

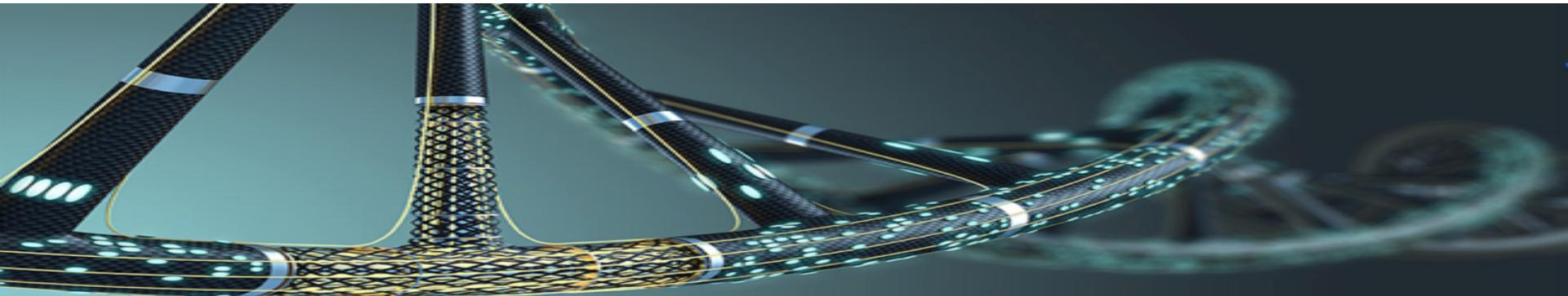


Майнор по биоинформатике

Семестр 2

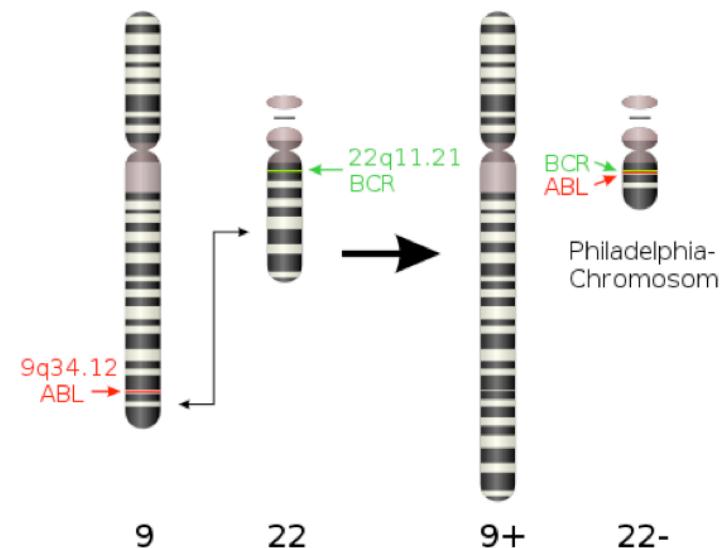
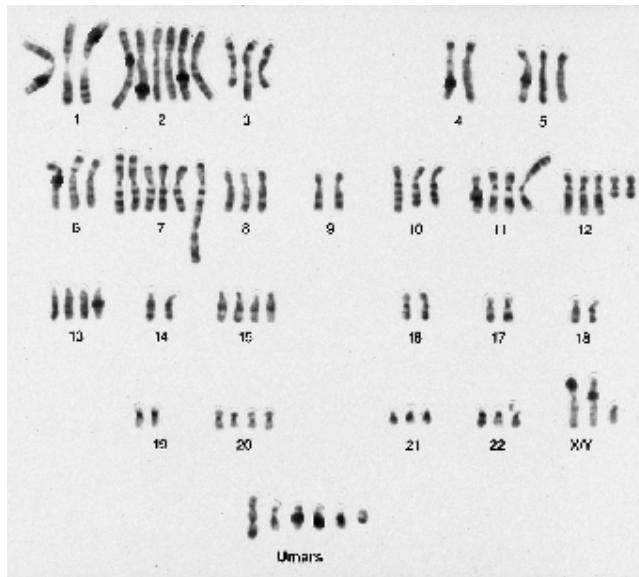
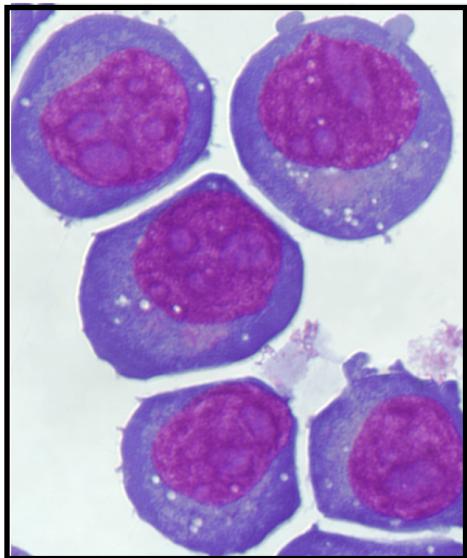
Лекция 6

Мария Попцова



Геномика рака

- Рак – болезнь генома



Lessons learned

➤ Heterogeneity within and across tumor types

➤ Перефразируя начало Анны Карениной:

Все нормальные геномы
похожи друг на друга,
каждый раковый геном
ненормален по-своему.

➤ Challenge in Treating Cancer

Every tumor is different

Every cancer patient is different

Международные консорциумные проекты по исследованию генетики раковых заболеваний

1-800-4-CANCER

Live Chat

Publications

Dictionary

ABOUT CANCER CANCER TYPES RESEARCH GRANTS & TRAINING NEWS & EVENTS

ABOUT NCI

search

Q

Home > About NCI > NCI Organization > CCG > Research > Structural Genomics



TCGA

[Program History](#) +[TCGA Cancers Selected for Study](#)[Publications by TCGA](#)[Using TCGA](#) +[Contact](#)

The Cancer Genome Atlas Program

The Cancer Genome Atlas (TCGA), a landmark [cancer genomics](#) program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between the National Cancer Institute and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions.

Over the next dozen years, TCGA generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data. The data, which has already lead to improvements in our ability to diagnose, treat, and prevent cancer, will remain [publicly available](#) for anyone in the research community to use.

NATIONAL CANCER INSTITUTE

THE CANCER GENOME ATLAS

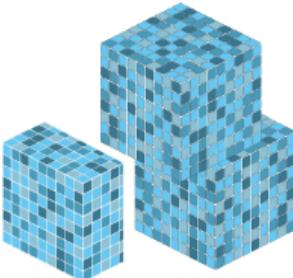
TCGA BY THE NUMBERS

TCGA produced over

2.5

PETABYTES

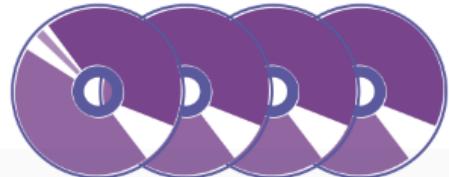
of data



To put this into perspective, **1 petabyte** of data
is equal to

212,000

DVDs



TCGA data describes



33

DIFFERENT
TUMOR TYPES

...including

10

RARE
CANCERS

...based on paired tumor and normal tissue sets
collected from



11,000

PATIENTS

...using

7

DIFFERENT
DATA TYPES





International
Cancer Genome
Consortium

Enter keywords

Search

Home About Us Cancer Genome Projects Committees and Working Groups Policies and Guidelines Media Publications

No cancer therapy is developed today without the genomic knowledge that ICGC provided to the world.

The ICGC, established in 2007, aimed to define the genomes of 25,000 primary untreated cancers (**the 25K Initiative**). The ICGC solved numerous data governance, ethical and logistical challenges to make global genomic data sharing for cancer possible, providing the international community with comprehensive genomic data for many cancer types.



The second ICGC Initiative, the Pan Cancer Analysis of Whole Genomes (**PCAWG**), known as the Pan-Cancer Project, defined similarities and differences between cancer types. Based on this information, the next project of the ICGC is addressing key clinical questions to Accelerate Research in Genomic Oncology (The ARGO Project). You can read more about the **ICGC-ARGO project** at www.icgc-argo.org.

To learn more about the **Pan-Cancer Project** and the **ICGC-ARGO Project** [watch this video](#).

[More about ICGC »](#)

[Cancer Projects](#)[Advanced Search](#)[Data Analysis](#)[DCC Data Releases](#)[Data Repositories](#)

Cancer genomics data sets visualization, analysis and download.

[Quick Search](#)[Search](#)

e.g. BRAF, KRAS G12D, DO35100, MU7870, FI998, apoptosis, Cancer Gene Census, imatinib, GO:0016049

[Advanced Search](#)[By donors](#)[By genes](#)[By mutations](#)

Data Release 28

March 27th, 2019

Cancer projects	86
Cancer primary sites	22
Donor with molecular data in DCC	22,330
Total Donors	24,289
Simple somatic mutations	81,782,588

 [Download Release](#)



Quick Search

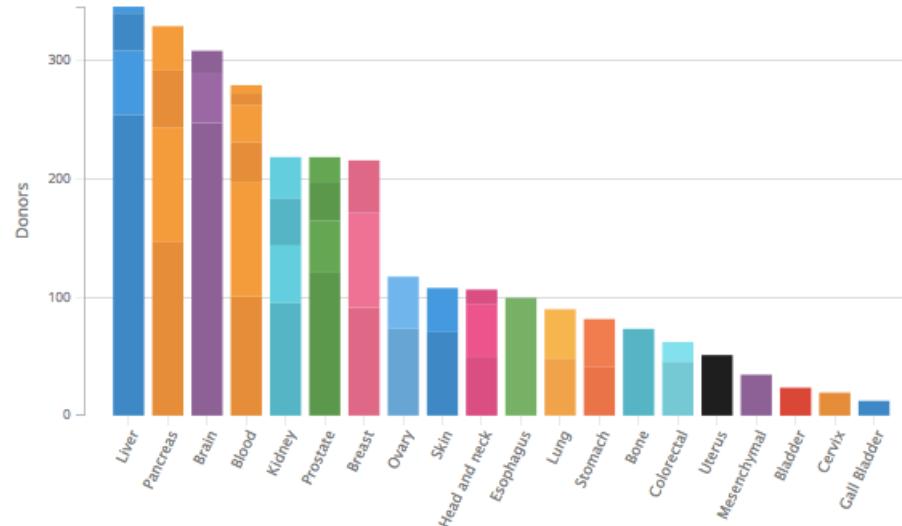
PCAWG - PANCAANCER ANALYSIS OF WHOLE GENOMES

The Pan-Cancer Analysis of Whole Genomes (PCAWG) study is an international collaboration to identify common patterns of mutation in more than 2,600 cancer whole genomes from the International Cancer Genome Consortium. Building upon previous work which examined cancer coding regions (Cancer Genome Atlas Research Network, The Cancer Genome Atlas Pan-Cancer analysis project, [Nat. Genet. 2013 45:1113](#), [Cell. 2018 Apr 5;173\(2\):283-285](#)), this project explored the nature and consequences of somatic and germline variations in both coding and non-coding regions, with specific emphasis on cis-regulatory sites, non-coding RNAs, and large-scale structural alterations.

In order to facilitate the comparison among diverse tumor types, all tumor and matched normal genomes have been subjected to a uniform set of alignment and variant calling algorithms, and must pass a rigorous set of quality control tests. The research activities have been coordinated by a series of working groups comprising more than 700 scientists.

