Module 4

Lecture 1

Introduction to Cancer

Sarah Groves, Ph.D.

Shayn Peirce-Cottler, Ph.D.



Pairs	
Student 1	Student 2
Yancy Martinez	Aryan Mhaskar
Owen Sines	Lauren McElfresh
Priya Shan	Conner Looney
Jake Carnell	Tahseen Azad
Delaney Broderi	Abhiram Ruthala
Jack Marchesi	Kaitlyn Cole

Pairs	
Student 1	Student 2
Asma Sanullah	Makayla Adams
Molly Janousek	Megan Sullivan
Ashley Hayden	Dane Wleklinski
Kidus Belay	Oge Ikem
Meliz Ozdemir	Eunela Manzano

Pairs	
Dani Folks	William Marschall
Mohamed Elfaki	Kai-Tang Tseng
Sara Elster	Lauren Kim
Tomas Daniel	Isabel Vikesland
Karinna Fink	Reagan Oswald
Roy Chen	Theresa Breckley

Pairs	
Vish Ramineni	Kat Horn
Christina Johnsc	Amelia Luongo
Grace Lee	Kiran Spencer
Zain Abed	Isabella Imburgia
Luke Friscola	Hayden Rue
Isabella Orlando	Christian Grosso

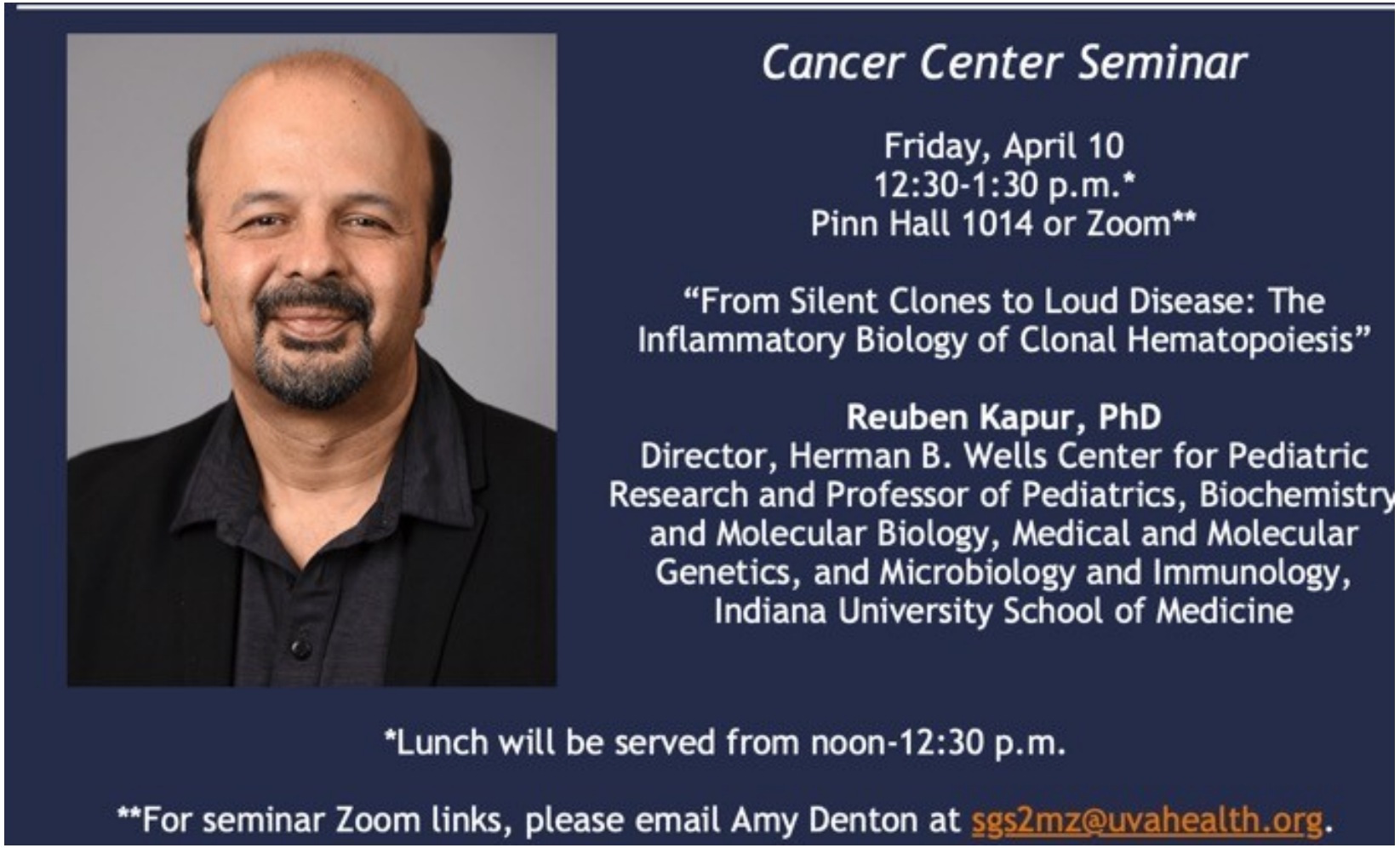
- Fork my repo
- Add your partner
- Confirm upstream remote!

https://github.com/smgroves/Module-4-Cancer

Learning Objectives for Module 3: Cancer

- *I can answer the question "what is cancer?" and explain what the **hallmarks** of cancer are.*
- *I can use **pandas and scikit-learn** to work with **RNA-seq data**, apply a machine learning method to a dataset, and validate the model.*
- *I understand how machine learning methods use **optimization** of a cost function to determine the best answer to my research question.*
- *I understand the difference between **supervised** and **unsupervised** learning methods and know when to use them.*
- *I know how to **avoid overfitting** and determine if my model is "good enough" using quantitative metrics.*

Cancer center seminar on Fridays!



Module 3 – Step-by-Step process for each set of pairs

Week 1

STEP 1: Learn about cancer (and a specific cancer type if desired!)

