BIMM 143

Cancer Genomics & Immunoinformatics

Lecture 18

Barry Grant

UC San Diego

http://thegrantlab.org/bimm143

Today's Menu

Cancer Genomics	Brief review of cancer fundamentals, What is cancer and what causes it?
Mining Cancer Genomic Data	**Hands-on analysis** to identify genomic changes in different cancers and identify new targets for therapy
Cancer Immunotherapy	**Hands-on analysis** to design personalized cancer vaccines and harness the patient's own immune system to fight cancer

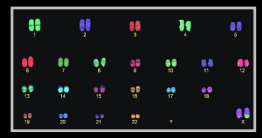
What is Cancer?

"Cancer is a name given to a collection of related diseases, where some of the body's cells begin to divide without stopping and spread into surrounding tissue"

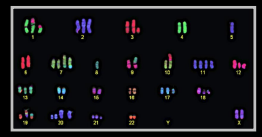
Source: https://www.cancer.gov NIH-NCI

Cancer is a disease of the Genome

- Caused by changes to genes that control the way our cells function, especially how they **grow and divide**.
- A major challenge in treating cancer is that every tumor is different: Each person's cancer has a unique combination of genetic changes (both "driver" & "passenger").
- As the cancer continues to grow, additional changes will occur.



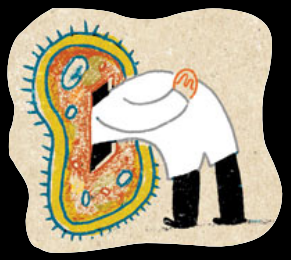
Healthy 46 chromosomes



Example cancer 59 chromosomes

Goals of Cancer Genome Research

- Identify changes in the genomes of tumors that drive cancer progression
- Identify new targets for therapy
- Select drugs based on the genomics of the tumor
- Provide early cancer detection and treatment response monitoring
- Utilize cancer specific mutations to derive neoantigen immunotherapy approaches



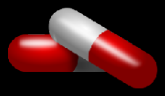
Finding Cancer Drivers

Motivation for adopting a genomics approach…

- Cancer is caused by mutations to specific genes

AGCT ⟶ AGAT

- Knowing which genes and proteins enables the development of **targeted treatments**

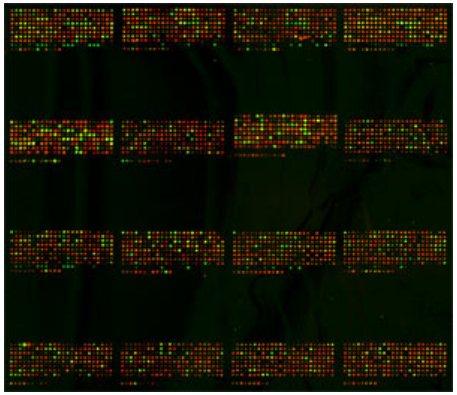


- 1st major Goal:
  Define ALL cancer genes!

