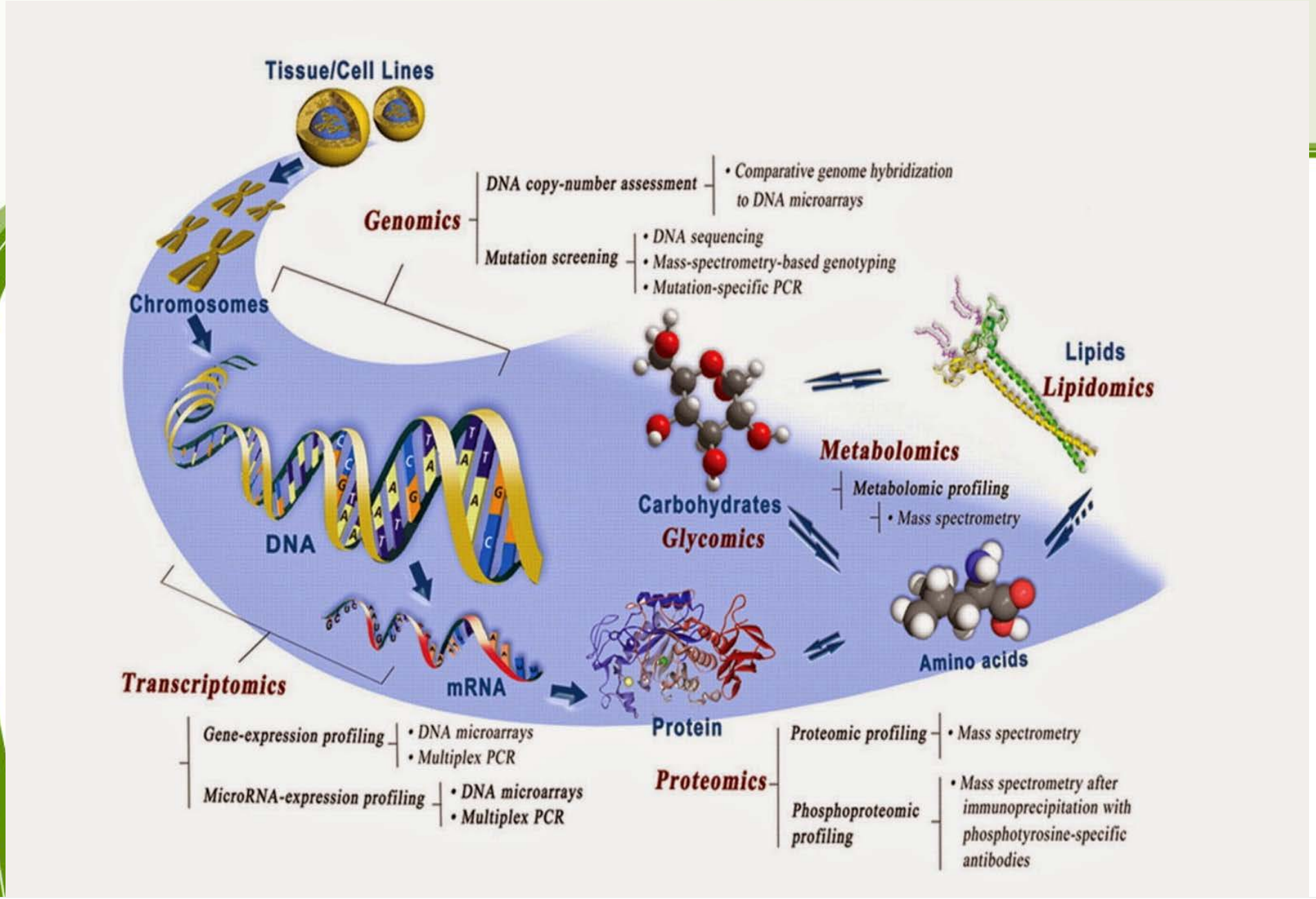
Omics technologies:

genomics, transcriptomics, metabolomics, databases, personalized medicine and big data

Vendula Pospíchalová, PhD
(pospich@sci.muni.cz)
Department of Experimental Biology
Animal Physiology and Immunology

Bi5599 Applied Biochemistry and Cell Biology Methods
2021-11-08



Schematic representation of omics technologies, their corresponding analysis targets, and assessment methods. Taken from Wu RD et al. JDR 2011; 90:561-572.

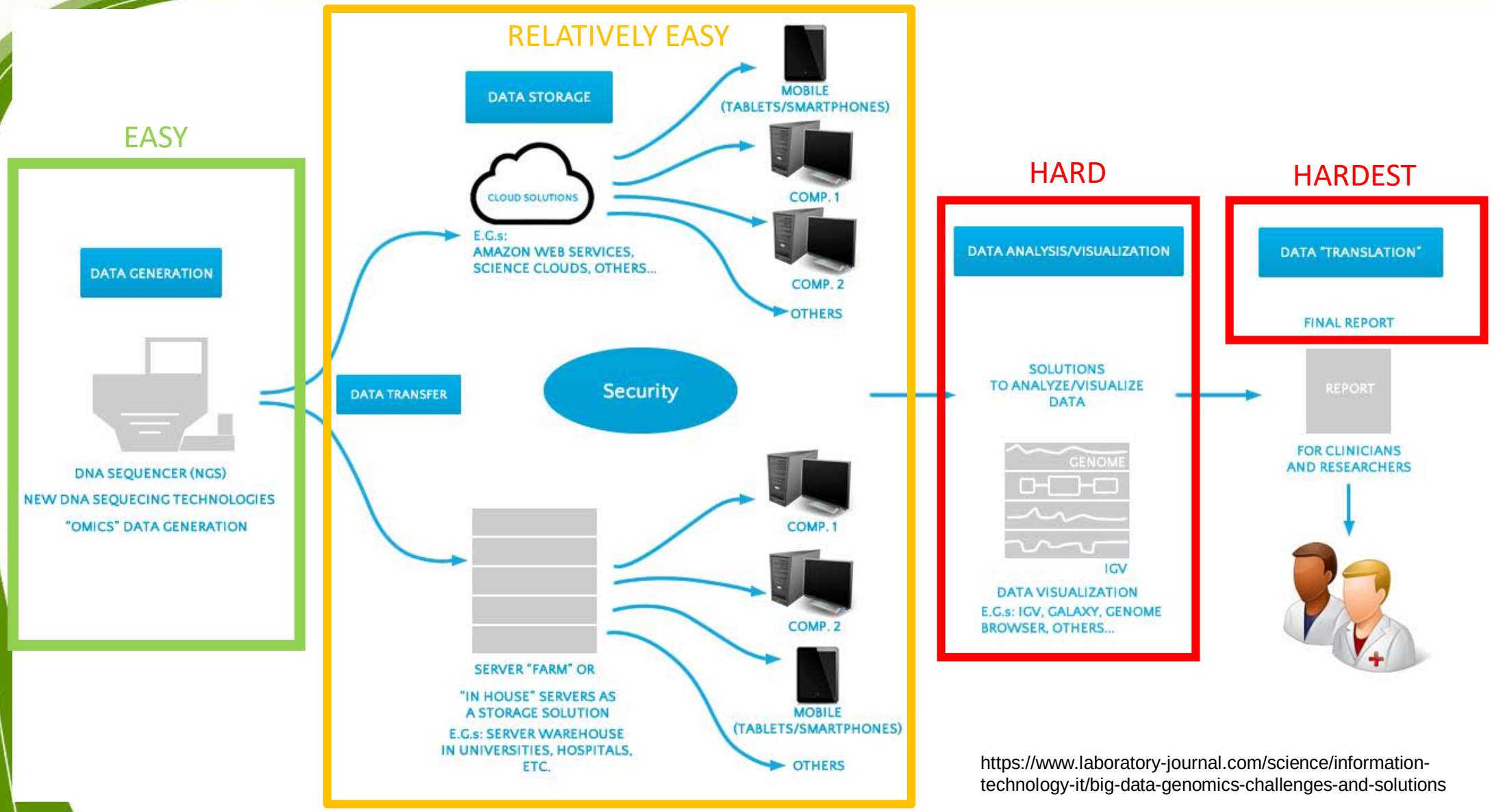
Contents

1. Introduction: what are –omics technologies + history
2. How does big data look and how to approach it
3. From –omics technologies to biomarkers and personalized medicine
4. Genomics: genomes vs exomes vs genotypes + DTC service
5. Cancer databases: COSMIC, TCGA and others
6. Transcriptomics: microarrays vs. RNA sequencing
7. Gene set analysis
8. Metabolomics
9. Cutting edge: single cell –omics and single cell multi-omics
10. Summary and take home messages

What are „-omics“ technologies

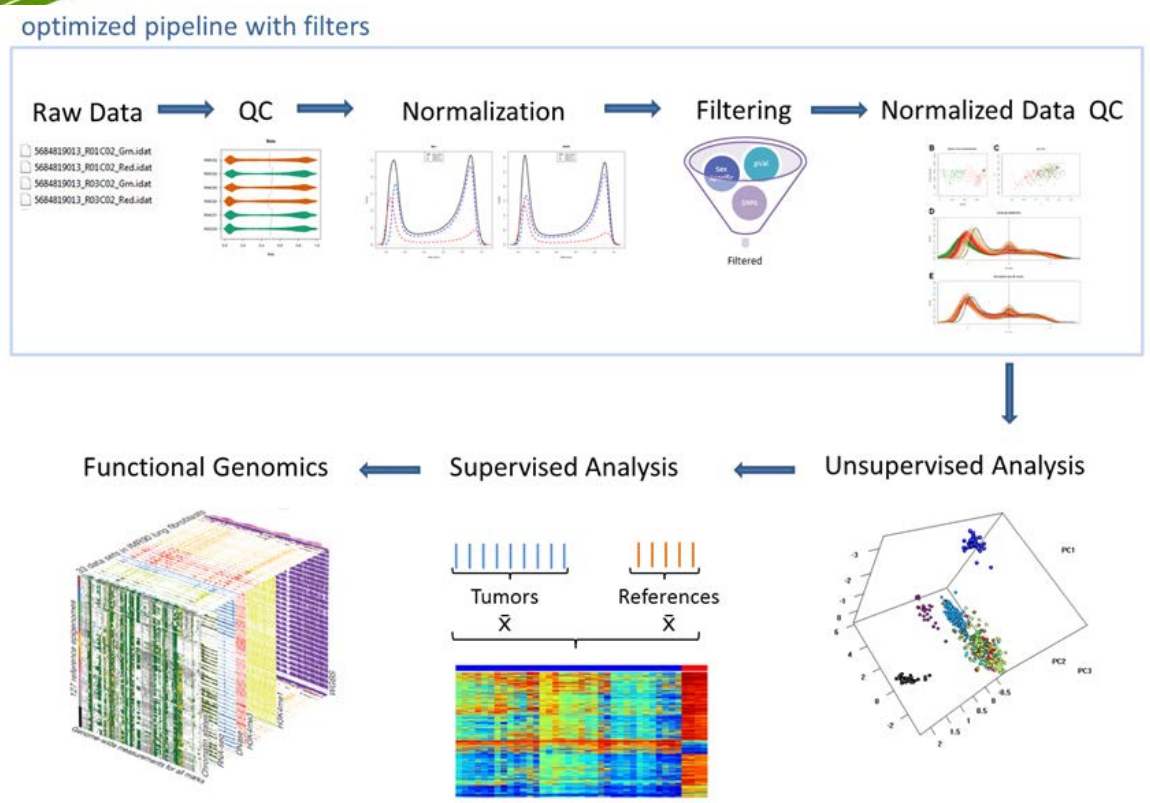
- **Omics** refers to a field of study in biology ending in **-omics**, such as genomics, proteomics or metabolomics
- The related suffix **-ome** is used to address the objects of study of such fields, such as the genome, proteome or metabolome
- **-ome** = many/collectivity or whole/all/complete in Greek
- **-omics** = study of large sets of biomolecules
- **High-throughput experimental technologies** characterized by **automation, miniaturized assays** and **large-scale data analysis**
- Analytic part of the experiment is usually much longer than the experiment itself – bioinformatics skills needed
- Raw data is the „gem“ but usually is in user unfriendly format
- **Interpreting** functional consequences of millions of discovered events is one of **the biggest challenges**

Big –omics data challenges



https://www.laboratory-journal.com/science/information-technology-it/big-data-genomics-challenges-and-solutions