STEP 2: Explore the data-set (you will be given one)

STEP 3: Ask a question related to the data set that you are curious about

Week 2

STEP 4: Visualize the data with **dimensionality reduction** (unsupervised)

STEP 5: Implement clustering, **regression**, or **classification** using Python/scikit-learn to answer the question (supervised)

Week 3

STEP 6: Verify and validate the answer using **ML metrics**

STEP 7: Draw conclusions and **reflect on the ethical implications**

Week 4

STEP 8: Communicate your question's background, methods, conclusions, limitations, implications, and future work

Due Thursday April 9 by class:
Cancer Hallmark ABM simulations activity

Due Friday April 10 midnight:
GitHub Check-in #1
Due Tuesday April 14 by class:
Video comprehension quiz

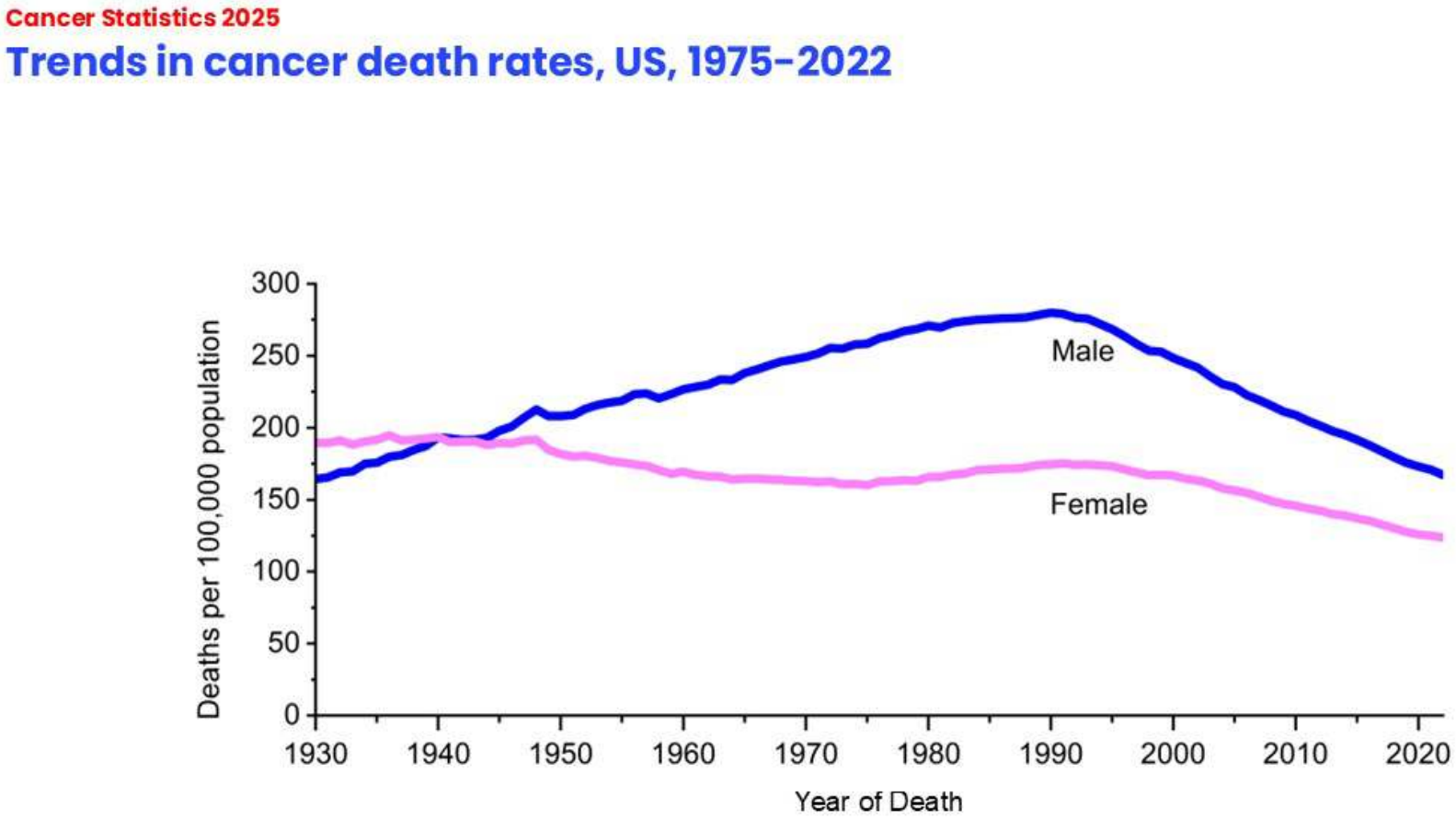
Due Friday April 17 midnight:
GitHub Check-in #2

Due Friday April 24 midnight:
GitHub Check-in #3

Due Tuesday May 5 midnight:
Final Project
Presentations: May 5 2pm (Final exam block)

The problem of cancer

- 1 in 3 people will get cancer in their lifetime
- 1/5 related to tobacco use, another 1/5 tied to obesity, alcohol, or infection
- #2 killer of Americans, accounting for 1 in 4 deaths in the US
- 200 billion dollars per year

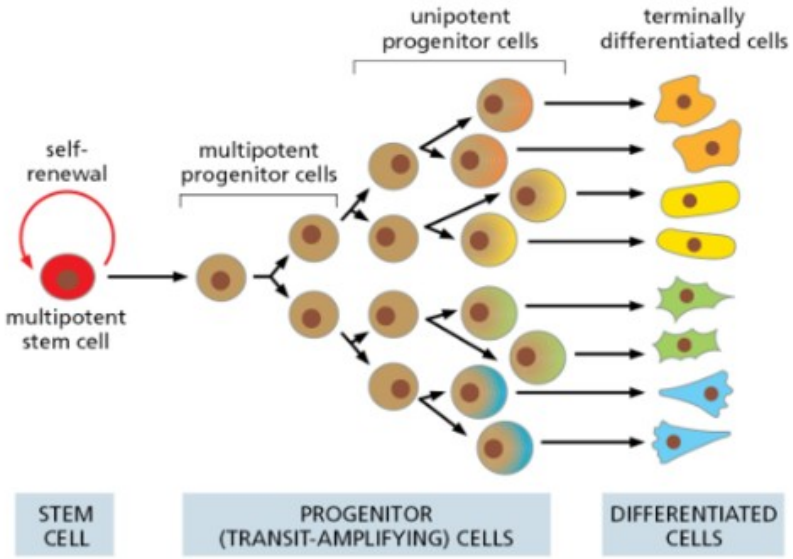


Let’s think about what normal cells do

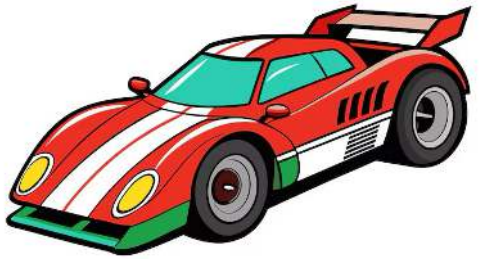
Normal cells start as stem cells
- This self-renewal is under tight control
- “Infinite” replicability, but checkpoint control
- No specialized functions yet, but strict lineage commitment

Genes for stem cell turn OFF and genes for specialized functions turn ON
- May have a few possible final identities (multipotent)

Differentiated cells: locked into one cell type that performs a particular job
- No division or tightly controlled
- Anchored in proper location
- One main cell “task”
- Can die by apoptosis









Core rules of normal cells



1. Grow only when told

"Gas Pedal"

Growth factors provide "permission" to grow, and without this signal, the cell doesn't divide

Ex: Growth factor → receptor activation → division genes turn ON → cell divides



2. Die when you should

"Brake pedal"

Programmed cell death (apoptosis) is triggered by cell damage

Ex: UV damage → cellular DNA has errors p53 activation → cell dies



3. Replicate a limited number of times

"Limited fuel tank"