Use A Cancer Genomics Approach

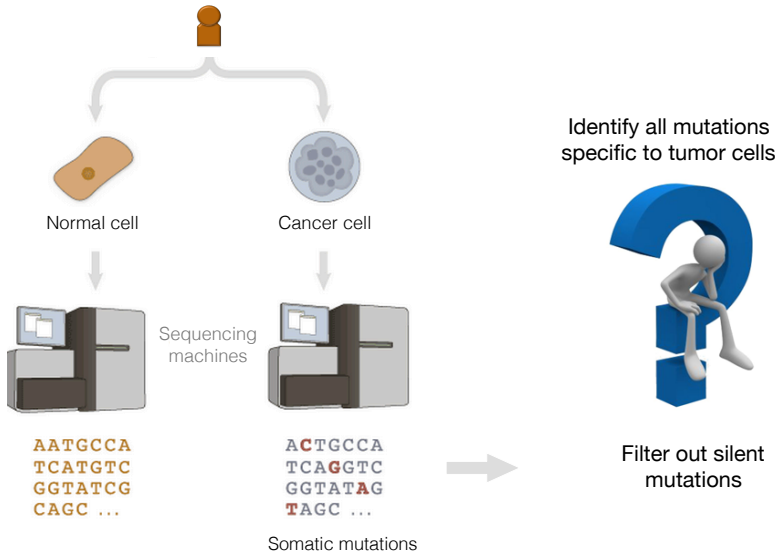
Arrays

Parallel Sequencing

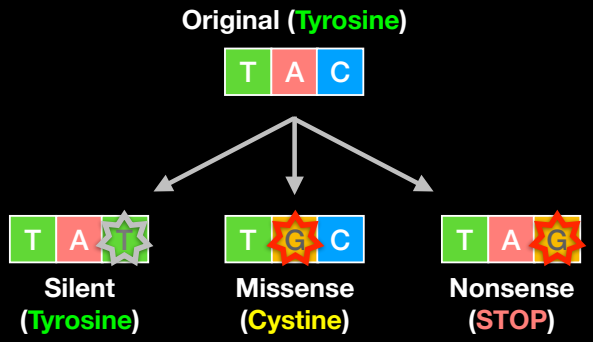


ACTCAGCCCCAGCGGAGGTGAAGGACGTCCTTCCCCAGGAGCCGGTGAGA
AGCGCAGTCGGGGGCACGGGGATGAGCTCAGGGGCCTCTAGAAAGATGTA
GCTGGGACCTCGGGAAGCCCTGGCCTCCAGGTAGTCTCAGGAGAGCTACT
CAGGGTCGGGCTTGGGGAGAGGAGGAGCGGGGGTGAGGCCAGCAGCAGGG
GACTGGACCTGGGAAGGGCTGGGCAGCAGAGACGACCCGACCCGCTAGAA
GGTGGGGTGGGGAGAGCATGTGGACTAGGAGCTAAGCCACAGCAGGACCC
CCACGAGTTGTCACTGTCATTTATCGAGCACCTACTGGGTGTCCCCAGTG
TCCTCAGATCTCCATAACTGGGAAGCCAGGGGCAGCGACACGGTAGCTAG
CCGTCGATTGGAGAACTTTAAAATGAGGACTGAATTAGCTCATAAATGGA
AAACGGCGCTTAAATGTGAGGTTAGAGCTTAGAATGTGAAGGGAGAATGA
GGAATGCGAGACTGGGACTGAGATGGAACCGGCGGTGGGGAGGGGGAGGG
GGTGTGGAATTTGAACCCCGGGAGAGAAAGATGGAATTTTGGCTATGGAG
GCCGACCTGGGGATGGGGAAATAAGAGAAGACCAGGAGGGAGTTAAATAG
GGAATGGGTTGGGGGCGGCTTGGTAACTGTTTGTGCTGGGATTAGGCTGT
TGCAGATAATGGAGCAAGGCTTGGAAGGCTAACCTGGGGTGGGGCCGGGT
TGGGGTCGGGCTGGGGGCGGGAGGAGTCCTCACTGGCGGTTGATTGACAG
TTTCTCCTTCCCCAGACTGGCCAATCACAGGCAGGAAGATGAAGGTTCTG
TGGGCTGCCCCGACCCGCTAGAAGGTGGGGTGGGGAGAGCATGTGGACTA
GGAGCTAAGCCACAGCAGGACCCCCACGAGTTGTCACTGTCATTTATCGA
GCACCTACTGGGTGTCCCCAGTGTCCTCAGATCTCCATAACTGGGAAGCC
AGGGGCAGCGAC

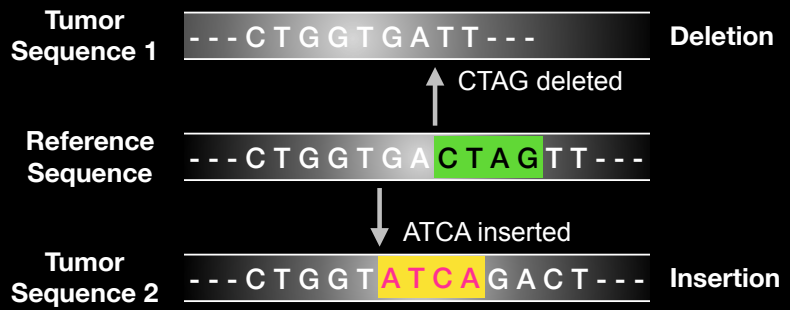
Finding Cancer Associated Mutations



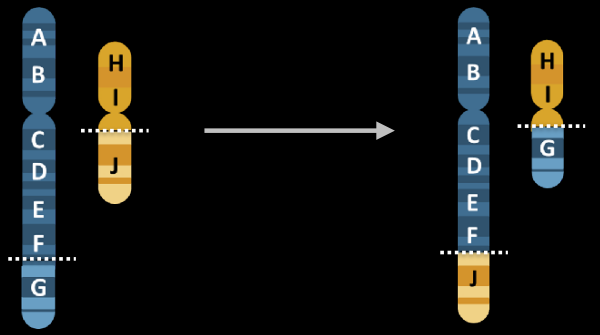
Mutations detected: Point mutations



Mutations detected: Indels



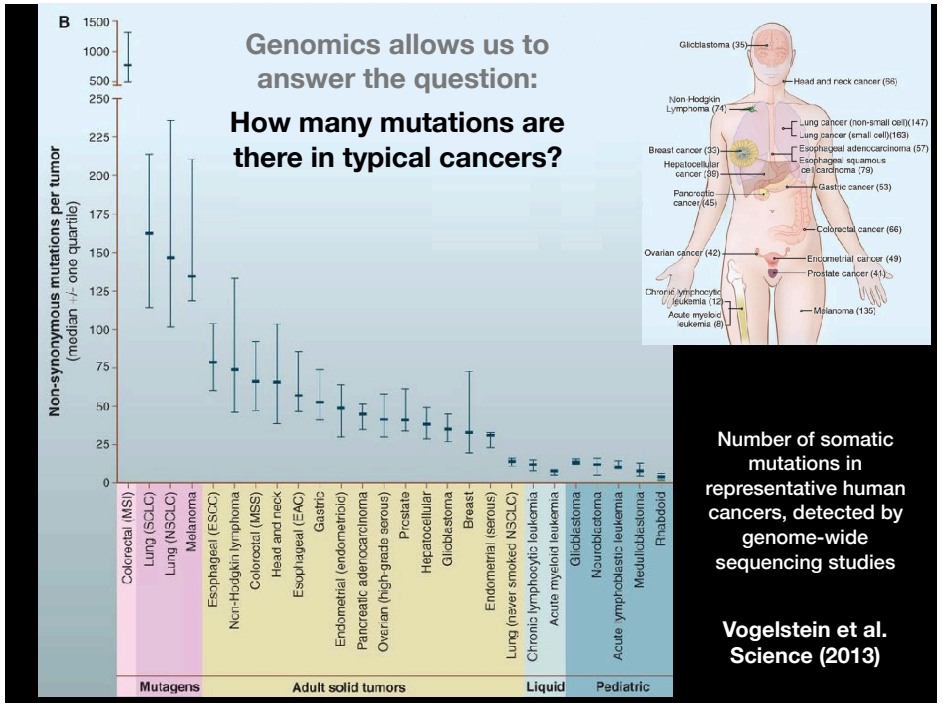
Mutations detected: Translocations



What can go wrong in cancer genomes?

