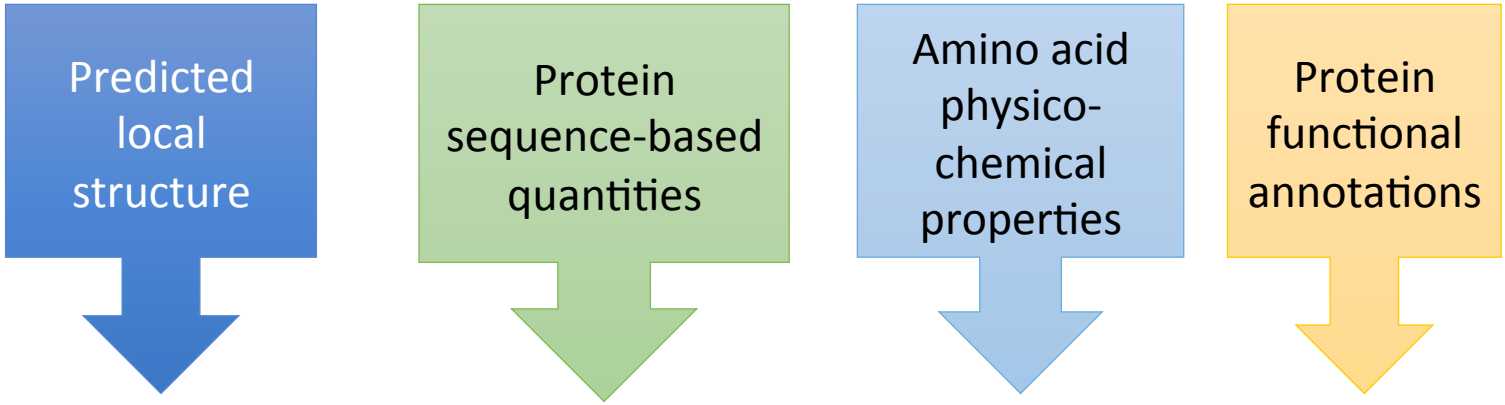
http://kgov.com/bel/20110304

Alter post-translational modification sites



Jerry et al. Breast Cancer Res 2002 4:91 doi:10.1186/bcr431

Features: Quantifying amino acid substitution impact



Mutation	Residue solvent accessibility	Relative Entropy of an MSA at amino acid position	Charge Change	Binding Site
TP53 S362A	Exposed	4.03	0	Yes
PIK3CA P539R	Buried	4.63	1	No
PIK3CA E545K	Exposed	4.66	2	No