Telomere shortening limits divisions (~50 max for a normal cell) and prevents endless replication



4. Stay where you belong

"Lane sensors"

Contact inhibition and tissue architecture keep cell from dividing when packed with neighbors & cells don't migrate without permission

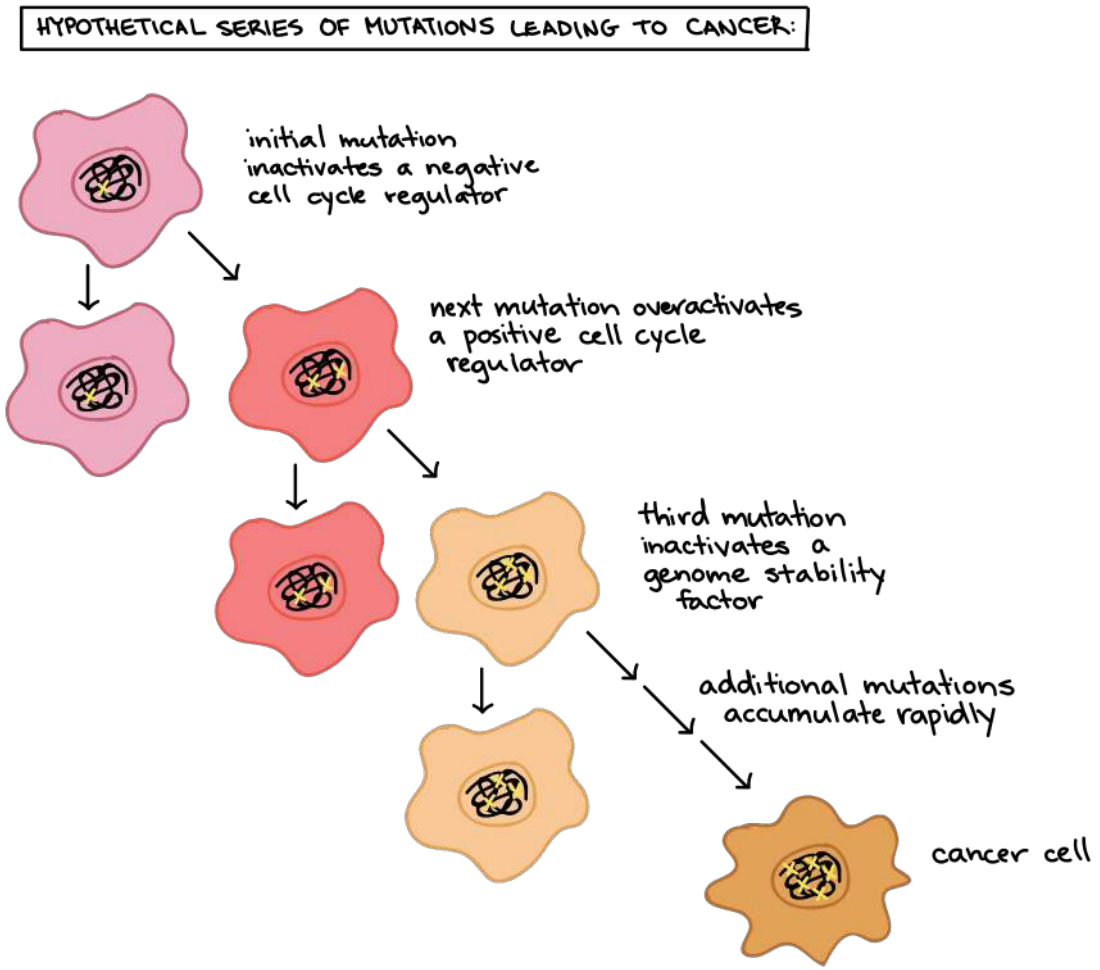


5. Maintain cell identity

"Carrier route for delivery truck"

Differentiated cells stay differentiated and genes for other cell types stay OFF

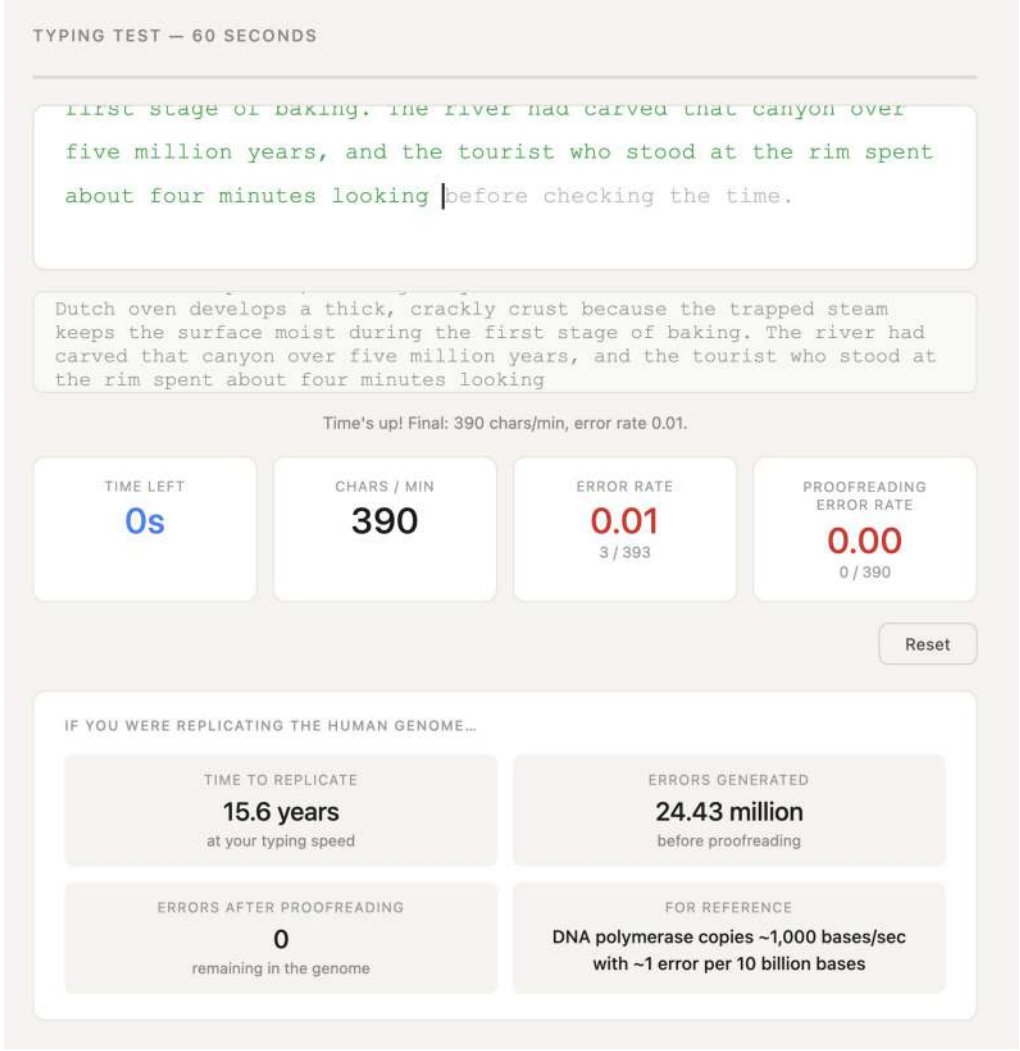
Cancer occurs when a cell breaks multiple rules via **mutations.**