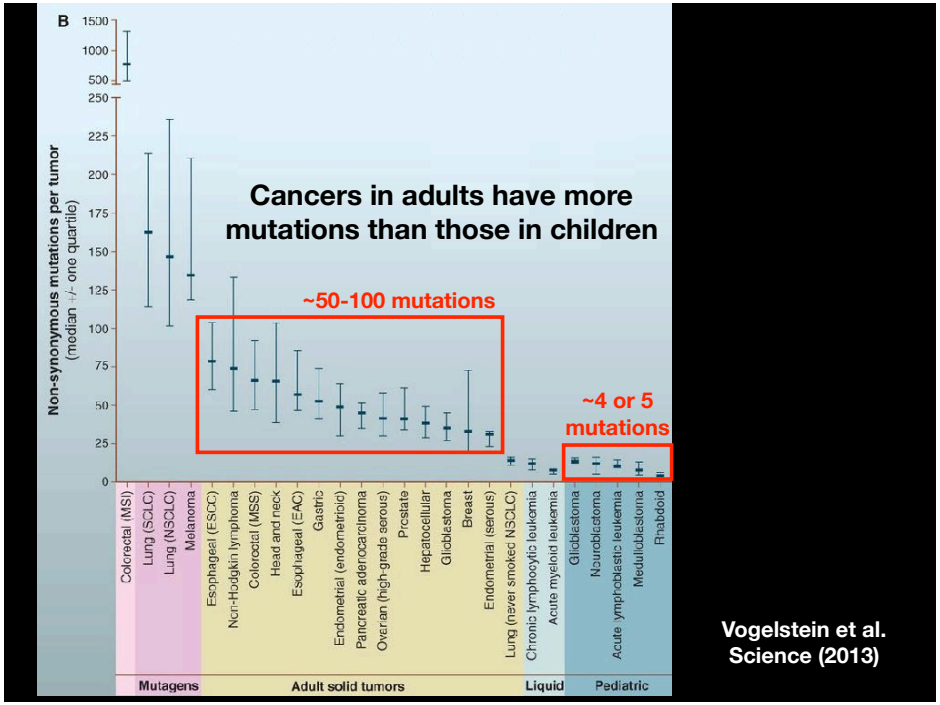
Type of change	Some common technology to study changes
DNA mutations	WGS, WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array, SNP array, WGS
DNA methylation	Methylation array, RRBS, WGBS
mRNA expression changes	mRNA expression array, RNA-seq
miRNA expression changes	miRNA expression array, miRNA-seq
Protein expression	Protein arrays, mass spectrometry

WGS = whole genome sequencing, WXS = whole exome sequencing
RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing

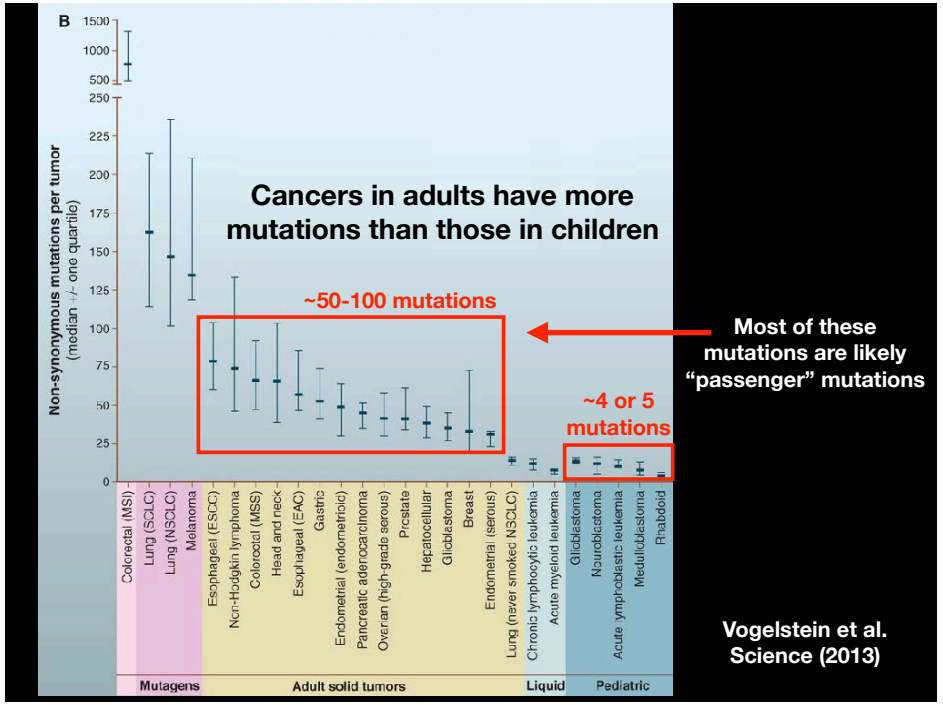


Number of somatic mutations in representative human cancers, detected by genome-wide sequencing studies

Vogelstein et al. Science (2013)

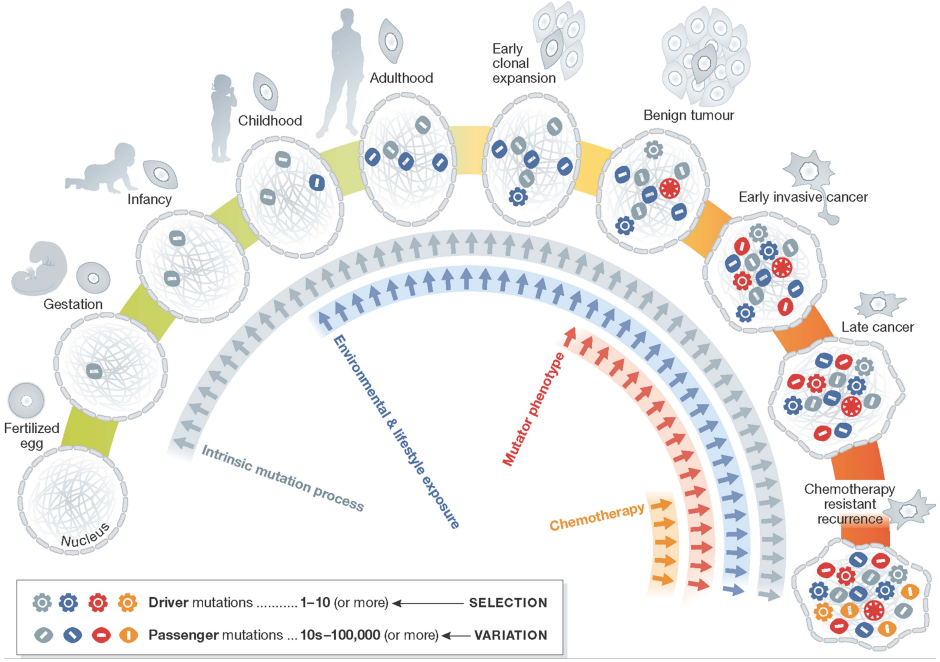


Vogelstein et al. Science (2013)