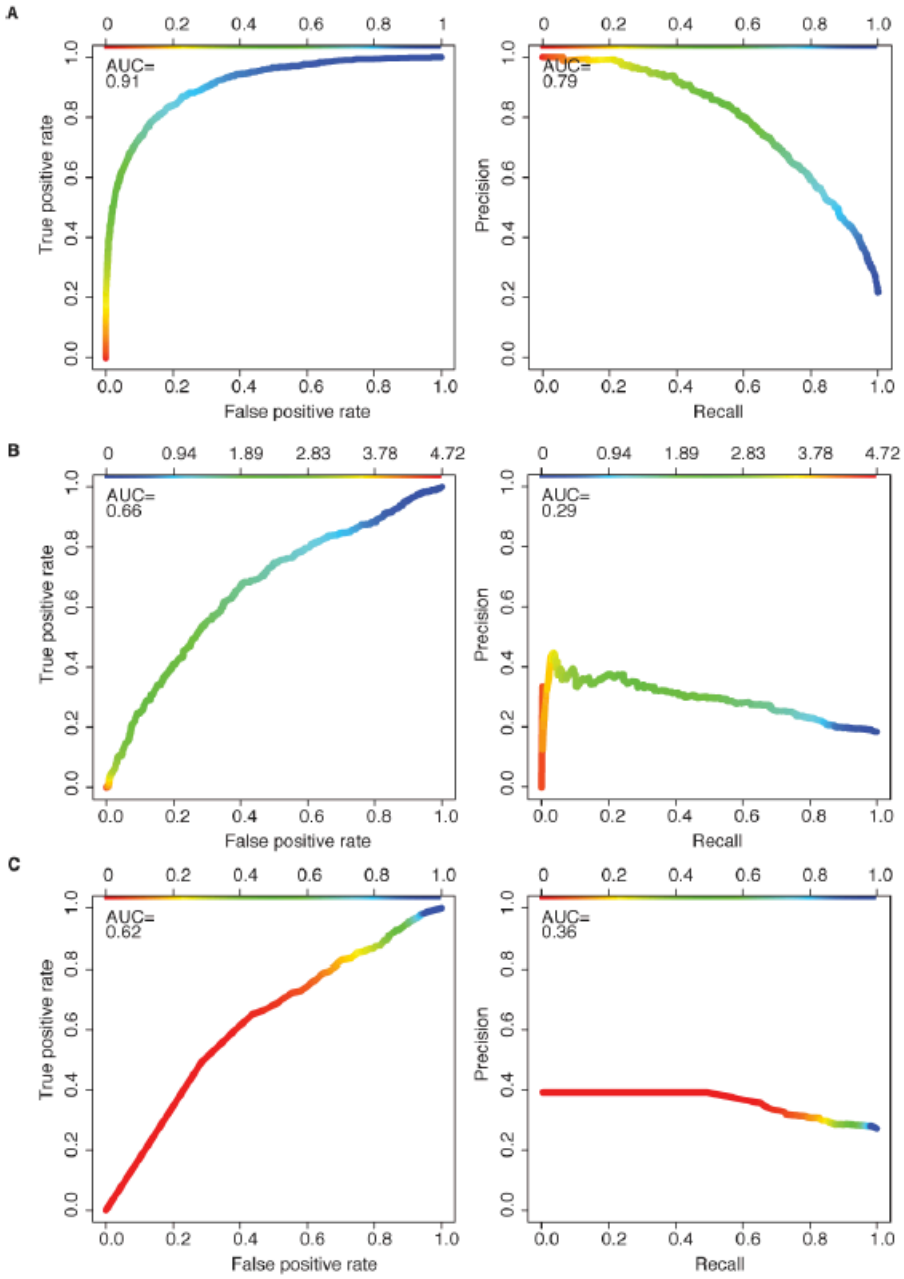
Machine learning algorithm: ***Random Forest***

- Ensemble classifier composed of multiple decision trees

- Each classifier trained on a subset of the training set, using a subset of the features

Breiman L. Random forest. Machine Learning 2001; 45: 5–32.,
Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Comput 1997; 9: 1545–88.

Comparison to SIFT & PolyPhen



CHASM

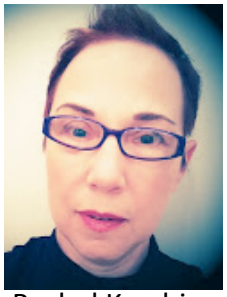
PolyPhen

SIFT

Left: ROC Curves
Right: Precision-Recall Curves

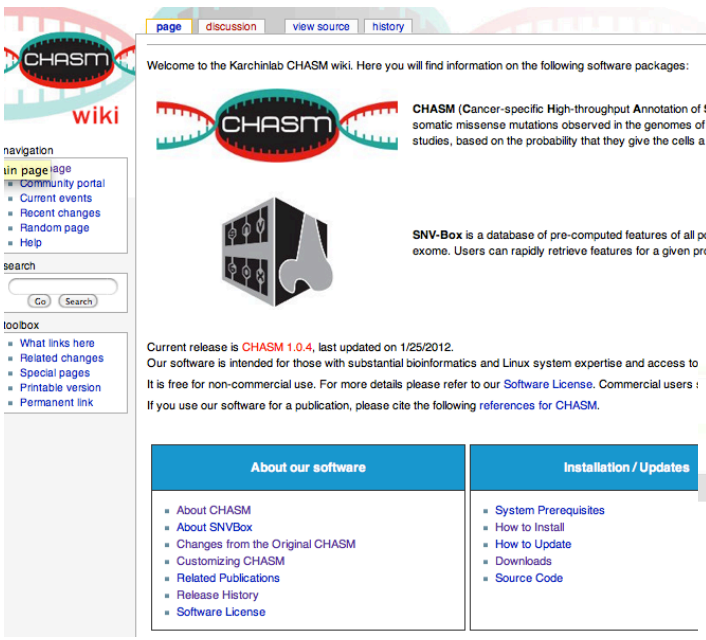
SIFT and PolyPhen score many of the synthetic passengers as likely to alter protein activity