- **3 billion** base pairs copied every cell division
- Human body has **37 trillion** cells (3.7×10^{13})
- Lifetime total divisions approximated at $\mathbf{10^{16}}$

https://bio1220.biosci.gatech.edu/death-01/why-and-how-does-cancer-affect-humans-and-other-animals/

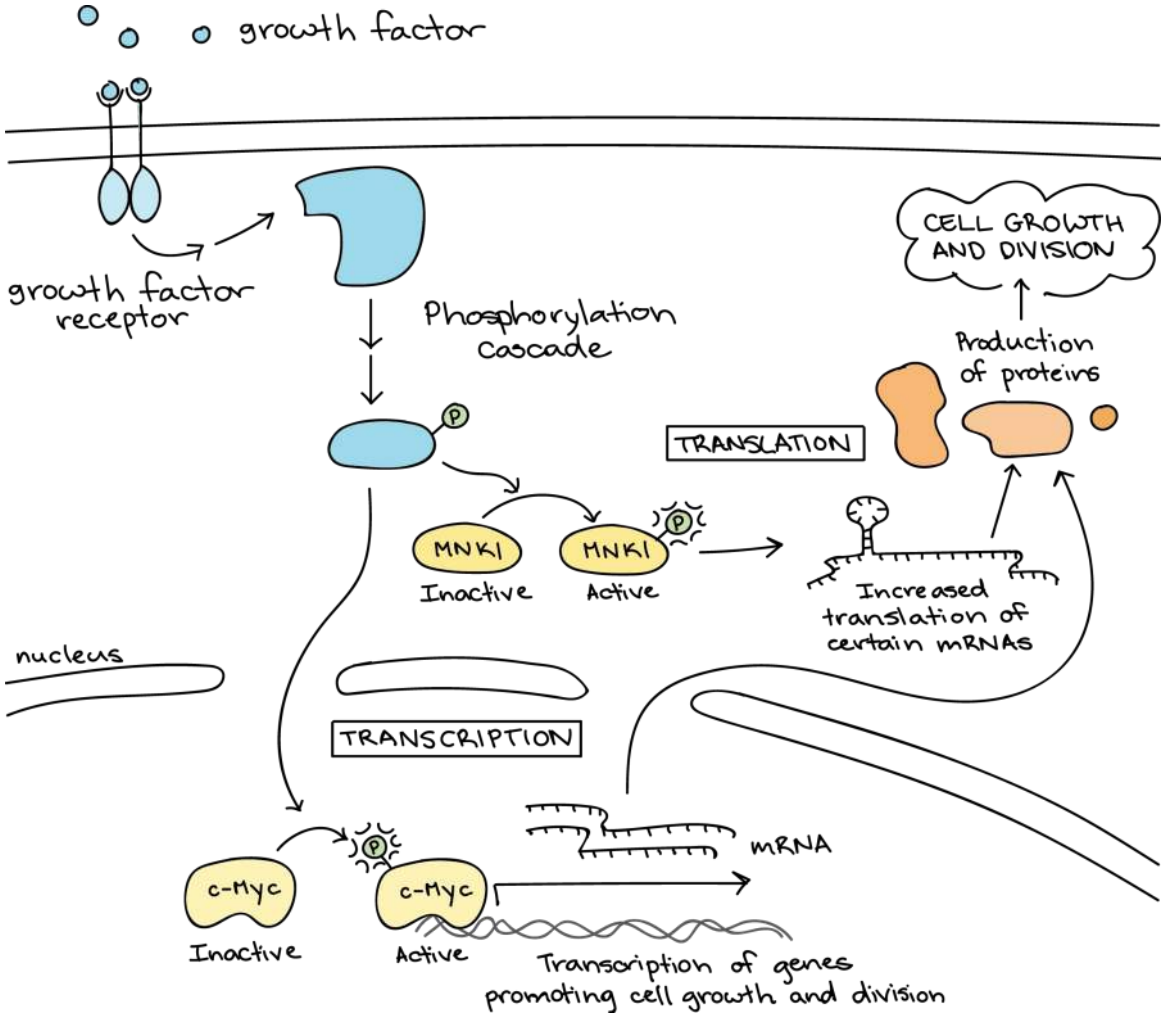
What if you were copying the genome?



- With proofreading and mismatch repair:
  - **1-3** mutations per cell division (**1 error per 10^9** bases)
  - = **10^7 total mutations** across the body over a lifetime
- Only need **4-10** mutations *in the right genes* to transform a normal cell to cancer

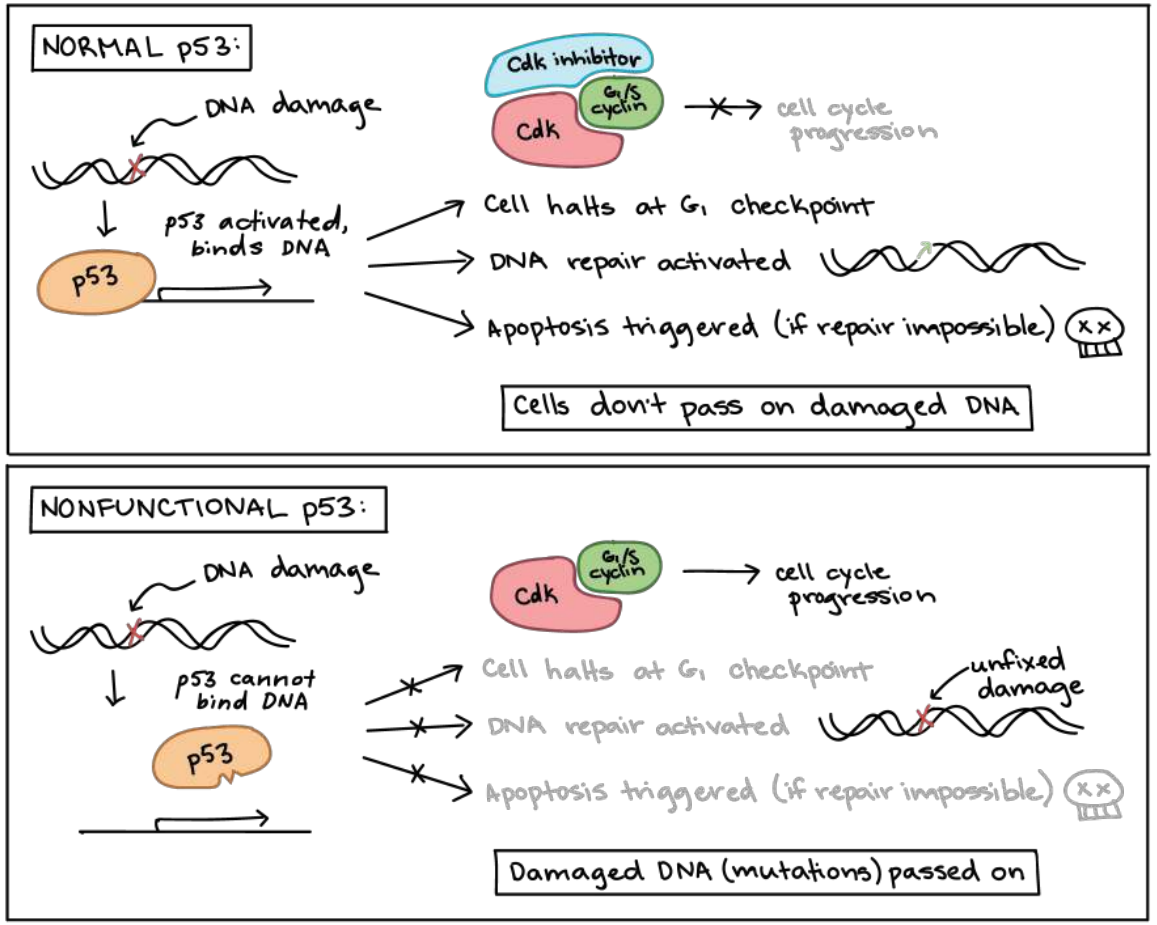
Gain-of-function mutations & overactivation of oncogenes