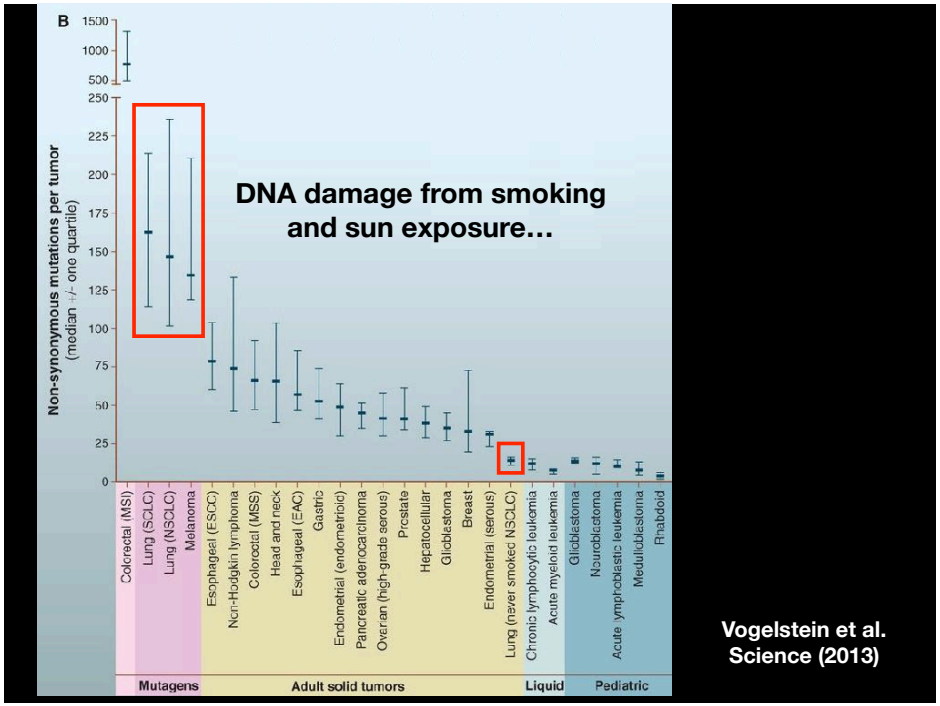


Most of these mutations are likely "passenger" mutations

Vogelstein et al. Science (2013)

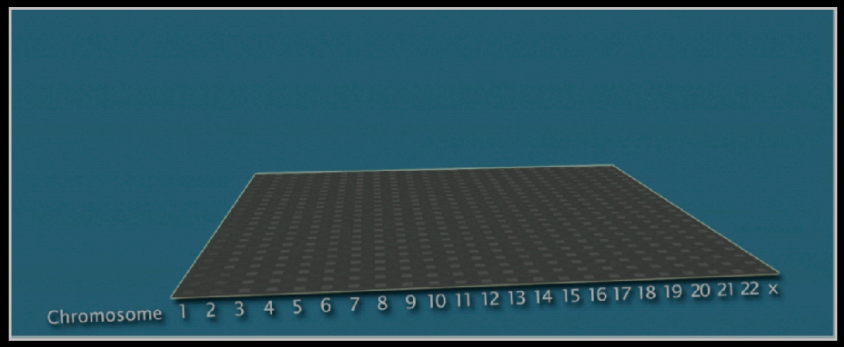


Mike Stratton. EMBO Molecular Medicine (2013)



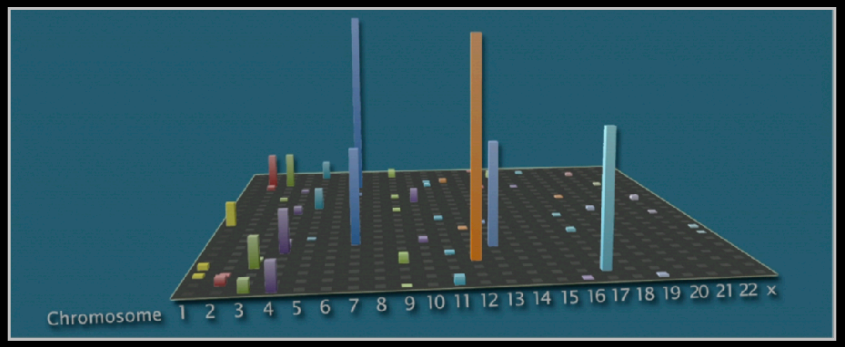
Vogelstein et al. Science (2013)

Genomic approaches can identify the genes most commonly mutated in cancer



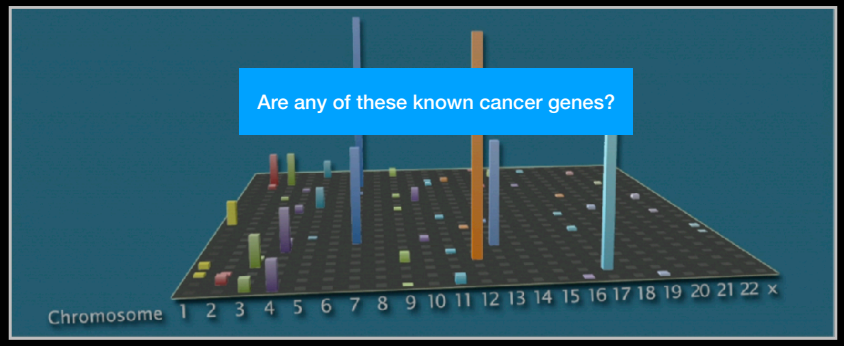
Arrange all genes in a matrix, ordered by chromosomes

