欢迎参加龙星计划“肿瘤信息学”短课

CCF龙星计划

加入CCF | 登录CCF

首页 | 关于龙星计划 | 课程表 | 课程承办申请 | 出版物 | 媒体报导 | 高层研讨会 | 联系方式

您的位置：首页 > 龙星计划 > 关于龙星计划

关于龙星计划

龙星计划委员会

动态

关于龙星计划

阅读量：5986　　系统管理员　　收藏本文

近年来，我国经济得到了持续发展，这意味着我国不仅在产业，而且在科学技术等方面要面临全球一体化的严峻挑战。在这激烈的竞争中，优秀人才是取得胜利的关键因素（在信息领域显得更为突出）。我们高兴的看到，海外一批中国留学生现已学有所成，在诸多信息科学前沿领域做出了重大贡献。

龙星计划就是组织一批在美国学术界已有成就、有一定地位的原中国留学生，不定期回国就某一领域，在中国各地大学，系统地讲授一门美国研究生课程（每门课程15-30课时）。同时，就所讲课程的学术领域、有关课题与国内科学家及研究生共同讨论研究。这对提高我国科研水平和培养优秀人才都将起着重要作用。

（1）龙星计划委员会（下称委员会）分为两部分，即海外部分和国内部分。分别由一位主任主持工作。委员都是国内某一技术领域专家。委员会负责遴选讲者、确定承办单位和课程设置等工作。 设在中国科学院计算技术研究所的龙星计划办公室为龙星计划顺利实施提供必要的支撑保障。

委员会每年征求授课讲者和承办单位，公布学术交流领域。各大学提出申请后由委员会进行选择。龙星计划每年评估学术交流活动情况，为下一年课程安排做参考。

（2）每年组织6-12人次回国讲学、短期工作。每次讲授一门研究生课程，计15-30小时。课程仅面向中国计算机学会会员招生，免收学费，食宿自理。CCF会员申请信息地址：http://www.ccf.org.cn/c/2017-02-22/582915.shtml 。

（3）承办单位负责组织学员及提供各种信息，以保证课程的圆满成功;负责给课程提供必要的设备、场地等;负责给讲者提供市内多方面的服务（如交通、住宿等）信息。

CCF龙星计划

肿瘤信息学

Cancer Biology: an informatics perspective

徐鹰

南方科技大学医学院

Main Topics

- Lecture 1 （12月2号）：肿瘤研究背景、及所需组学数据
- Lecture 2 （12月3号）：肿瘤的演化框架：化学稳态失衡、维持平衡的代谢重编程、表观调控及基因突变的作用
- Lecture 3 （12月4号）：肿瘤的各种特征及底层原因、肿瘤演化框架在不同器官及微环境的应用

Questions To Study

- What is cancer?
- What drives a cancer to start, progress, metastasize?
- Why cancer characteristics tend to be organ-specific?
- What dictate age-dependent cancer occurrence rates?
- What drives drug resistance by cancer cells?
- Why metastasized cancers behave differently from their primary counterparts?

Format

- Lectures: 9:00 – 12:00pm each day, December 2 – 4, 2025
- Reading: handouts posted online at

sysbio.med.sustech.edu.cn/学术活动.html#cn-downloads

https://sysbio.med.sustech.edu.cn/

- Teaching Assistants
  - Ms. Jing Yan (严婧)
  - Mr Bocheng Shi(石博诚)
  - Ms. Yinghua Zhao (赵英华)

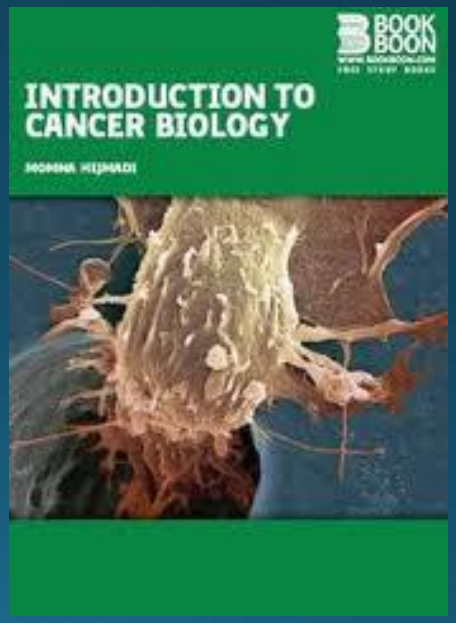
yanjer
福建 厦门
扫一扫上面的二维码图案，加我为朋友。

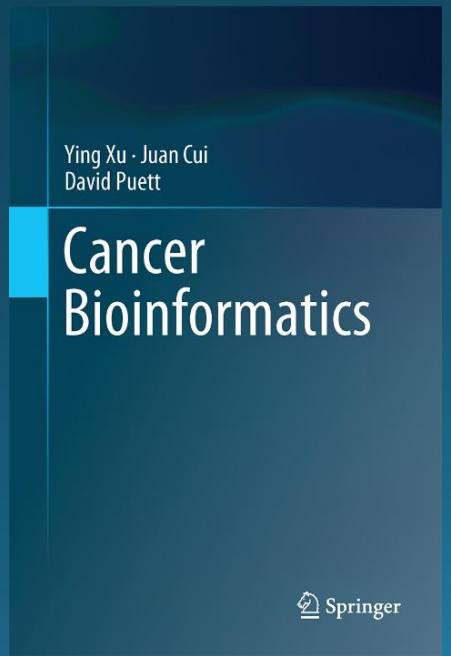
碳化锑先生
吉巴
扫一扫上面的二维码图案，加我为朋友。

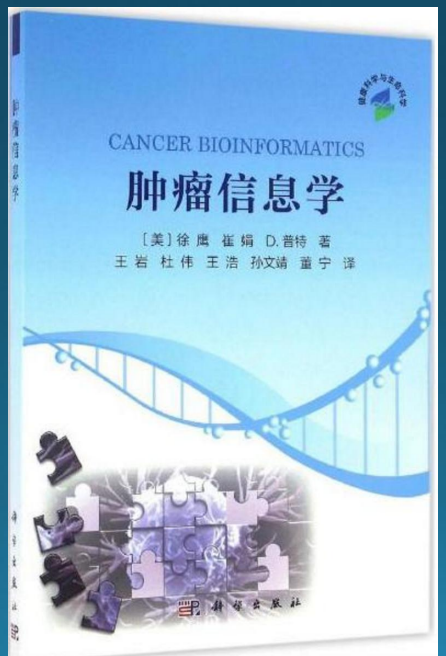
sukura.
广东 深圳

Reading Material and References

- All lecture and reading material are downloadable from online
- "Introduction to cancer biology" (a free on-line book), by Momna Hejmadi
- "Cancer Bioinformatics", Springer (2014)







Expectation of the Students

- Read the short book "Introduction to cancer biology" and the suggested literature
- Active participation in in-class discussions and <u>no private in-class conversations</u>
- Have your phones switched off and absolutely no phone calls in class!

Lecture I

Introduction to Cancer Biology:

Study of cancer from an informatics perspective

What is Cancer According to Literature

A cancer is often defined as a collection of cells that grow uncontrollably by disregarding the rules imposed on normal cells/tissues, which can invade and colonize other tissues

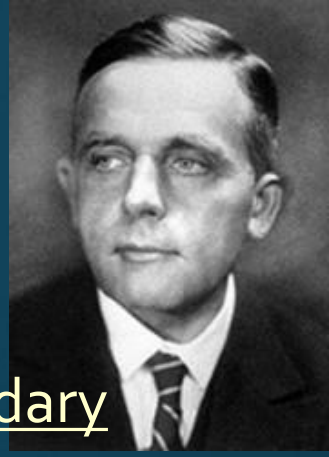
While the definition is simple, we do not have a detailed understanding about what causes a cancer and what drives the disease.

Yet, as we will learn later, this definition may not capture the true essence of the disease.

A Historical Perspective

- A case of breast cancer was identified and clearly documented by Egyptian physician Imhotep 4,50 years ago
- It is Greek physician who named the disease *kartinos* 2,200 years ago, Greek word for crab, now coming down to us as *cancer*
- Multiple theories were developed in the past 1,000 years, particularly since 19th century when surgeries, along with anesthesia and antibiotics, were widely used to treat human illness
- The first major breakthrough in understanding of cancer at the molecular level is the observation by German biochemist Dr. Otto Warburg that cancer cells tend to use glycolytic fermentation pathway regardless of the level of available O2.

A Historical Perspective



- In 1960's, Warburg stated: Cancer … has countless secondary causes; but there is only one prime cause, (which) is the replacement of respiration of oxygen in normal body cells by a fermentation of sugar.

- He went on to further state: … the de-differentiation of life takes place in cancer development. The highly differentiated cells are transformed into non-oxygen-breathing fermenting cells, which have lost all their body functions and retain only the now useless property of growth … What remains are growing machines that destroy the body in which they grow

Otto Warburg 1931
Nobel prize winner

A Historical Perspective

- The first oncogenic virus was discovered by Peyton Rous of Rockefeller Institute in 1916 (received Nobel Prize in 1966)
- Rous excised a sarcoma in a chicken, ground and injected the soluble filtrate into chickens; then a sarcoma would develop
- After years of intensive research, the transmissible agent was identified as the Rous sarcoma virus (*RSV*)
- The actual oncogenic element in the retroviral genome was a mutated gene *SRC*

A Historical Perspective

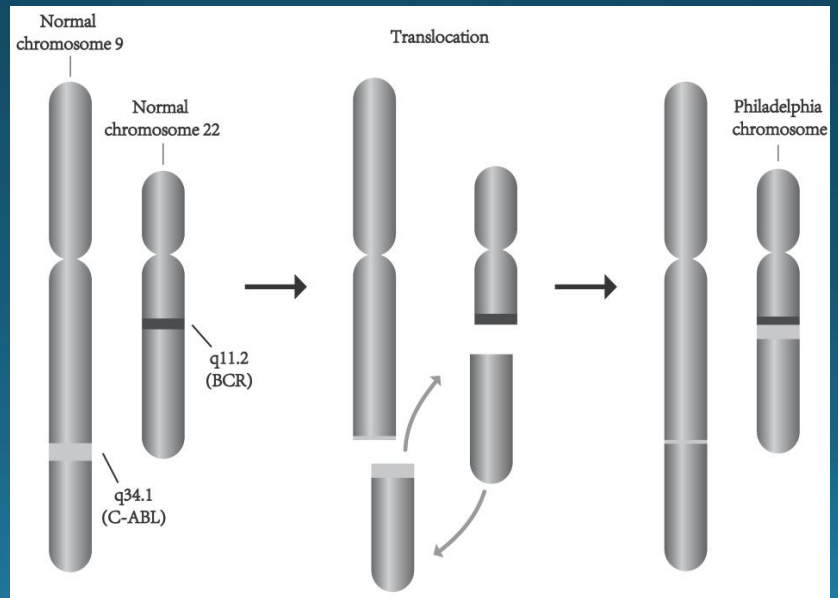
- The concept of oncogene was coined in 1969 by NIH scientists, George Todaro and Robert Heubner
- The first confirmed oncogene was discovered in 1970 and was termed SRC by Dr. Steve Martin
- For demonstrating over-expression in human SRC can transform normal cells to cancer cells, Bishop and Varmus received Nobel Prize in 1989， 开始了肿瘤是基因突变结果的时代

Bishop and Varmus, 1989
Nobel prize winners

A Historical Perspective

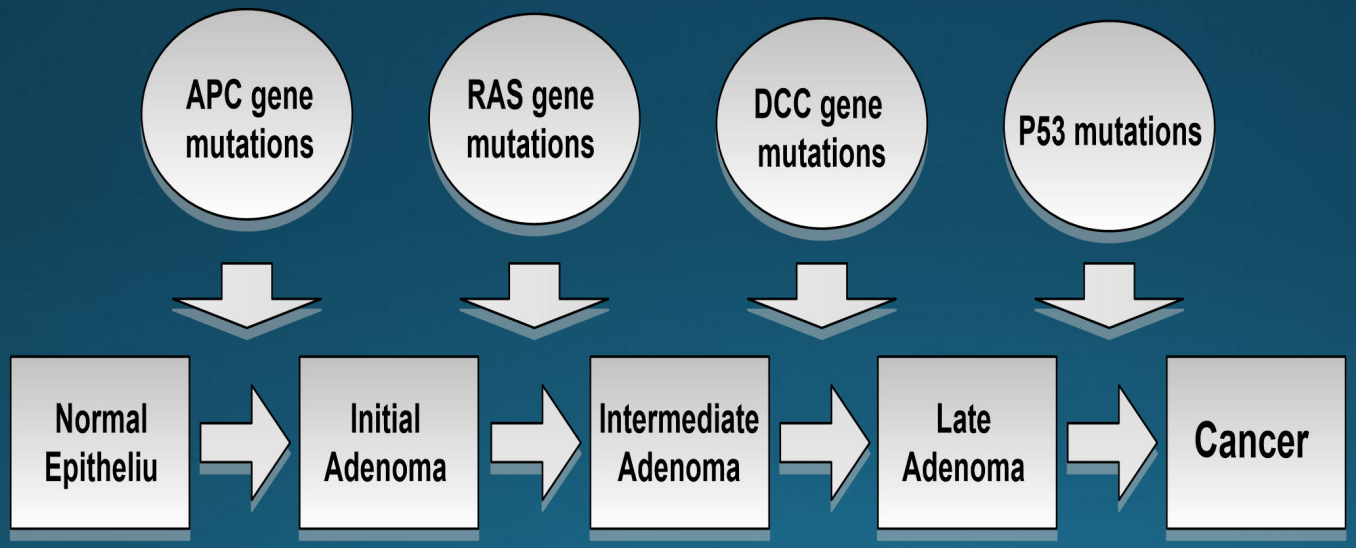
- The discovery of (proto)oncogenes by Bishop and Varmus and tumor suppressor genes by AG Knudson, both in 1970s, laid a foundation for the now popular theory that cancer is the result of genomic mutations. This theory has dominated the thinking in cancer research for ~40 years

- Philadelphia chromosome is believed to be the cause of CML, a type of blood cancer



A Historical Perspective

- The first mutation-driver model of cancer was proposed in 1990 by Fearon and Vogelstein based on the observation that vast majority of colorectal cancers have mutations in the APC gene