- Gas pedal gets stuck on
- Growth/division promoting genes
  - **Growth factors**: stimulate cells to grow and divide (e.g. EGF)
  - **RTKs**: receptors that receive growth signals (e.g. EGFR, HER2)
  - **Signal transducers**: pass growth signals from receptors to the nucleus (e.g. RAS, BRAF)
  - **Transcription factors** that activate genes necessary for cell division (e.g. MYC, FOS, JUN)
  - **Apoptosis blockers** (e.g. BCL-2)

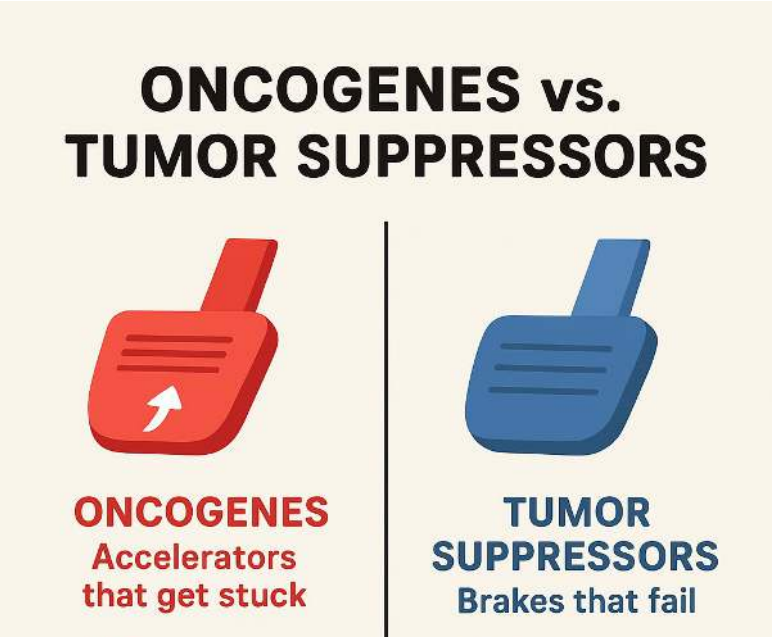


Inactivating mutations in tumor suppressors

- Brake pedal broken
- Normally suppress growth or maintain overall genetic stability
  - **Cell cycle regulators**: checkpoint control genes that prevent division if there are issues (e.g. p53, Rb1)
  - **Apoptosis inducers**: Trigger programmed cell death if cell is beyond repair
  - **DNA repair**: these fix errors in DNA before a cell divides (e.g. BRCA1/2)



We only need 4-10 mutations to cause cancer!!
But they need to be in the "right genes"



Gain of function/ overactivation

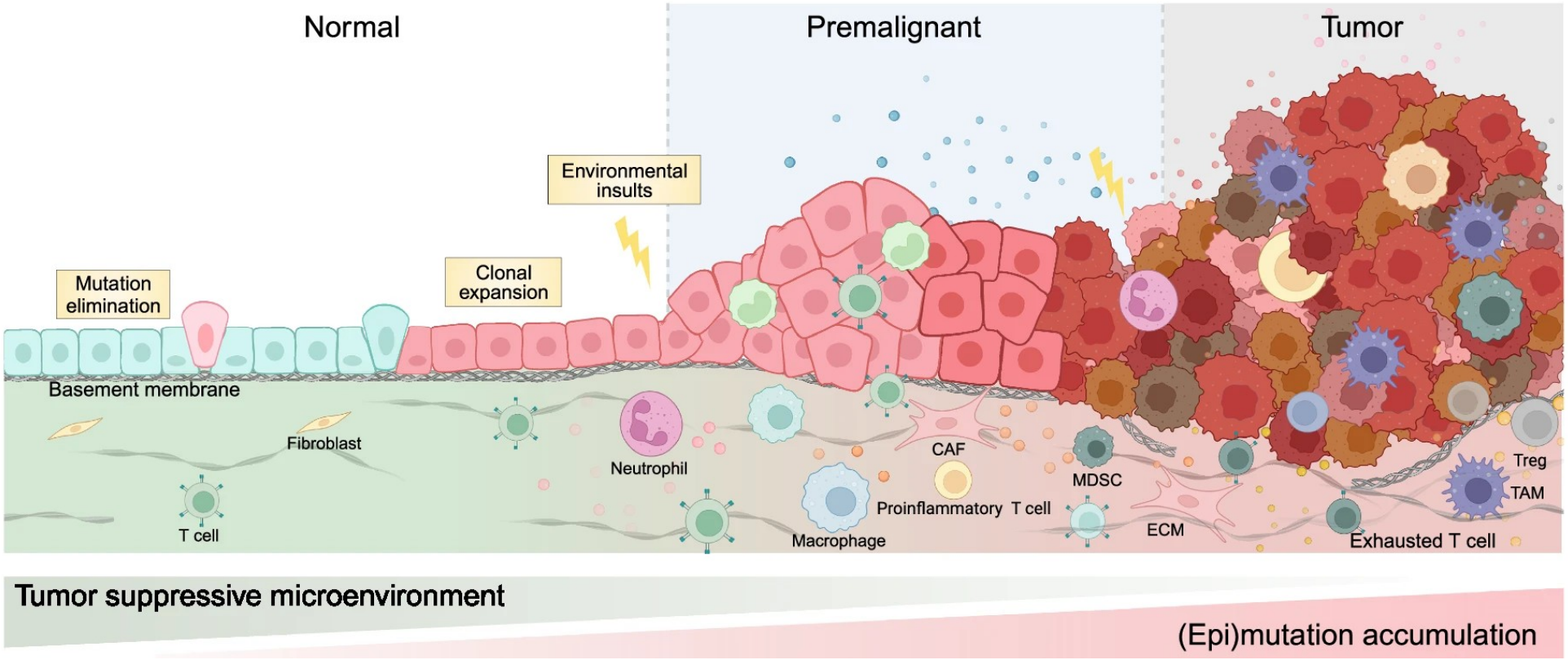
Loss of function/ inactivation

- Humans have **~20,000 protein-coding genes**
- **~723 genes** are listed as having evidence for a role in cancer
  - ~250–300 are tumor suppressor genes