Carter et al Cancer Research 2009



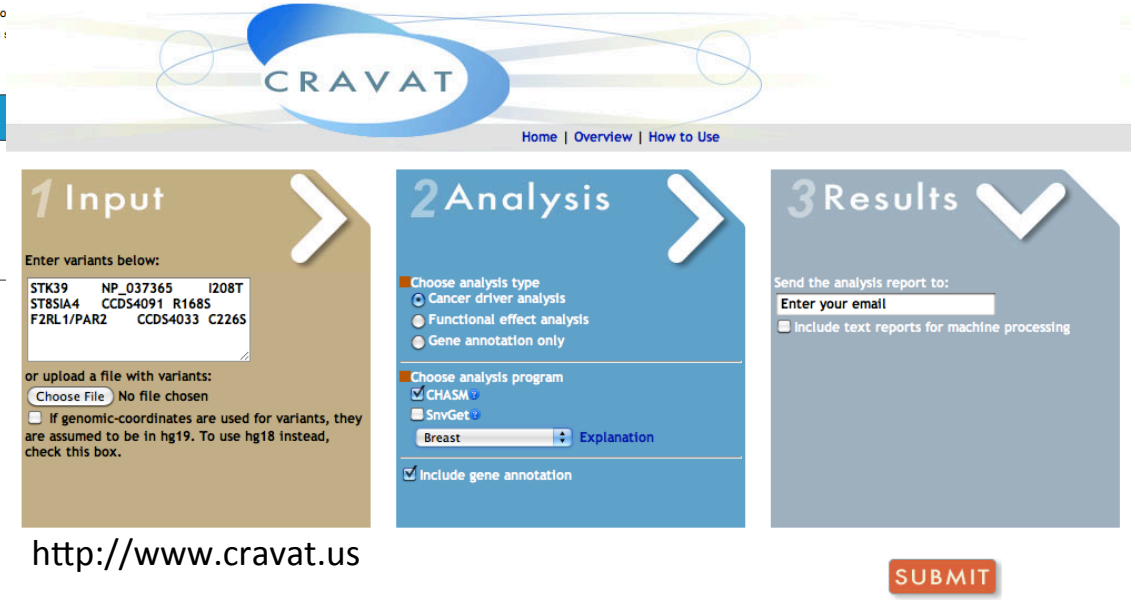
Rachel Karchin

Cancer-specific High-throughput Annotation of Somatic Mutations (CHASM)



http://wiki.chasmsoftware.org

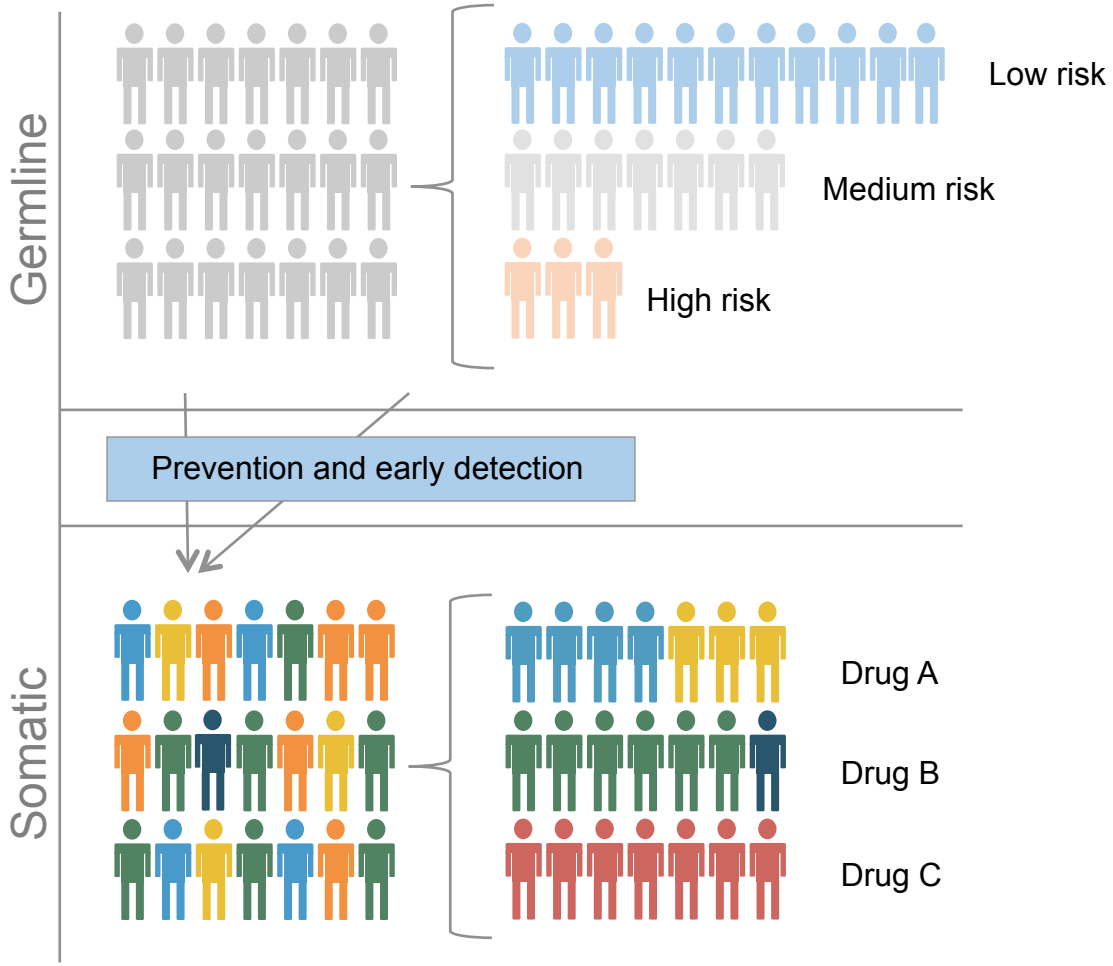
- Random Forest Classifier
- 85 Features
- ~14000 Training examples
- ROC AUC = 0.91