Only skilled bioinformaticians can process raw data



ARTICLE

doi:10.1038/nature10166

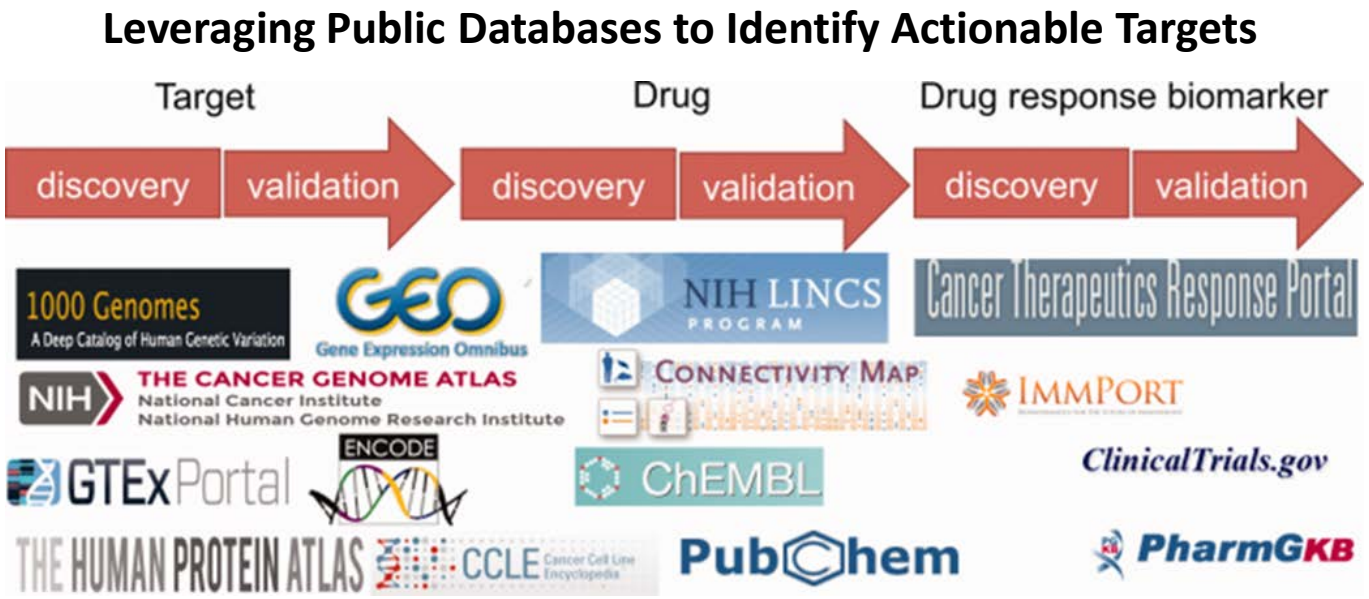
Integrated genomic analyses of ovarian carcinoma

The Cancer Genome Atlas Research Network*

Integrated analyses of –omics studies are only possible in large consortia (*e.g.* TCGA). Subsequently, the authorlist of such articles can more than 2 pages long with substaintial part of the authors being bioinformaticians. Among the reviewers, bioinformaticians are also necessary, *etc.*

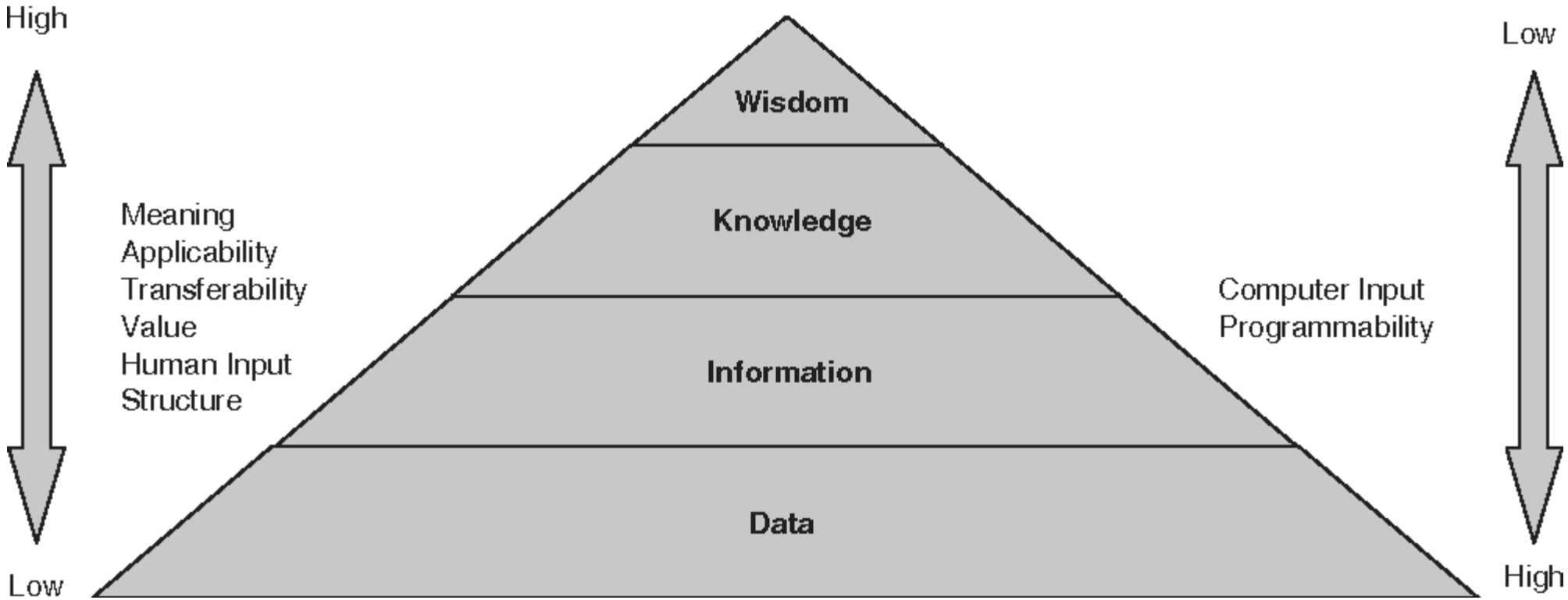
Data sharing policy

- The concepts of data sharing and open data are becoming increasingly important in science
- Funding bodies, journals and societies are now encouraging or mandating data sharing (usually the raw data)
- Sharing data publicly is an important way of improving reproducibility and showing that researchers are confident in their work
- Studies with raw data shared in a repository also receive more citations than those without publicly available data
- But raw -omics data
  are hard to analyse, so many platforms gather the publicly available data, thoroughly analyze it, curate it and share it in a user friendly format



DIKW pyramide

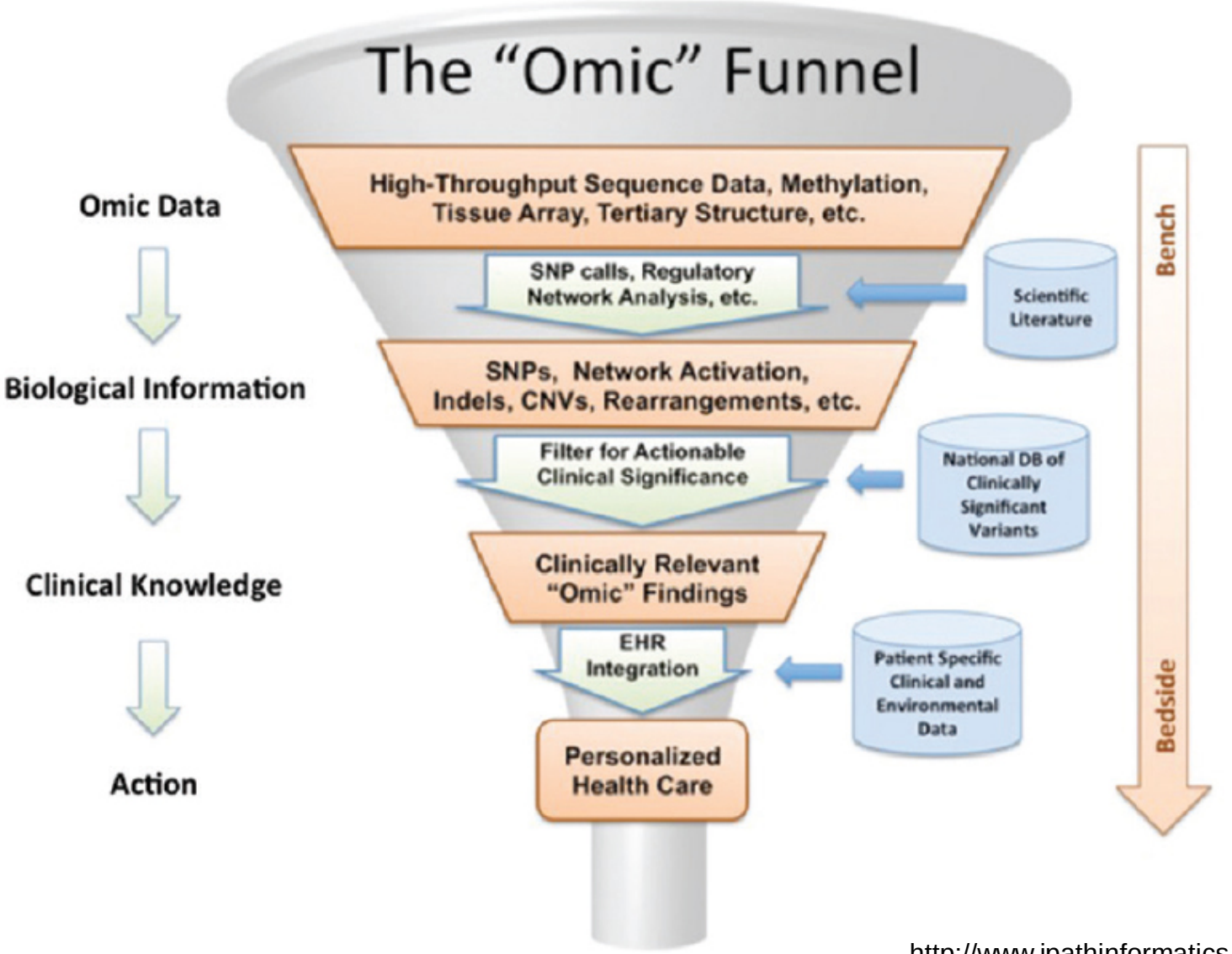
„Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.“ – Clifford Stoll



Jennifer RowleyPublished 2007 in J. Information Science DOI:10.1177/0165551506070706

What is the aim of OMICS technologies

In Genomics:
Sequence of 3 billions letters in .txt file
↓
information where the individual's genome varies from reference sequence
↓
CYP2C9 or *TPMT* genotype, which has known pharmacogenomic associations
↓
individualize the dose of a new warfarin prescription



The DIKW pyramid metaphor:

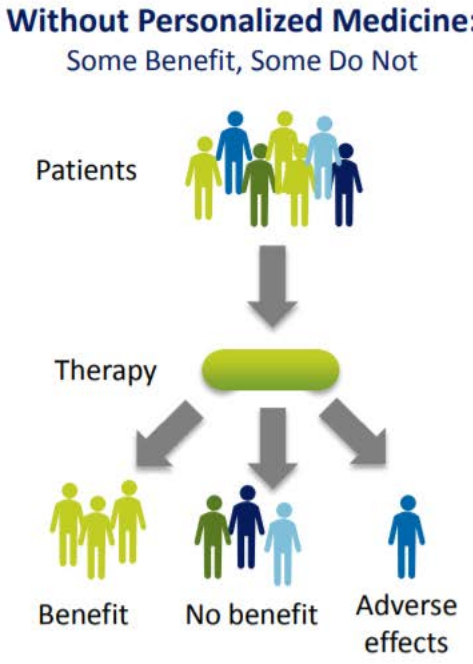
"know-nothing"
(Data)
↓
"know-what"
(Information)
↓
"know-how"
(Knowledge)
↓
"Know-why"
(Wisdom)

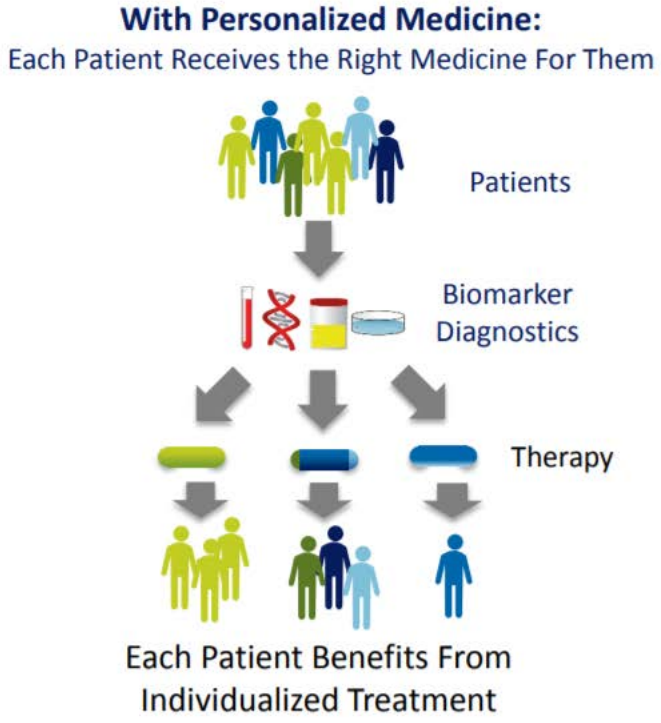
Zeleny (2005)

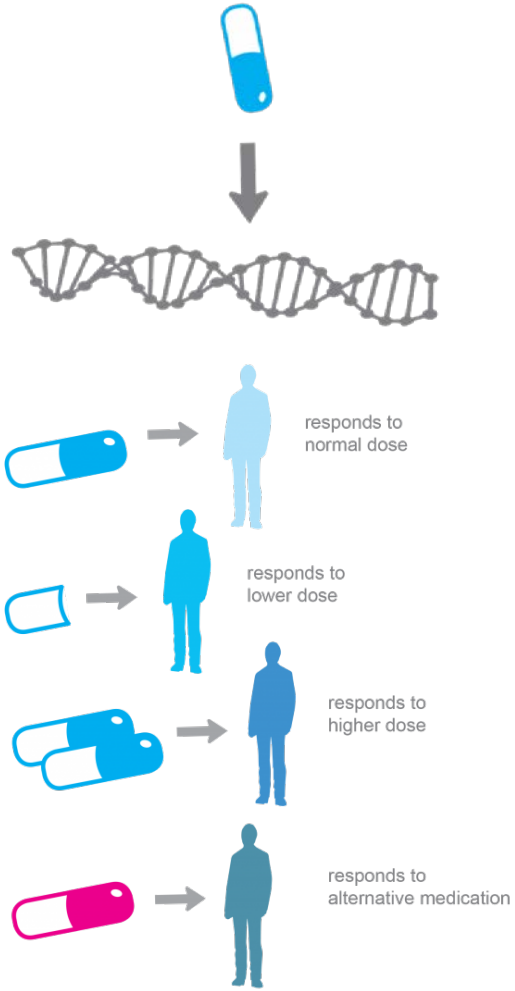
http://www.jpathinformatics.org/text.asp?2015/6/1/46/163985

What is personalized health care?