Gene (and what it normally does)	ONC	TS
P53: if DNA damage, cell halts at G1 checkpoint (or damage is fixed)		

Cancer development progresses over years and years

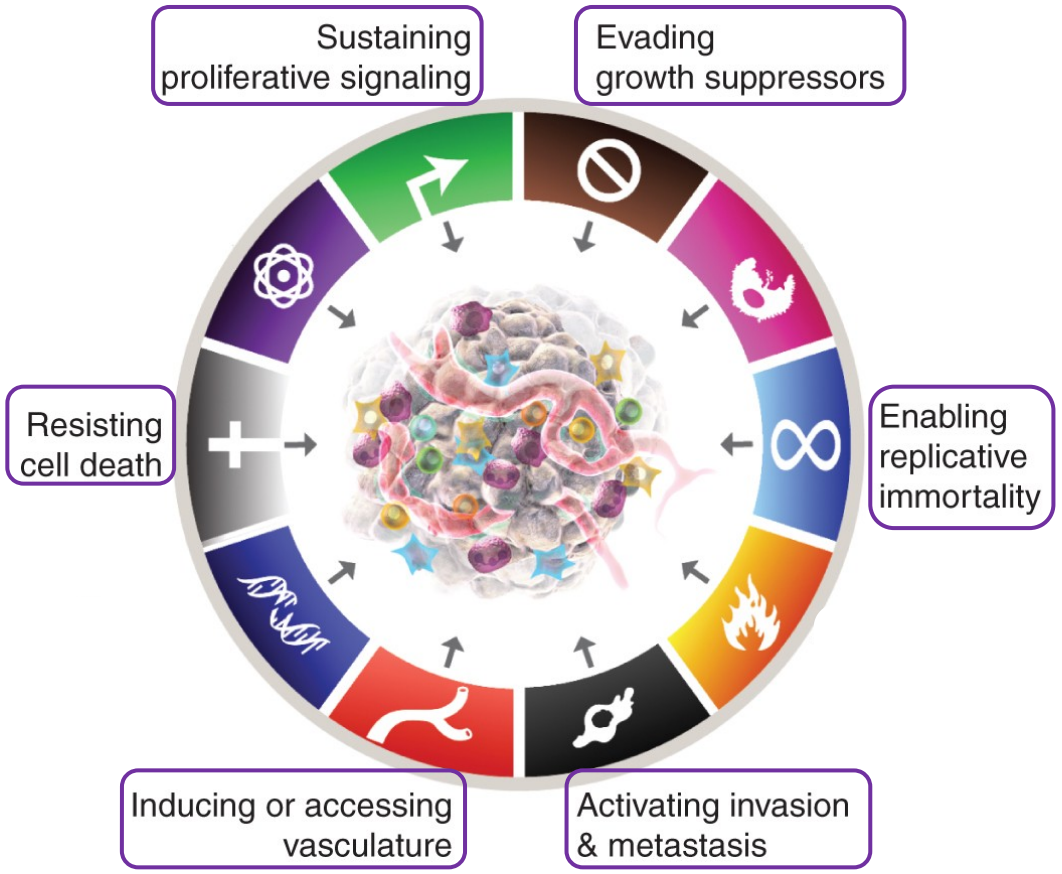
Cells accumulate mutations from the environment + inherited + random errors



"Tumor development proceeds via a process formally **analogous to Darwinian evolution**, in which a succession of genetic changes, each conferring one or another type of growth advantage, leads to the **progressive conversion** of normal human cells into cancer cells." – Hallmarks of Cancer

https://doi.org/10.1038/s41392-024-01848-7

Mutations in oncogenes and tumor suppressors lead to cancer cells with certain “hallmarks”



These are the ways cancer cells break the “rules”

Rule 1: Grow when you are told
Rule 2: Die when you should
Rule 3: Replicate a limited number of times
Rule 4: Stay where you belong
Rule 5: Maintain cell identity

https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2025-cancer-facts-figures.html

Hanahan and Weinberg, 2000; 2011; Hanahan, 2022

Breaking rule #1: grow only when told

Sustaining proliferative signaling

HOW:
Cancer cells produce their own growth signals (autocrine signaling) OR hyperactivate growth factor receptors
MECHANISM/GENES INVOLVED:

- Mutate RAS/KRAS/etc to always be on
- EGFR pathway upregulated

EXAMPLES:

- Pancreatic cancer ~90% have KRAS mutations
- Breast cancers often have HER2 upregulated (~30%)

Evading growth suppression

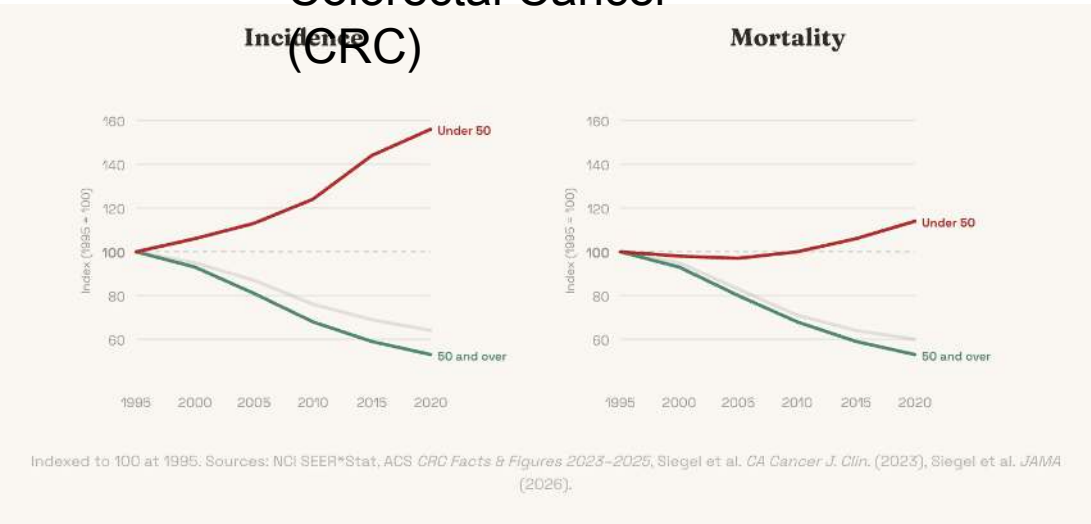
Reprogramming energy metabolism

https://www.figma.com/board/RyPLxKcCUJ3g46xwpIoNsv/The-Hallmarks-of-Cancer?node-id=0-1&t=kuFV2fuiY1HE2mLT-1