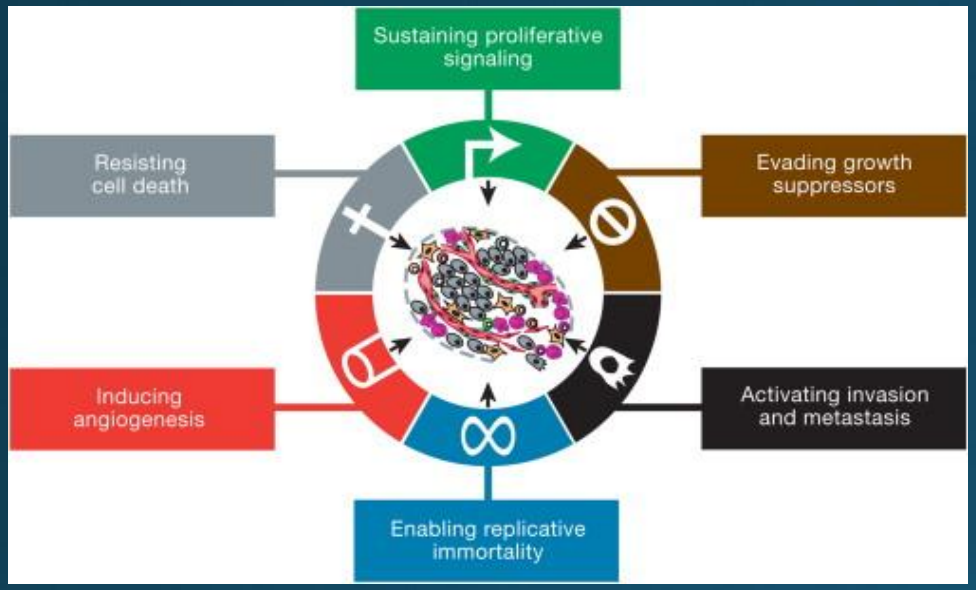
Bert Vogelstein

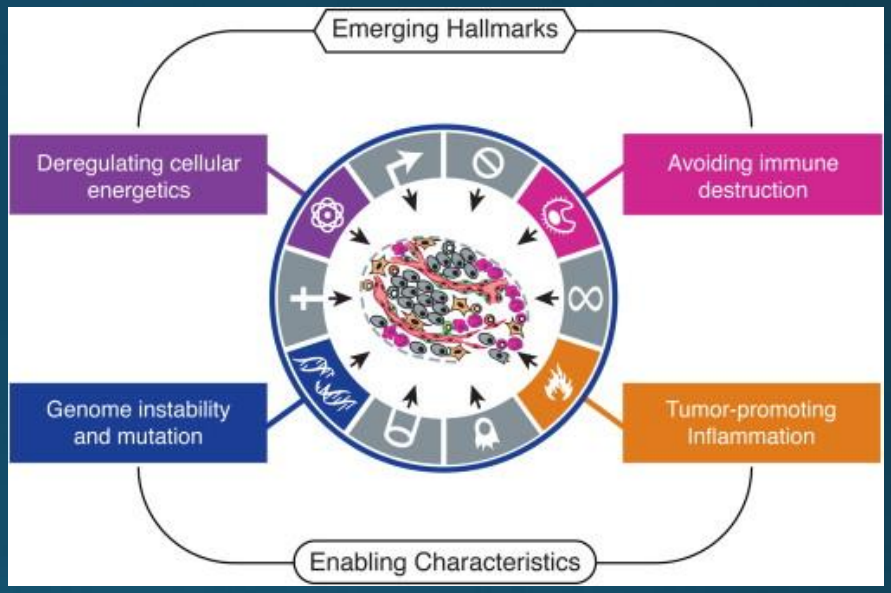
Phenotypic Characteristics of Cancer

- While substantial amount of information has been generated about cancer (over 1 million research articles), it remains largely unclear what really constitutes a cancer at the cellular and tissue level!

- Hanahan and Weinberg published two seminal papers "The Hallmarks of Cancer" and "The Hallmarks of Cancer: the next generation", which for the first time defines the distinguishing molecular level characteristics of a cancer

Hallmarks of Cancer



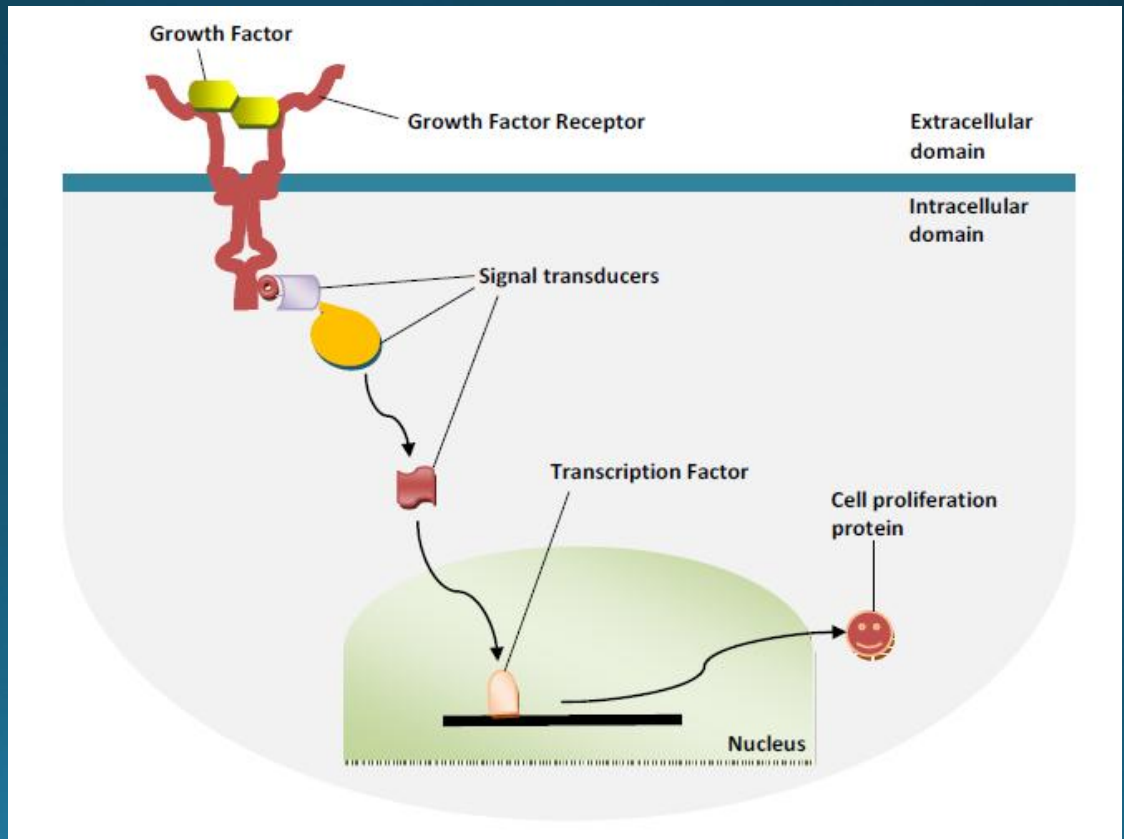


- Sustained proliferative signaling
- Evading growth suppressors
- Resisting cell death
- Enabling replicative immortality

- Inducing angiogenesis
- Activating invasion/metastasis
- Reprogramming energy metabolism
- Avoiding immune destruction

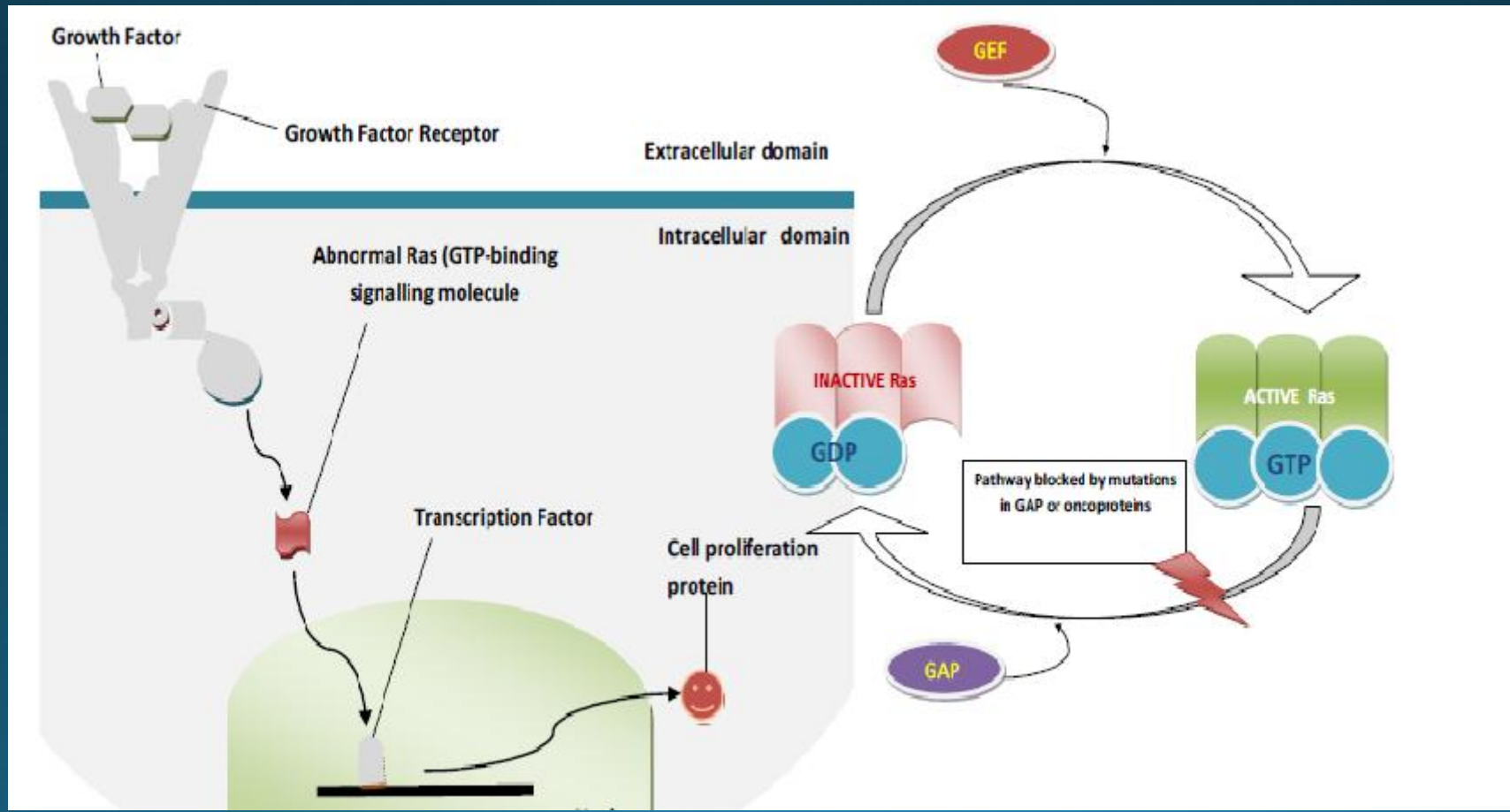
I: Sustained Proliferative Signaling

- Normal cells will proliferate only when they receive "growth signals"
- Cancer cells, for "unknown" reasons, grow without necessarily having such signals



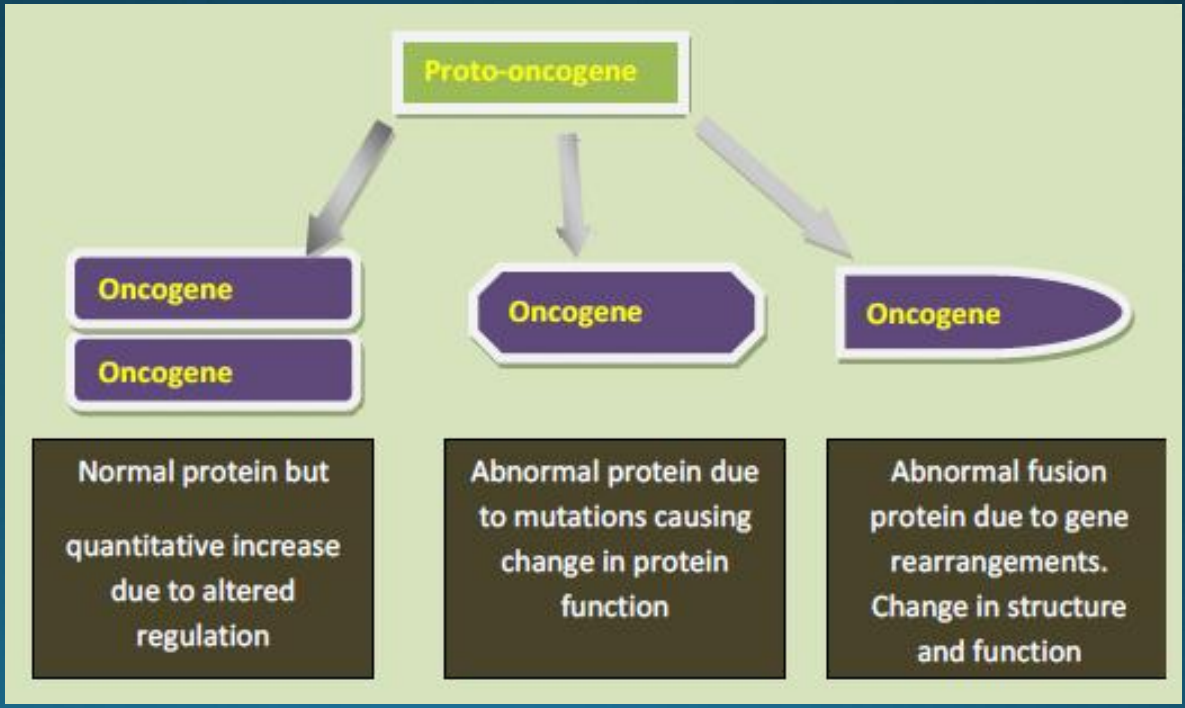
Sustained Proliferative Signaling

- Abnormal signaling by Ras (as an oncogene)



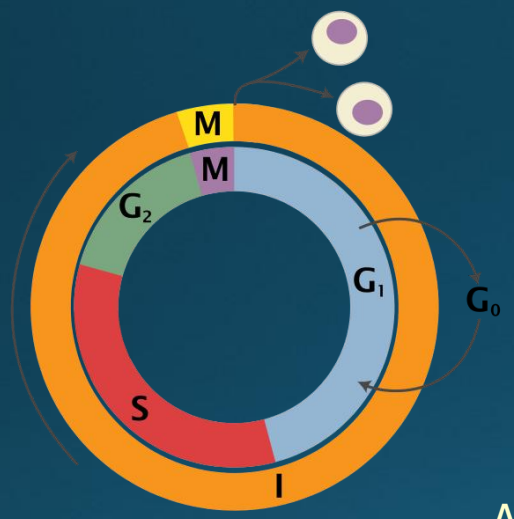
Oncogenes

- Oncogenes are genes whose over-expression or mutations can lead to cancer
- Hundreds of oncogenes have been identified
- Different cancers may have their own main oncogenes