Personalized medicine, sometimes referred to as *precision* or *individualized* medicine, is an emerging field of medicine that uses diagnostic tools to identify specific biological markers, often genetic, to help assess which medical treatments and procedures will be best for each patient.

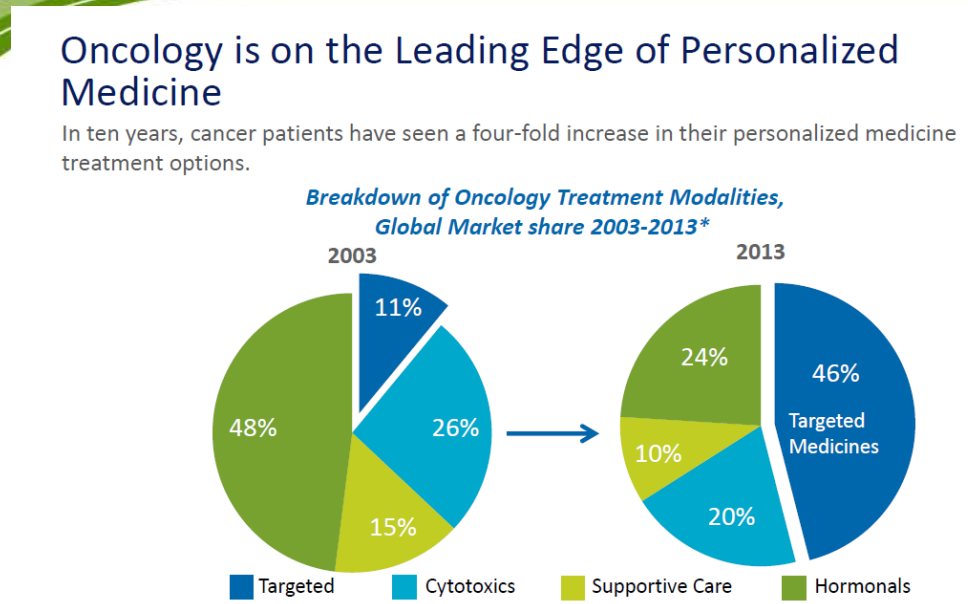


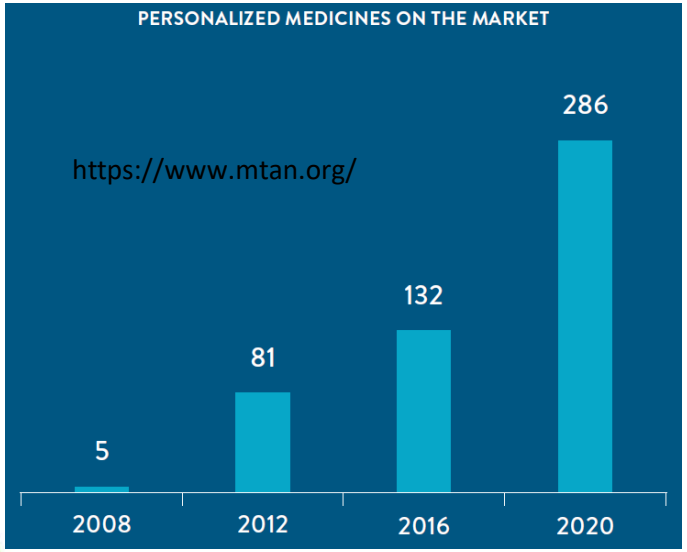


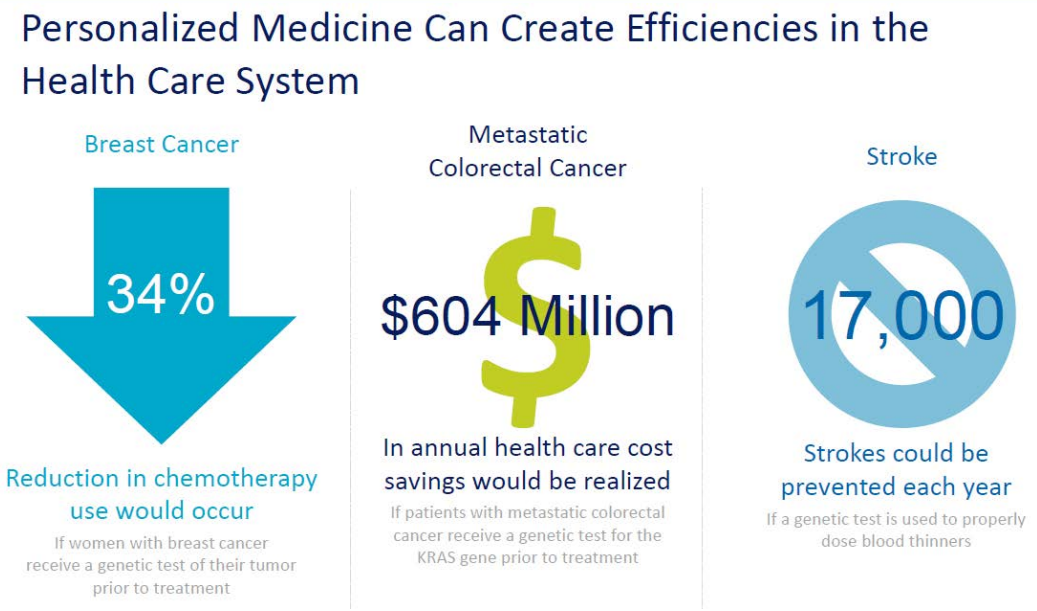


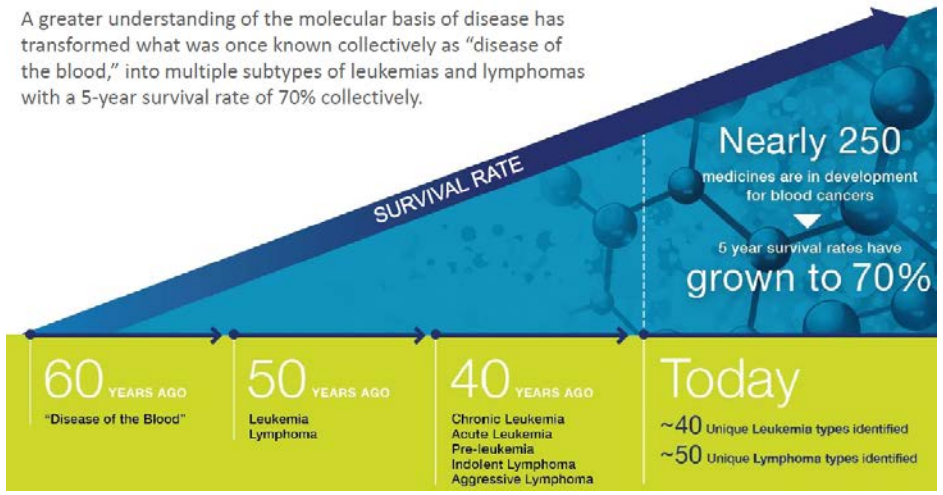
https://pharma.bayer.com/en/research-and-development/research-focus/oncology/personalized-medicine/index.php

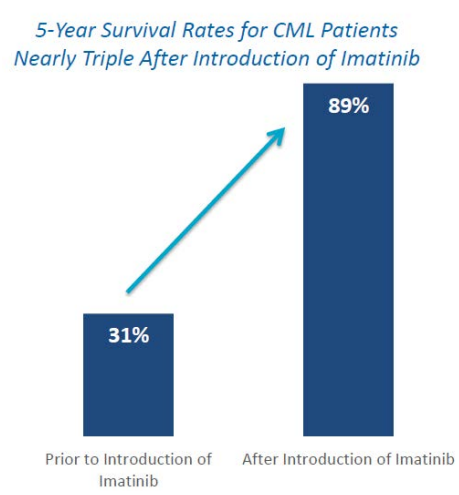
Value of personalized medicine









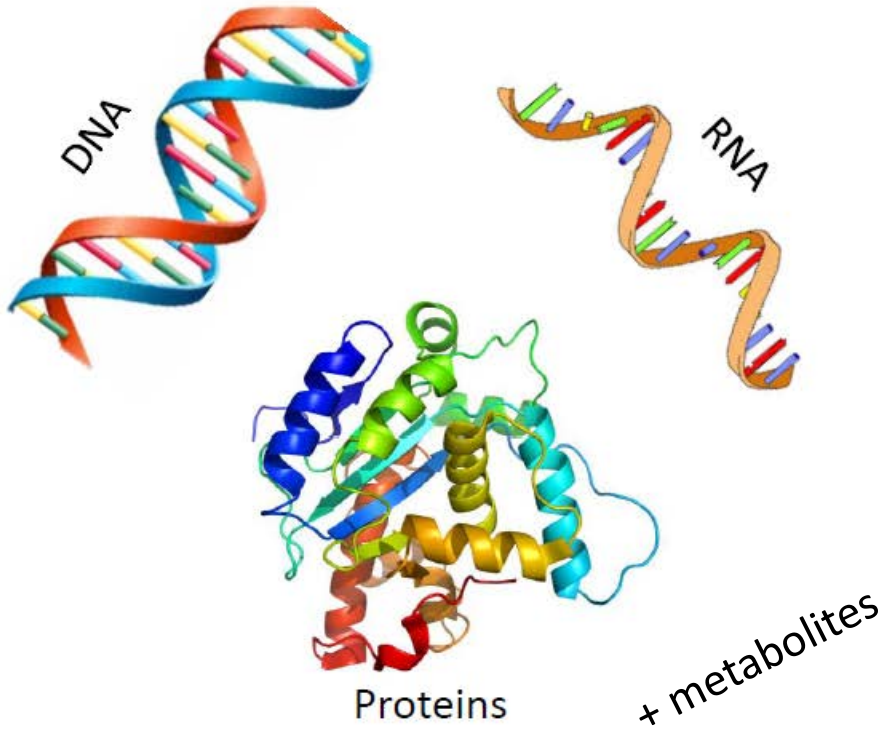


And many more examples, see http://www.personalizedmedicinecoalition.org for more detailed information on PM

What is a biomarker?

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, disease processes, or biological responses to a therapeutic intervention. Biomarkers can be used to reduce uncertainty and guide clinical care.

Molecular Biomarkers Can Include:



Biomarkers Help Inform Medical Decisions:

→ Prevention measures?
→ Which diagnosis?
→ Treat or don't treat?
→ What dose?

How Do You Detect a Biomarker?