Identifying genes most commonly mutated in cancer



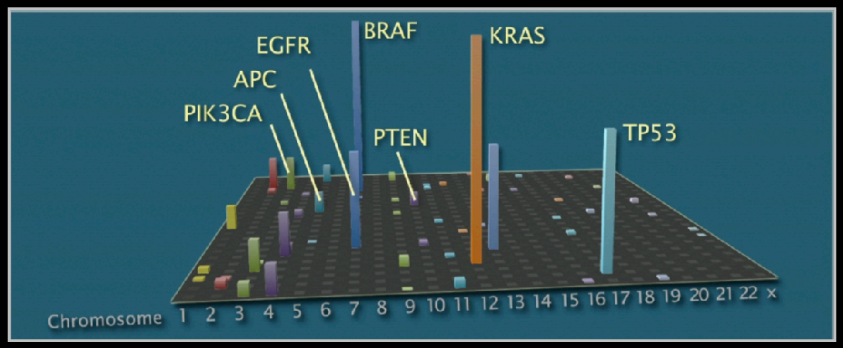
Add all data together to see which genes are most often mutated

Identifying genes most commonly mutated in cancer



Add all data together to see which genes are most often mutated

Identifying genes most commonly mutated in cancer



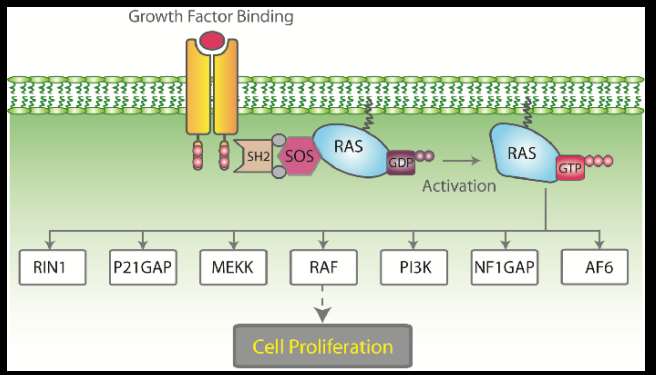
Many are famous porto-oncogenes, many others are new cancer genes!

Three Main Types of Cancer Genes:

- Oncogenes, such as **Ras**, normally function to accelerate cell division and growth. They can be mutated to act like stuck gas pedals.
- Tumor suppressor genes, such as **p53** normal act like breaks. Mutations can cause these breaks to fail.
- DNA repair genes, such as **BRCA1** & **2**, normally function to fix minor damage to DNA when it replicates. When these genes are mutated, DNA damage can accumulate and lead to cancer.

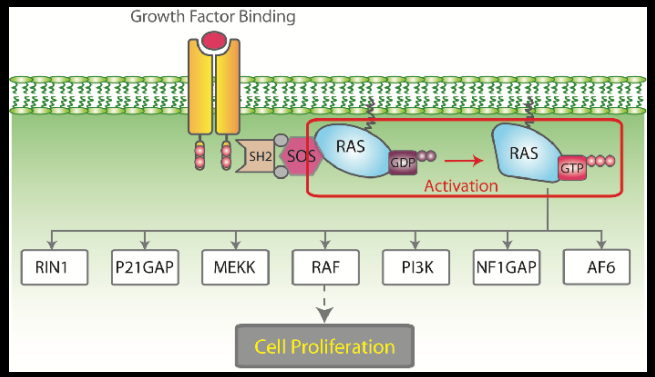
Cell growth and survival genes

Many participate in signaling pathways that promote cell proliferation (E.G. EGFR, Ras, BRAF, MEK etc.)



Cell growth and survival genes

Many participate in signaling pathways that promote cell proliferation (E.G. EGFR, Ras, BRAF, MEK etc.)



Regulators of Cell Cycle and Cell Death

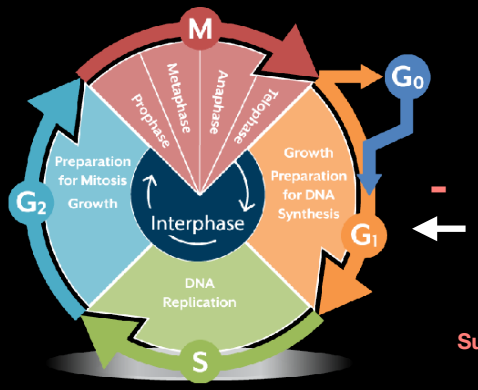
Some stimulate the cell cycle

Cyclin D1

+

CDK4

Oncogenes



Some inhibit the cell cycle

-

P53

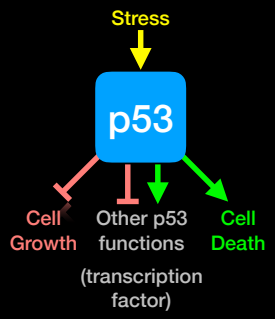
RB

Suppressor genes

p53 Regulates Cell Division

Probably the most famous cancer gene that is mutated in about half of all tumors. Often called the '*guardian of the genome*'

- p53 normally shuts down cell division when a cell is stressed (e.g. by DNA damage)
- When DNA is damaged, p53 activates genes that stop cell growth or trigger the cell to die.
- Thus, p53 guards against changes to cells that might lead to tumor formation.
- It appears necessary to inactivate p53 to develop many forms of cancer.



Do it Yourself!

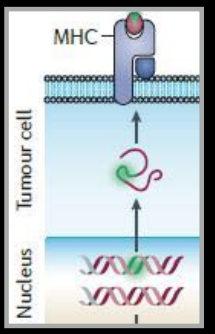
Hands-on time!