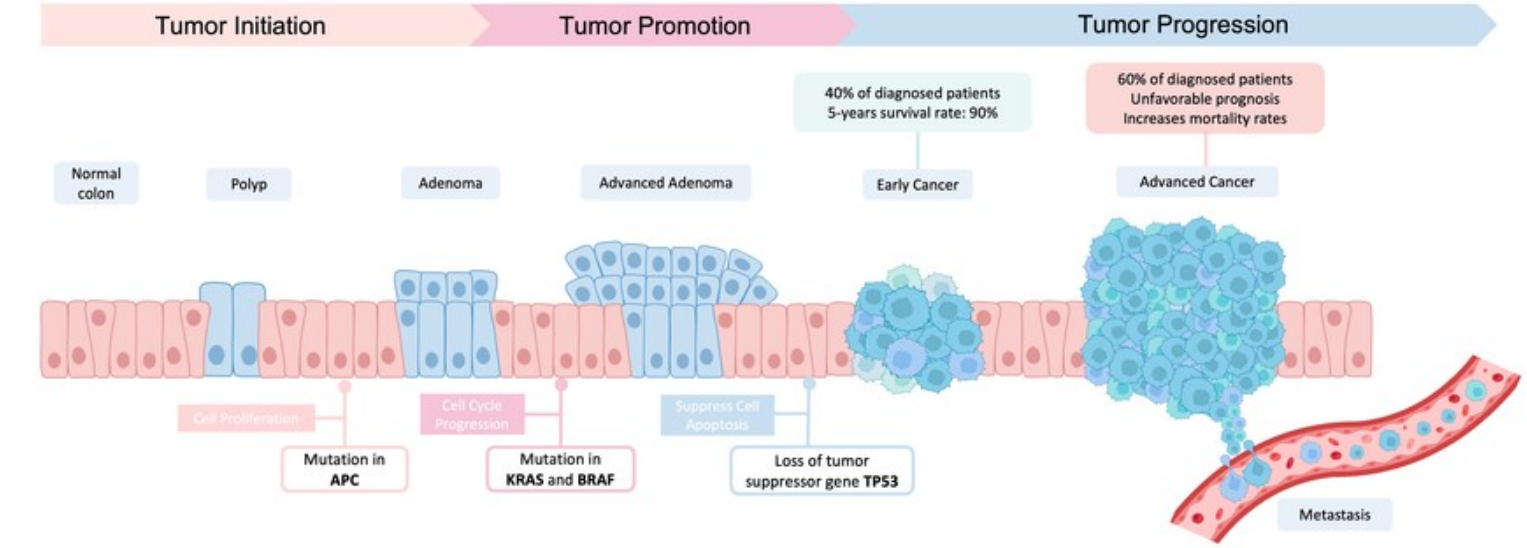
Senescent cells promote growth

Exploring the hallmarks: everything is connected!

Colorectal Cancer (CRC)







- Try running with **only one of the hallmarks "active."**
  - What happens when you only turn on "sustained growth signals" – and what keeps the cell from dividing?
  - If the cell can only "resist apoptosis," does it eventually spread?
- Try **different combinations**
  - How many hallmarks do you need for the cancer to grow?
  - What about for it to metastasize (reach new sites)?
- Once you feel comfortable with the first version, try the advanced version (extra six hallmarks added)

https://pmc.ncbi.nlm.nih.gov/articles/PMC11714384/#bib14

https://www.hankgreen.com/crc

Tumor staging (clinical data)

TNM Staging

T = Primary Tumor

TX Primary tumor cannot be evaluated
T0 No evidence of primary tumor
Tis Carcinoma in situ (early cancer that has not spread to neighboring tissue)
T1, T2, T3, T4 Size and/or extent of the primary tumor

N = Regional Lymph Nodes

NX Regional lymph nodes cannot be evaluated
N0 No regional lymph node involvement
N1, N2, N3 Involvement of regional lymph nodes (number and/or extent of spread)

M = Distant Metastasis

MX Distant metastasis cannot be evaluated
M0 No distant metastasis
M1 Distant metastasis

STAGE

- Stage 0: Carcinoma in situ
- Stage I, Stage II, and Stage III: Higher numbers indicate more extensive disease: greater tumor size, and/or spread of the cancer to nearby lymph nodes and/or organs adjacent to the primary tumor.
- Stage IV The cancer has spread to another organ.