VCF-format files representing somatic variant calls for single-nucleotide variants, small indels, structural variants and copy number variants can be downloaded using the links in the table at the right. In addition, we provide aligned BAM files for download or cloud-based access via the [Cancer Genome Collaboratory](#) and [Amazon Web Services](#). Note that data sets that may contain germline SNPs are controlled tier and require credentials provided by the ICGC Data Access Committee and/or the TCGA dbGaP Data Access

Donor Distribution by Primary Site
48 projects and 20 primary sites



 **2,793** Donors

 **71,709** Files

 **795.14 TB**

PCAWG

Data Type	# Donors	# Files	Format	Size
SGV	2,715	8,505	VCF	517.27 GB
StGV	2,715	5,668	VCF	7.29 GB
Aligned Reads	2,793	12,169	BAM	794.42 TB
Simple Somatic Mutations	2,715	25,501	VCF	189.99 GB
Copy Number Somatic Mutations	2,715	5,671	VCF	132.62 MB
Structural Somatic Mutations	2,715	14,195	VCF	1.61 GB

Available data as of Jan 24, 2020

Recent News

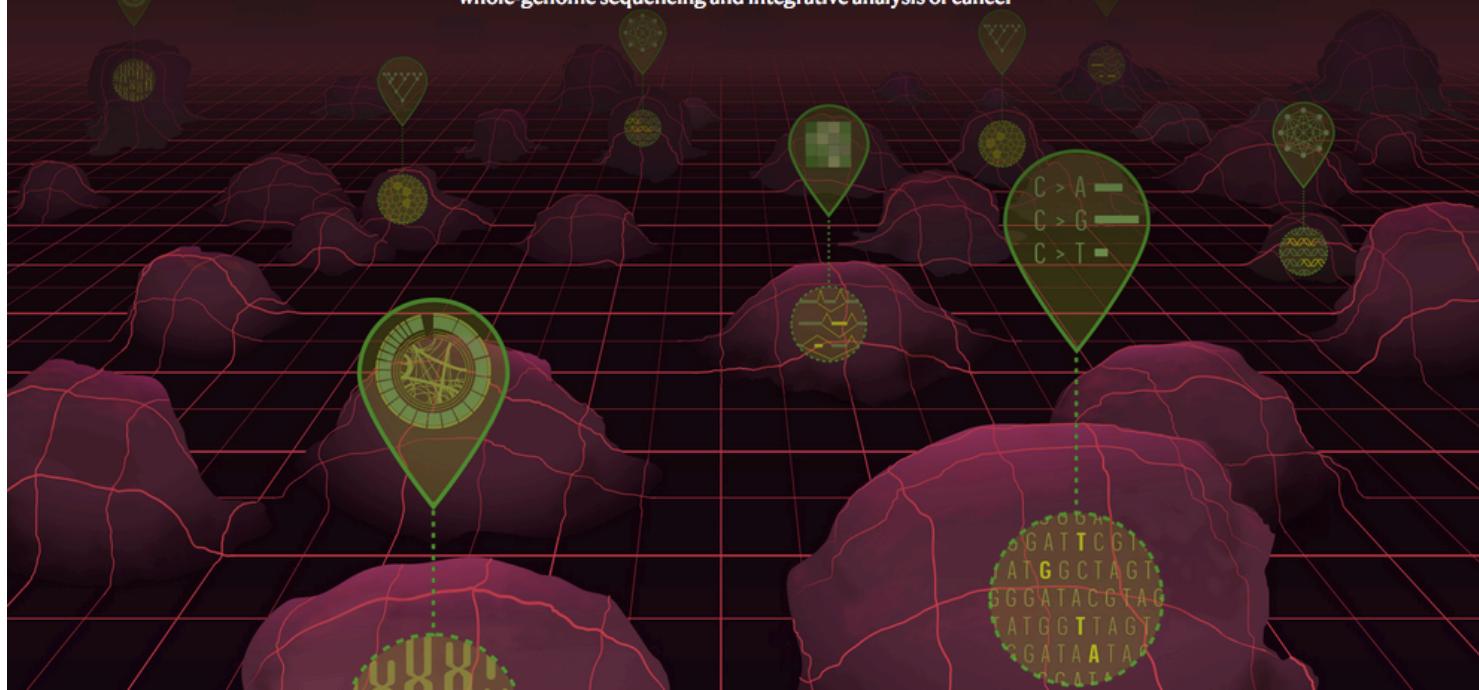
NEWS AND VIEWS · 05 FEBRUARY 2020

Global genomics project unravels cancer's complexity at unprecedented scale

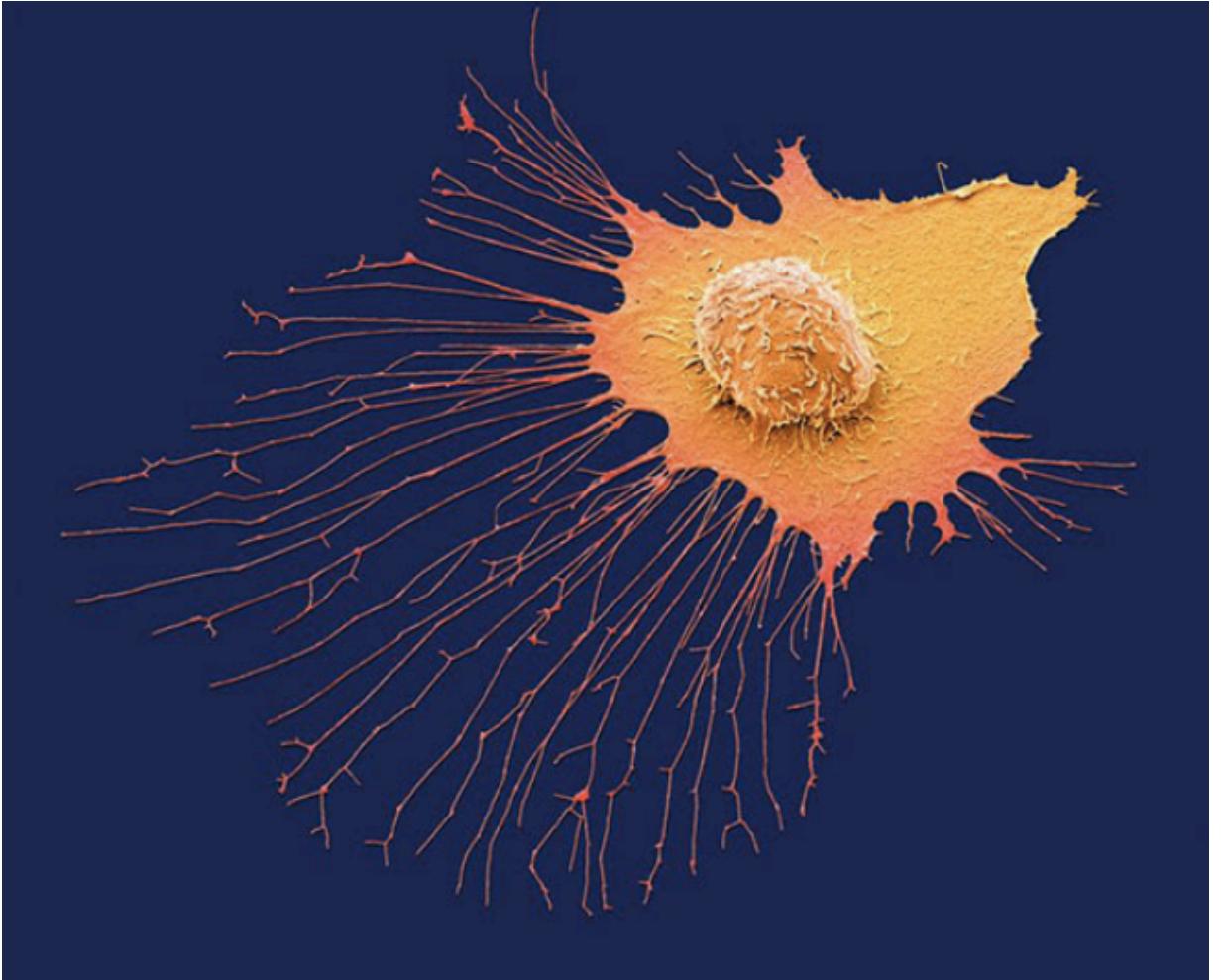
A massive international effort has yielded multifaceted studies of more than 2,600 tumours from 38 tissues, generating a wealth of insights into the genetic basis of cancer.

Pan-Cancer Analysis of Whole Genomes

A collection of research and related content from the ICGC/TCGA consortium on whole-genome sequencing and integrative analysis of cancer



<https://www.nature.com/immersive/d42859-020-00001-y/index.html>

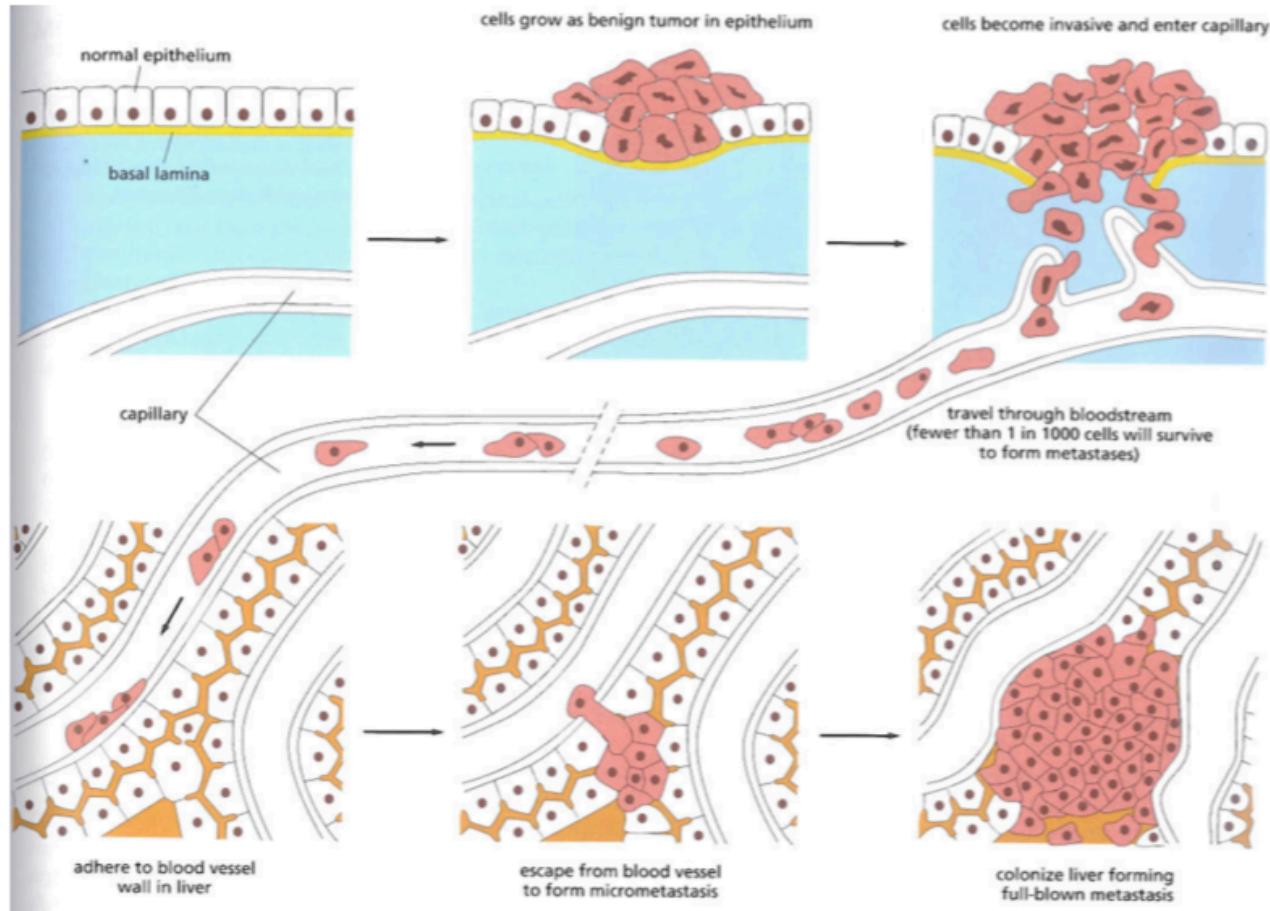


A migrating breast cancer cell

Types of cancers

- Carcinomas: cancers arising from epithelial cells
- Sarcomas: cancers arising from connective tissue or muscle cells
- Leukemias and lymphomas: cancers derived from white blood cells and their precursors