- Diagnostics
  - Blood draw
  - Microscopic analysis
  - Gene sequencing
  - Biopsy
  - Protein analysis

Source: National Cancer Institute, "NCI Dictionary of Cancer Terms" (accessed M

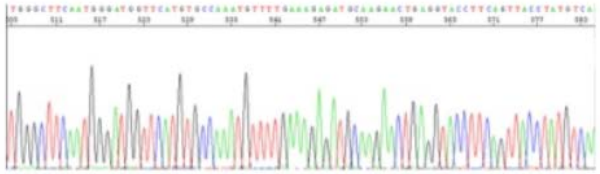
-OMICS technologies and their integration is crucial for biomarker discovery and validation

History of „-omics“ technologies

- Genome – central part of all –omics technologies
- NGS = next generation sequencing

Sanger VS NGS

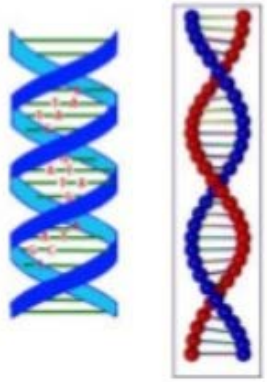
	Bases	Genes
Human Genome	3.3×10^9	~20,000



Sequencing of the human genome using Sanger technology took more than a decade and cost an estimated $70 million dollars

In 3 days (one run), Illumina HiSeq 4000 is able to produce $1,680 \times 10^9$ bases for ~$32,000

Brief History of DNA Sequencing



1953: Discovery of DNA structure by Watson and Crick

1973: First sequence of 24 bases published

1977: Sanger sequencing method published

1982: GenBank started

1987: 1st automated sequencer: Applied Biosystems Prism 373 (up to 600 bases)

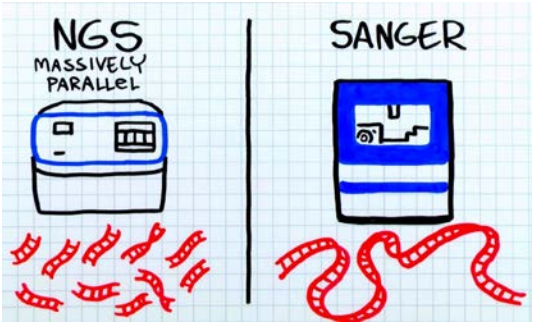
1996: First Capillary sequencer: ABI310

2000-2003: Human Genome Sequenced

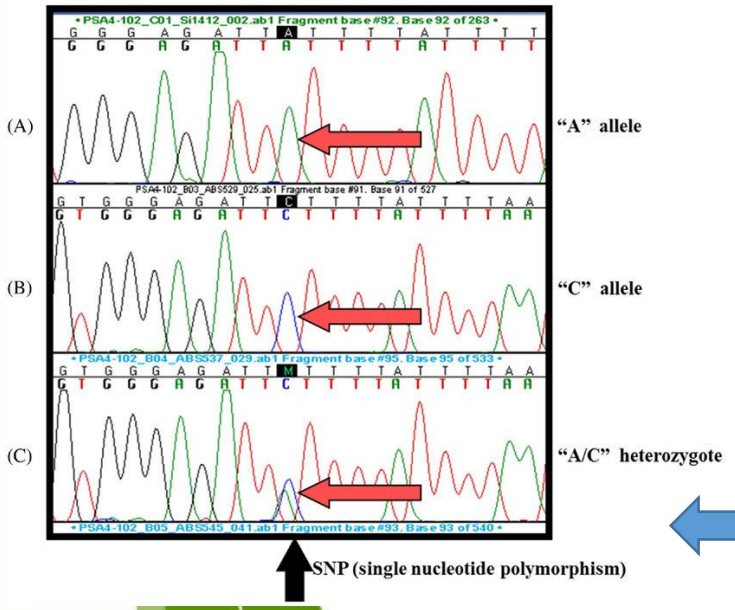
2005- : First NGS sequencers 454 Life Sciences, Solexa/Illumina, Helicos, Ion Torrent

Sanger vs next generation sequencing

- Sanger sequencing
- https://www.youtube.com/watch?v=e2G5zx-OJIw



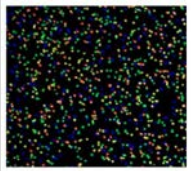
- Next generation sequencing (Illumina is shown as an example)
- https://www.youtube.com/watch?v=9YxExTSwgPM



Sanger Sequencing

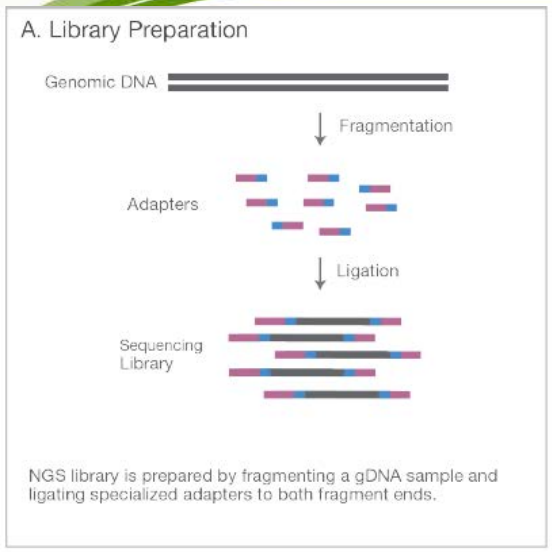
Advantages	Disadvantages
Lowest error rate (1.5%)	High cost per base
Long read length (~750 bp)	Long time to generate data
Can target a primer	Need for cloning
Used to confirm NGS results	Amount of data per run
Seeing is believing	

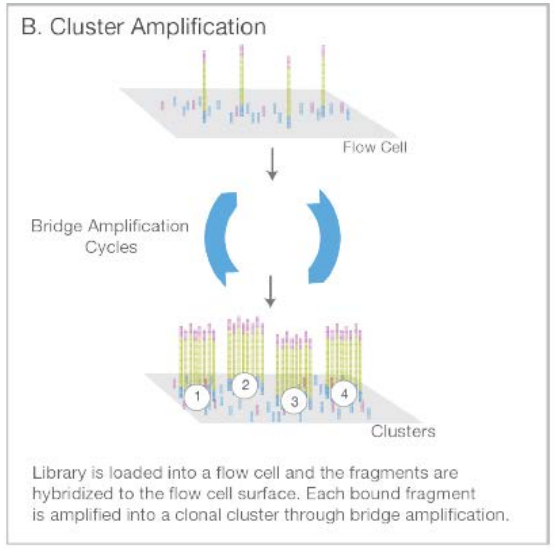
Illumina Sequencing

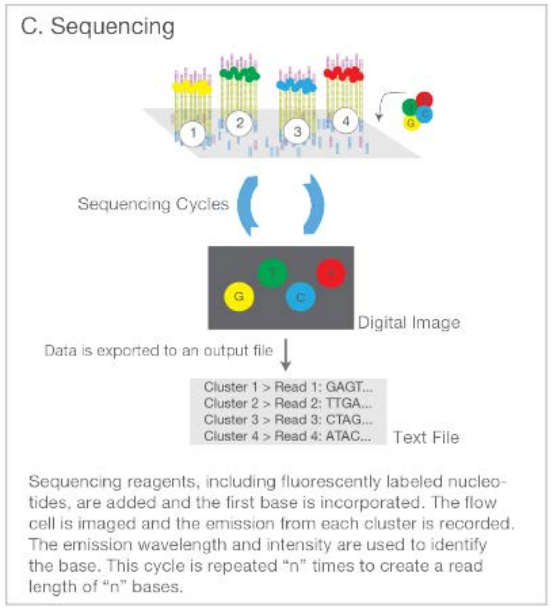
Advantages	Disadvantages
Low error rate	Must run at very large scale
Lowest cost per base	Short read length (50-75 bp)
Tons of data	Runs take multiple days
 An image of hundreds of extended molecules	High startup costs
	De Novo assembly difficult