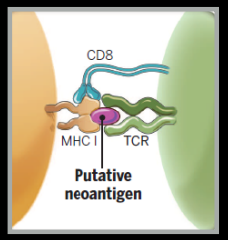
https://bioboot.github.io/bimm143_W18/lectures/#18

Part 1 Only Please

Cancer Immunotherapy

- Cancers genomes accumulate mutations
- Mutations in coding regions are translated in mutated protein sequences
- Mutated peptides can be presented as epitopes on **MHC** to **T cells**





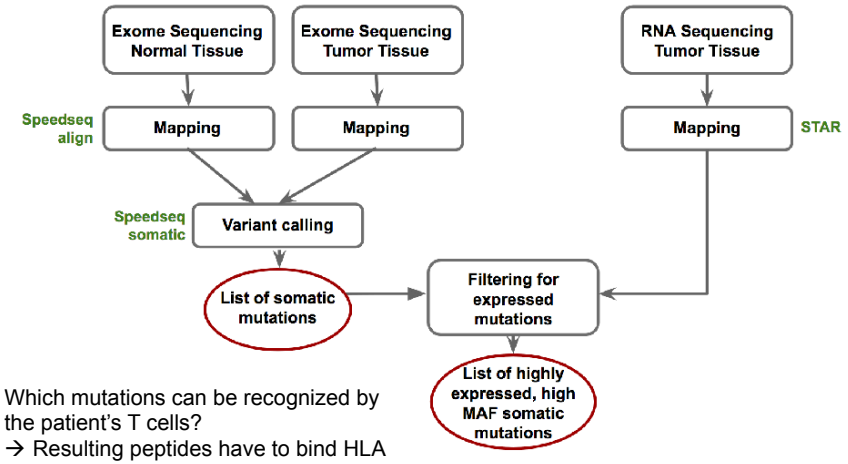
- **Neoepitopes** are presumably recognized by tumor-infiltrating lymphocytes (TILs)
- **Neoepitopes** are highly tumor-specific!

Coulie et al, Nat Rev Cancer. 2014 Feb;14(2):135-46
Schumacher & Schreiber, Science. 2015 Apr 3;348(6230):69-74

- **Vaccination**: Introduce or boost an immune response against a specific target (antigen)
- Cancer cells contain non-self antigens that *could* be recognized by T cells, but the presence of cancer means this mechanism has failed, typically by the tumor suppressing immune responses
- **Checkpoint blockade treatments**: Block immune suppressive mechanisms to boost T cell immune responses against cancer cells.
- **Problem**: Checkpoint blockade is unspecific, and will also boost unwanted autoimmune responses
- **Personalized Cancer Immunotherapy**: Boost anti-tumor response with vaccine containing peptides corresponding to cancer mutations that can be recognized by T cells.

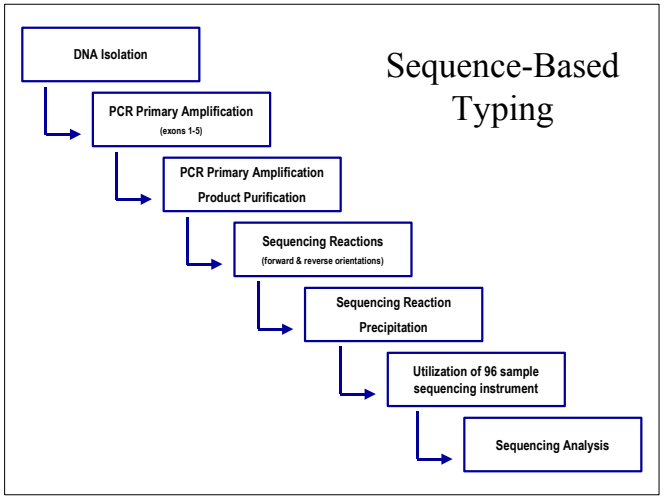
Q. How can such a vaccine be designed?

DNA and RNA sequencing identifies tumor specific somatic mutations

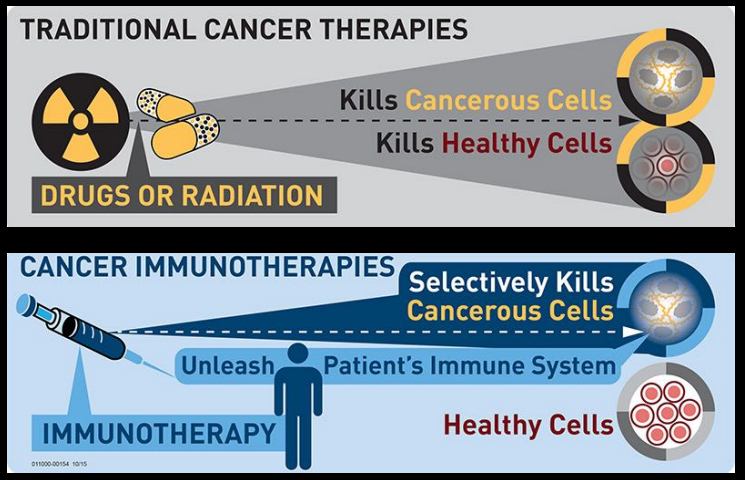


Which mutations can be recognized by the patient's T cells?
→ Resulting peptides have to bind HLA molecules of the patient

HLA Typing: Targeted sequencing of HLA locus



•http://www.ashi-hla.org/publicationfiles/ASHI_Quarterly/25_2_2001/highthrusbt3.htm



Do it Yourself!

Hands-on time!

https://bioboot.github.io/bimm143_W18/lectures/#18

Part 2: Designing a personalized cancer vaccine

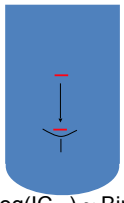
Bonus Slides
(For Reference)

Slide from: **Bjoern Peters** (LIAI)

Measuring and predicting MHC:peptide binding

Experimental Basis: MHC Binding Assay

List of peptides with allele specific binding affinity



Sequence	IC_{50}
QIVTMFEAL	3.6
LKGPDIYKG	308
NFCNLTSAF	50,000
AQSQCRTFR	38,000
CTYAGPFGM	143
CFGNTAVAK	50,000
...	