http://www.cravat.us

Precision Cancer Medicine

- Identify individuals at risk
- Preventative measures and screening for early detection
- Patient stratification for prognostic or treatment purposes



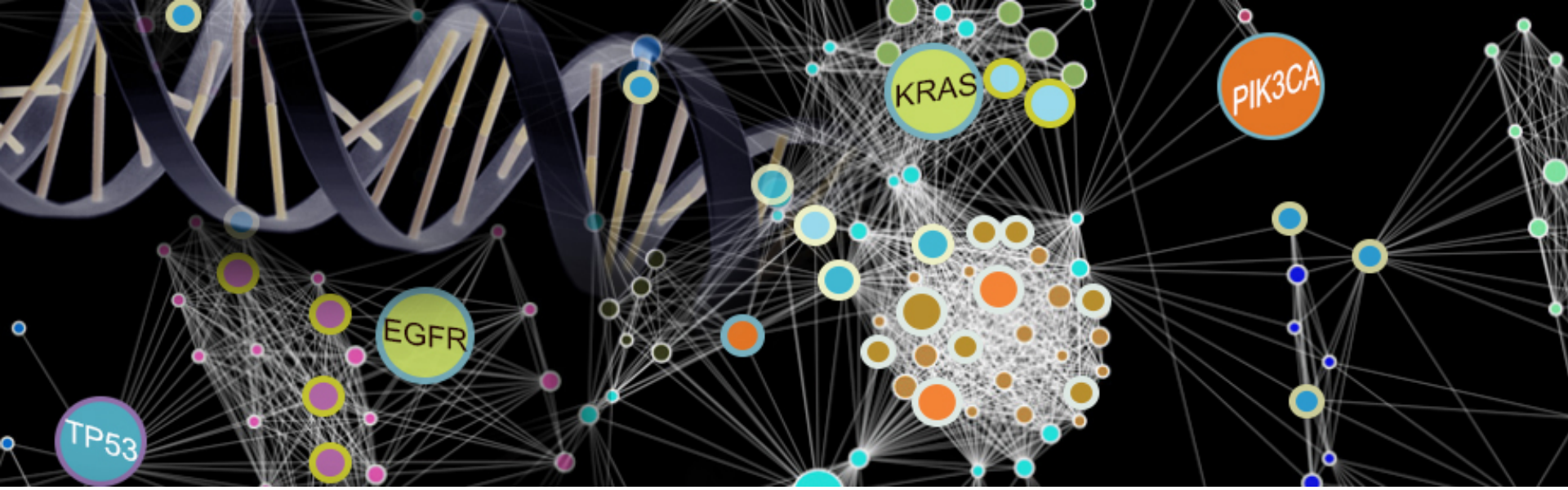
Conclusions

Mutation burden is more than just the number of mutations in a tumor

The majority of mutations in cancer are passengers – it is important to know which are drivers

Cancer evolution generates heterogeneity in many ways

Quantifying these factors is necessary to develop effective treatment strategies



THANK YOU!