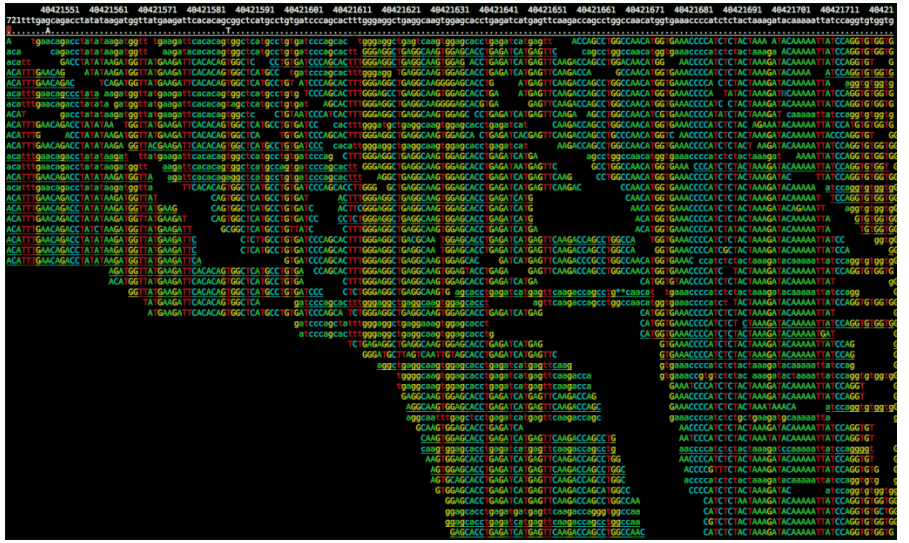
https://slideplayer.com/slide/5799907/

Illumina NGS overview

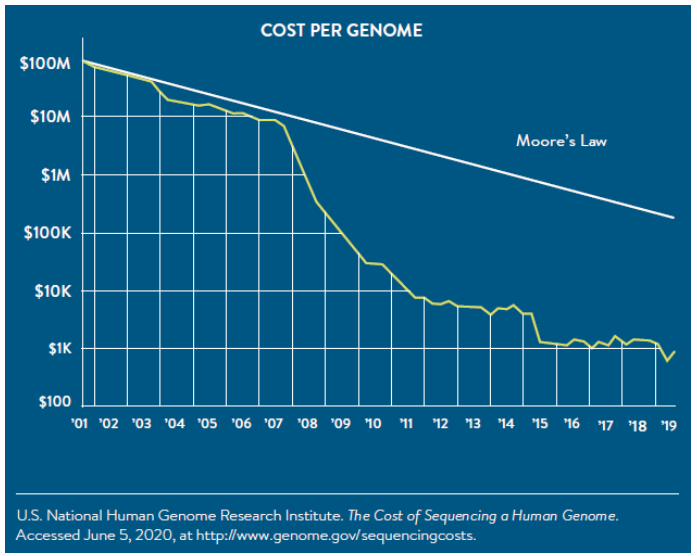




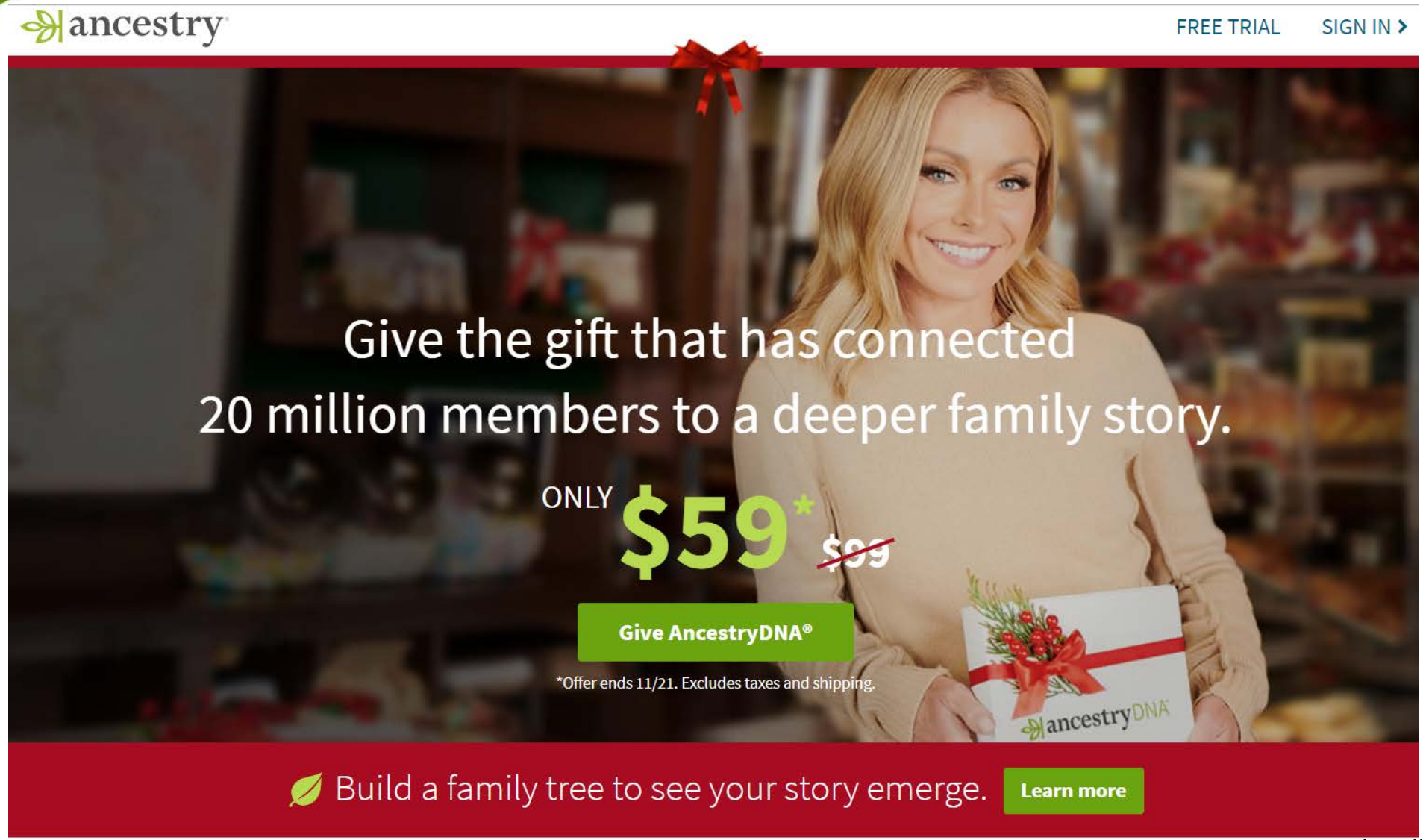




- www.illumina.com/technology/next-generation-sequencing.html



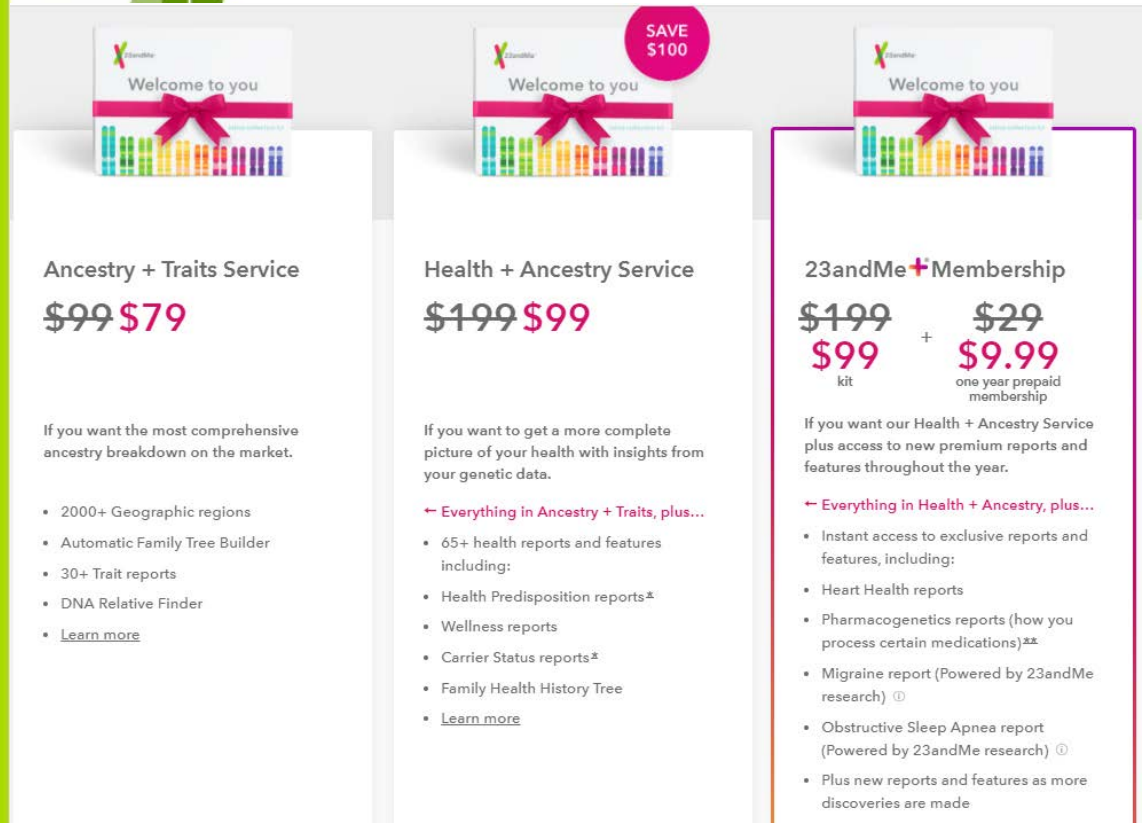
DTC (direct-to-customer) genetic testing



http://ancestry.com/

Genotyping vs Sequencing

- **Genotyping** - determining which genetic variants an individual possesses through a variety of different methods, especially genotyping chips (based mostly on SNPs – single nucleotide polymorphisms)
  - cheap, but require prior identification of the variants of interest

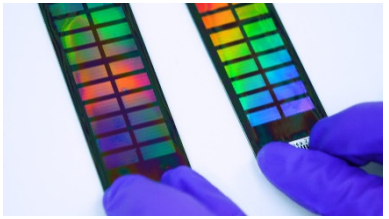


Methods

We use genotyping technology to look at specific genetic variants in the genome that can be most informative about an individual's health and ancestry.

Unlike sequencing which analyses all nucleotides in a gene to identify changes, genotyping detects specific known variants within the genome. 23andMe uses a custom Illumina HumanOmniExpress-24 format chip that analyses approximately half a million variants. This custom chip has been designed to include variants:

- In medically relevant genes
- Involved in drug metabolism, efficacy and side effects
- With known disease associations
- Associated with traits
- Used to assign genetic ancestry and ethnicity

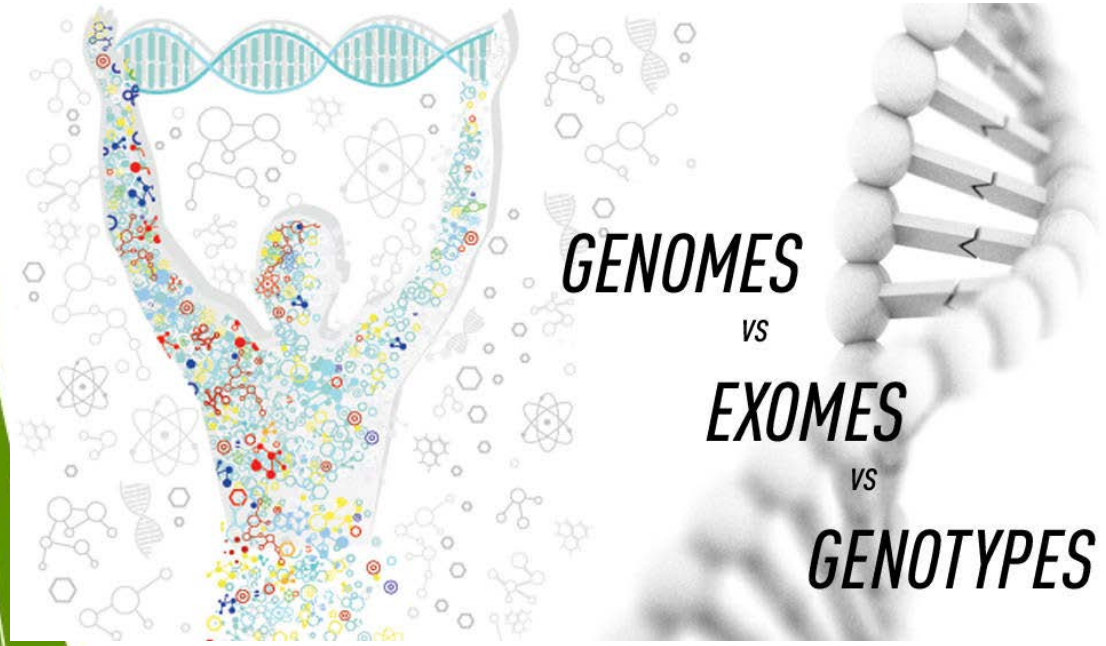


https://www.23andme.com/

How SNP genotyping works

- https://www.youtube.com/watch?v=Naona1y_I2U

- For more information see YouTube Channel Useful Genetics: https://www.youtube.com/channel/UCtXCrx28msMBQ-vFUIOIReA

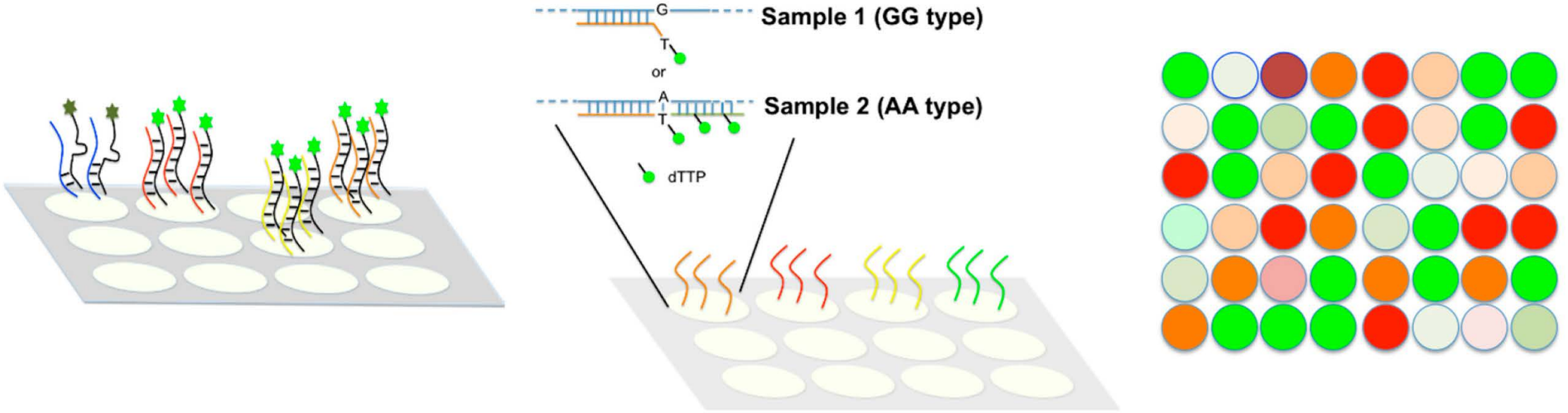


https://www.jax.org/news-and-insights/jax-blog/2016/september/genomes-versus-exomes-versus-genotypes

SNP - **Single nucleotide polymorphisms**

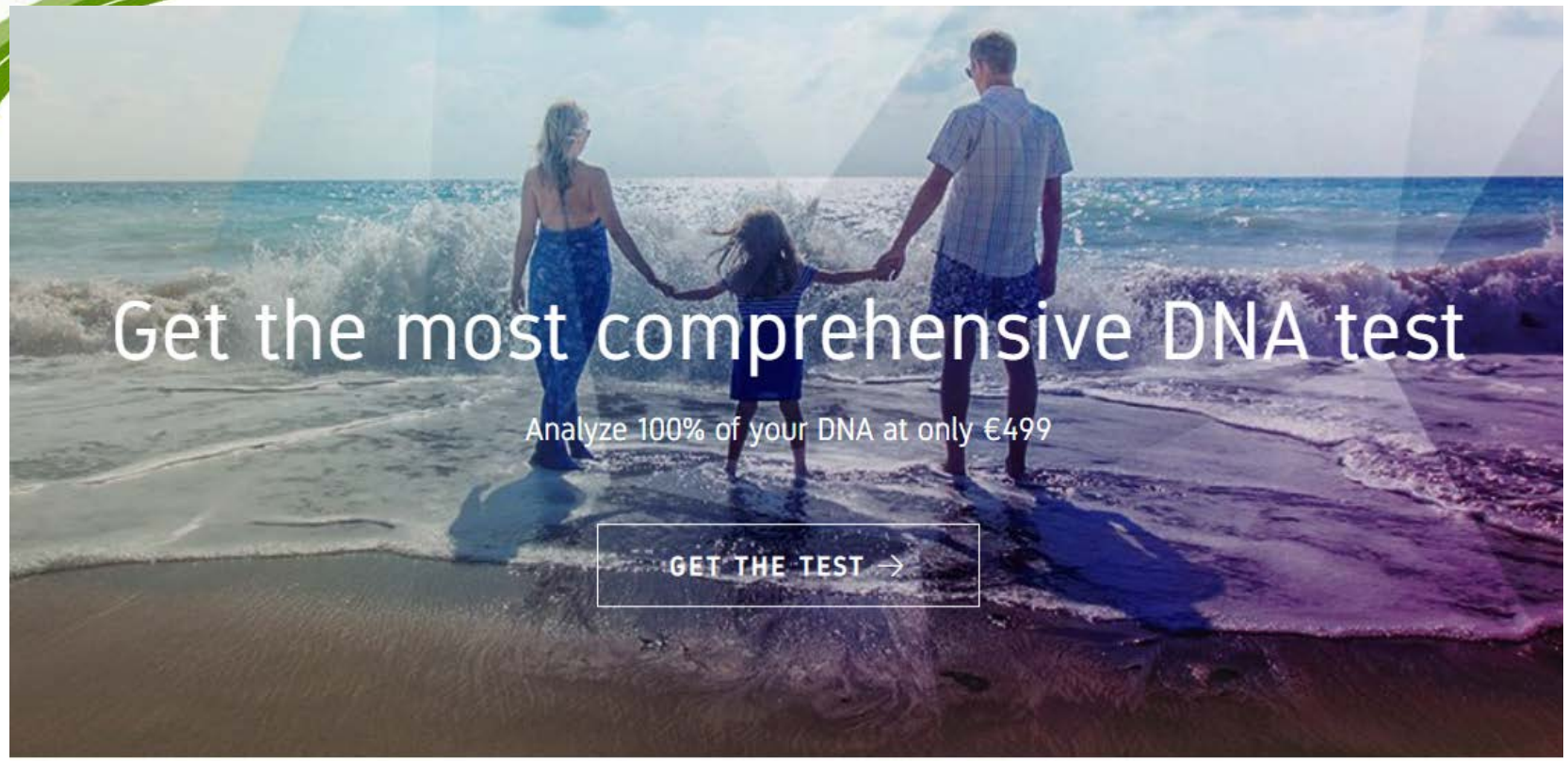
- the most common type of genetic variation
- occur almost once in every 1,000 nucleotides on average, 4 to 5 million SNPs in a person's genome
- may be unique or occur in many individuals; scientists have found more than 100 million SNPs in populations around the world
- most commonly in non-coding DNA
- can act as biological markers, helping locate genes associated with disease
- most SNPs have no effect on health or development
- some SNPs have proven to be very important in the study of human health.
- may help predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases.
- SNPs can also be used to track the inheritance of disease genes within families