Metastasis



Нахождение всех мутаций, ассоциированных с раком

- Секвенирование генома опухолевой клетки и неопухолевой клетки у одного пациента
- Выделить мутации, специфичные для опухолевой клетки
- Исключить молчаше мутации (изменяют нуклеотидные последовательности генов, но не аминокислотные)

140 cancer genes

Oncogenes

Normally stimulate growth: Dominant

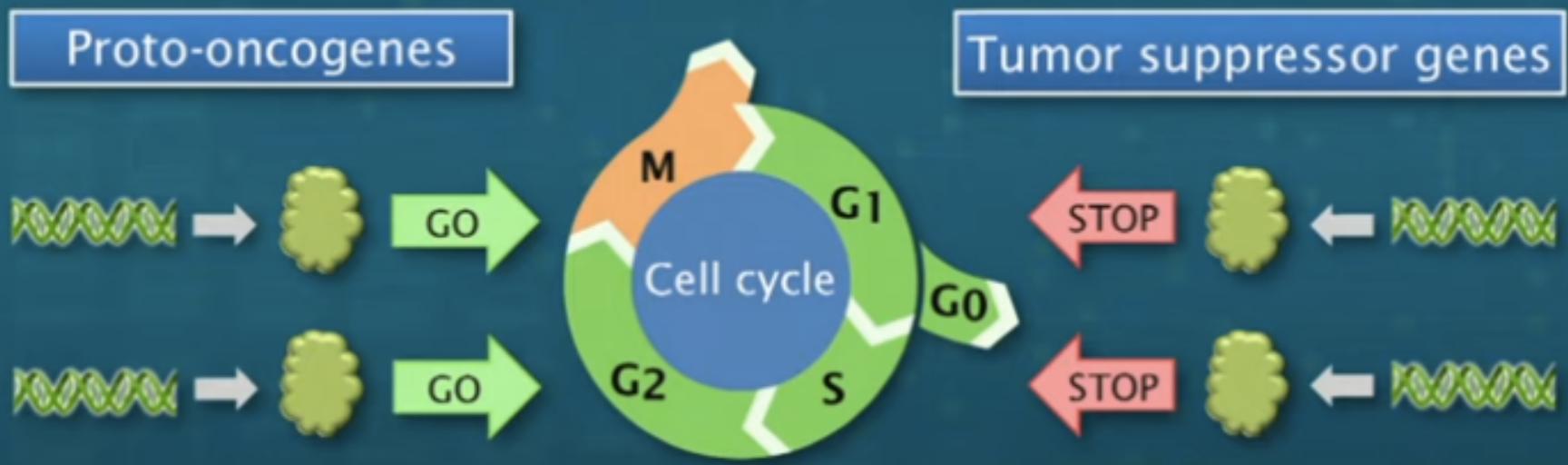


Suppressor genes

Normally inhibit growth: Recessive

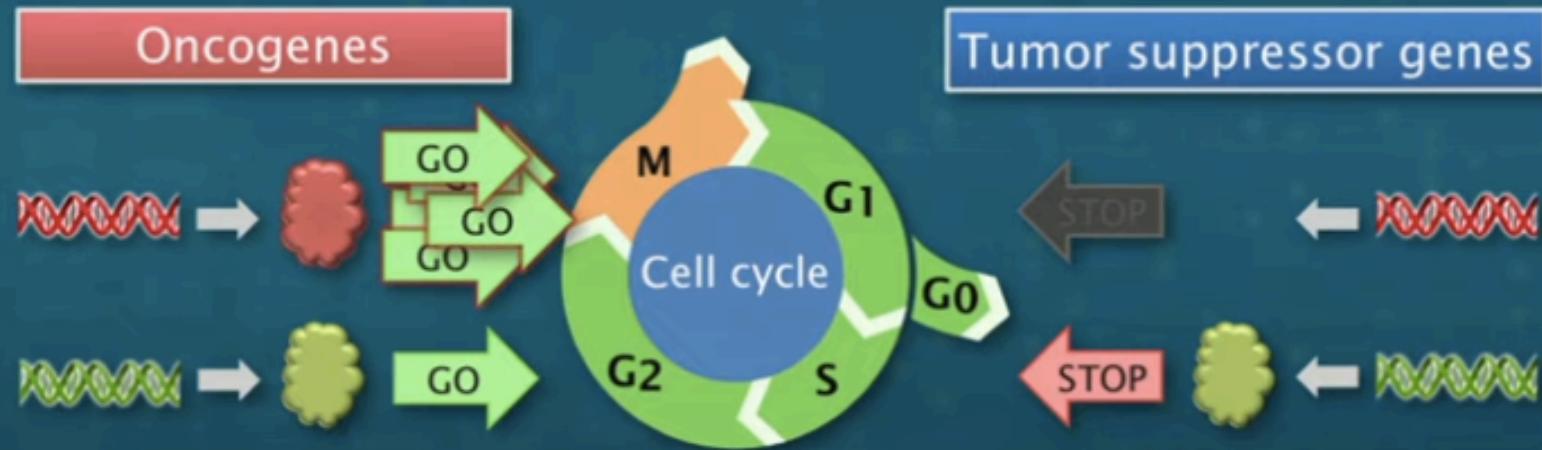


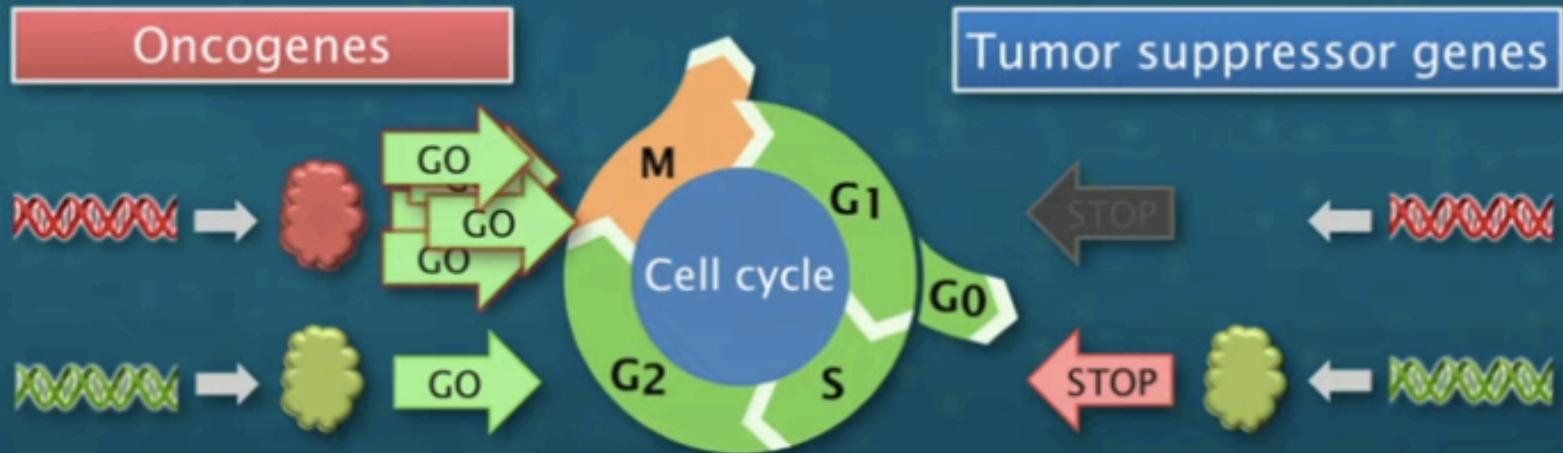
Dominant vs recessive



G1 – growth, S-DNA synthesis, G2 growth and preparation for mitosis, M – mitosis (cell division)