II: Evading Growth Suppressors

- Like growth signals, there are anti-growth signals to stop cells from division

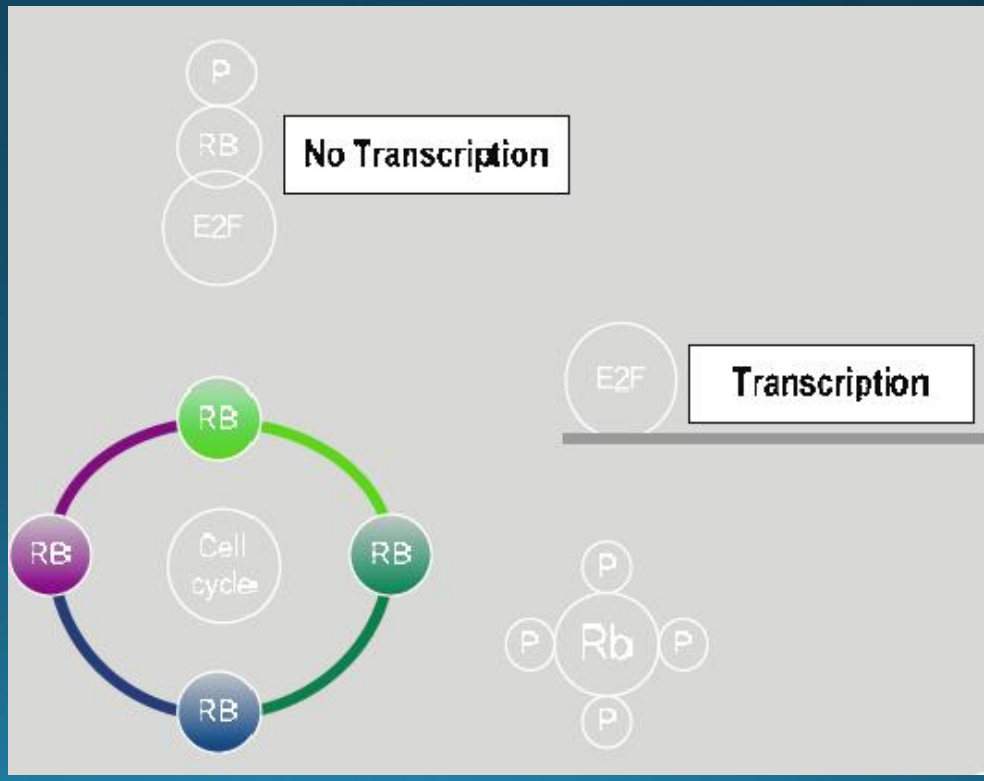


Cell cycle

Anti-growth signals can force dividing cells into the quiescent phase (G0) of the cell cycle

Evading Growth Suppressors

- RB (retinoblastoma) protein is one such anti-growth protein, which binds to the regulators of the cell cycle
- P53 is another anti-growth protein
- Cancer cells somehow have learned to by-pass the anti-growth mechanism through having mutations or repression of proteins like RB and P53



Tumor Suppressor Genes

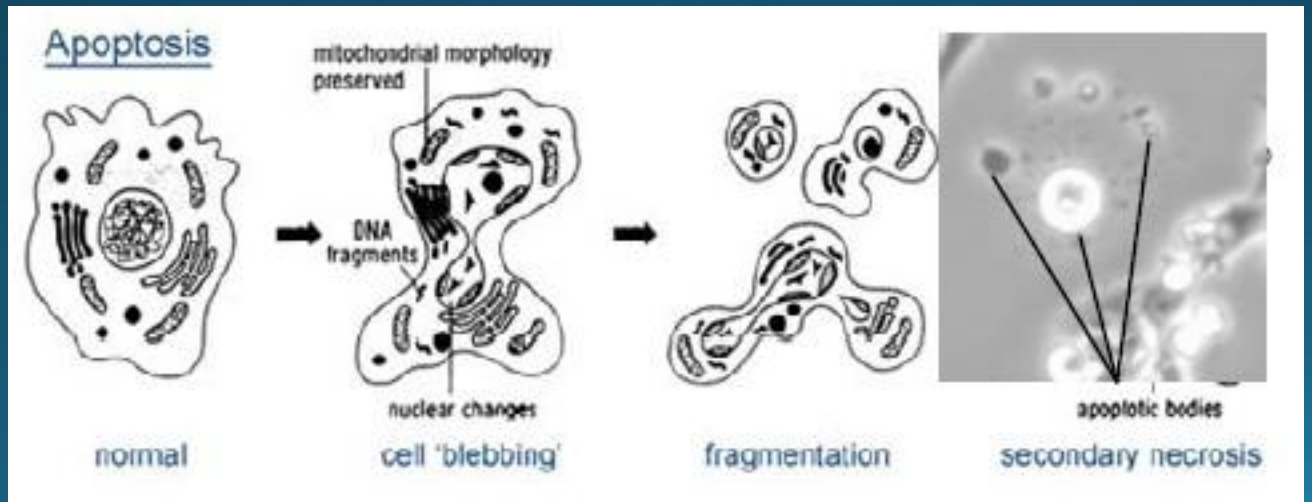
- Genes that encode proteins capable of inhibiting cell division, like RB, are called tumor suppressor genes
- In cancer genomes, multiple tumor suppressor genes may have loss-of-function mutations
- A few hundred tumor suppressor genes have been identified for different cancers

Later we will learn that there could be a fundamentally different way to look at oncogenic and tumor suppressor mutations!

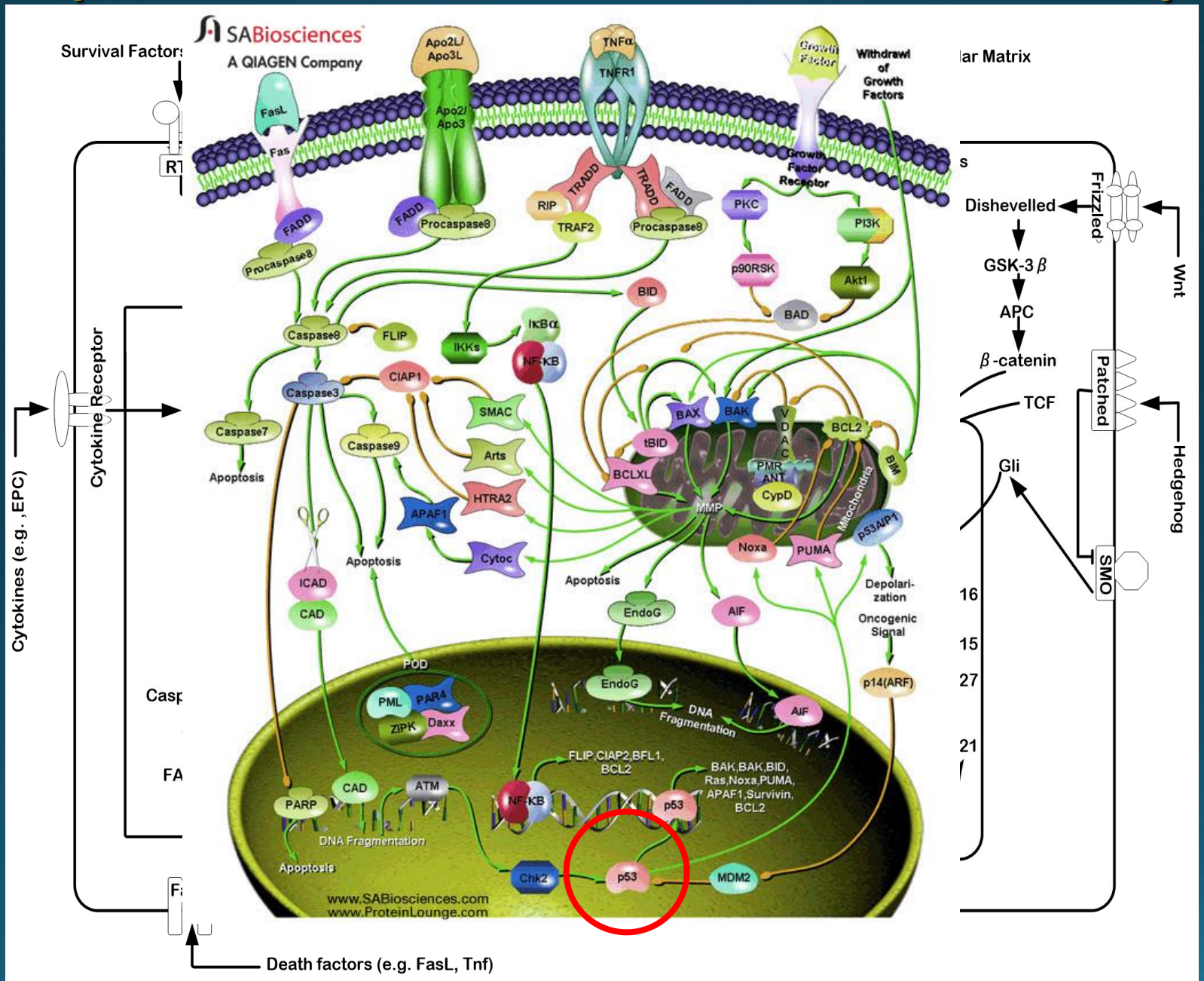
III: Resisting Cell Death

- A cell constantly surveys its internal state including access to oxygen and nutrients, integrity of its genome and balance of its cell cycle pathways

- If malfunction or damage is detected, the cell activates cell death (apoptotic) pathway to kill itself



Apoptosis and Associated Pathways



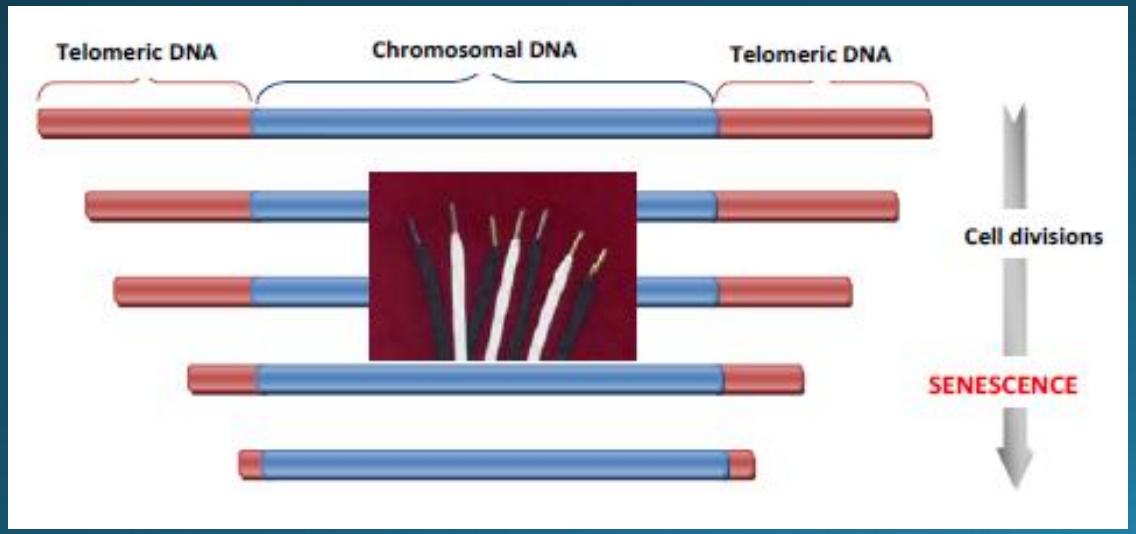
Resisting Cell Death

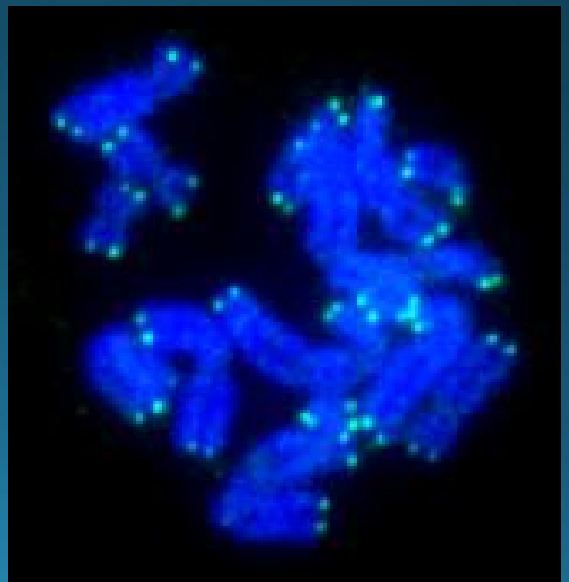
- Cancer cells all learned to by-pass the apoptosis process to avoid to be killed
- They use (at least) two pathways to avoid apoptosis by impairing the sensing of and signaling about abnormal internal status or the execution apoptosis
- One main mechanism is through having mutations in the main regulator, p53, of apoptosis
  - 50% of the cancer genomes have mutations in p53

IV: Enabling Replicative Immortality

- Normal human cells can divide 60-70 times and then reaches the end of its natural life

- Cells all keep a biological clock that keeps track of their ages





Enabling Replicative Immortality

- Cancer cells learned to protect their telomeres, so they do not get shortened when cells divide, hence making cancer cells immortal
- They use telomerase to add to the ends of telomeres after each division to maintain their lengths
- This is an encoded mechanism in human cells but has been used only by embryonic stem cells

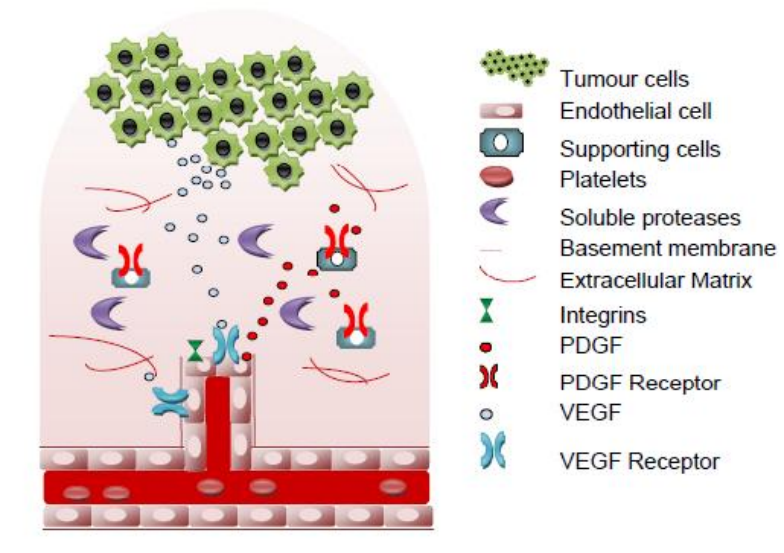
V: Inducing Angiogenesis

- All cells, healthy or diseased, need oxygen and nutrients, which can be provided only through blood vessels
- Human bodies are well designed so every cell is within 100 um of a capillary
- Cancer cells need additional blood supply to support its rapid growth

Inducing Angiogenesis

- Angiogenesis is a process that grows new blood vessels from the existing ones, which is used during wound healing or menstruation

- Cancer cells learned to send out angiogenic signals to endothelial cells lining nearby vessels to grow new vessels