How SNP genotyping works



There are two types of microarray commonly used in multiplexing SNP analysis: allele-specific oligonucleotide (ASO) hybridization and allele-specific primer (ASP) extension. (**A**) ASO hybridization: The allele-specific oligonucleotide for every SNP is synthesized and separately immobilized onto the glass plate. Fluorescence labeled targets containing SNP sites are produced from a PCR reaction and plotted separately into each well to conduct the hybridization reaction. The mismatched base pair between target and oligonucleotide can decrease the binding strength with the fluorescence-labeled target removed after a stringent washing. A fluorescence signal is detected on a perfectly matched base pair; (**B**) Allele-specific primer (ASP) extension: The specific primer for SNP location is designed and separately immobilized onto a microarray. A different fluorescence labeled dNTP is individually used in an extension reaction. The extended fragment showing fluorescence signal can only be found when the 3′ end of primer pair is perfectly matched (AA type in this case) in contrast to the mismatched primer pair (GG type in this case); (**C**) The SNP genotype can be determined according to fluorescent intensity from the products/target DNA. **https://doi.org/10.3390/microarrays4040570**

DTC genome sequencing as popular demand

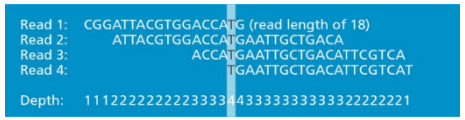


WHAT YOU GET

Dante Labs analyzes 100% of your DNA, so that we can give you reports on predispositions on any genetic disease.
You will receive easy reports for you and your doctor, as well as raw data to explore.

www.dantelabs.com

Coverage (or depth) in sequencing



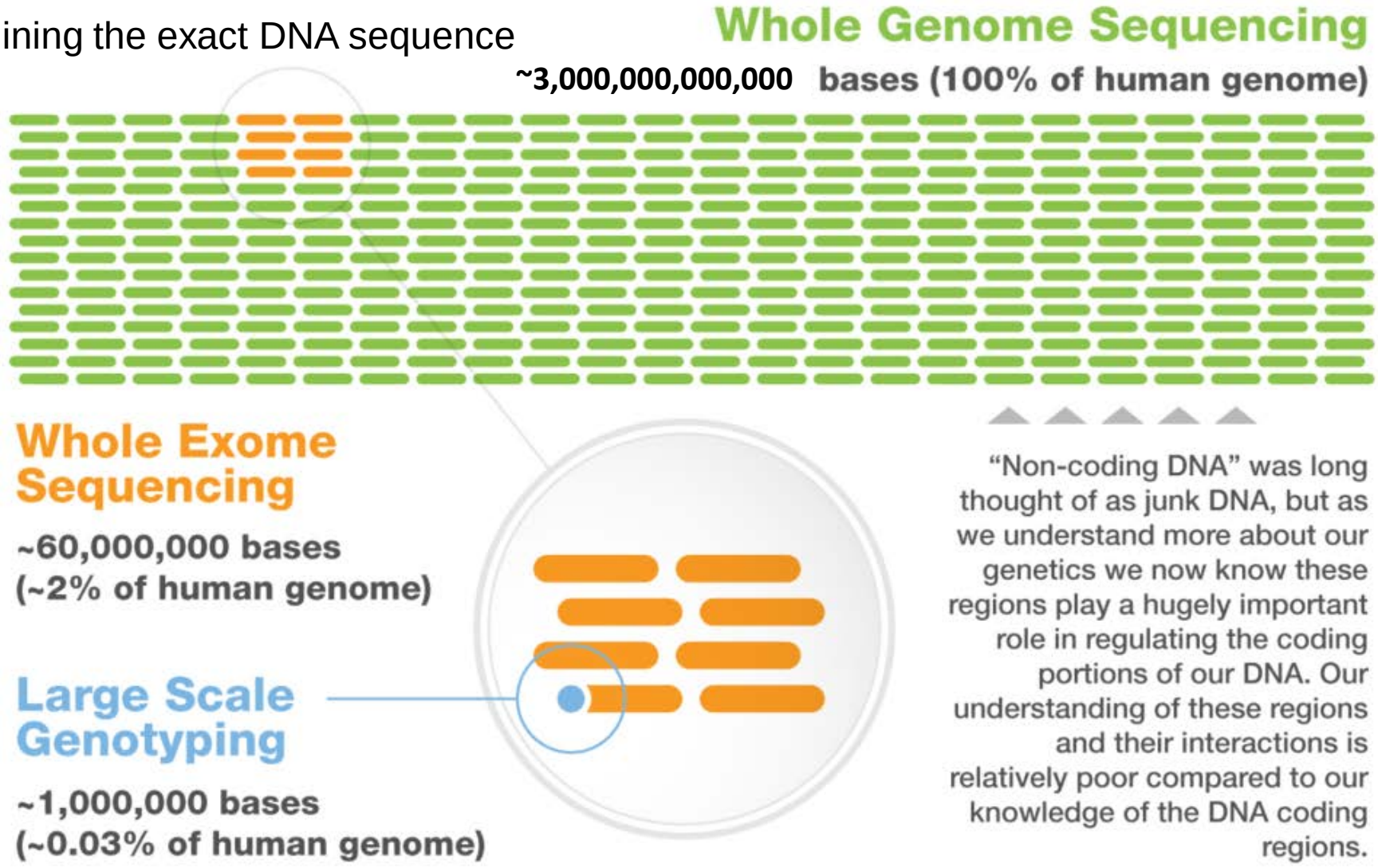
My Full DNA: Whole Genome Sequencing with mtDNA

€449.00 EUR ~~€850.00 EUR~~

YOU SAVE €401.00 EUR

Sequencing – WGS and WES

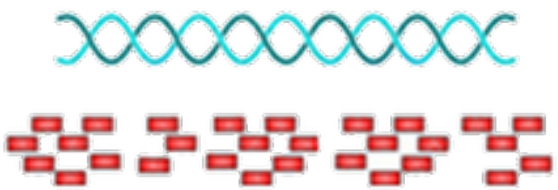
- Determining the exact DNA sequence



https://www.mygenefood.com/finding-best-dna-test-genotype-sequence/

Genomes vs exomes vs genotypes

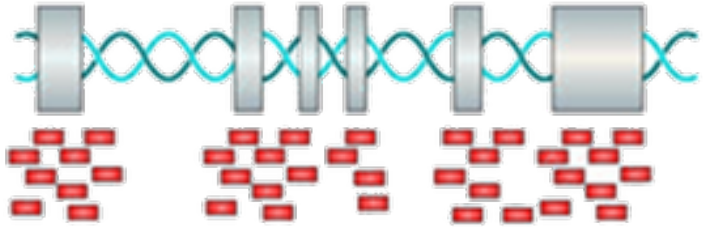
WGS
Whole genome sequencing



- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

- Results are sometimes challenging to interpret

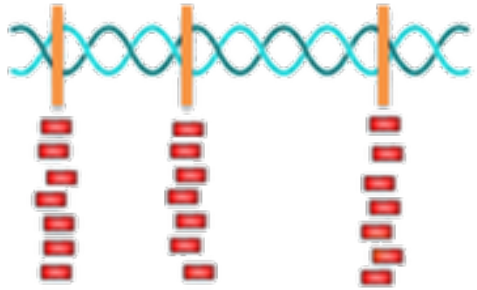
WES
Whole exome sequencing



- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

- Good alternative to WGS in terms of clinical use

Hotspot sequencing = Targeted sequencing



- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

- Most sensitive – able to detect rare tumor cells in a biopsy

https://2wordspm.wordpress.com/2017/10/30/ngs-%EA%B2%80%EC%82%AC-whole-genome-exome-targeted-sequencing-%EB%B9%84%EA%B5%90/

What to expect

- Genetic testing provided by most of the companies is moreless for fun (ancestry, health and wellness, nutrigenetics, skincare, sports,…)

https://dantelabs.com/products/livelonger-genome-test

- More expensive, and complete, sequencing like the one provided by Illumina can be used for medical investigation
- Do not expect your genome sequencing to tell you how long is your life expectation, whether you are likely to get cancer and so on
- So far our knowledge on the “implication” of the genome are quite limited
- What we can already do in health care is to look at the genome once you have been diagnosed a specific ailment and look for specific genes that would make one cure more effective than another (this has become normal practice in some form of cancer cure)

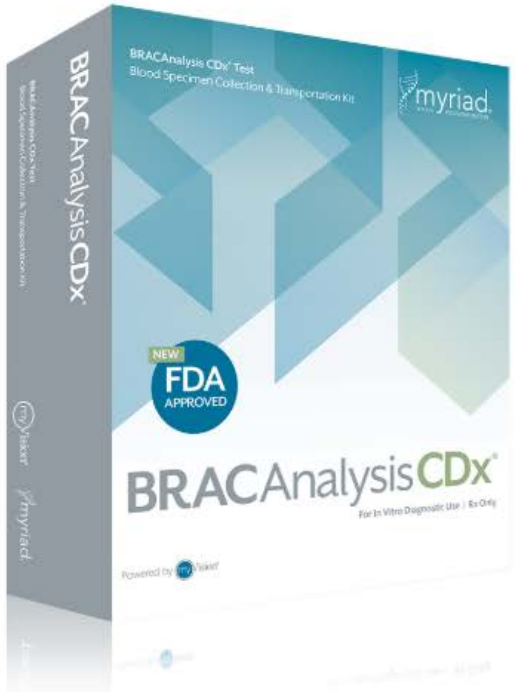
Based on http://sites.ieee.org/futuredirections/2017/12/26/did-you-get-your-genome-sequenced-for-christmas/

Example of genetic testing in clinical practise

- *BRCA* genes testing for PARP inhibitor treatment

BRACAnalysis CDx® Ovarian Cancer

Overview



Mutations in *BRCA1* or *BRCA2* cause Hereditary Breast and Ovarian Syndrome (HBOC). Now mutations in the *BRCA1* and *BRCA2* genes provide an indication for treatment with Lynparza™ (olaparib) for patients with ovarian cancer. Specifically, BRACAnalysis CDx® is the only FDA-approved laboratory developed test approved to be used to inform treatment decisions for the PARP inhibitor, Lynparza. A positive BRACAnalysis CDx result in patients with ovarian cancer is also associated with enhanced progression-free survival (PFS) from Zejula™ (niraparib) maintenance therapy.[1,2,3]

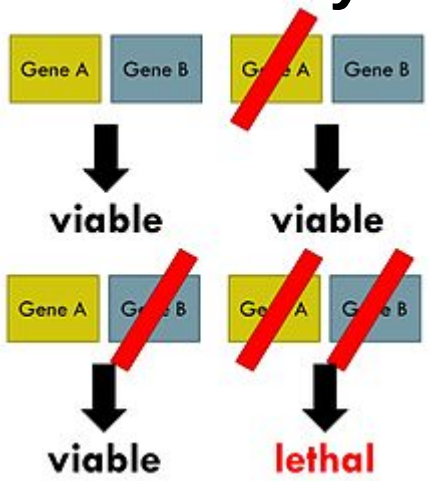
Learn More | Order BRACAnalysis CDx

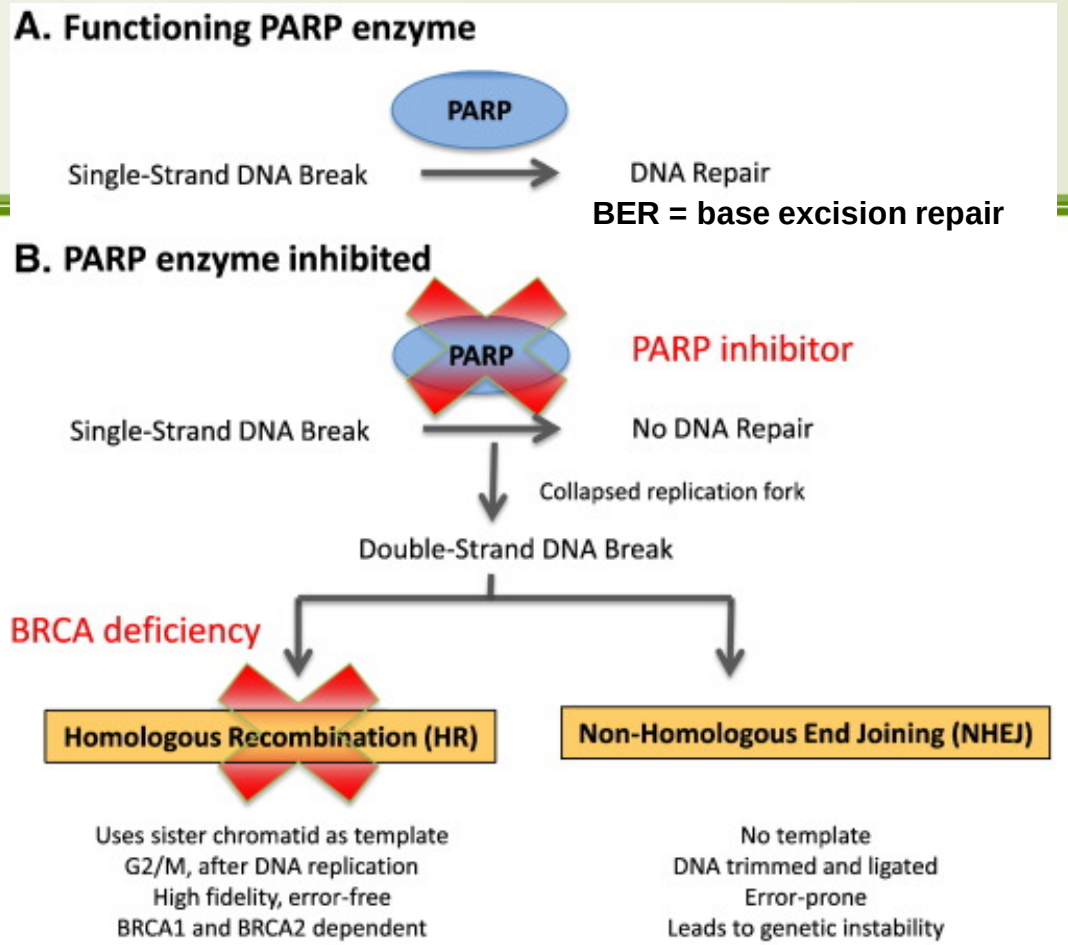
- More info: https://www.youtube.com/watch?v=ilwMGRH276M

PARP inhibitors

- In December 2014, the drug olaparib (Lynparza) became the first of a new class of treatments known as PARP (poly(ADP-ribosa)polymerase) inhibitors to be licensed for clinical use, heralding in a new era for personalised, targeted treatment—and turning the promise of 'synthetic lethality' into reality.