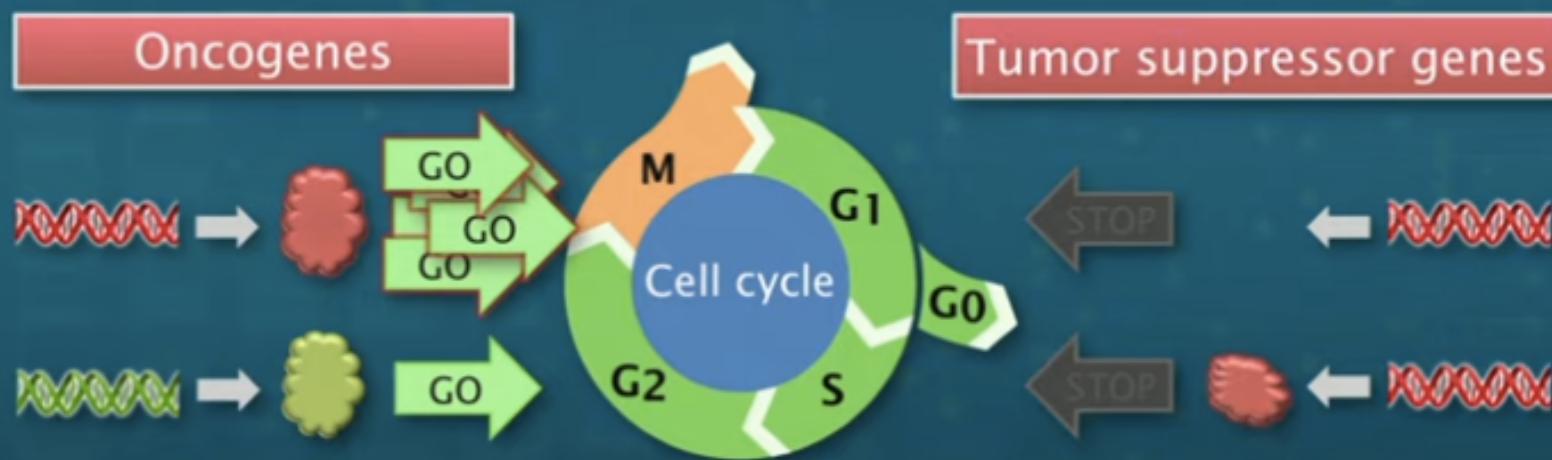
Mutation of oncogene = GO signal

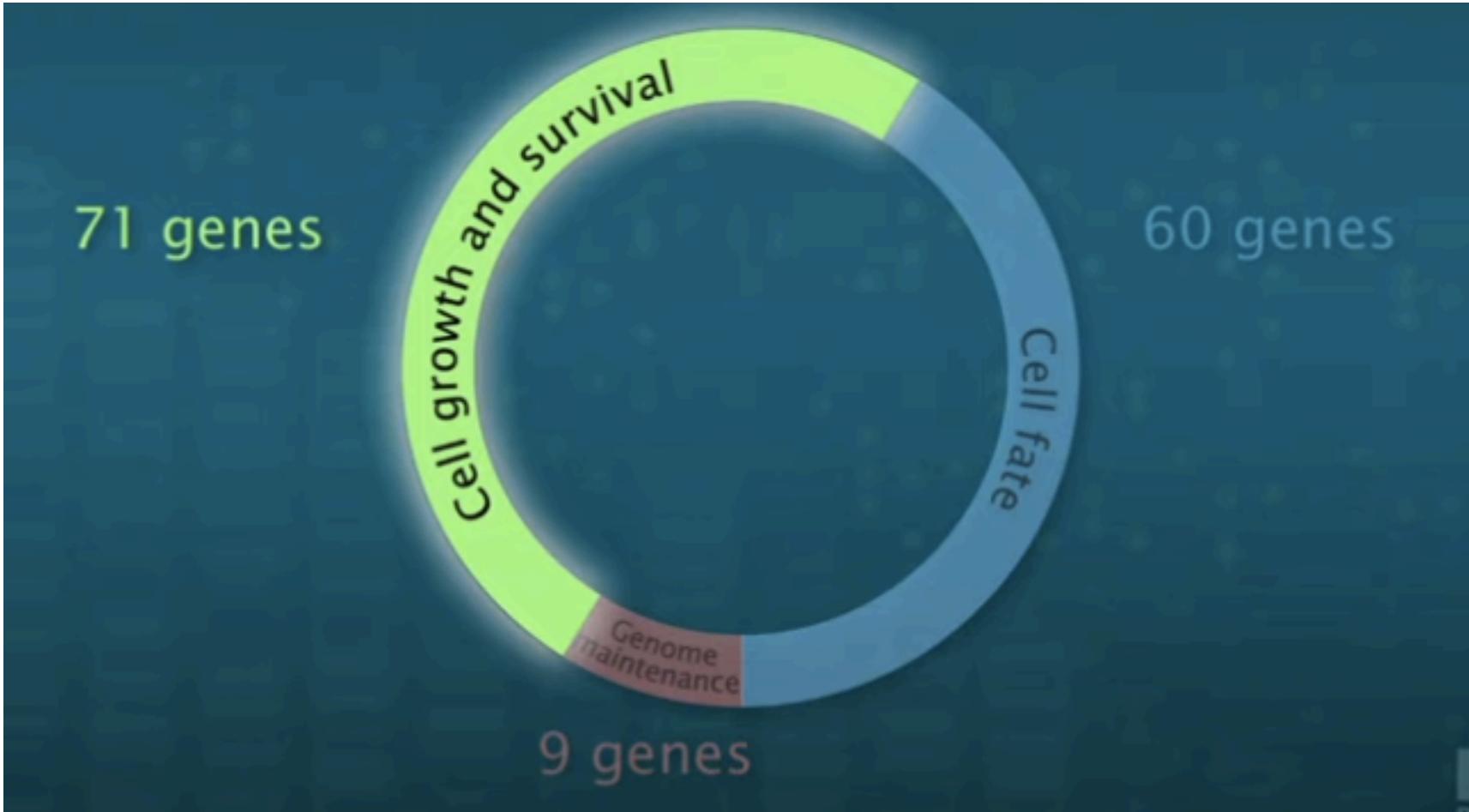




One copy of tumor suppressor gene is mutated – cell cycle is OK

Two copies of tumor suppressor genes are mutated – cell cycle corrupted





Cell Growth and Survival Genes

Growth factors



Cell surface receptor molecules

EGFR



Many participate in the signaling pathway that promotes cell growth.

membrane

Signaling cascade molecules

RAS

BRAF

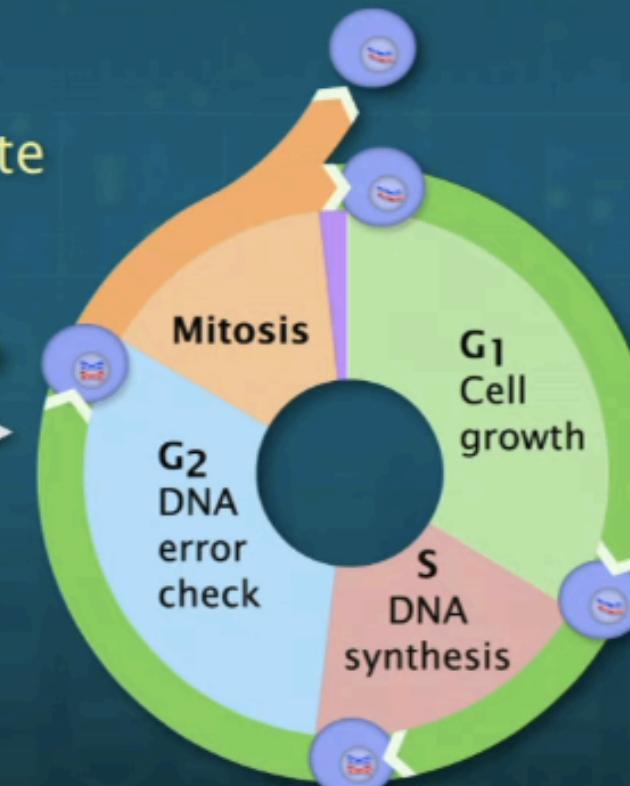
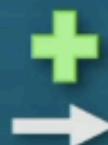
MEK