Inducing Angiogenesis

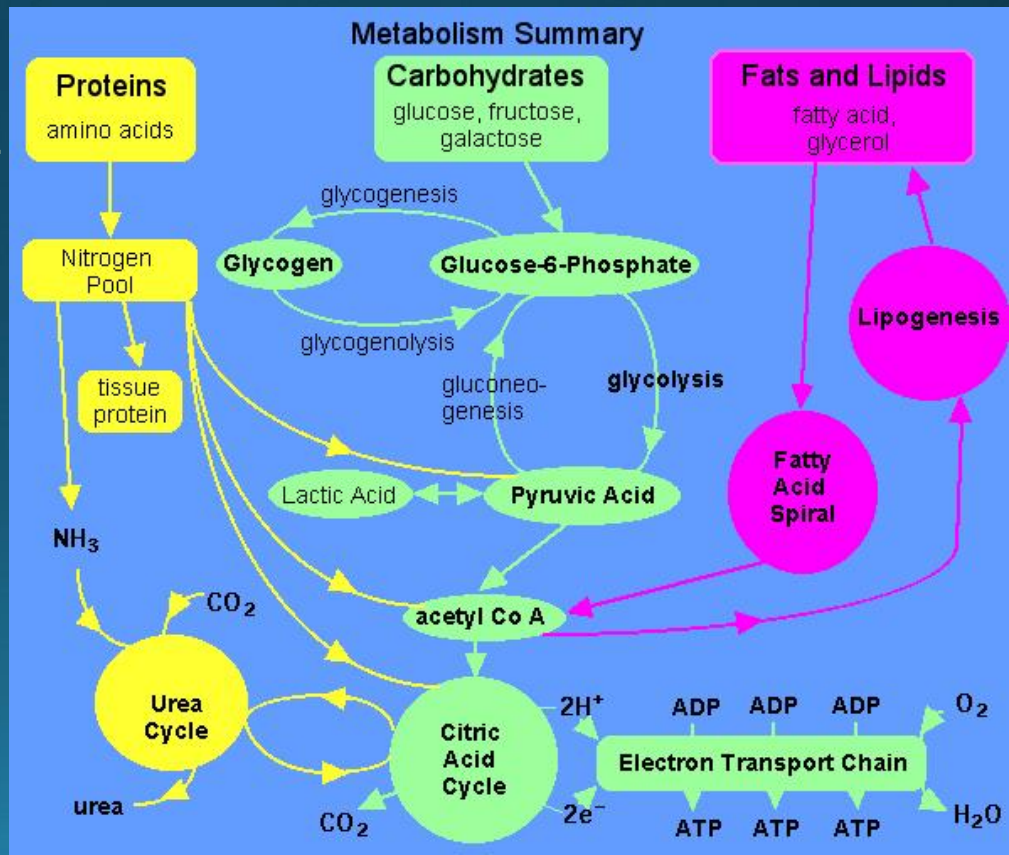
- Without angiogenesis, a cancer will not be able to grow big nor able to spread to other parts of the body
- Cancer becomes dangerous only after it starts to have its own blood vessels
- One type of cancer treatment is to kill cells with messy blood vessels

VI: Reprogrammed Energy Metabolism

- Human cells have multiple ways to convert nutrients to energy (ATP)

- Oxidation of pyruvate through utilization of electron transport chain requires oxygen

- Glycolysis does not require oxygen and uses lactate as the electron receiver



Reprogrammed Energy Metabolism

- Otto Warburg observed in 1927 that cancer cells use glycolytic fermentation in addition to oxidation of pyruvate regardless of the O2 level (Warburg effect)
- Oxidation of pyruvate is by far the most efficient energy metabolism per glucose
- It seems that all cancers utilize suboptimal energy metabolisms, which may be a key reason for their explosive growth – a paradox

Reprogrammed Energy Metabolism

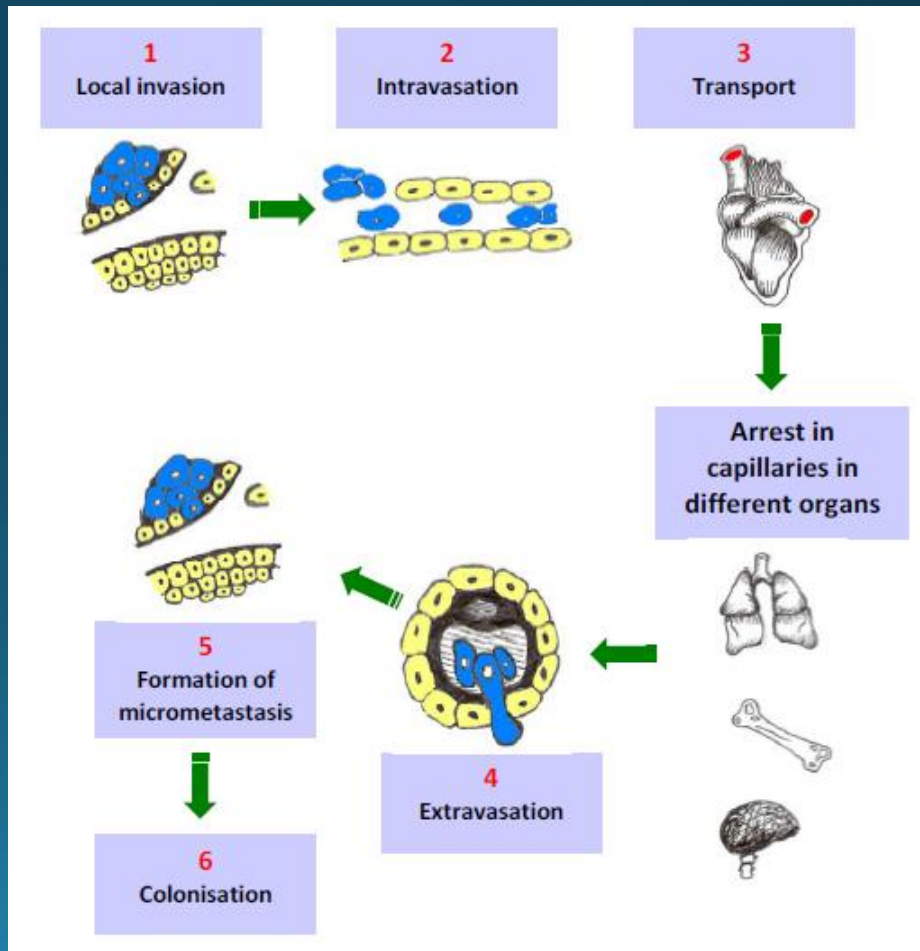
- While Warburg effect was widely observed in cancer tissues, no generally accepted explanation has been developed
- This remains to be one of the most intriguing issues related to cancer development
- We will present a model to answer the question

VII: Avoiding Immune Destruction

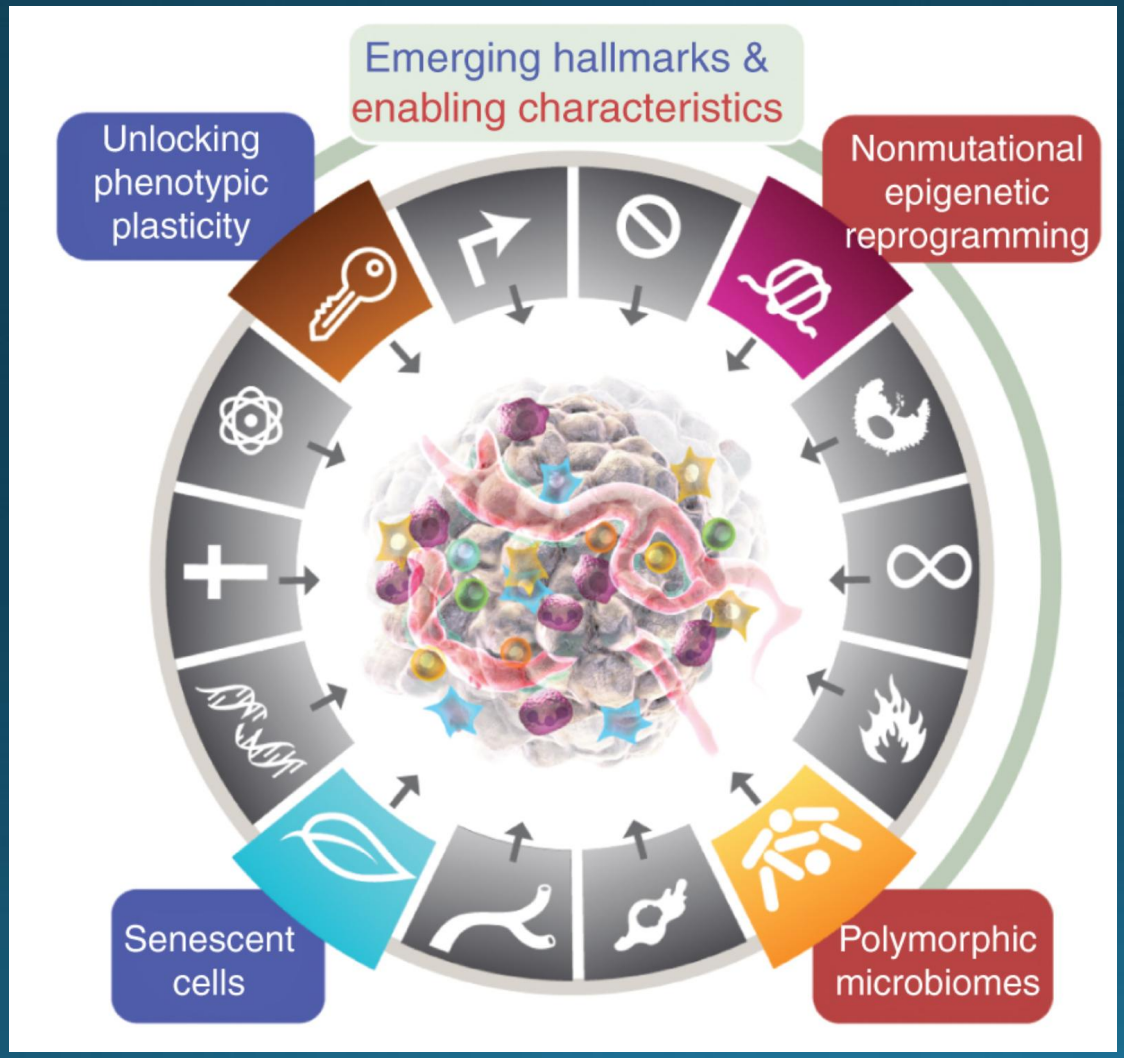
- Cancer development and immunity are linked at the root
  - Cancer is often considered as a wound that will not heal
  - Immune system responds to two things: (a) invasion of pathogens, and (b) tissue damages
- Cancer studies using immune-deficient mice may not necessarily lead to cancer related discoveries; the same can be said about cell-based cancer research

VIII: Activating Invasion/Metastasis

- What makes cancer really deadly is its ability to invade neighboring cells and spread to other organs
- Over 93% of the cancer related deaths are due to metastatic cancers
- … and yet, our previous view about metastatic cancers may not be correct!



Hallmarks of Cancer: new dimensions



Cancer discovery (2022)

Hallmarks of Cancer: new dimensions

- Unlocking phenotypic plasticity: cancer cells can de-differentiate, differentiate and trans-differentiate to various cell types
- Senescent cells: Cell senescence is irreversible and tend to release various molecules that trigger immune responses. Cancer cells seem to be able to switch between senescence and proliferation state
- Non-mutational epigenomic reprogramming: tumor microenvironment can trigger abnormal changes in folded DNA structures, altering encoded transcription programs
- Polymorphic microbiomes: There are interactions between gut microbiota and cancer tissues

Cancer Hallmarks

- The first two papers have been widely used as the guiding framework in cancer research
- They have clearly captured some of the key phenotypic characteristics of cancers in general
- They represent the state of the art in understanding of cancers
- BUT they did not touch on the root issue of cancer: what drive cancers to evolve?
- Issues discussed in the third paper are not cancer specific

Genome Instability

- Genome instabilities are common in cancer cells, and they are considered a "trademark" for these cells.

- It is widely believed that sporadic tumors (non-familial ones) are originated due to the accumulation of several genetic errors (mainstream thinking but it may not be correct)

- Studies of cancer genomes, to learn about “driver mutations” have been popular but also disappointing as not many new insights have been gained about cancer initiation and development

Mutations in Other Diseases

Science. Author manuscript; available in PMC 2014 Feb 3.
Published in final edited form as:
Science. 2013 Jul 5; 341(6141): 1237758.
doi: 10.1126/science.1237758

PMCID: PMC3909954
NIHMSID: NIHMS546313

Somatic Mutation, Genomic Variation, and Neurological Disease

Annapurna Poduri,[1,2] Gilad D. Evrony,[3,4] Xuyu Cai,[3,4] and Christopher A. Walsh[2,3,4,*]

- … Increasingly, somatic mutations are being identified in diseases other than cancer, including neurodevelopmental diseases. Somatic mutations can arise during the course of prenatal brain development and cause neurological disease, resulting in brain malformations associated with epilepsy and intellectual disability.

Mutations in Other Diseases

nature genetics

Home | Current issue | Comment | Research | Archive ▾ | Authors & referees ▾ | About the journal ▾

home ▸ archive ▸ issue ▸ letter ▸ full text

NATURE GENETICS | LETTER

日本語要約

Somatic mutations of t
associated gene *PARI*
human malignancies

nature genetics

Home | Current issue | Comment | Research | Archive ▾ | Authors & referees ▾ | About the journal ▾

home ▸ archive ▸ issue ▸ letter ▸ abstract

NATURE GENETICS | LETTER

日本語要約

Genome-wide association identifies multiple
ulcerative colitis susceptibility loci

- The authors stated: N
  ubiquitin ligase, is th
  disease. In a search f
  as a frequently targe

Mutations in Other Diseases

Gan To Kagaku Ryoho. 2000 Mar;27(3):335-40.

[Genome analyses for precancerous lesions in the gastrointestinal tract .

[Article in Japanese]

Sowa M[1], Nakata B.

Author information

Journ Annu Diabetol Hotel Dieu. 1997:25-31.

Detection and prevalence of mitochondrial genome mutations in diabetes .

[Article in French]

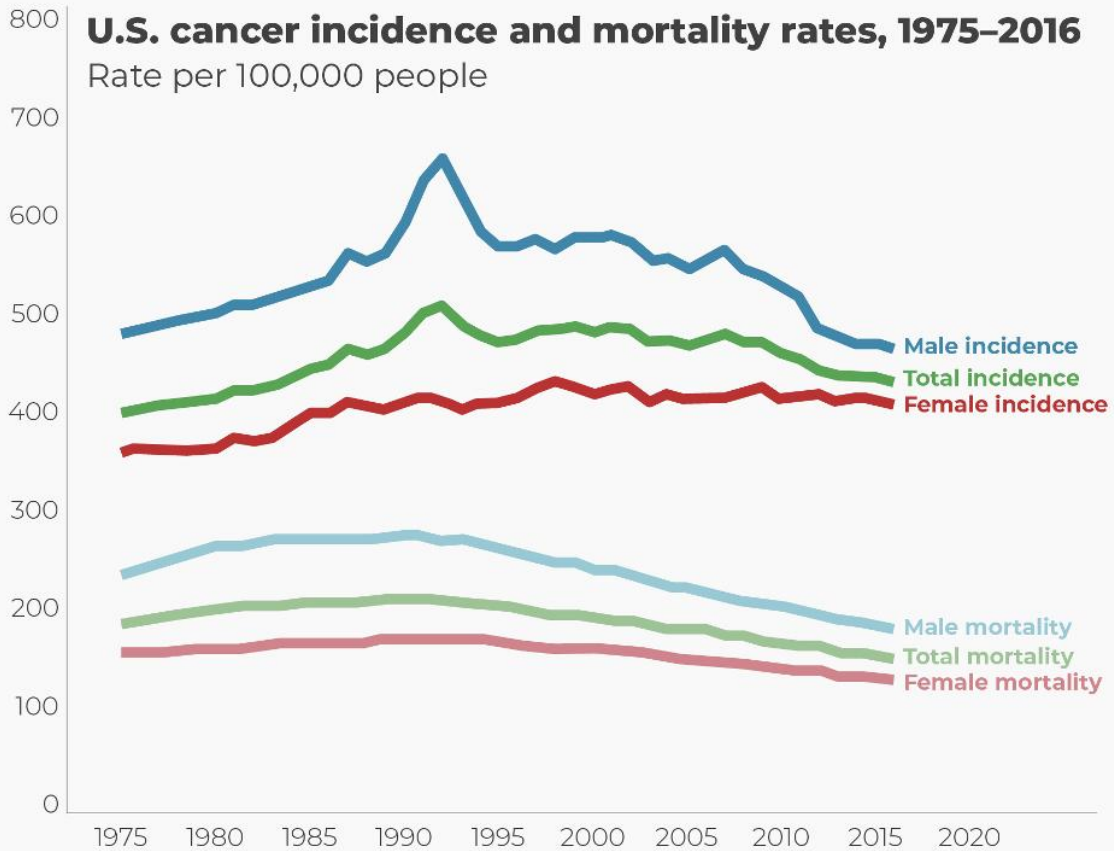
Paquis-Flucklinger V[1], Vialettes B, Canivet B, Freychet P, Hieronimus S, Vague P, Saunières A, Desnuelle C.