Synthetic lethality concept





C. Deficiency in HR and BER together lead to synthetic lethality

Condition	HR	BER	Outcome
Normal cells	+	+	Viable
BRCA deficient	-	+	Viable
Normal cells, PARP inhibitor	+	-	Viable
BRCA deficient, PARP inhibitor	-	-	Cell Death

More info on PARPi:
https://www.youtube.com/watch?v=mgW30YyaJz4

https://doi.org/10.1016/j.ygyno.2015.02.017

The Present and Future of Genome Sequencing

- Genomics England - **100,000 patients** with rare diseases, their families, and cancer patients
- Precision Medicine Initiative (PMI) **1-million-volunteer** health study, data including genetics and lifestyle factors
- GenomeAsia **100K** - genomic data for Asian populations
- … a many more initiatives

https://labiotech.eu/features/genome-sequencing-review-projects/

- How to handle such huge amount of data and the ethical implications?
- In the US, the Genetic Information Nondiscrimination Act (2008) but mostly no act in other countries and somewhat grey legal position in Europe

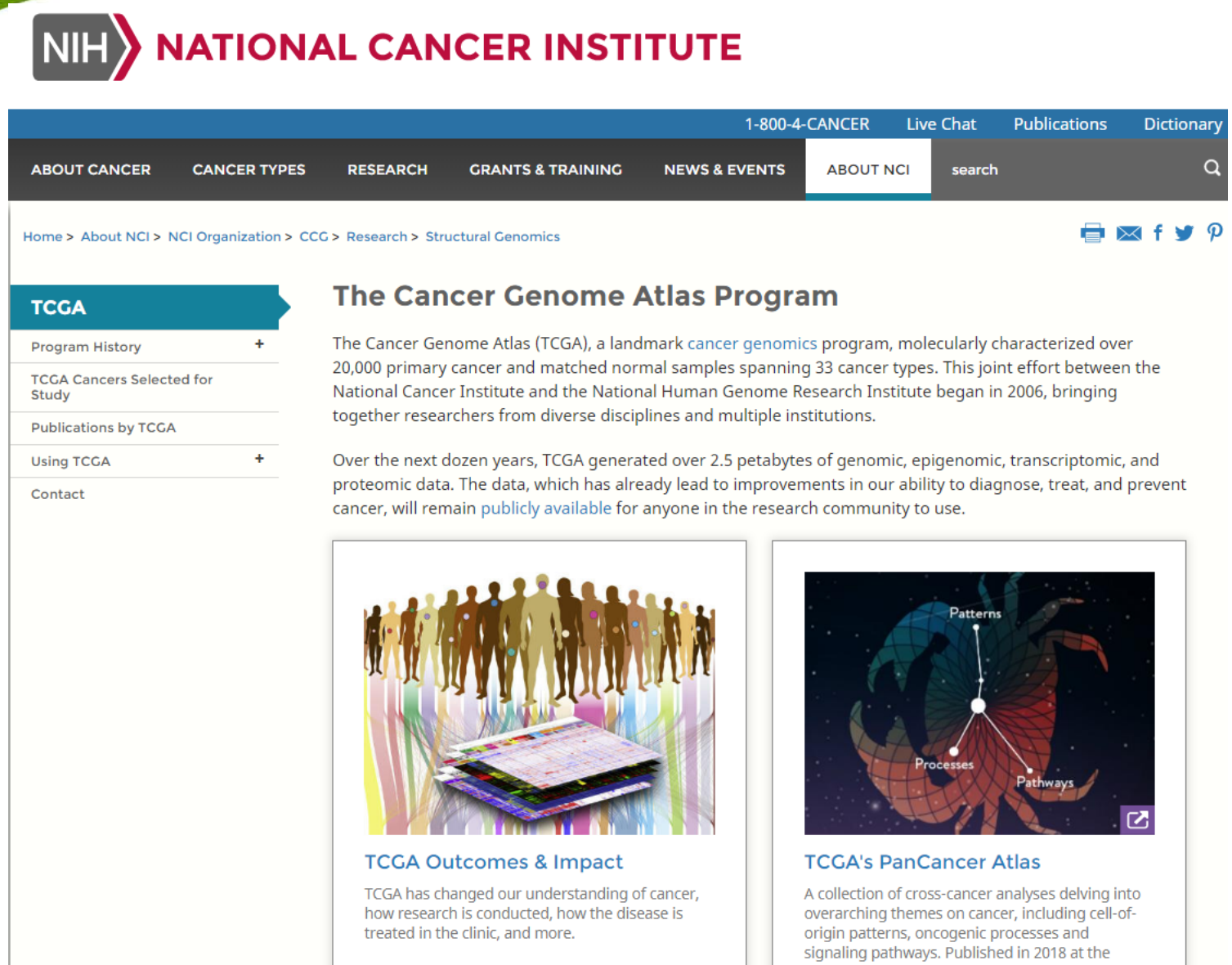

Time for
a BREAK

COSMIC: Cataloque of Somatic Mutations in Cancer

https://cancer.sanger.ac.uk/cosmic/

- How to use COSMIC database: https://www.youtube.com/watch?v=2FD5RabgK6o, https://www.youtube.com/watch?v=k477uAiKx74

TCGA: The Cancer Genome Atlas



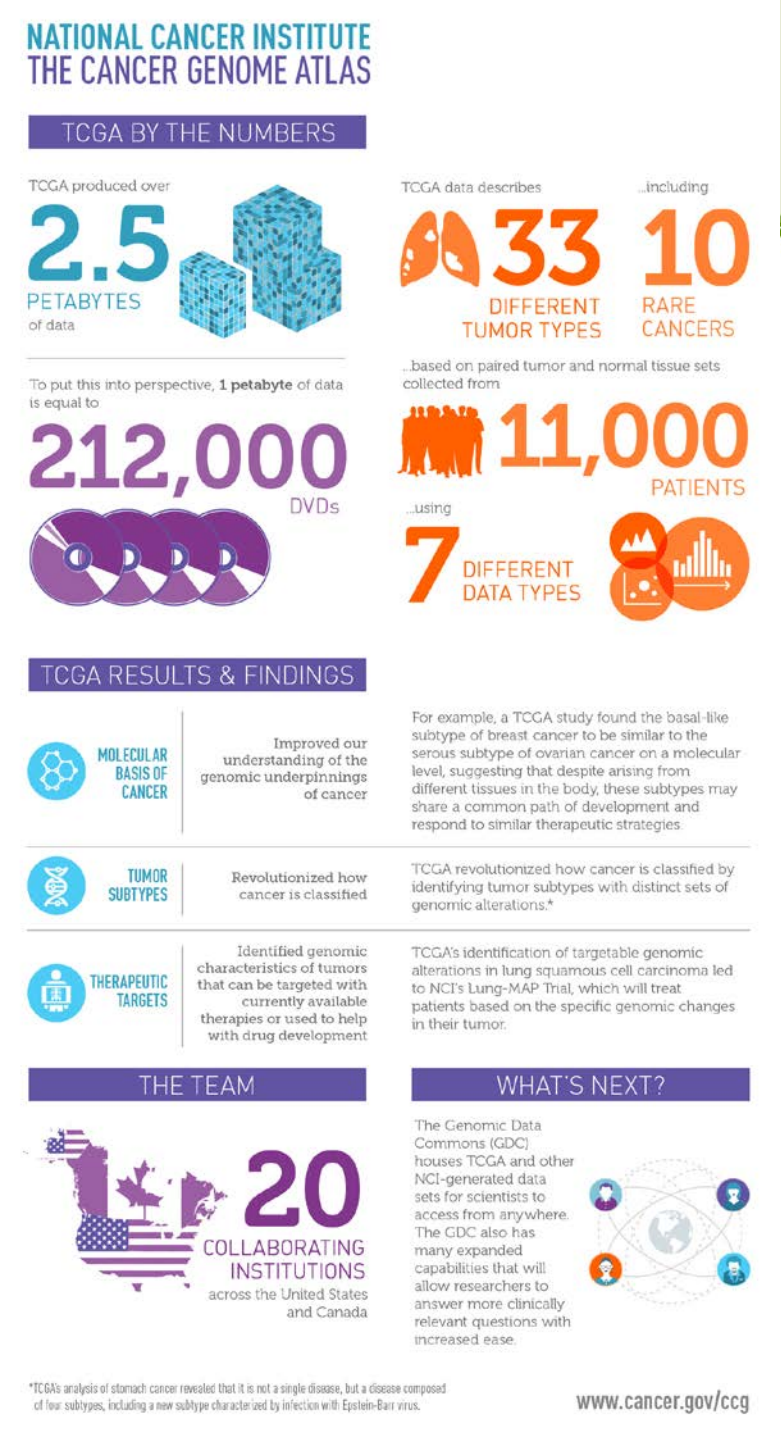
https://www.youtube.com/watch?time_continue=249&v=epsZjJ_A1y4

https://cancergenome.nih.gov/

TCGA: Overview

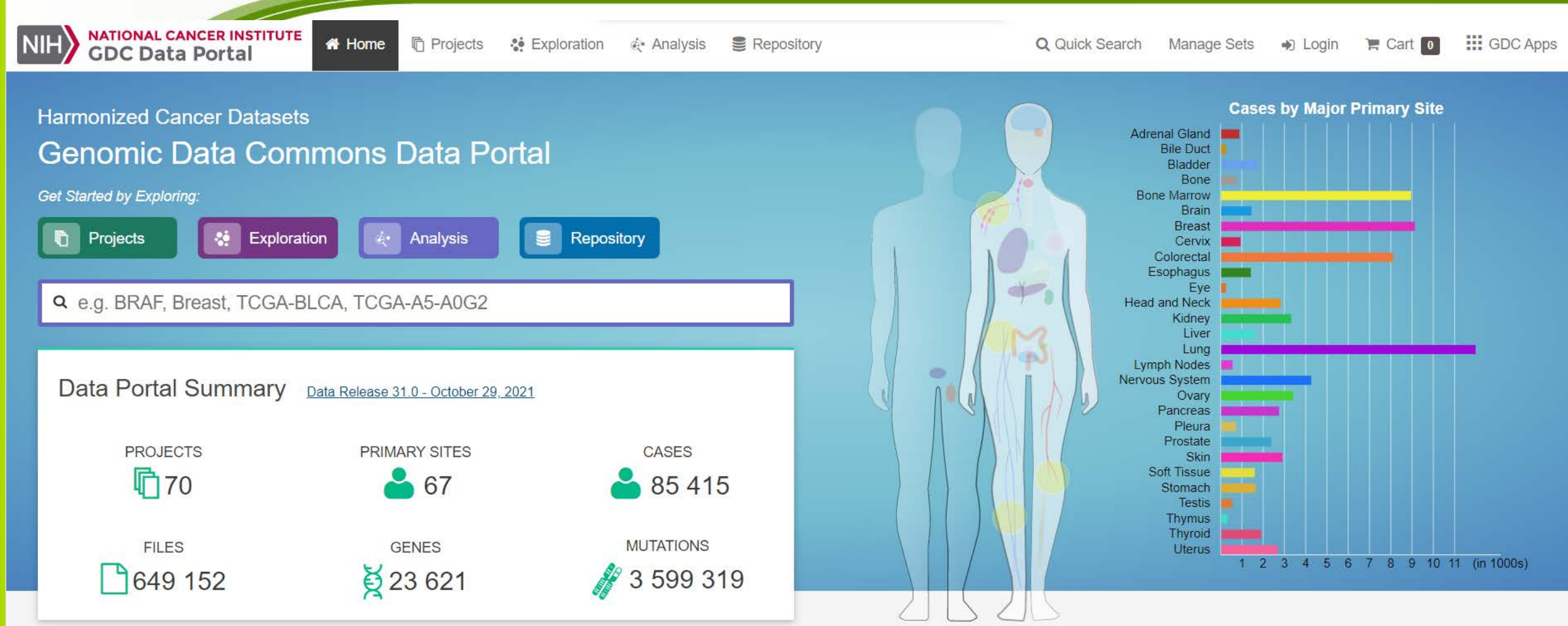
- Initiated in 2005
- A joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI).
- 27 participating Institutes in US and Canada.
- The overarching goal of TCGA is to improve our ability to diagnose, treat and prevent cancer, through the application of genome analysis technologies, including large-scale genome sequencing.
- The Cancer Genome Atlas Network have published more than 20 papers since the project began