Cell proliferation

Regulators of Cell Cycle and Cell Death

Some stimulate
the cell cycle

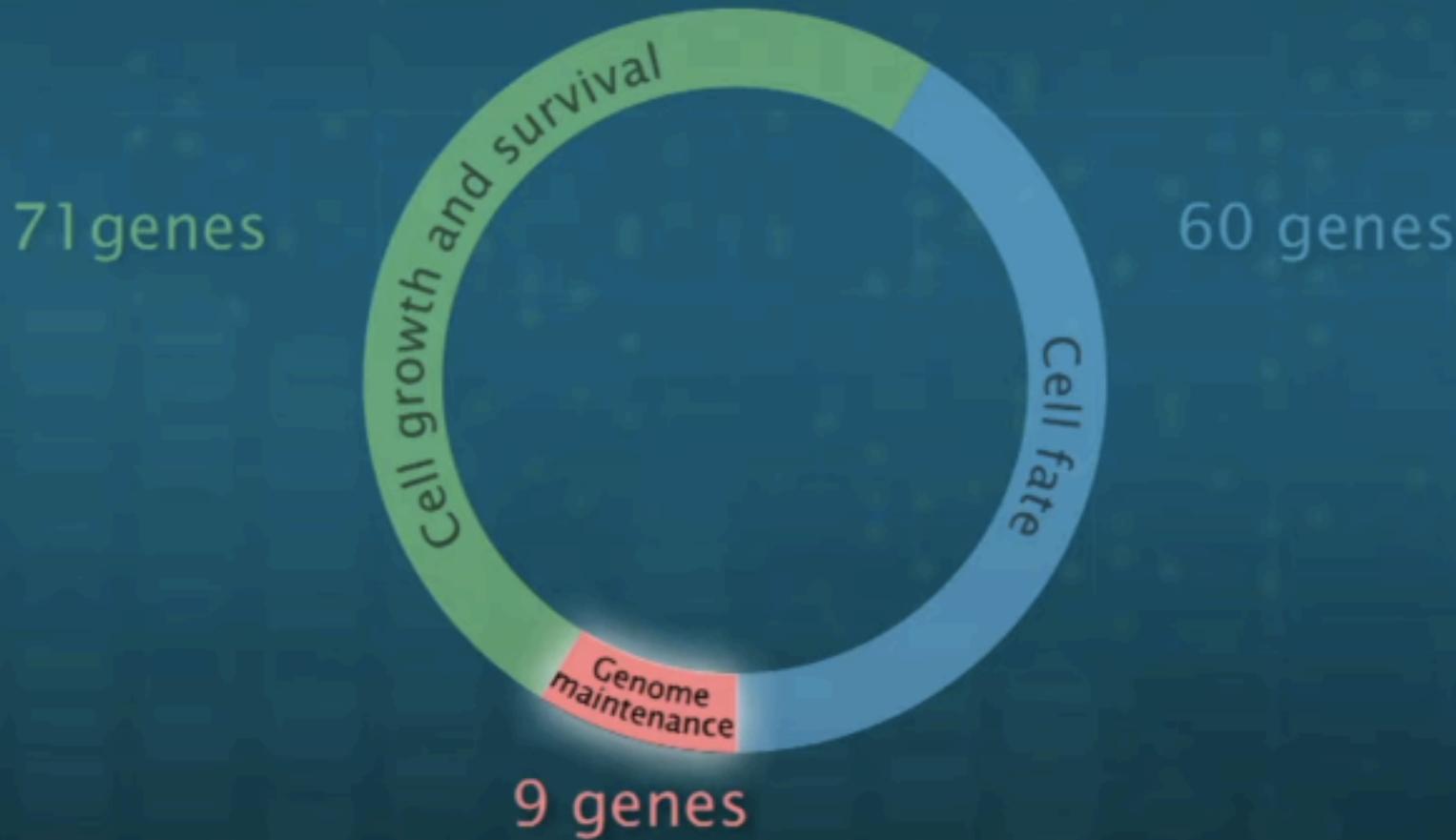
- Cyclin D1
- CDK4



Some inhibit
the cell cycle

- P53
- RB

Functions of the 140 Cancer Genes



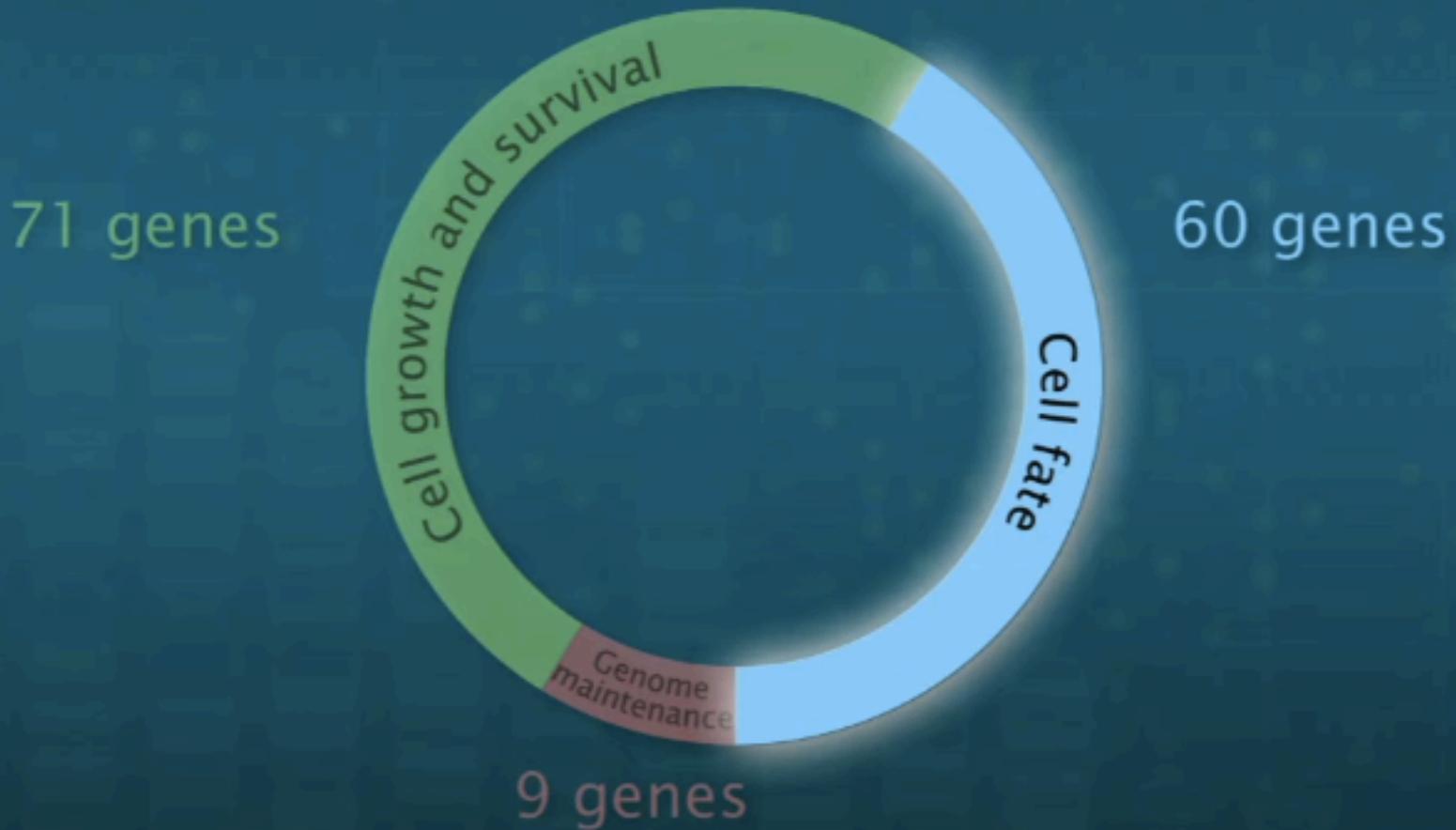
Examples of Genome Maintenance Genes

DNA proofreading genes

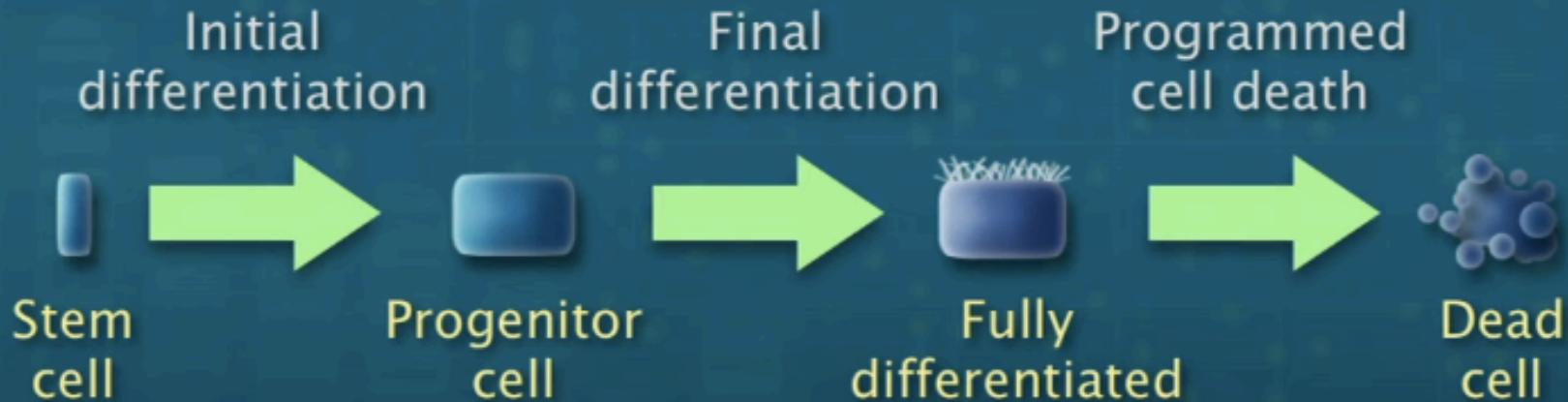


DNA repair genes

Functions of the 140 Cancer Genes

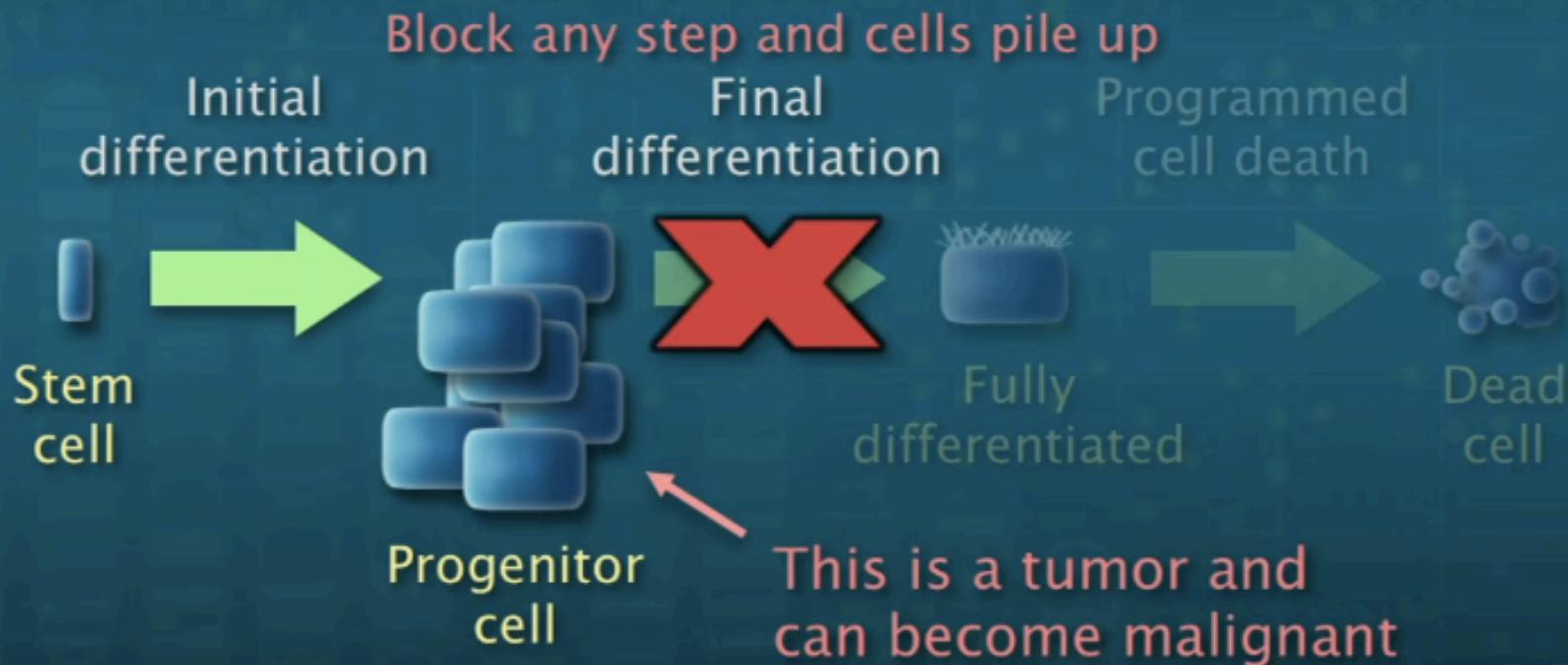


How Can Mutations in Cell Fate Genes Cause Cancer?



Triggering cell death requires differentiation

How Can Mutations in Cell Fate Genes Cause Cancer?

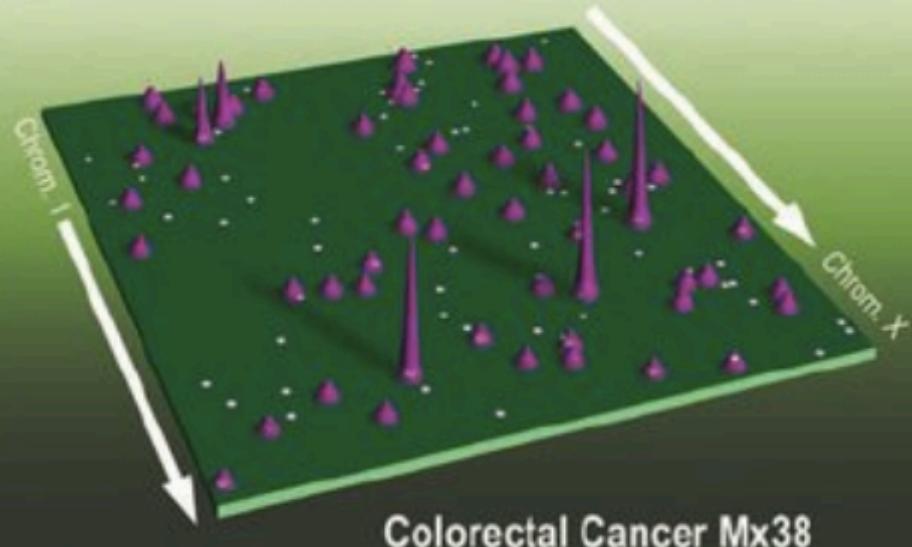


Driver and Passenger Mutations

- Driver mutations
 - Causally implicated in oncogenesis
 - Gives growth advantage to cancer cells
 - positively selected in the microenvironment of the tissue –
 - E.g., mutations that de-activate tumor suppressor genes
- Passenger mutations
 - Somatic mutations with no functional consequences
 - Does not give growth advantage to cancer cells