Ann Lab Med. 2015 Jan; 35(1): 1–14. PMCID: PMC4272938
Published online 2014 Dec 8. doi: 10.3343/alm.2015.35.1.1

Mitochondrial DNA Aberrations and Pathophysiological Implications in Hematopoietic Diseases, Chronic Inflammatory Diseases, and Cancers

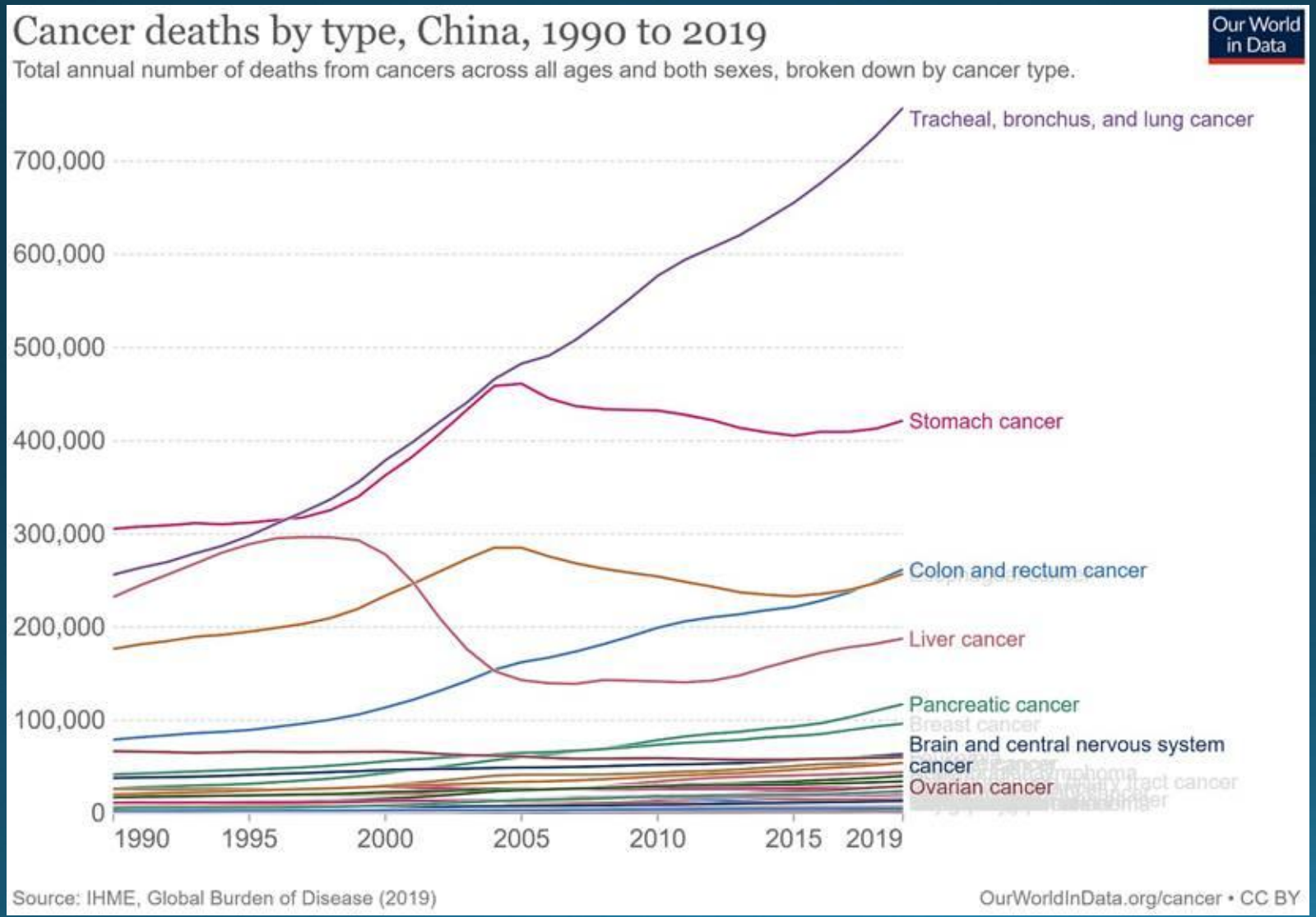
Hye-Ran Kim, Ph.D.,[1,2,*] Stephanie Jane Won, B.S.,[3,*] Claire Fabian, Ph.D.,[4] Min-Gu Kang, M.D.,[1,2] Michael Szardenings, Ph.D.,[4] and Myung-Geun Shin, M.D.[1,2,5]

肿瘤死亡率降低的一个重要原因是早期发现，早期治疗



Source: Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal, "Cancer Statistics, 2020," CA: A Cancer Journal for Clinicians 70, no. 1 (January/February 2020): 15, Figure 2.

Cancer Death Rates in China



Fighting Against Cancer

- The US congress declared "War on cancer" in 1971, based on a recommendation from President Richard Nixon, which was renewed three times, now called "moonshot"
- Hundreds of billions of dollars have been invested into cancer research worldwide but our abilities in treating cancer have not advanced substantially
- The war on cancer, after almost forty years, must be deemed a failure with a few notable exceptions

James Watson, NY Time, 2009

Some Thoughts

- The essence of cancer is an evolutionary problem, i.e., the tissue has to evolve to adapt to an increasingly more challenging microenvironment but virtually no studies have been done on what stresses the diseased cells must overcome

- Traditional cancer studies are highly reductionist in nature, which aim to identify "bad parts" in a functional system; in reality the initial challenges encountered by the functional system may not be bad parts but instead some fundamental balances among different ingredients of the system might have been altered

Bacterial Evolution

- E. coli under persistent ethanol stress becomes low in ATP production due to membrane leakages resulted from oxidation by ethanol, leading to reduced biosynthesis of large biomolecules such as phospholipids, forming a vicious cycle generally resulting in progressively reduced ATP generation and ultimately death.
- Some cells adapt to the stressor by generating and selecting mutations in genes encoding major ATP-consuming proteins, and utilize the saved ATPs towards biosynthesis of phospholipids that are used to repair the damaged membrane, leading to an increased level of ATP production and forming a positive cycle towards full recovery of ATP production and cellular functions

Physicochemical Conditions
in healthy vs. cancer tissues & cells

健康细胞

- 细胞浆 pH: 6.8 – 7.2.
- 细胞间液 pH: 7.3 – 7.5
- 细胞内及外主要电解质浓度:
  - 钠离子: 10 : 140
  - 钾离子: 145 : 4
  - 钙离子: 1 : 15,000
- 细胞外膜电势: 70 mV
- 线粒体膜电势: 140 mV

肿瘤细胞

- 细胞浆 pH: 7.2 – 7.5 .
- 细胞间液 pH: 6.4 – 6.6
- 细胞内及外主要电解质浓度:
  - 钠离子: 60 : 140
  - 钾离子: 145 : 5
  - 钙离子: 1 : 15,000
- 细胞外膜电势: 27 mV
- 线粒体膜电势: 210 – 280 mV

Physiological vs Pathological Conditions

- These differences in the basic chemical and physical conditions fundamentally changed the biology
- Namely, cancer biology is fundamentally different from normal biology.

- 我们需要新的思维方式及分析框架、技术来研究肿瘤生物学、及其它疾病生物学
- 从肿瘤大数据中挖掘肿瘤演化信息

Omic Data Collected on Cancer Tissues

- Cancer transcriptomic data
- Cancer genomic data
- Cancer epigenomic data
- Cancer metabolomic data
- Proteomic data
- …….

The hope is that by mining these omic data, we can start to see the big and the whole picture of cancer development as an evolutionary process

Information from the Omic Data

- Substantial amount of information is to be uncovered from the cancer omic data that has been generated through large consortia such as
  - TCGA, the largest cancer tissue omic database,
    - UALCAN: a front portal https://ualcan.path.uab.edu/analysis.html
  - GEO, the most comprehensive transcriptomic database for diseases in general, and
  - GTEx, the largest transcriptomic database for normal human tissues

Differentially Expressed Genes

- Consider two sets of samples
  - one being colon cancer tissues and the other being adjacent non-cancerous tissues so one can study genes possibly involved in cancer formation and progression
  - one being colon cancer samples with drug resistance and the other without drug resistance
- We are interested in finding out if a gene is differentially expressed between the two sets of samples

Differentially Expressed Genes

- T-test is a widely used statistic for assessing if the expressions: X1, …., Xn of gene X in one set of n samples is differentially expressed from the expressions: Y1, …, Yn in another set of samples
  - one diseased set *versus* control set
- $T(X,Y) = \frac{\bar{X}-\bar{Y}}{s\sqrt{2}}\sqrt{n}$, where $\bar{X}$ and $\bar{Y}$ are the means of {X1, …., Xn } and {Y1, …, Yn }, and S is the standard deviation.
- Consider a set of 10 cancer samples and a set of 10 matching control samples. If $T(X,Y) = 2.9$, then the statistical significance for the observation that gene X is differentially expressed is 0.005

TABLE of CRITICAL VALUES for STUDENT'S *t* DISTRIBUTIONS

Column headings denote probabilities (α) ***above*** tabulated values.

d.f.	0.40	0.25	0.10	0.05	0.04	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	7.916	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	3.320	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	2.605	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.104	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	1.948	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	1.928	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	1.869	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.539	[illegible]	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.518	[illegible]	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.255	0.681	1.306	1.688	1.802	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.255	0.681	1.305	1.687	1.800	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.255	0.681	1.304	1.686	1.799	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.255	0.681	1.304	1.685	1.798	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	1.781	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.254	0.678	1.292	1.664	1.773	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.254	0.677	1.290	1.660	1.769	1.984	2.081	2.364	2.626	2.871	3.174	3.390
120	0.254	0.677	1.289	1.658	1.766	1.980	2.076	2.358	2.617	2.860	3.160	3.373
140	0.254	0.676	1.288	1.656	1.763	1.977	2.073	2.353	2.611	2.852	3.149	3.361
160	0.254	0.676	1.287	1.654	1.762	1.975	2.071	2.350	2.607	2.847	3.142	3.352
180	0.254	0.676	1.286	1.653	1.761	1.973	2.069	2.347	2.603	2.842	3.136	3.345
200	0.254	0.676	1.286	1.653	1.760	1.972	2.067	2.345	2.601	2.838	3.131	3.340
250	0.254	0.675	1.285	1.651	1.758	1.969	2.065	2.341	2.596	2.832	3.123	3.330
inf	0.253	0.674	1.282	1.645	1.751	1.960	2.054	2.326	2.576	2.807	3.090	3.290

Estimate the statistical significance of a predicted differentially expressed gene

Differentially Expressed Genes

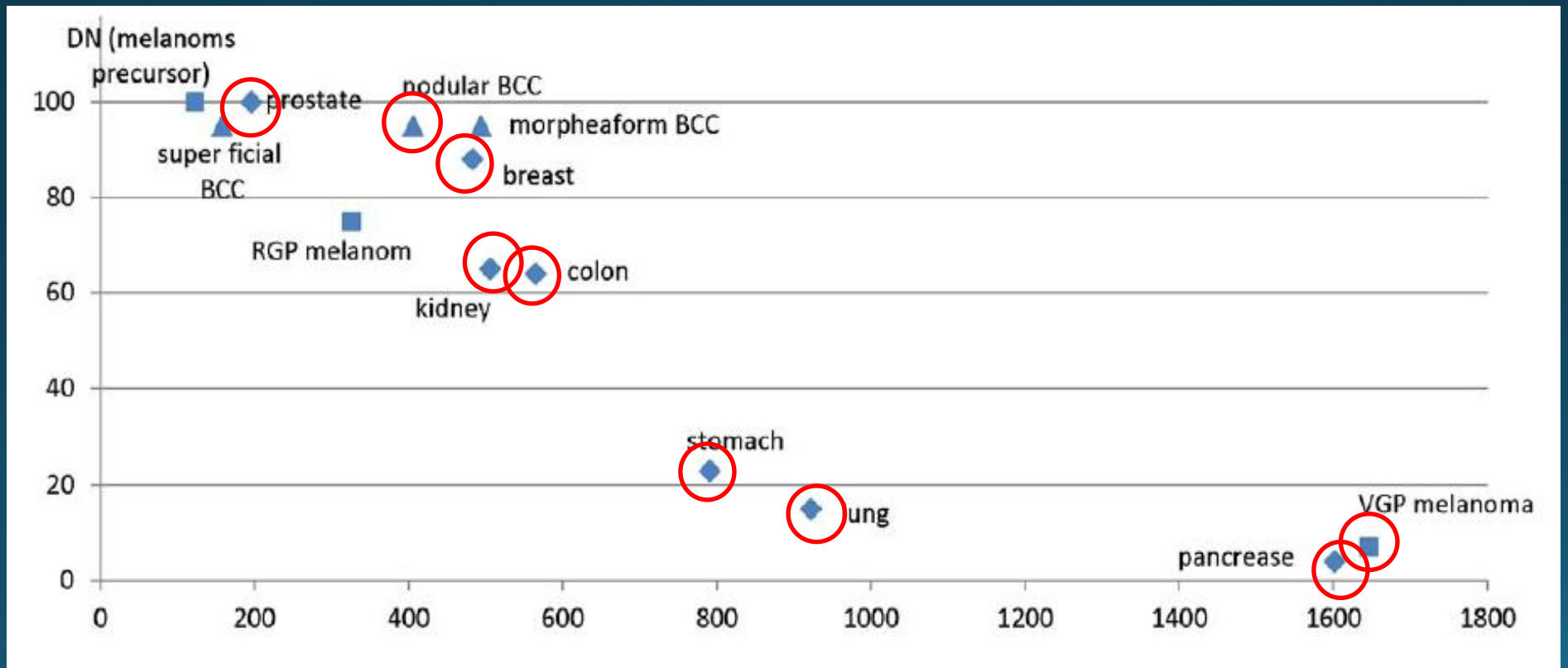
- Using this test, one can assess if a specific gene is over-expressed (up-regulated if $\bar{X} > \bar{Y}$) or under-expressed (down-regulated if $\bar{X} < \bar{Y}$) in one set of samples *versus* another