$\log(IC_{50})$ ~ Binding free Energy

low IC_{50} → high affinity

Impossible to measure all peptides

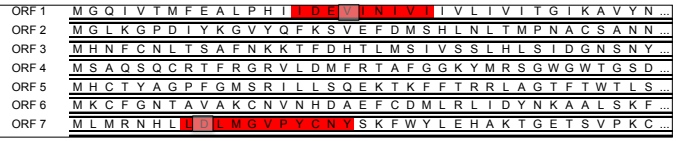
→ Predict binding peptides using machine learning

Find function F_i in F_i (Sequence) ≈ Affinity

$F_1, F_2, F_3, \ldots$

Many different approaches (ANN, SVM, HMM, LP, ...)

T cell epitope mapping



Slide from: **Bjoern Peters** (LIAI)

Calculate scoring matrix from affinities

Machine learning PSSM = Minimize the difference between predicted and measured binding affinities by varying the matrix values

N peptides with measured binding affinities

log (IC50)	Peptide
0.50	FQPQNGSFI
0.72	ISVANKIYM
2.37	RVYEALYYV
3.42	FQPQSGQFI
3.46	LYEKVKSQL
4.07	FKSVEFDMS
4.18	FQPQNGQFH
4.24	VLMLPVWFL
4.39	YMTLGQVVF
4.40	EDVKNAVGV
4.90	VFYEQMKRF
...	

	HLA A*0201								
	1	2	3	4	5	6	7	8	9
A	-0.3	0.8	-0.3	-0.3	-0.2	-0.3	0.0	0.0	-0.9
C	0.2	0.9	0.0	0.3	-0.5	-0.1	0.1	0.2	0.4
D	0.8	0.9	-0.4	-0.3	0.3	0.2	0.4	0.3	0.6
E	0.6	-0.4	0.7	-0.2	0.1	-0.4	-0.2	-0.2	-0.5
F	-1.3	0.5	-0.5	0.1	-0.1	0.0	-0.3	-0.4	-0.8
G	-0.2	0.1	0.3	-0.1	0.0	0.4	0.3	-0.1	0.2
H	1.1	0.9	-0.1	0.4	0.1	0.2	0.0	0.2	0.8
I	-0.4	-0.7	-0.4	0.1	-0.1	-0.4	-0.5	0.5	-1.4
K	-0.3	0.0	1.1	0.1	0.1	0.6	0.9	0.2	0.9
L	0.0	-1.9	-0.4	-0.2	0.0	-0.2	0.0	-0.1	-1.1
M	-0.7	-1.2	-0.7	0.2	-0.6	0.0	0.0	0.0	-0.8
N	-0.1	0.3	0.1	0.3	-0.1	-0.3	0.0	0.2	0.7
P	1.2	0.5	0.6	-0.3	0.4	0.0	-0.4	-0.5	0.7
Q	0.4	-1.1	0.0	-0.1	0.4	-0.2	-0.3	0.2	0.7
R	-0.2	0.9	1.0	0.3	0.1	0.4	0.7	0.0	0.9
S	-0.3	0.1	0.1	-0.4	0.1	0.3	-0.2	-0.1	0.2
T	-0.2	-0.5	0.1	0.4	0.1	-0.5	0.2	0.0	-0.1
V	-0.1	-0.9	-0.1	0.2	0.0	-0.3	0.1	0.1	-1.9
W	0.0	0.7	-0.5	-0.2	-0.1	0.2	-0.3	-0.1	0.4
Y	-0.3	0.2	-0.6	0.2	0.0	0.4	-0.4	-0.3	0.8
Offset:	4.3								

Your Turn

Read and share your thoughts on the following class *Readings*

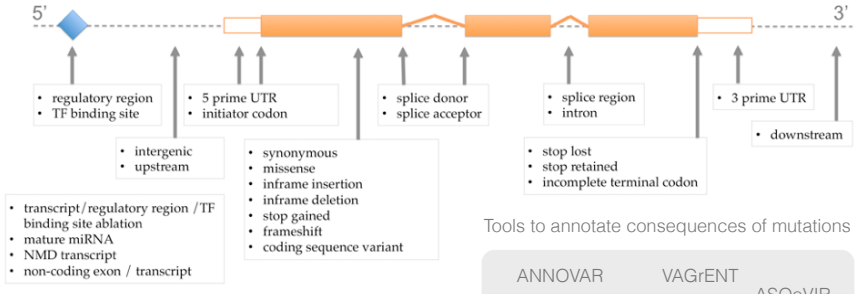
- Calling cancer's bluff with neoantigen vaccines
- Can genomics help detect early cancer and monitor treatment effectiveness?
- The increasing cost of cancer therapies

https://bioboot.github.io/bimm194_W18/readings/

1. Predict consequences of mutations

ACTGCCTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCCATGCTACTGCATCTCGGGTTAACTC
GACGTTTTTCATGCATGTGTGCACCCCAATATATATGCAACTTTTGTGCACCTCTGTCACGCGCGAGTTGGCA
CTGTCGCCCCTGTGTGCATGTGCACTGTCTCTCGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCGCTT
AACCCGTACTCCCATGCTACTGCATCTCGGGTTAACTCGACGTTTTGCATGCATGTGTGCACCCCAATATATA
TGCAACTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGTCTCTCGA

Map mutations into genome annotations to predict its possible effect



Tools to annotate consequences of mutations

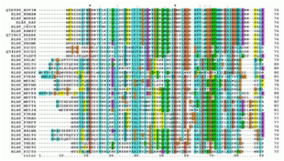
ANNOVAR VAGrENT Ensembl VEP ASOoVIR snpEff annTools

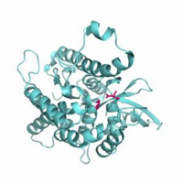
2. Assess the functional impact of nsSNVs

nsSNVs = non-synonymos Single Nucleotide Variant (missense)

```
ATC GAA GCA CGT
Met Glu Ala Gly

ATC GAC GCA CGT
Met Asp Ala Gly
```