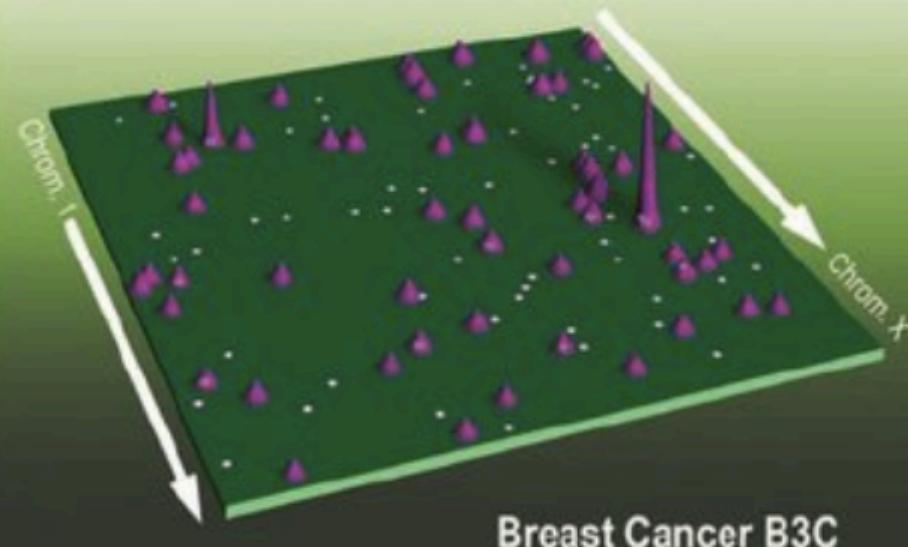
Cancer genome landscapes

A



Colorectal Cancer Mx38

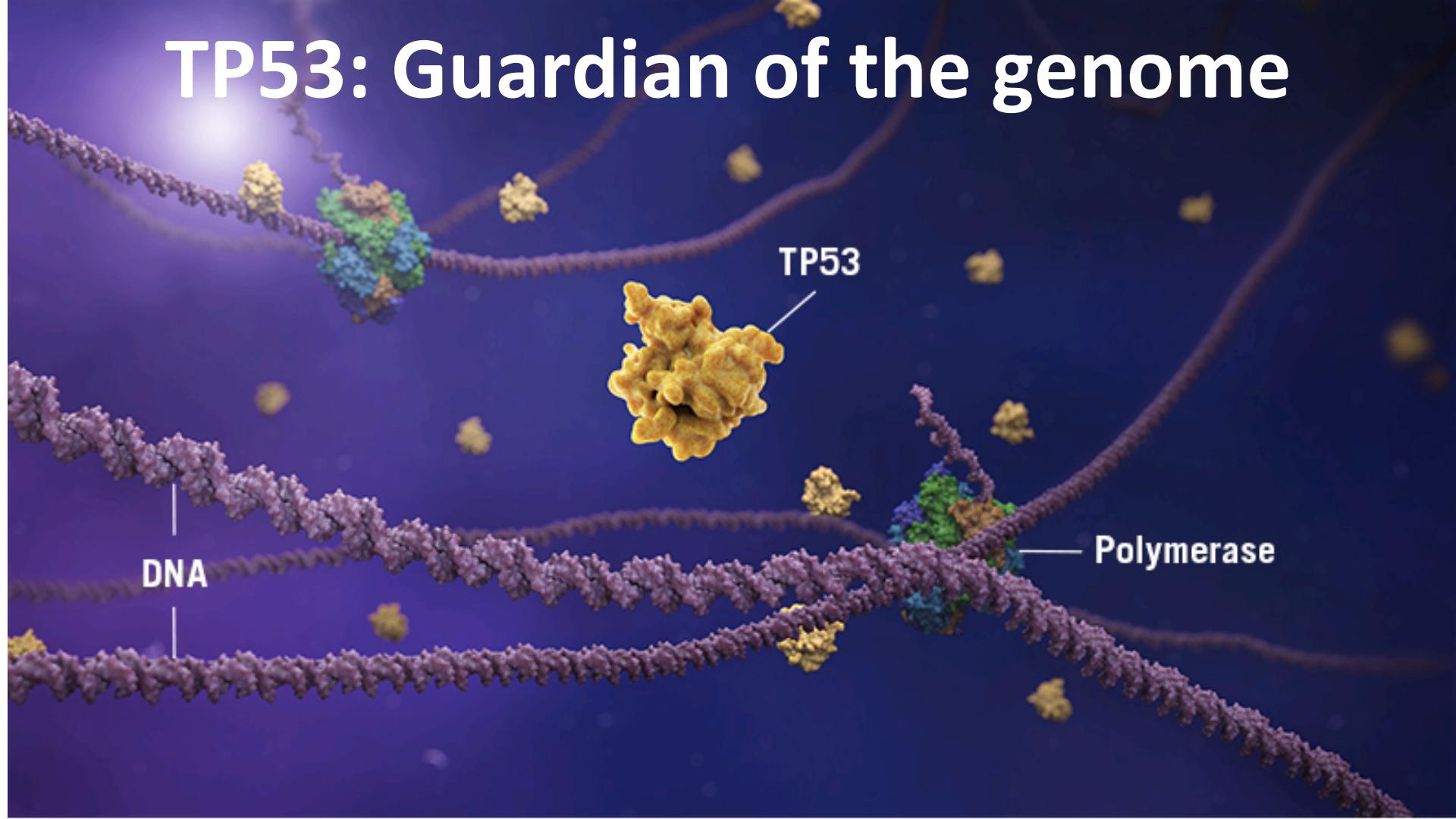
B



Breast Cancer B3C

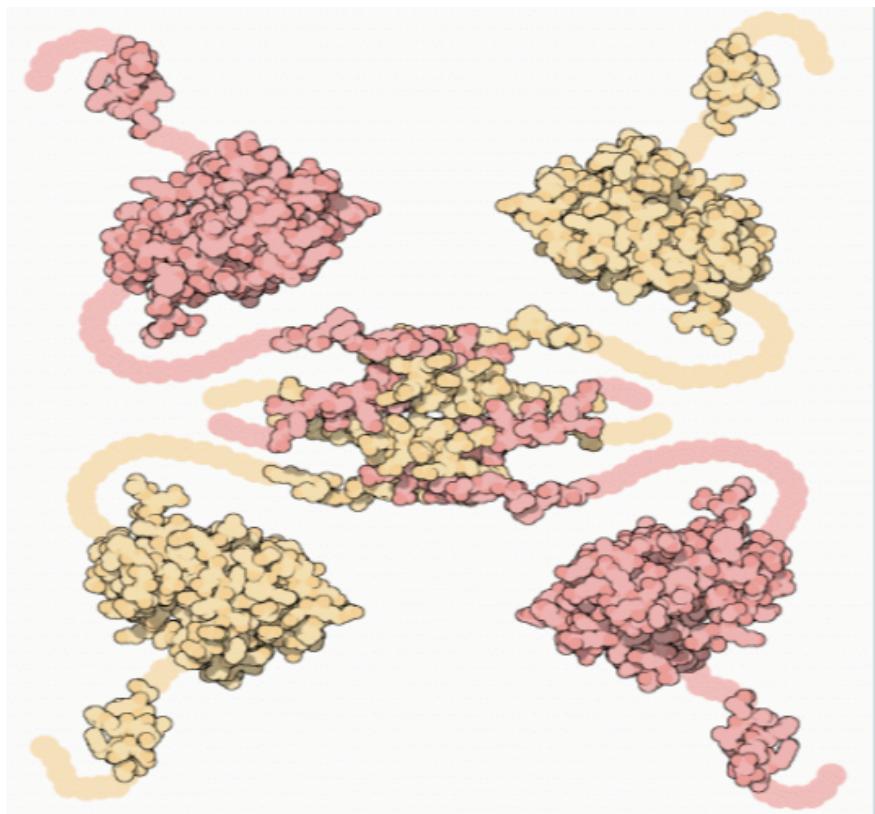
The dots represent genes that were somatically mutated

TP53: Guardian of the genome

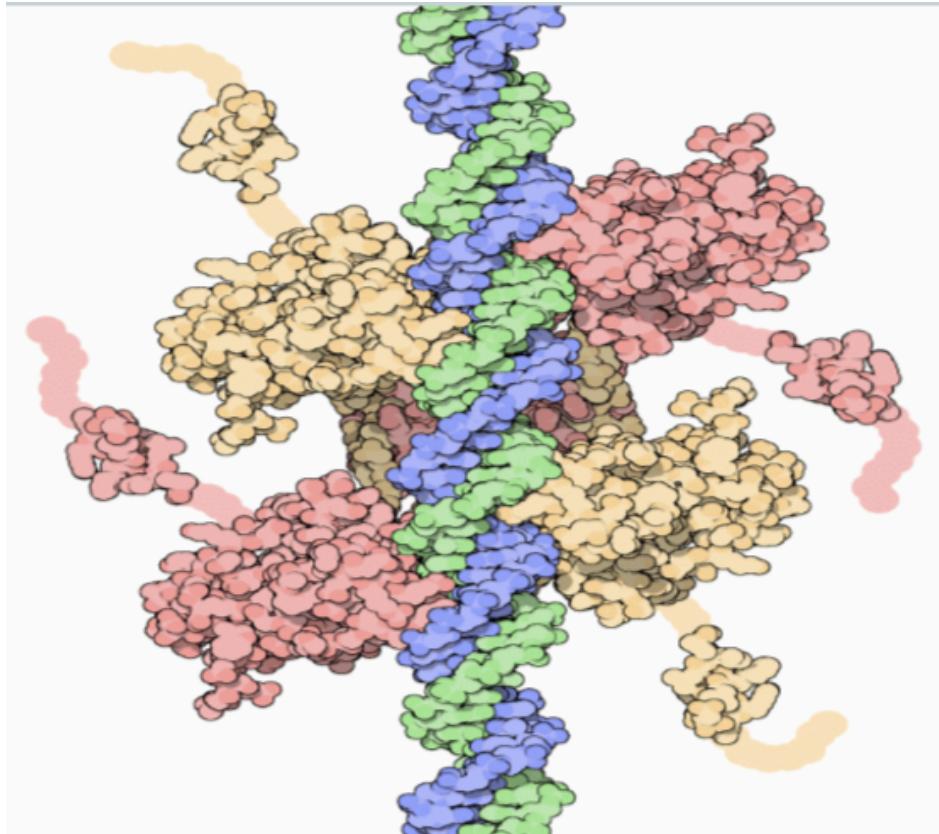


TP 53

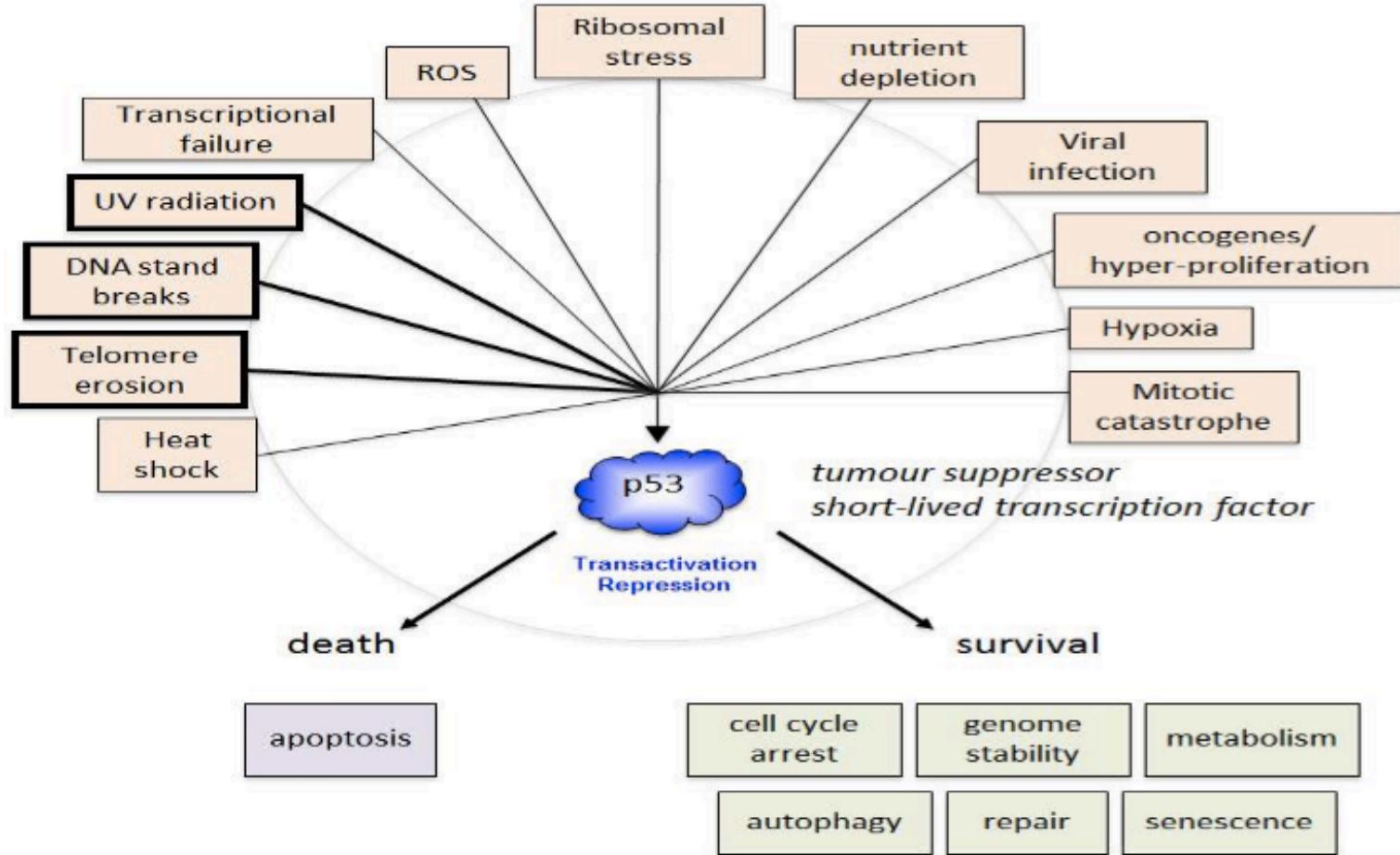
in tetramer form



bound to DNA



Overview of Central Role of p53



Gene fusions in Cancer

The Cancer Genome Project website lists at least 326 genes that have been shown to form gene translocations in cancer

<http://www.sanger.ac.uk/genetics/CGP/Census/>

Census

GRCh38 · COSMIC v90

- Overview
- Cancer Gene Census
- Breakdown
- Abbreviations

[Reset page](#)

Overview

The Cancer Gene Census (CGC) is an ongoing effort to catalogue those genes which contain mutations that have been causally implicated in cancer and explain how dysfunction of these genes drives cancer. The content, the structure, and the curation process of the Cancer Gene Census was described and published in [Nature Reviews Cancer](#).

The census is not static, instead it is updated when new evidence comes to light. In particular we are grateful to Felix Mitelman and his colleagues in providing information on more genes involved in uncommon translocations in leukaemias and lymphomas. Currently, more than 1% of all human genes are implicated via mutation in cancer. Of these, approximately 90% contain somatic mutations in cancer, 20% bear germline mutations that predispose an individual to cancer and 10% show both somatic and germline mutations.

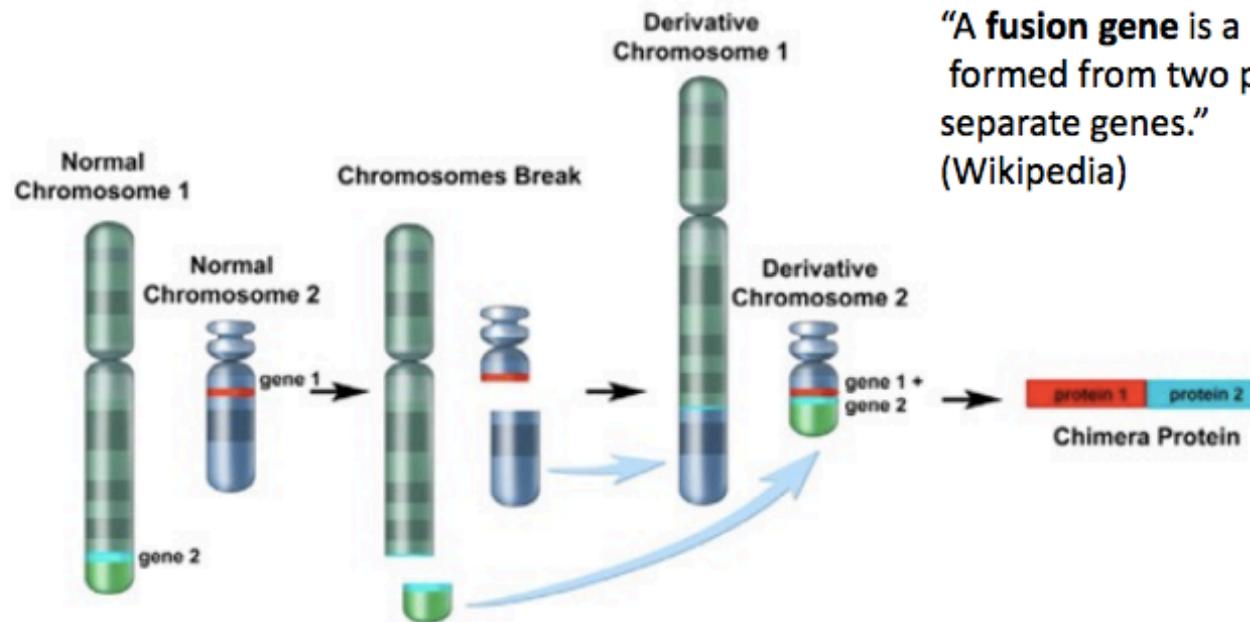
Census tiers

Genes in the Cancer Gene Census are divided into two groups, or tiers.