- If one wants to be conserved, one can require the average change is at least, say, 1.5 or 2 fold: if $\bar{X}/\bar{Y} > 1.5$ of 2.0

- Typically a few hundreds to a few thousands of genes are differentially expressed in cancer samples *versus* adjacent control samples for different cancer types

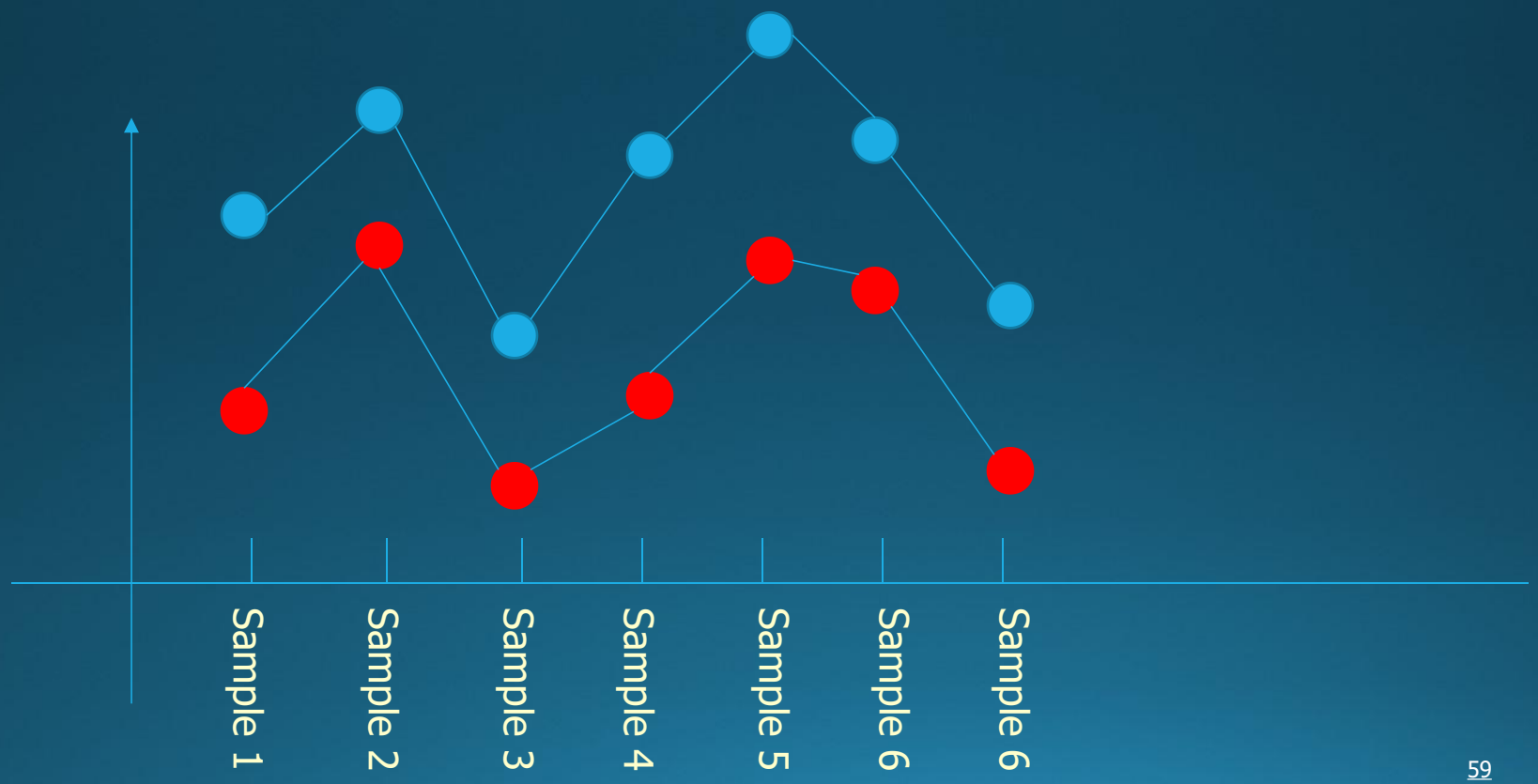
Differentially Expressed Genes

- By comparing the average number of differentially expressed genes for each cancer type and its five-year survival rate, one can get the following



Co-Expressed Genes

- Certain genes may show coordinated expression patterns across different samples, which are referred as co-expressed genes



Co-Expressed Genes

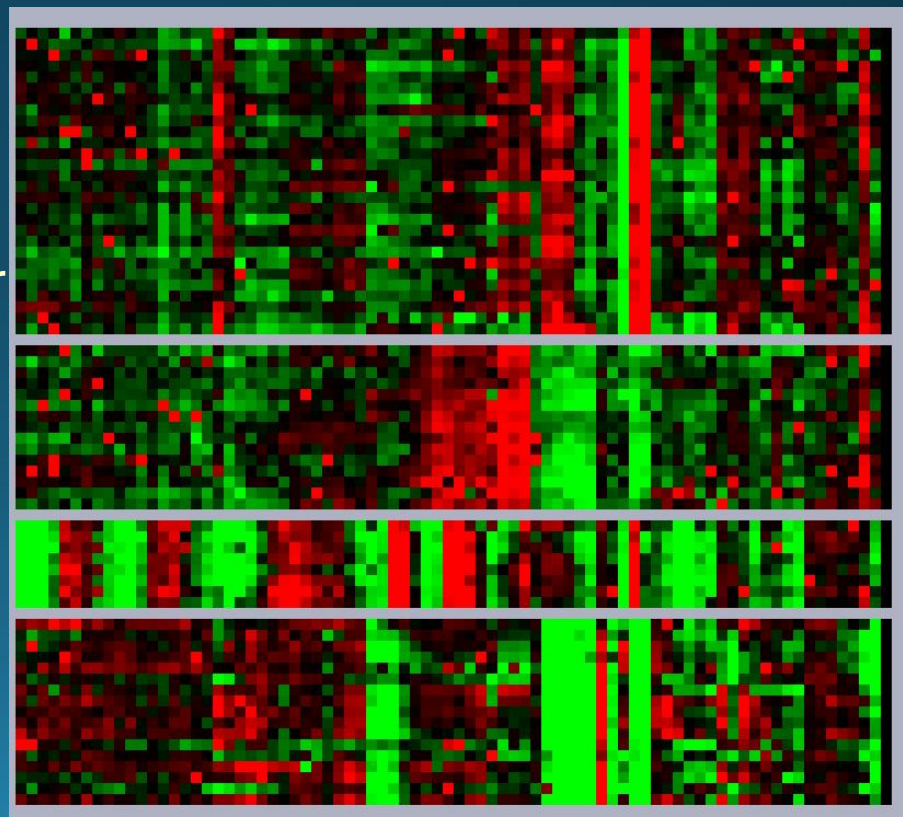
- Consider two genes X and Y, and their expression levels in *n* samples: $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$. The correlation coefficient between two expression patterns is measured using

$$CC(X,Y) = \frac{\sum((X_i - \bar{X}) * (Y_i - \bar{Y}))}{\sqrt{\sum(X_i - \bar{X})^2 * \sum(Y_i - \bar{Y})^2}}$$

- The two genes are called highly positively correlated if CC(X, Y) = 1; highly negatively correlated if CC(X, Y) = -1; not correlated if CC(X, Y) = 0

Co-Expressed Genes

- All co-expressed genes in a set of samples (e.g., colon cancer or E. coli treated with ethanol) can be identified using a clustering method.
- The figure shows 80 genes falling into 4 clusters across 110 colon cancer samples (column)
- Red means up-regulation; green for down-regulation and black for no changes between cancer and matching control



Classification Analysis of Samples

- Given one set of primary cancer samples without metastasis and another set of primary cancer samples which have been metastasized to a distant location, can we possibly find a set of genes whose expression patterns distinguish these two sets?

- If we can do this, we can possibly predict if a given cancer sample (with gene-expression data) has already metastasized or not.

- If we apply this idea to multiple cancer types, we can potentially derive the common set of genes or pathways that are essential to metastasis (a good project problem).

Classification Analysis of Samples

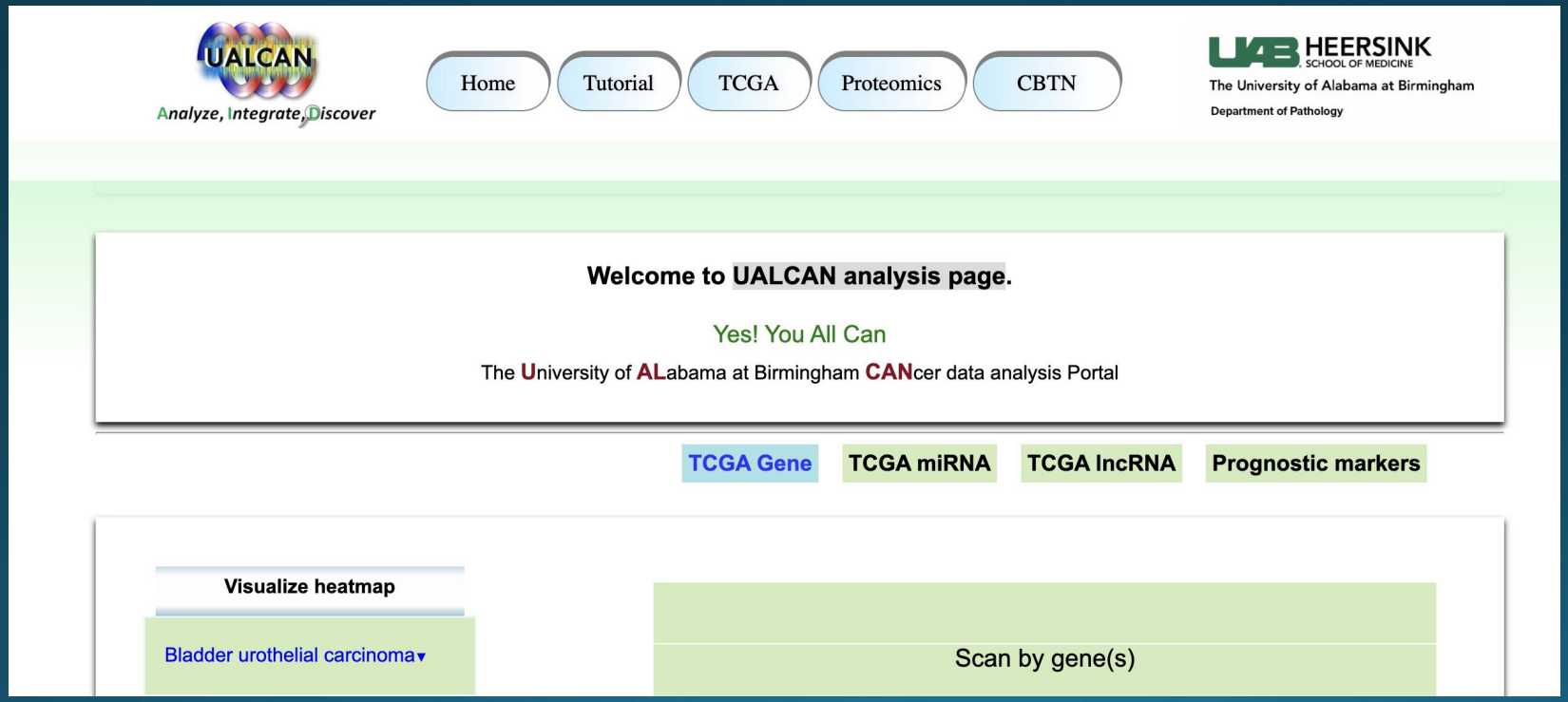
- Cancer biology problems that can potentially be solved using this type of technique:
  - Key differences among cancers of the same type but of different grades
  - Distinguishing characteristics among primary cancers of the same type but spread to different organs
  - Common characteristics among "slow growing" cancers as well as among "very fast growing" cancers
  - Distinguishing characteristics between pediatric cancers and adult cancers of the same types
  - Why certain organs do not or rarely develop cancers?

All are possible project problems.

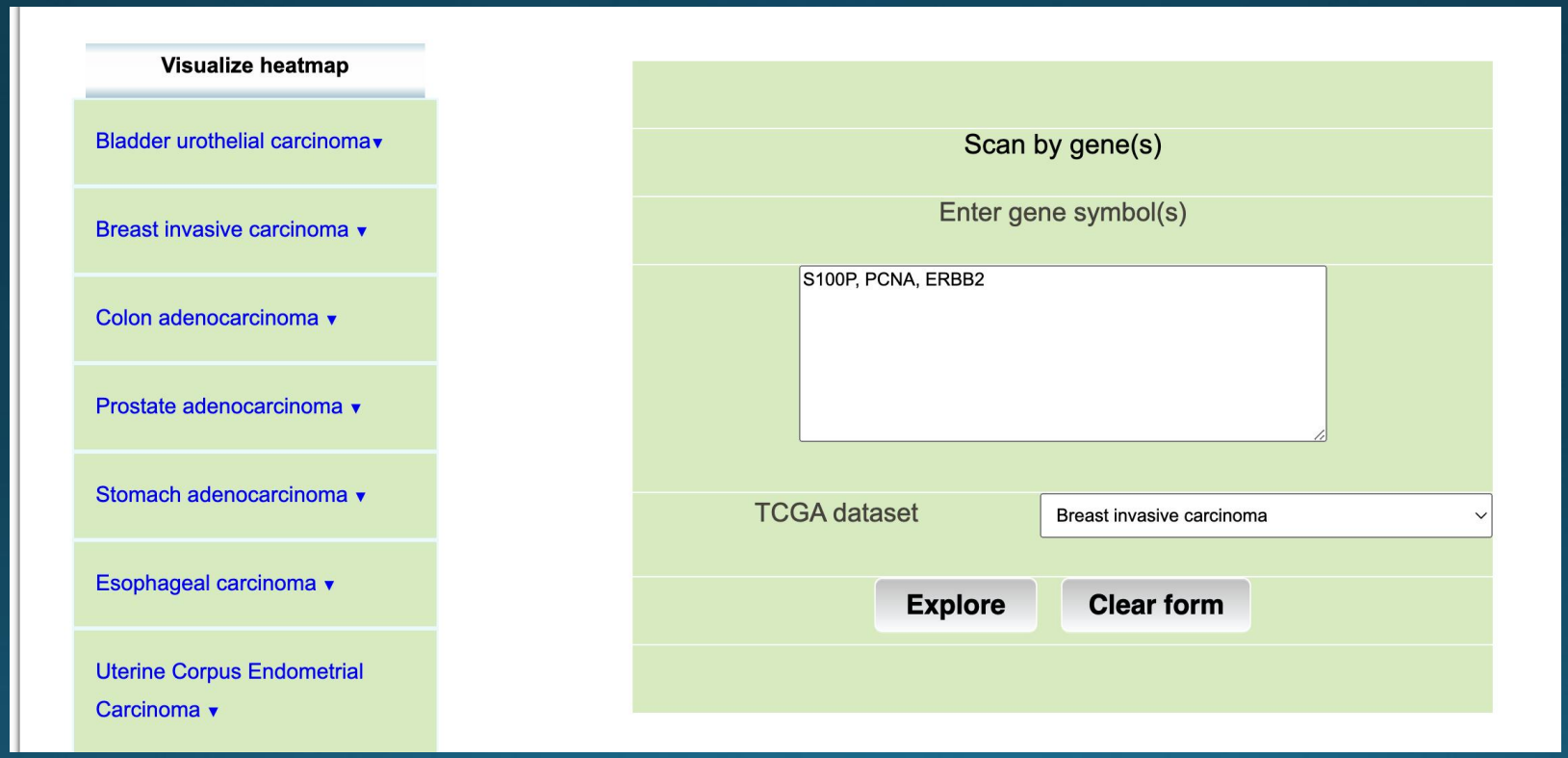
UALCAN Analysis Page :