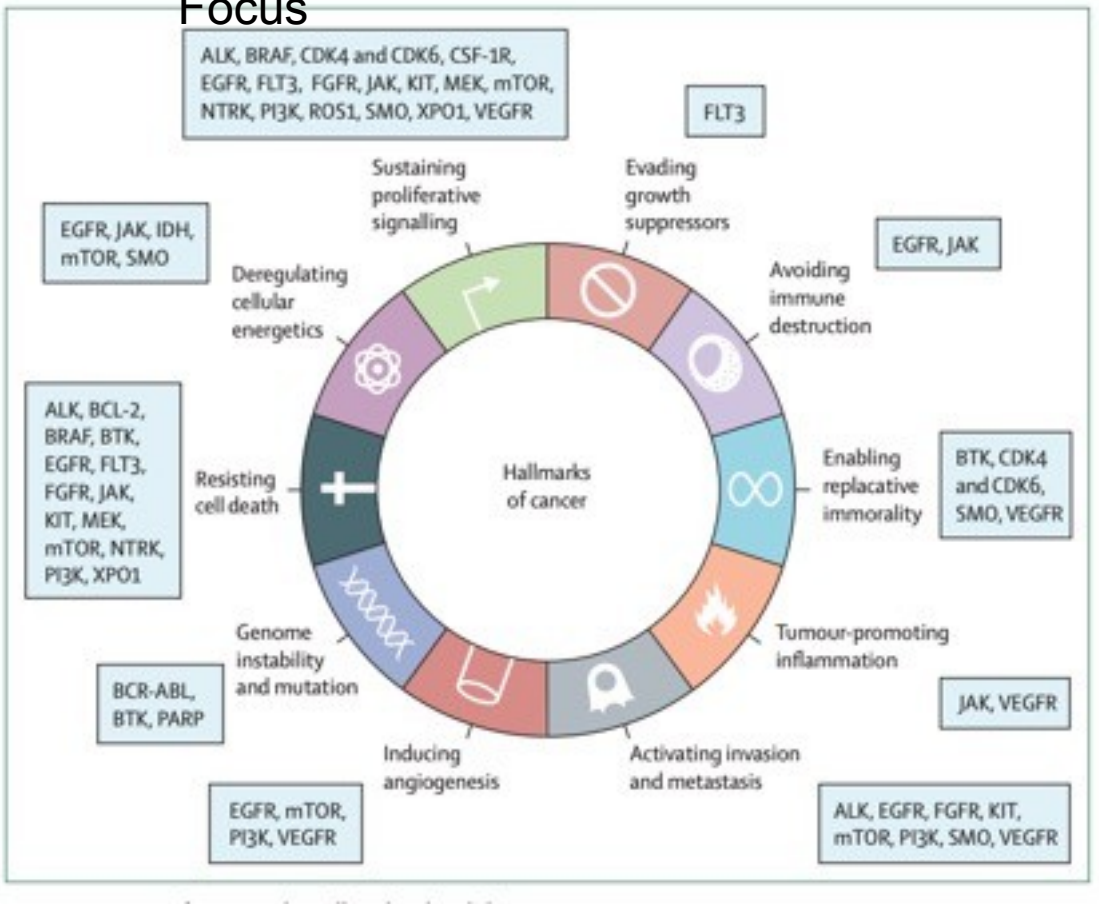
SUMMARY STAGE

- In situ: early cancer present only in the layer of cells in which it began
- Localized: limited to the organ in which it began
- Regional: spread to nearby lymph nodes or organs and tissues.
- Distant: spread to distant organs or distant lymph nodes.
- Unknown

Treatment options

- **Chemotherapy**: kills fast growing cells
- **Radiotherapy**: radiation applied to specific organ/region containing tumor
  - Causes DNA damage to kill proliferative cells
- **Targeted therapy**: Targeting a specific deregulated or mutated gene, like growth factor receptors that are overactive
  - **Hormone therapy**: targeting hormones that drive tumor growth
  - **Epigenetic therapy**: targeting epigenetic mechanisms like DNA methylation/gene silencing
- **Immunotherapy**: Enlisting immune cells to help kill the tumor

Targeted therapies by Hallmark Focus



https://doi.org/10.1016/S0140-6736(20)30164-1

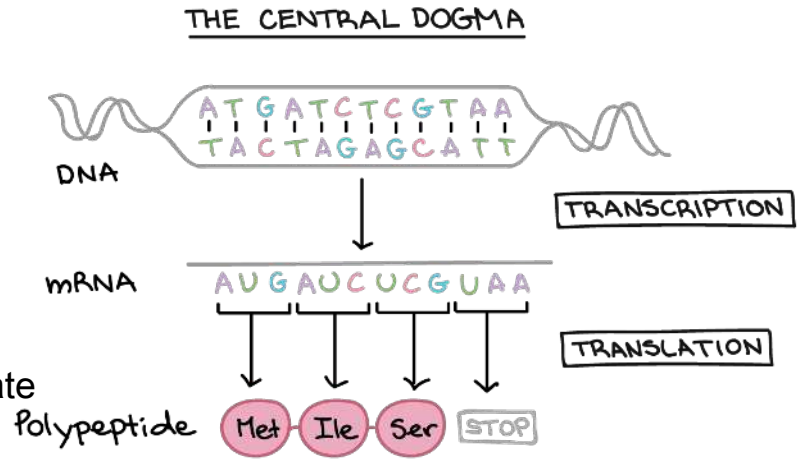
Preview of what's coming… how do we know which rules are broken in different cancers?

- We'll use RNAseq data (gene expression data)
- Mutations in oncogenes and tumor suppressors → abnormal gene expression patterns
- Hallmarks organize data into "gene expression modules" = related genes that tend to be correlated or anti-correlated
  - Often upregulated (turned UP) or turned OFF altogether

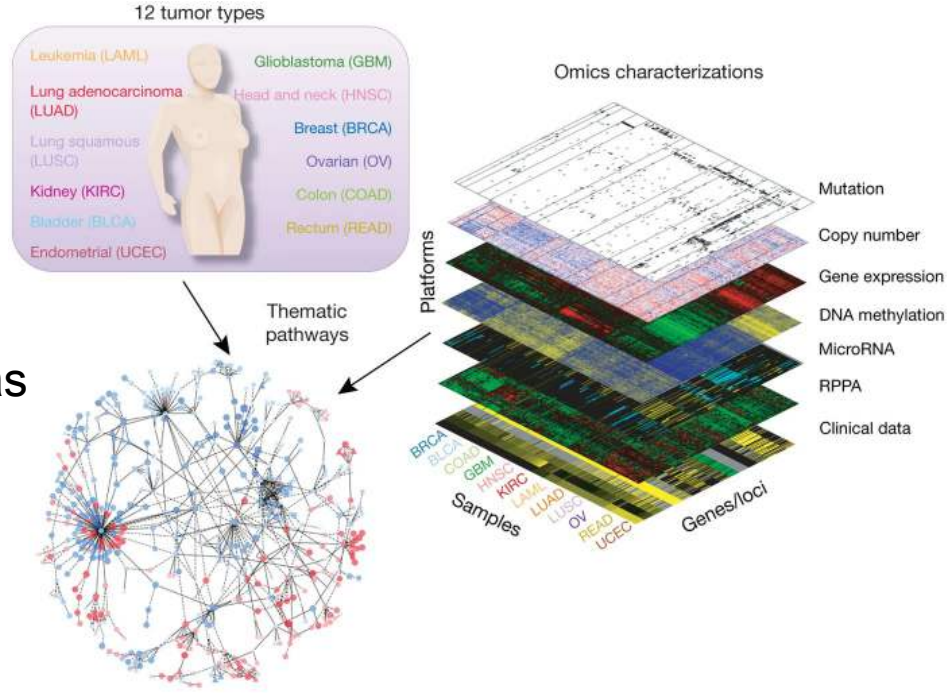
Mutations on this level (genome)

Affect "cell identity" on RNA level (transcriptome) "Live feed"

Proteins are made and accumulate to act as little machines "Highlights reel"



The Cancer Genome Atlas (TCGA)



Deciding on a project goal: choices to make!

- What **hallmark** of cancer will you focus on?
- Do you want to focus on a **single cancer type** or look for patterns **across cancer types**?
- Start to think about what kind of question you want to ask. Example question “scaffolds”:
  - Is hallmark X predictive of survival in Y cancer type?
  - Is hallmark X related to hallmark Y in Z cancer type?
  - Can different cancer types be distinguished by X hallmark alone?
  - Are there distinct groups of cancer types with hallmark X expression that respond to treatment in different ways?
  - Does patient history of X predict expression of hallmark Y in Z cancer?
  - Does expression of X hallmark predict Y tumor stage or Z treatment response?

Due Friday by midnight:
First GitHub Check-in with disease background, data background, and first pass at a question/project goal