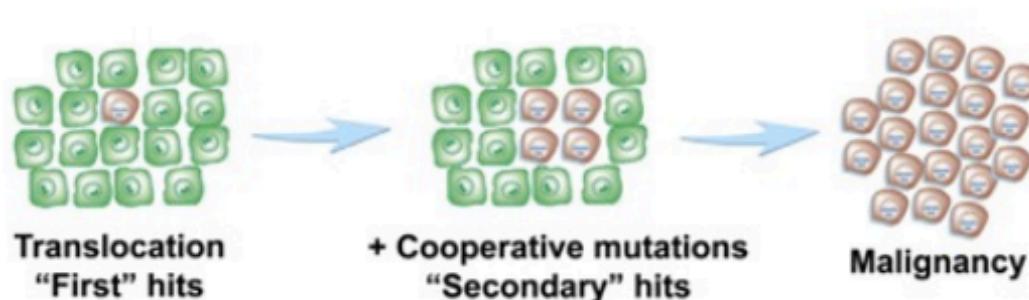
Tier 1

To be classified into Tier 1, a gene must possess a documented activity relevant to cancer, along with evidence of mutations in cancer which change the activity of the gene product in a way that promotes oncogenic transformation. We also consider the existence of somatic mutation patterns across cancer samples gathered in COSMIC. For instance, tumour suppressor genes often show a broad range of inactivating mutations and dominant oncogenes usually demonstrate well defined hotspots of missense mutations. Genes involved in oncogenic fusions are included in Tier 1 when changes to their function caused by the fusion drives oncogenic transformation, or in cases when they provide regulatory elements to their partners (e.g. active promoter or dimerisation domain).

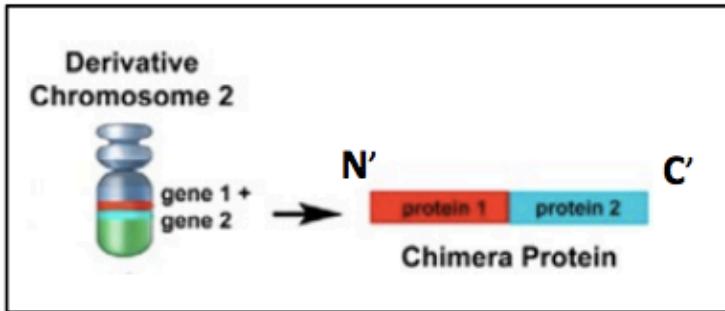
Fusion genes



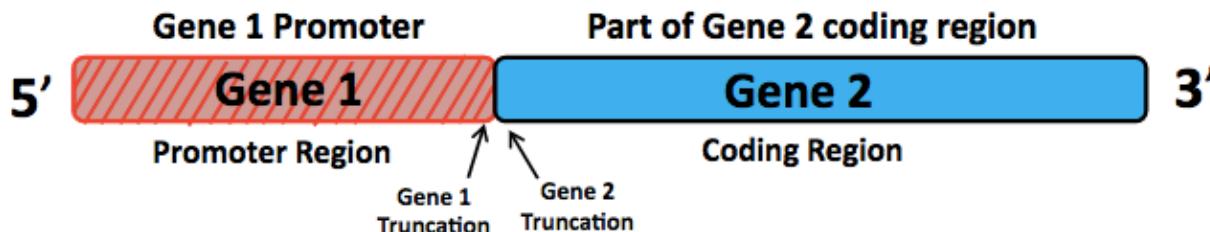
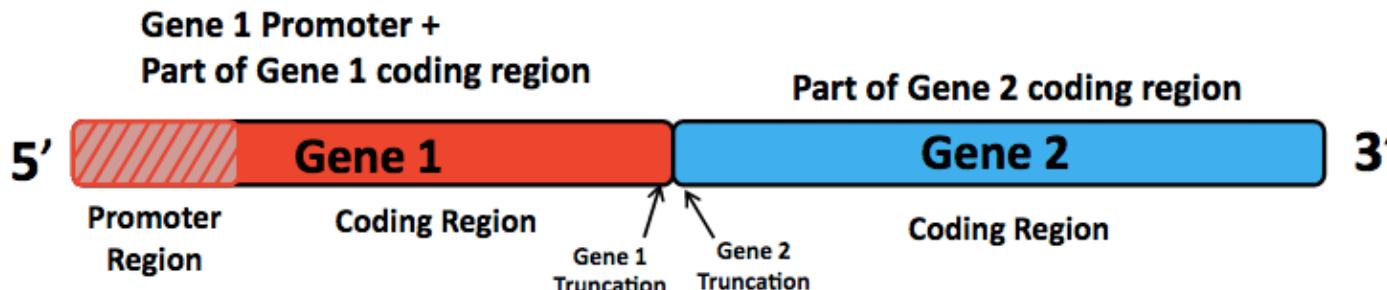
"A **fusion gene** is a hybrid gene formed from two previously separate genes."
(Wikipedia)



Fusion genes result in aberrant gene expression

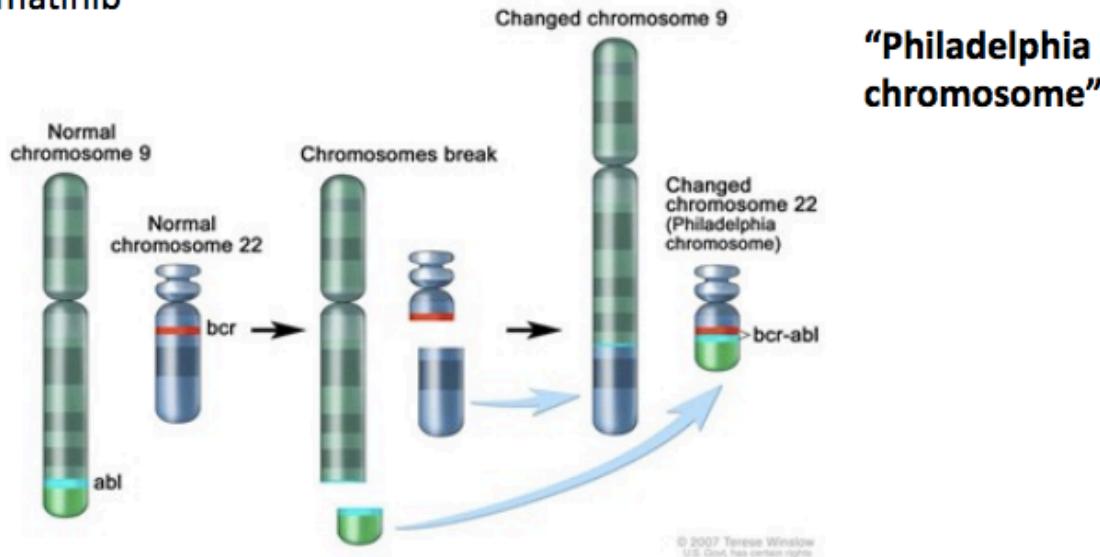


- Gene 2 expression and transcriptional regulation is now dictated by the Gene 1 promoter and all its regulatory units
- If Gene 1 has a highly active promoter region, Gene 2 will be overexpressed



Fusion genes in cancer: Bcr-Abl (CML and ALL) t(9;22)

- Poster child for fusion genes in cancer due to the development of the drug Imatinib



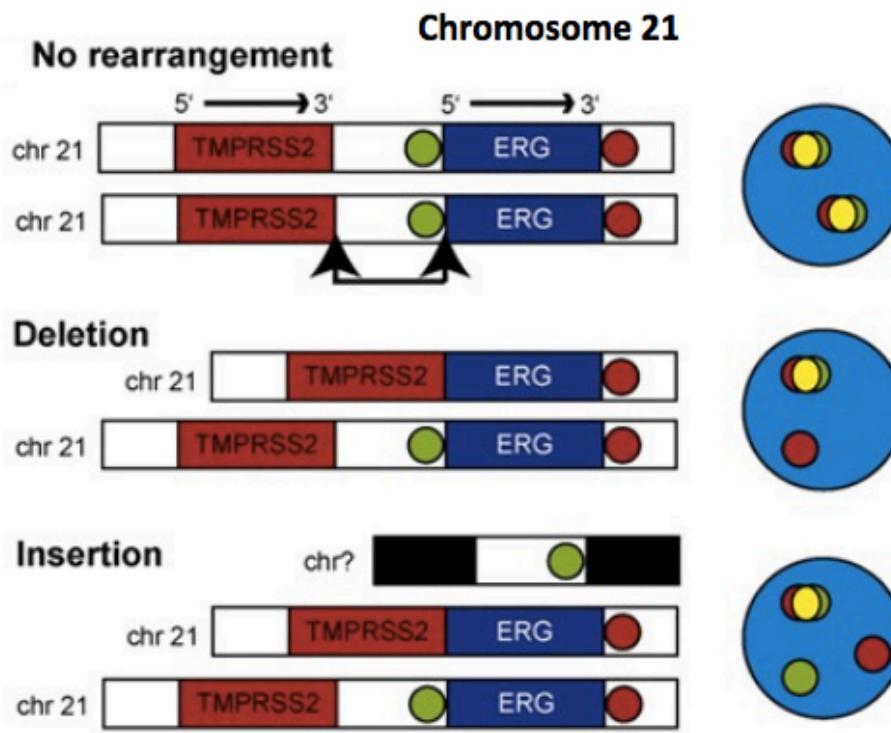
Prevalence of Bcr-Abl in leukemia patients

CML (Adults) 90% (Children) CML rare in children

ALL (Adults) 25-30% (Children) 2-10%

Fusion genes in cancer: TMPPRSS2-ERG (prostate cancer)

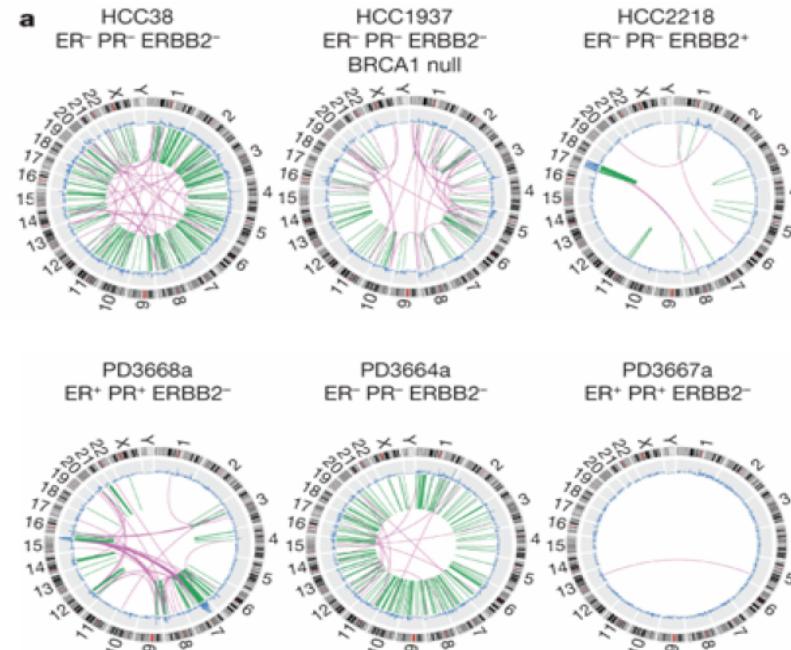
- The TMPPRSS2-ERG fusion gene is present in approximately 50% of prostate cancer patients