a front portal of TCGA

- https://ualcan.path.uab.edu/analysis.html



UALCAN Analysis Page



UALCAN Analysis Page

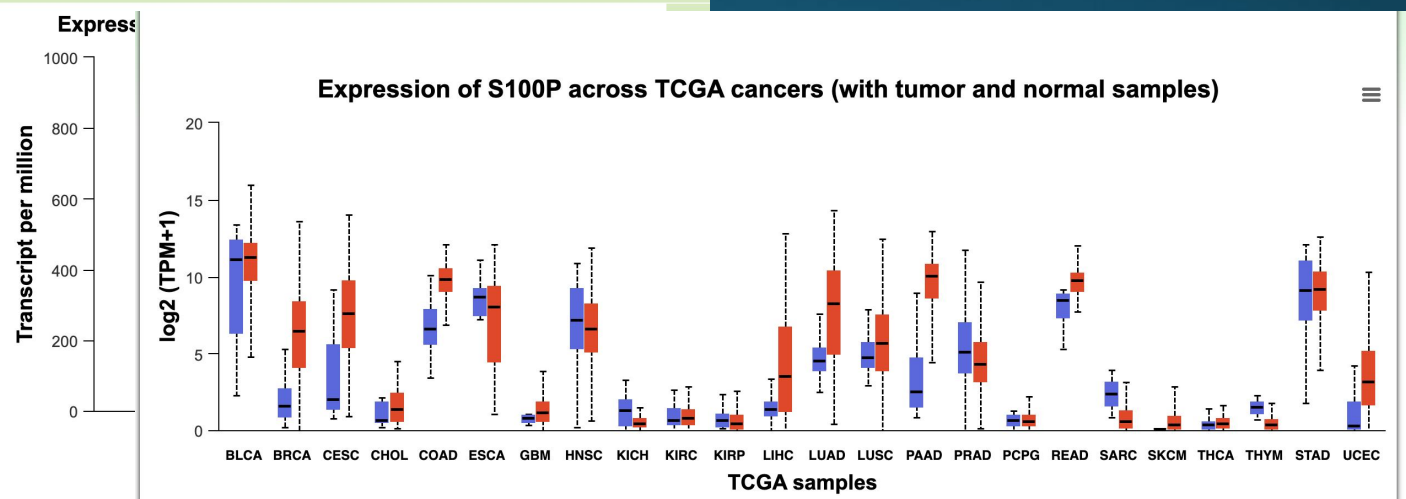
Show all gene expression in same page | Heatmap for query genes

Input gene	Links for analysis	External links
S100P	Expression, Survival, Methylation, Correlation, Pan-cancer view	HPRD, GeneCards, TargetScan, PubMed-Cancer, HumanProteinAtlas, OpenTargets, GTEx, iPATH, Drugbank
PCNA	Expression, Survival, Methylation, Correlation, Pan-cancer view	HPRD, GeneCards, TargetScan, PubMed-Cancer, HumanProteinAtlas, OpenTargets, GTEx, iPATH, Drugbank
ERBB2	Expression, Survival, Methylation, Correlation, Pan-cancer view	HPRD, GeneCards, TargetScan, PubMed-Cancer, HumanProteinAtlas, OpenTargets, GTEx, iPATH, Drugbank

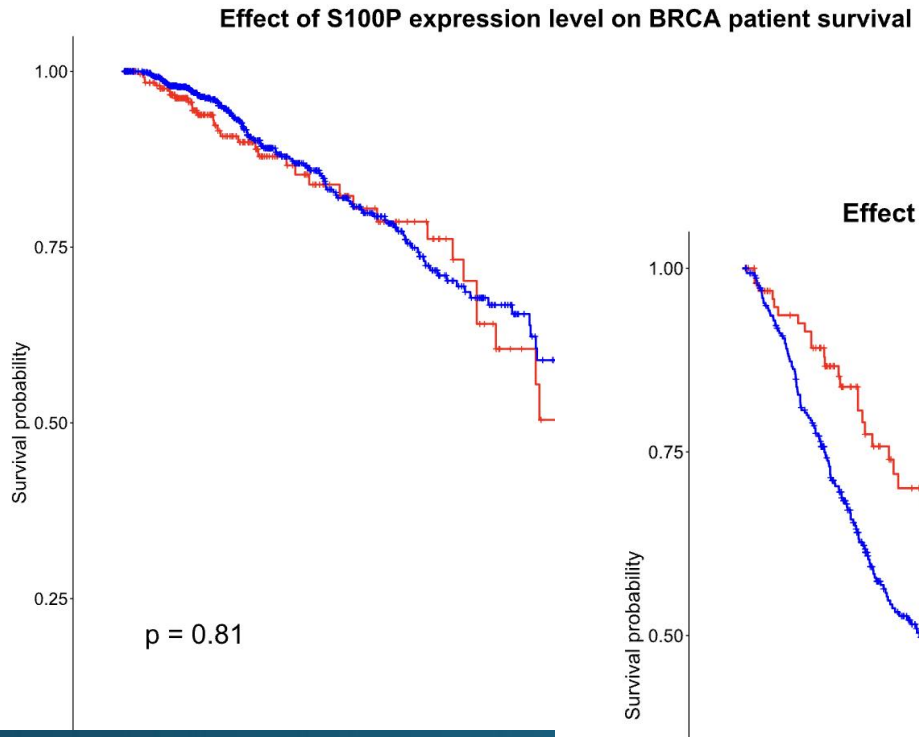
Select cancer

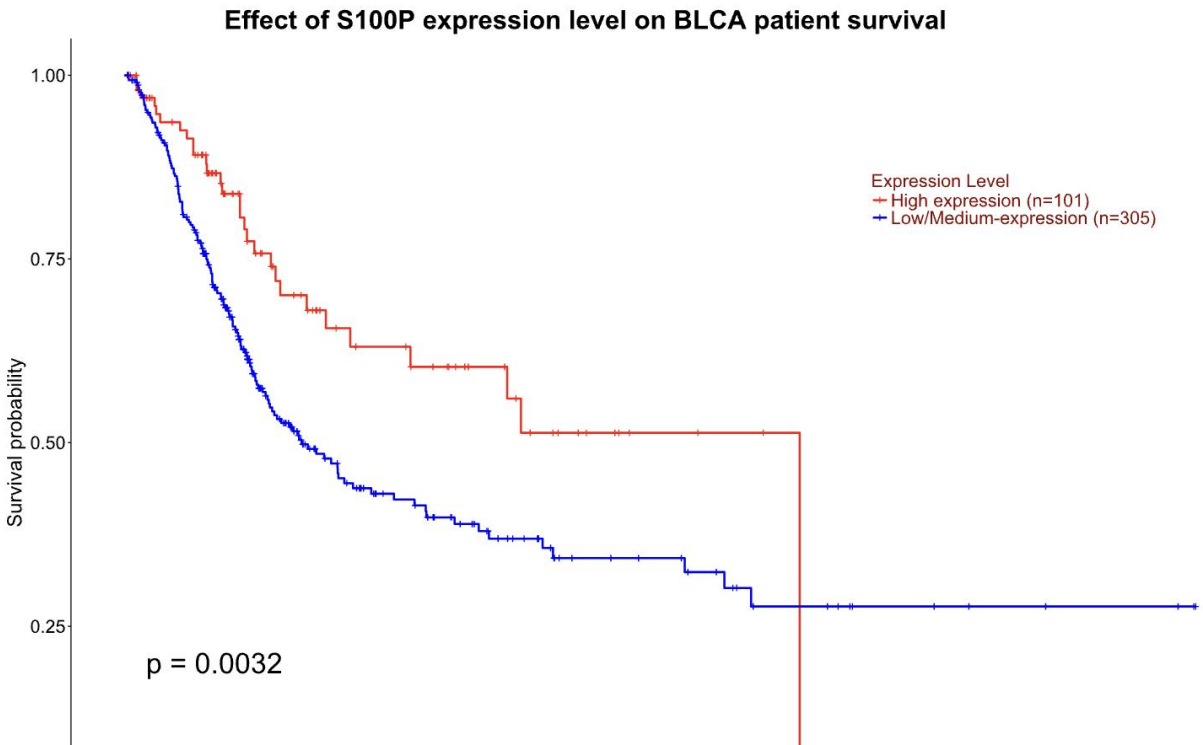
- Adrenocortical carcinoma
- Bladder urothelial carcinoma
- Brain lower grade glioma
- Breast invasive carcinoma
- Metastatic breast cancer
- Cervical squamous cell carcinoma
- Cholangiocarcinoma
- Colon adenocarcinoma

S100P expression based on Sample types



UALCAN Analysis Page





UALCAN Analysis Page

Genes positively correlated with S100P in BRCA

Gene	Pearson-CC	Visualize	Links	
SPTBN2	0.57	Show plot	GEx Profile	Survival Profile
C11orf80	0.43	Show plot	GEx Profile	Survival Profile
PRODH	0.36	Show plot	GEx Profile	Survival Profile
FXYD3	0.36	Show plot	GEx Profile	Survival Profile
SLC26A2	0.34	Show plot	GEx Profile	Survival Profile

UALCAN Analysis Page

DRUG RELATIONS

Drug Relations

Show 10 entries

Search

DRUGBANK ID	NAME	DRUG GROUP	PHARMACOLOGICAL ACTION?	ACTIONS	DETAILS
DB01003	Cromoglicic acid	approved	unknown	antagonist	Details

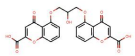
Showing 1 to 1 of 1 entries

1

- Identification
- Pharmacology
- Interactions
- Products
- Categories
- Chemical Identifiers
- References
- Clinical Trials
- Pharmacoeconomics
- Properties
- Spectra
- Targets (1)

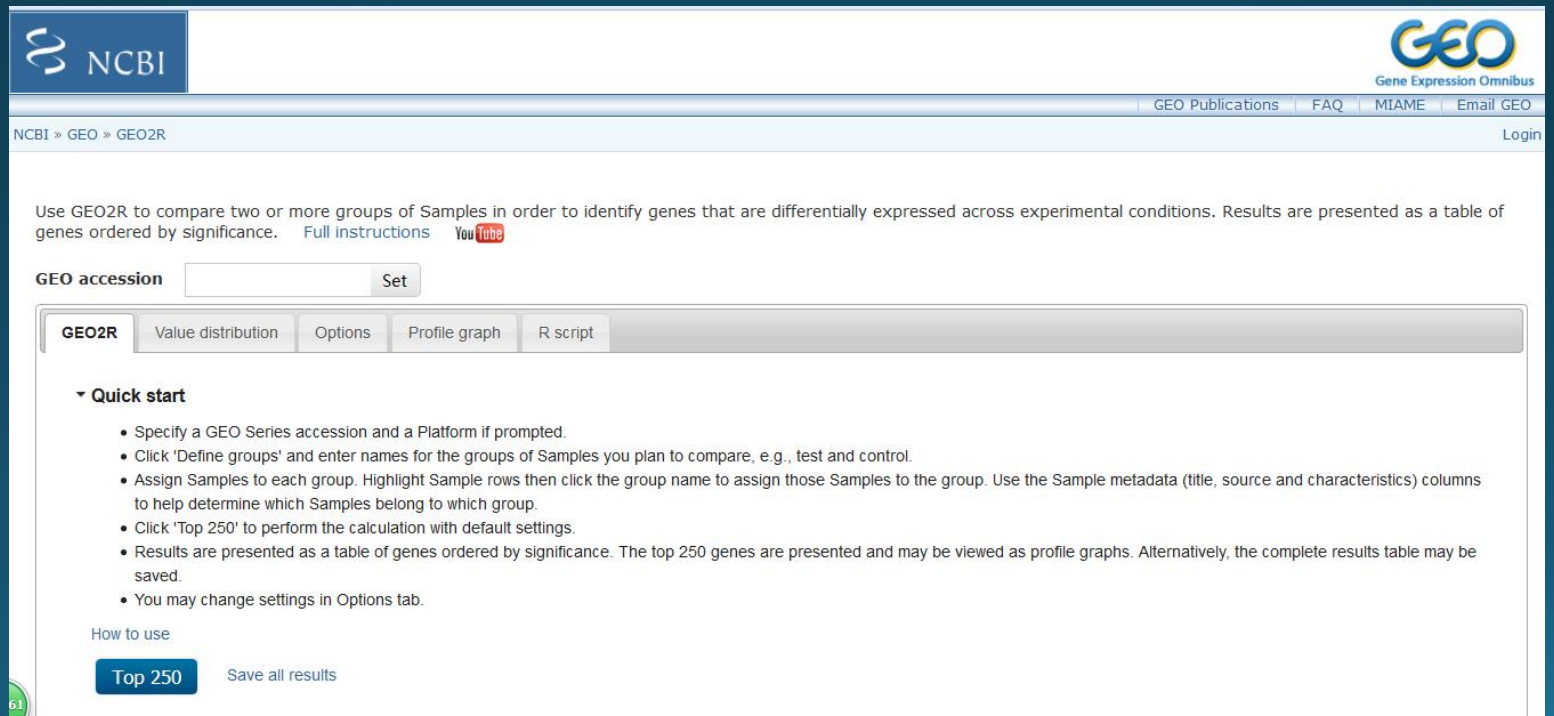
Cromoglicic acid

Star

Summary	**Cromoglicic acid** is a medication used to treat asthma, allergic reactions of the eyes and nose, as well as other mast cell reactions.		
Brand Names	*Gastrocrom, Nalcrom, Nasalcrom*		
Generic Name	Cromoglicic acid	**DrugBank Accession Number**	DB01003
Background	A chromone complex that acts by inhibiting the release of chemical mediators from sensitized mast cells. It is used in the prophylactic treatment of both allergic and exercise-induced asthma, but does not affect an established asthmatic attack.		
Type	Small Molecule	**Groups**	Approved
Structure	 3D Download Similar Structures	**Weight**	Average: 468.3665 Monoisotopic: 468.069261354
		Chemical Formula	$C_{23}H_{16}O_{11}$
Synonyms Show All Synonyms	Acide Cromoglicique, Acido Cromoglicico, Acidum Cromoglicicum, Cromoglicate, Cromoglicic acid, Cromoglycate, Cromoglycic acid, Cromolyn		

Differential Gene Detection using GEO2r

- GEO2r is an on-line tool for detection of differentially expressed genes between two sets of given samples
- http://www.ncbi.nlm.nih.gov/geo/geo2r/



Pathway Enrichment

- One can examine if a set of up-regulated (or down-regulated) genes statistically enrich a specific pathway

- The basic idea: consider a specific pathway P with K genes out of the 20,000 genes encoded in the human genome, and a set M of up-regulated genes in cancer *versus* controls. We consider P is enriched by the up-regulated genes if

$$|M \cap P| / |M| >> K/20{,}000$$

- There are fancier ways to more accurately assess the level of "enrichment" such as Kolmogorov–Smirnov statistic

Biological Pathways and Networks

- Metabolic pathway: a series of enzymatic reactions that produce a specific product

- Regulatory networks: pathways that regulate a cell's behaviors, including transcription, translation, degradation, motility, ...