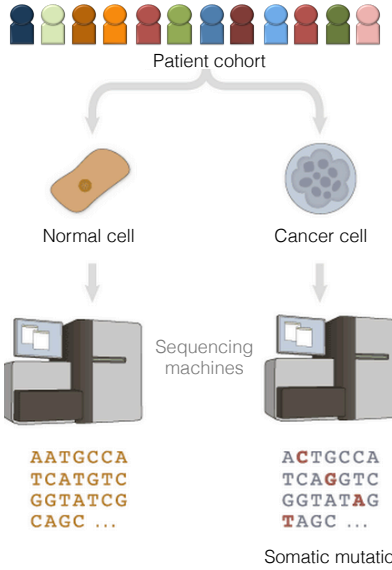




Computational methods to assess the functional impact of nsSNVs

MutationTaster, LogRe, Condel, MutPred, SNPs&GO, CanPredict, PolyPhen2, CHASM, SNPeffect, SIFT, MutationAssessor, PMut, transFIC

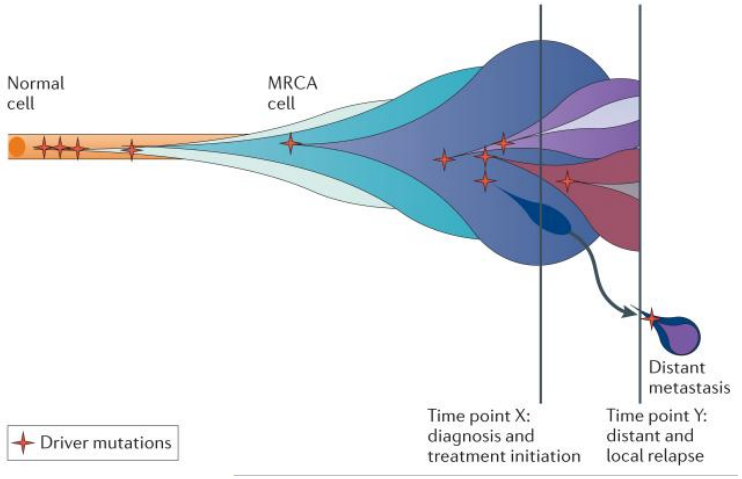
3. Identify cancer drivers from somatic mutations



Which mutations are cancer drivers?

Find signals of selection across tumors

Cancer is an evolutionary process



Yates and Campbell et al, Nat Rev Genet 2012

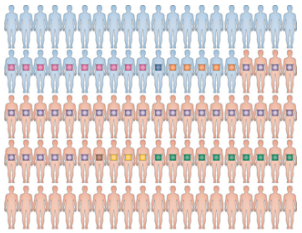
How to differentiate drivers from passengers?

```
ACTGCCTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCCATGCTACTGC
ATCTCGGGTTAACTCGACGTTTTTCATGCATGTGTGCACCCCAATATATATGCAACTT
TTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGT
CTCTCGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCC
ATGCTACTGCATCTCGGGTTAACTCGACGTTTTGCATGCATGTGTGCACCCCAATATA
TATGCAACTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCA
TGTGCACTGTCTCTCGAGTTTTGCATGCATGTGTGCACTGTGCACCTCTGTTACGTCT
```

How to differentiate drivers from passengers?

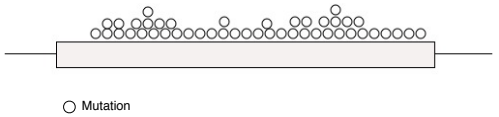
ACTG**C**CTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCCATGCTACTGC
ATCTCGGGTTAACTCGACGTTTT**T**CATGCATGTGTGCACCCCAATATATATGCA**A**CTT
TTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGT
CTC**T**CGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCC
ATGCTACTGCATCTCGGGTTAACTCGACGTTTT**G**CATGCATGTGTGCACCCCAATATA
TATGCA**A**CTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCA
TGTGCACTGTCTC**T**CGAGTTTT**G**CATGCATGTGTGCACTGTGCACCTCTGTTACGTCT

Find signals of positive selection across tumour re-sequenced genomes



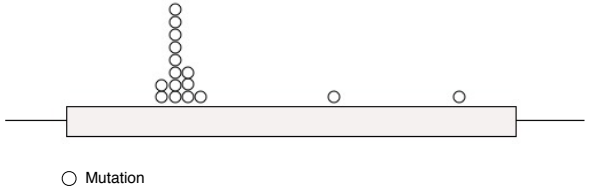
Signals of positive selection

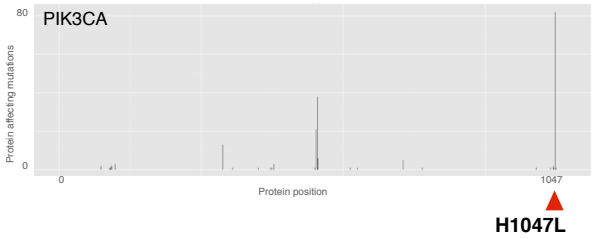
Recurrence **MuSiC-SMG / MutSigCV**



Identify genes mutated more frequently than background mutation rate

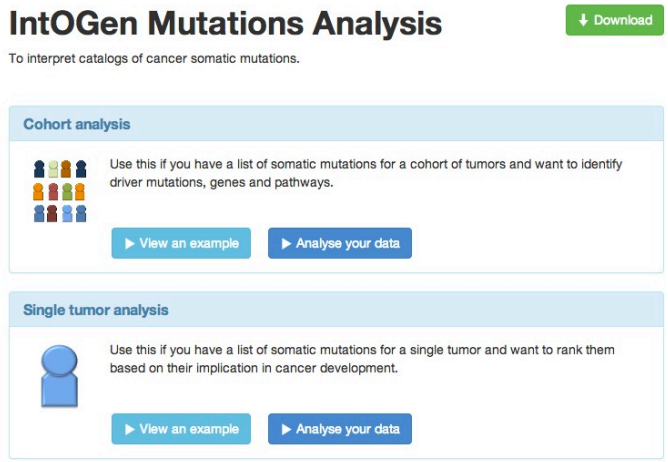
Mutation clustering **OncodriveCLUST**





PIK3CA is recurrently mutated in the same residue in breast tumours

http://www.intogen.org/mutations/analysis



Gonzalez-Perez et al, Nature Methods 2013