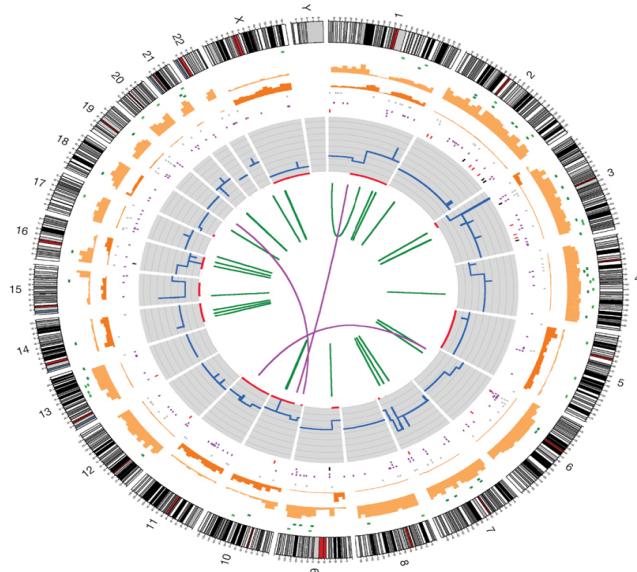
Complex landscapes of somatic rearrangement in human breast cancer genomes

2,166 confirmed somatic rearrangements were identified among the 24 cancers

Somatic rearrangements observed in six of the twenty-four breast cancer samples screened.



A comprehensive catalogue of somatic mutations from a human cancer genome

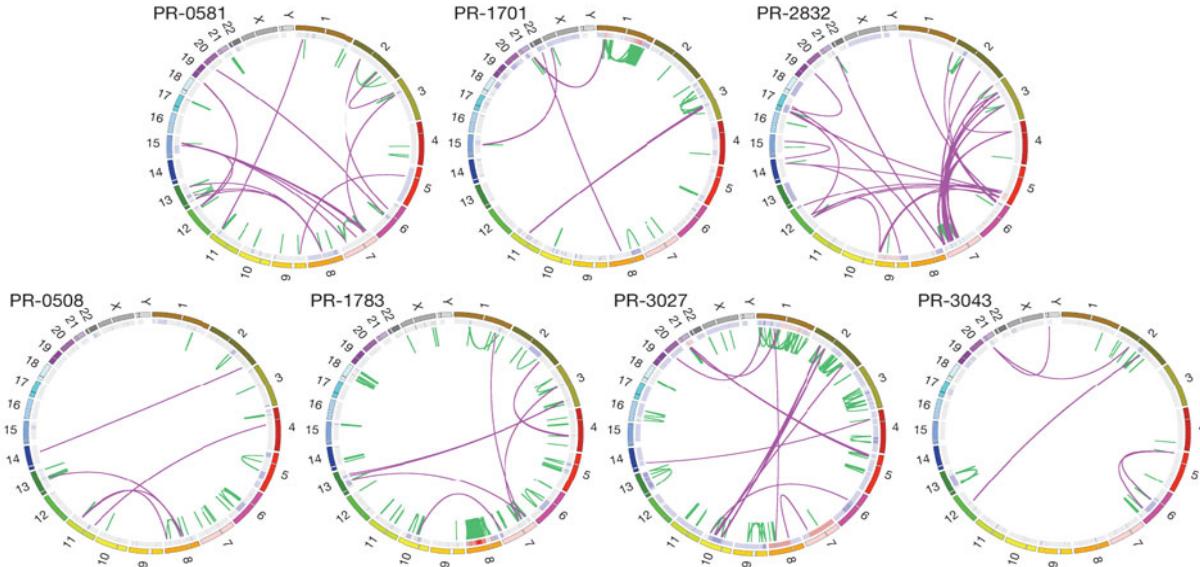


- sequenced the genomes of a malignant melanoma and a lymphoblastoid cell line from the same person
- 33,345 somatic base substitutions.
- 680 small deletions and 303 small insertions.
- 51 somatic rearrangements

The genomic complexity of primary human prostate cancer

M F. Berger *et al.* *Nature* **470**, 214-220 (2011)

11 institutions:
Harvard, MIT,
Yale, Cornell,
PCA, etc.



Graphical representation of seven prostate cancer genomes (Complete sequencing of primary human prostate cancers and their paired normal counterparts.)

nature

Cancer and Immortality

The Nobel Prize in Physiology or Medicine 2009



© The Nobel Foundation. Photo: U. Montan
Elizabeth H. Blackburn
Prize share: 1/3



© The Nobel Foundation. Photo: U. Montan
Carol W. Greider
Prize share: 1/3

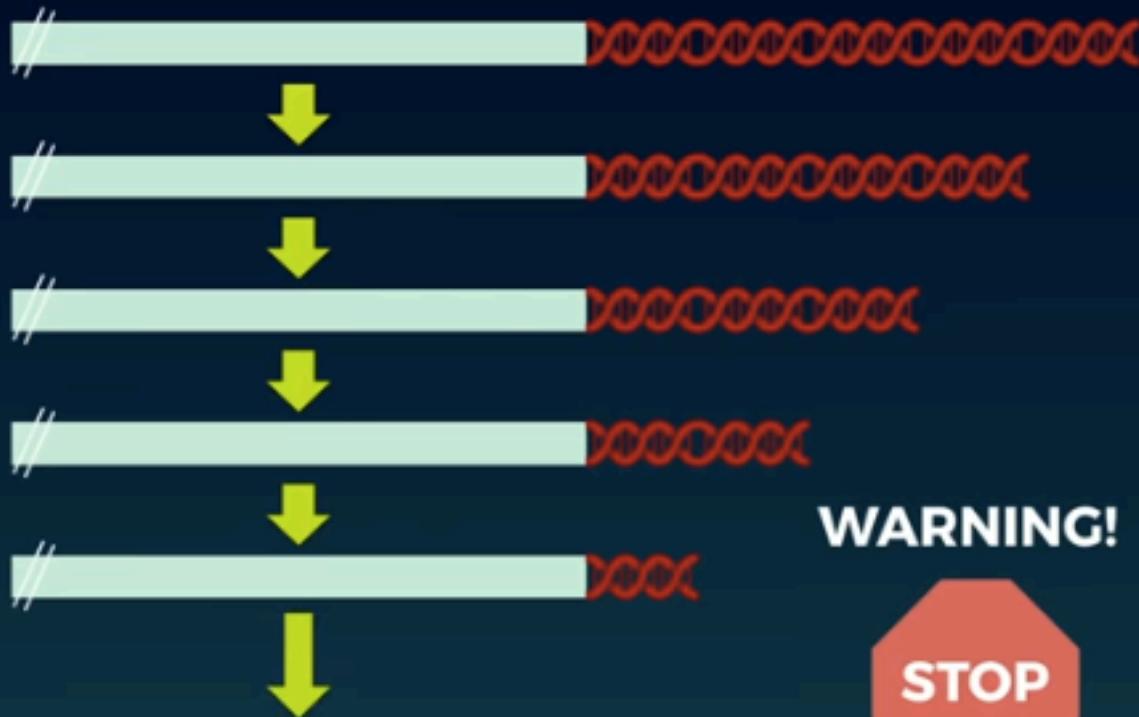


© The Nobel Foundation. Photo: U. Montan
Jack W. Szostak
Prize share: 1/3

for the discovery
of how
chromosomes
are protected by
telomeres and
the enzyme
telomerase.

Chromosomes replicate

cell divisions ↓



Cells stop dividing

WARNING!



**Telomeres
progressively
shorten**

Tetrahymena chromosomes

cell divisions



Plenty of
telomerase

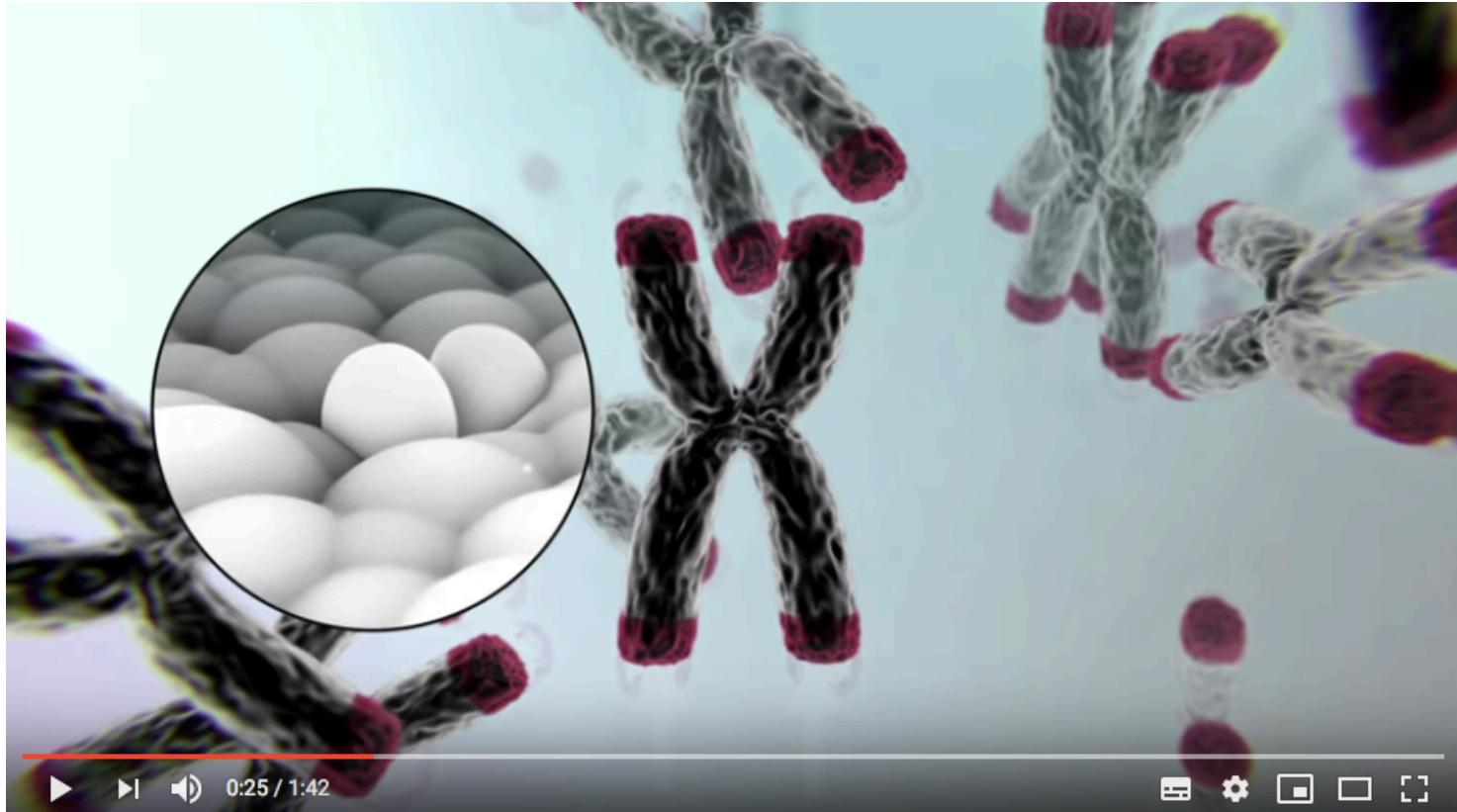


Tetrahymena are immortal

HeLa - The immortal cells of Henrietta Lacks



Stopping Cancer at the Starting Line



<https://www.youtube.com/watch?v=3LpZzcCBLe0>