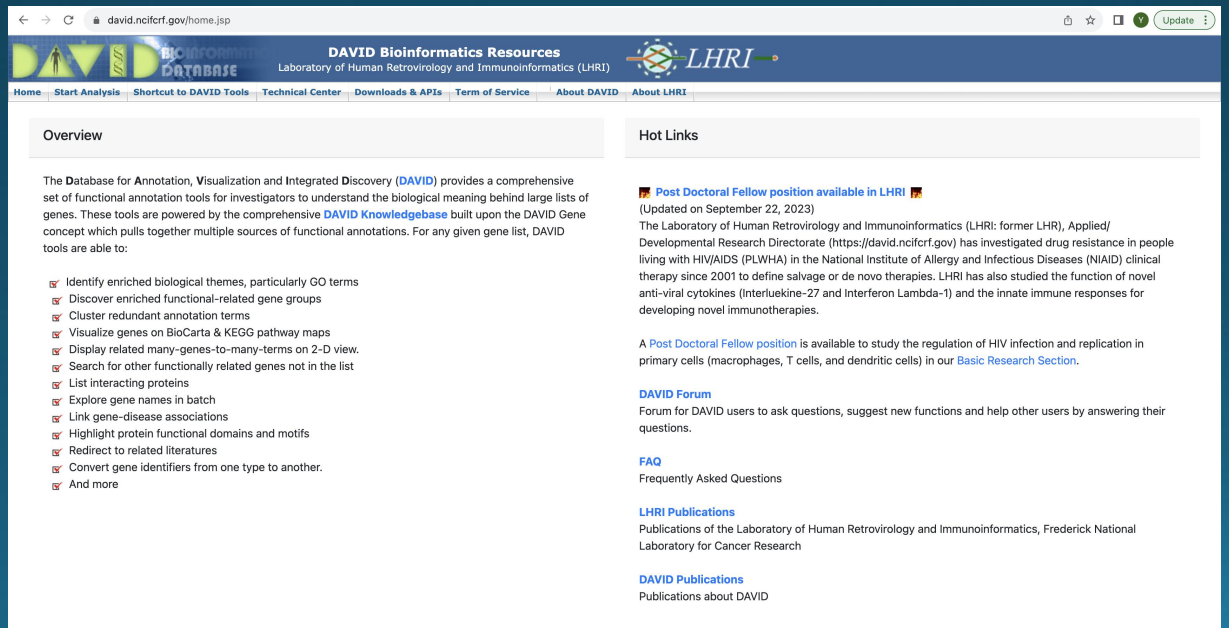
- Signal transduction pathway and networks: cellular processes that recognize extra- or intra-cellular signals and induce appropriate cellular responses

Widely Used Pathway Databases

- Gene Ontology: https://www.geneontology.org/
- KEGG: https://www.genome.jp/kegg/
- BioCyc: https://www.biocyc.org/
- Reactome: https://reactome.org/

Pathway Enrichment Analysis

- DAVID (https://david.ncifcrf.gov/home.jsp) is a popular tool that can inform which pathways in KEGG, REACTOME or other pathway databases are enriched by up- or down-regulated genes using a statistical approach
- … hence providing a way to organize gene-level data to pathway level information and helping to simplify data analysis



Pathway Enrichment Analysis

- Step 1: click "Start analysis"
- Step 2: paste a gene list onto "Paste a List" under "Upload"
- Step 3: select "OFFICIAL_GENE_SYMBOL"
- Step 4: select "gene list"
- Step 5: click on "submit"
- Step 6: answer "OK in the popup window
- Step 7: select "Homo sapiens" as the background
- Step 8: select " Functional Annotation Chart"
- Step 9: select "Pathway"

Activity Levels of Pathways

- Each (metabolic) pathway has one rate-limiting enzyme, whose gene-expression changes can reflect the overall activity level change of the pathway
- E.g., the rate-limiting enzyme of glycolysis is PFKL; hence this gene can be used as the "signature" of the pathway
- This is true virtually for all pathways or more generally "activities" such as various types of stresses

Signature Genes of Cellular States

- Hypoxia: HIF genes
- ROS: a combination of multiple ROS related genes
- Oxidative stress:
- Different types of inflammation: various cytokines and associated proteins
- Lactic acidity:
- ER stress:
- Mitochondrial stress:
- .....

Quantitative Relationships

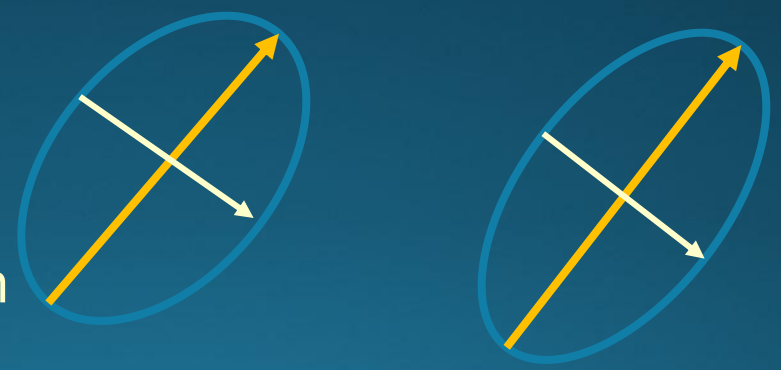
- We can use the expressions of a group of genes to reflect the levels of cellular states, called signature genes
  - Hypoxia, alkalosis, ATP level ..

- Hence cellular states, the levels of pathway activities, gene expressions can be naturally linked through statistical analyses

Pathway/Activity Level Association

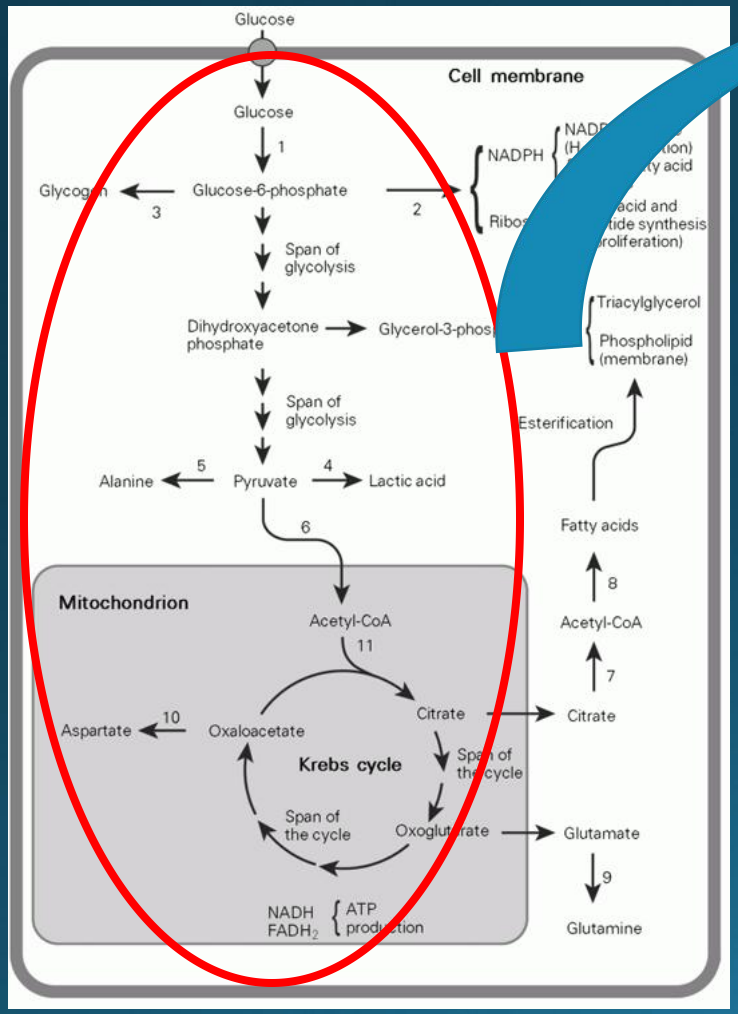
- The essence is: How to compare to two datasets in terms of their similarity?
- There are techniques to compare two datasets in terms of their major axes
- Principle component analyses
- Using such techniques, one can compare gene sets just like individual genes to infer association and even causal relations

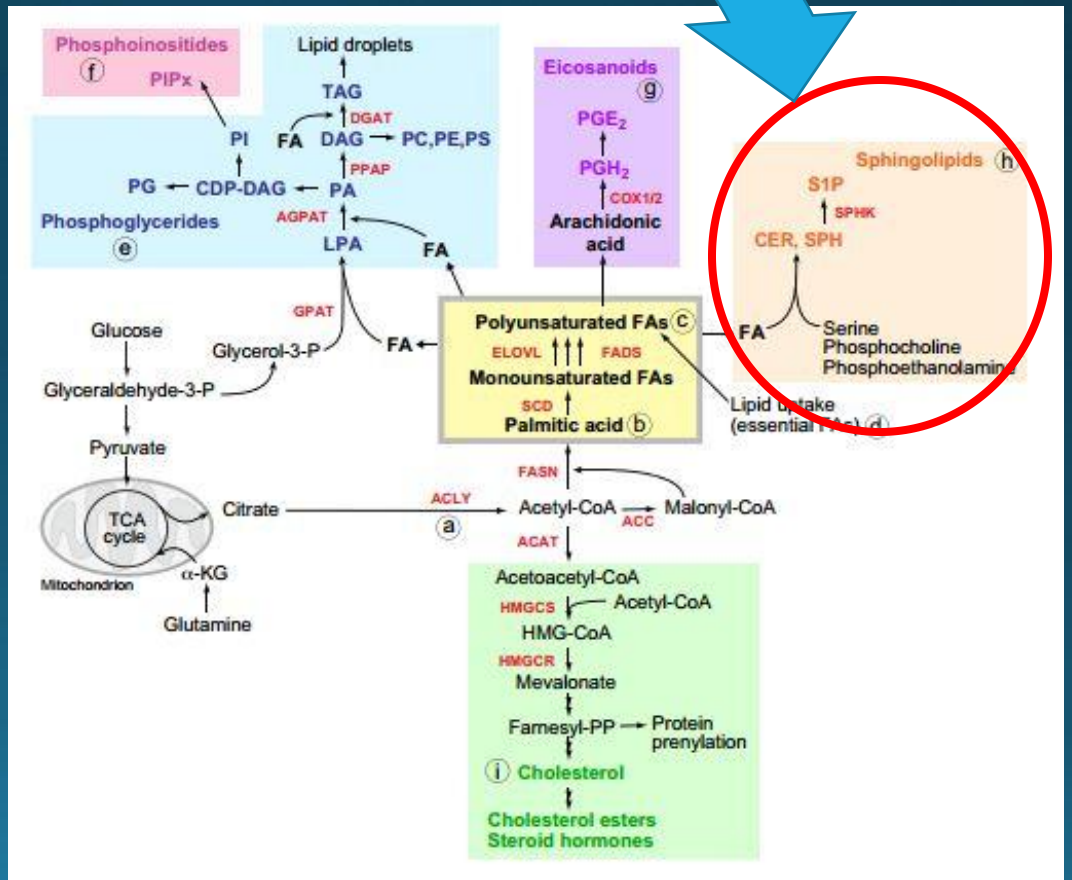
Not strongly related



Strongly related

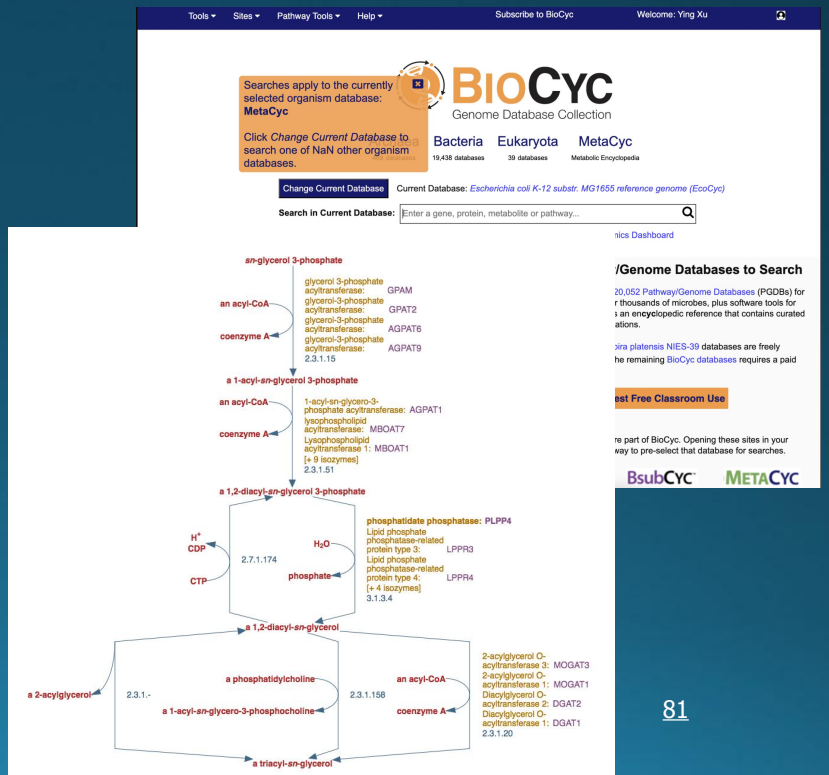
Inference of Functional Relationships Among Pathways





Causal Relationship

- Underlying the metabolic networks are chemical reactions catalyzed by enzymes
  - BioCyc is one such database for metabolic networks and the underlying chemical reactions
- Chemical reactions provide a natural direction for activities that are statically related
- Deep learning-based causal inference analyses



Take-Home Message

- Large quantities of cancer omic data may contain possibly all the information regarding
  - The origin of a cancer
  - The reasons for similar behaviors of different cancer types
  - The reasons for distinct properties of individual cancers
  - Hints about how we can possibly design more effect approaches to detect and treat cancer
- It takes some guidance and techniques to uncover all the information hidden in the omic data

Homework

- Reading Chapters 1, 2, 3 of the textbook
- Reading "Hallmarks of cancer" (2000) , "Hallmarks of cancer: the next generation" (2011) by Hanahan and Weinberg
- Reading "Hallmarks of cancer: new dimensions" (2022), Hanahan
- Reading Stehelin D, Varmus HE, Bishop JM and Vogt PK . (1976b). *Nature*, **260**, 170–173.