(https://tcga-data.nci.nih.gov/docs/publications/)



TCGA Data Portal

https://portal.gdc.cancer.gov/



GDC Applications

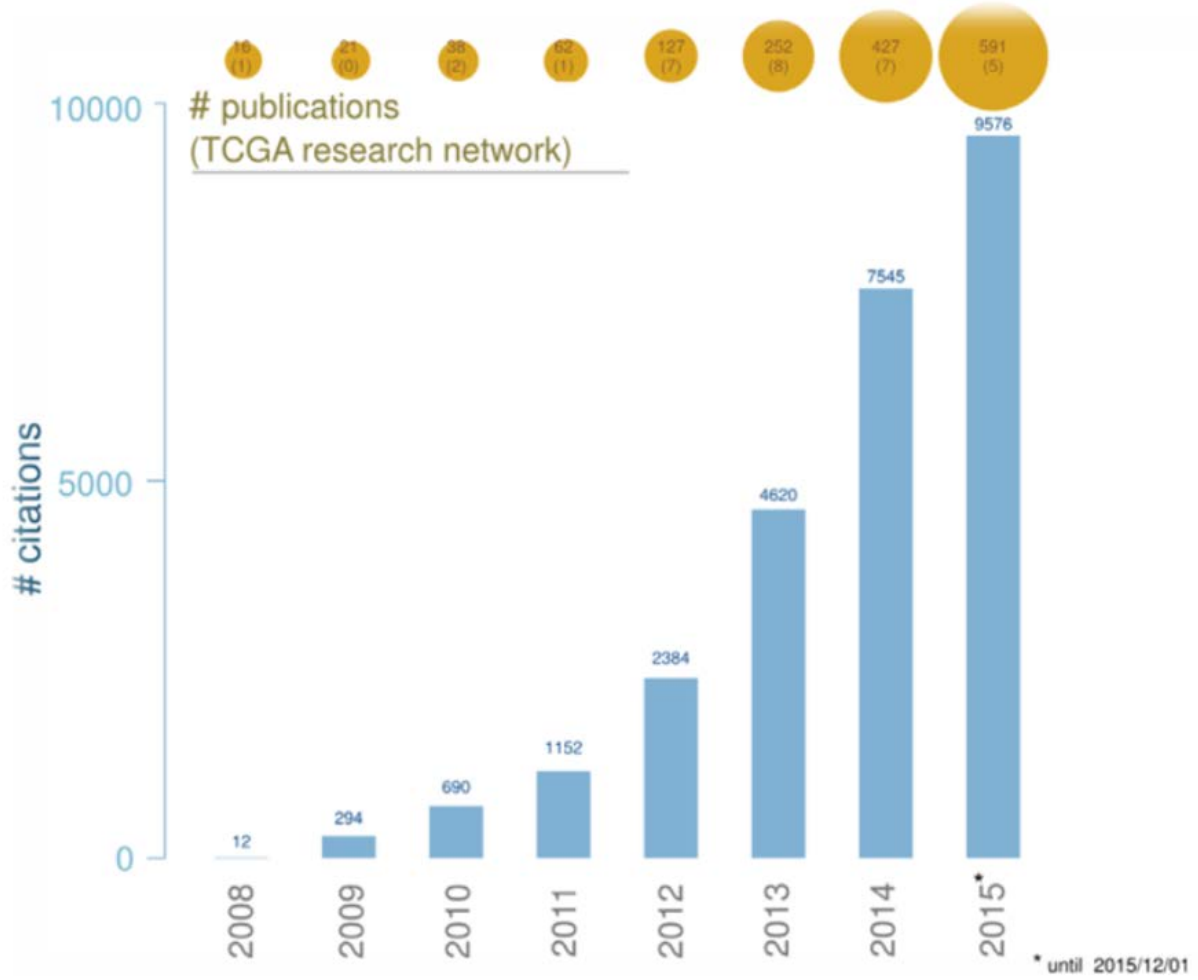
The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

TCGA: A Valuable Resource for Research Community

TCGA Data Types

- Clinical data
- DNA sequencing
- miRNA sequencing
- Protein expression
- mRNA sequencing
- Total RNA sequencing
- Array-based expression
- DNA methylation
- Copy number variations

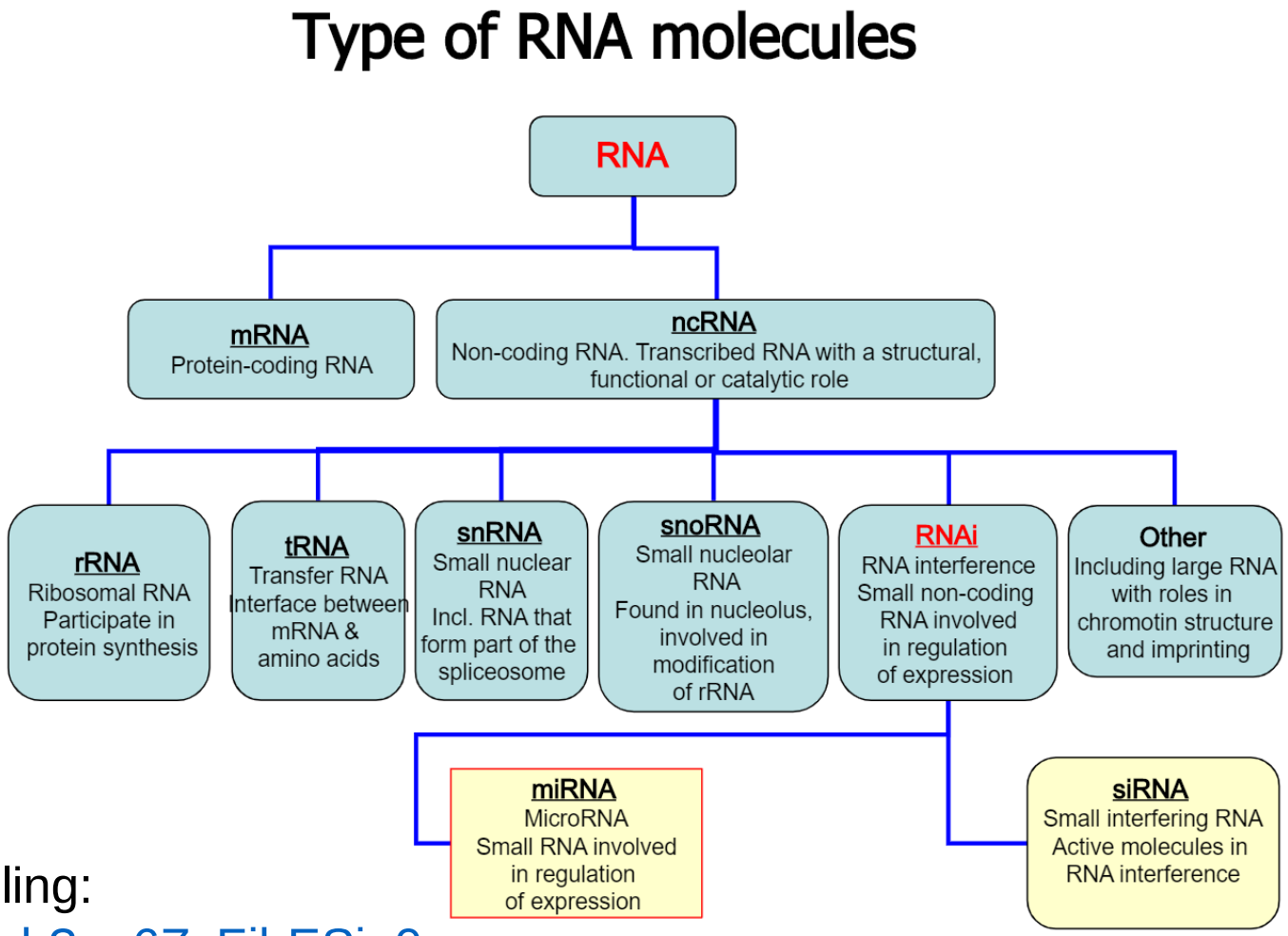
+ Computational tools



How to use TCGA: https://www.youtube.com/playlist?list=PL-hYJ1isbXhURdasc-RmwDRLhrHzdzKtN

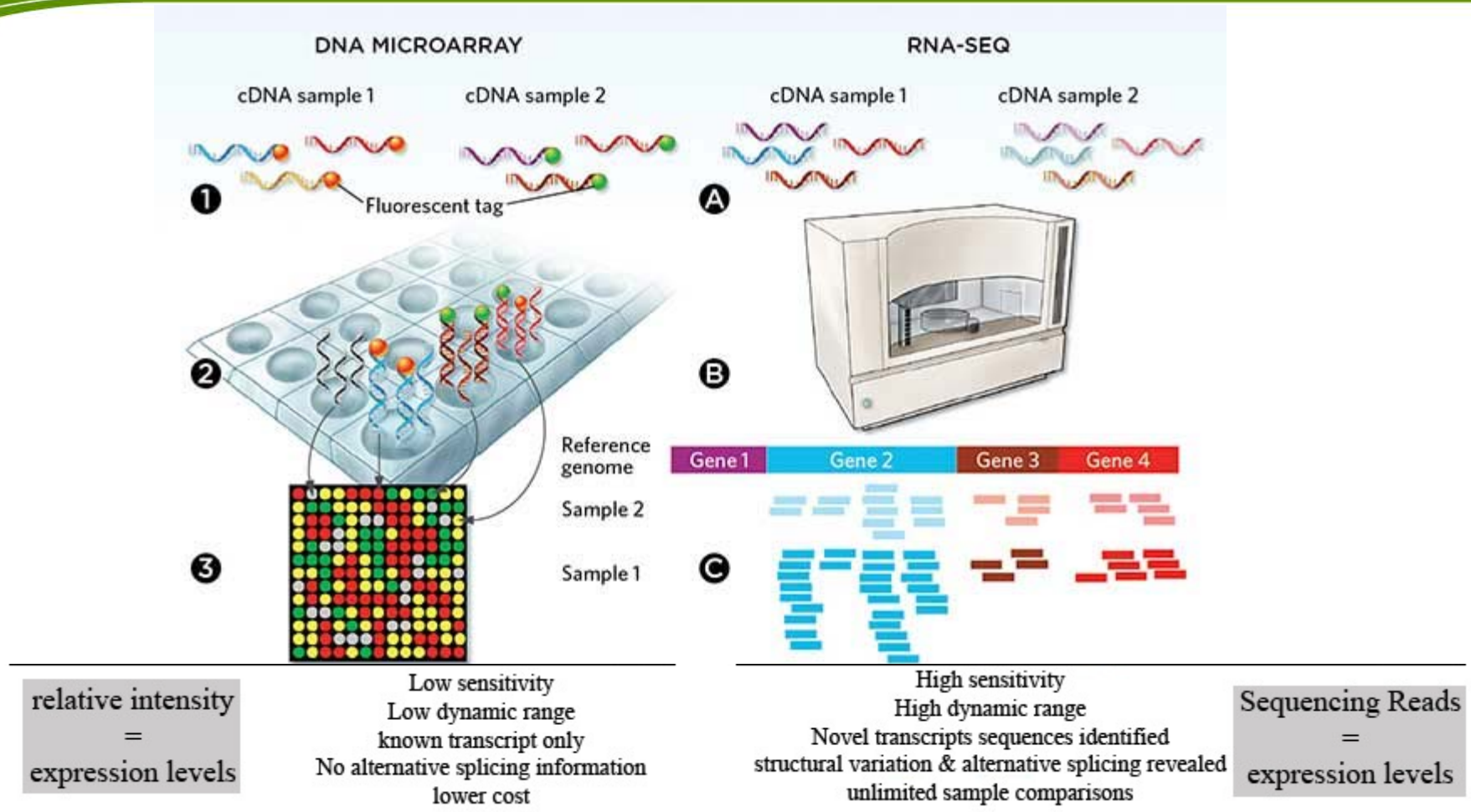
Transcriptomics

- Study of transcriptome, the sum of all RNA transcripts
- Two most widely studies types of RNA
  - mRNA - transcriptome or the expressed genes. Usually contains genes with poly A tail.
  - miRNA - Small non-coding RNA (containing about 21-25 nucleotides), important in gene regulation.



- Array-based Expression Profiling:
- https://www.youtube.com/watch?v=6ZzFihESjp0

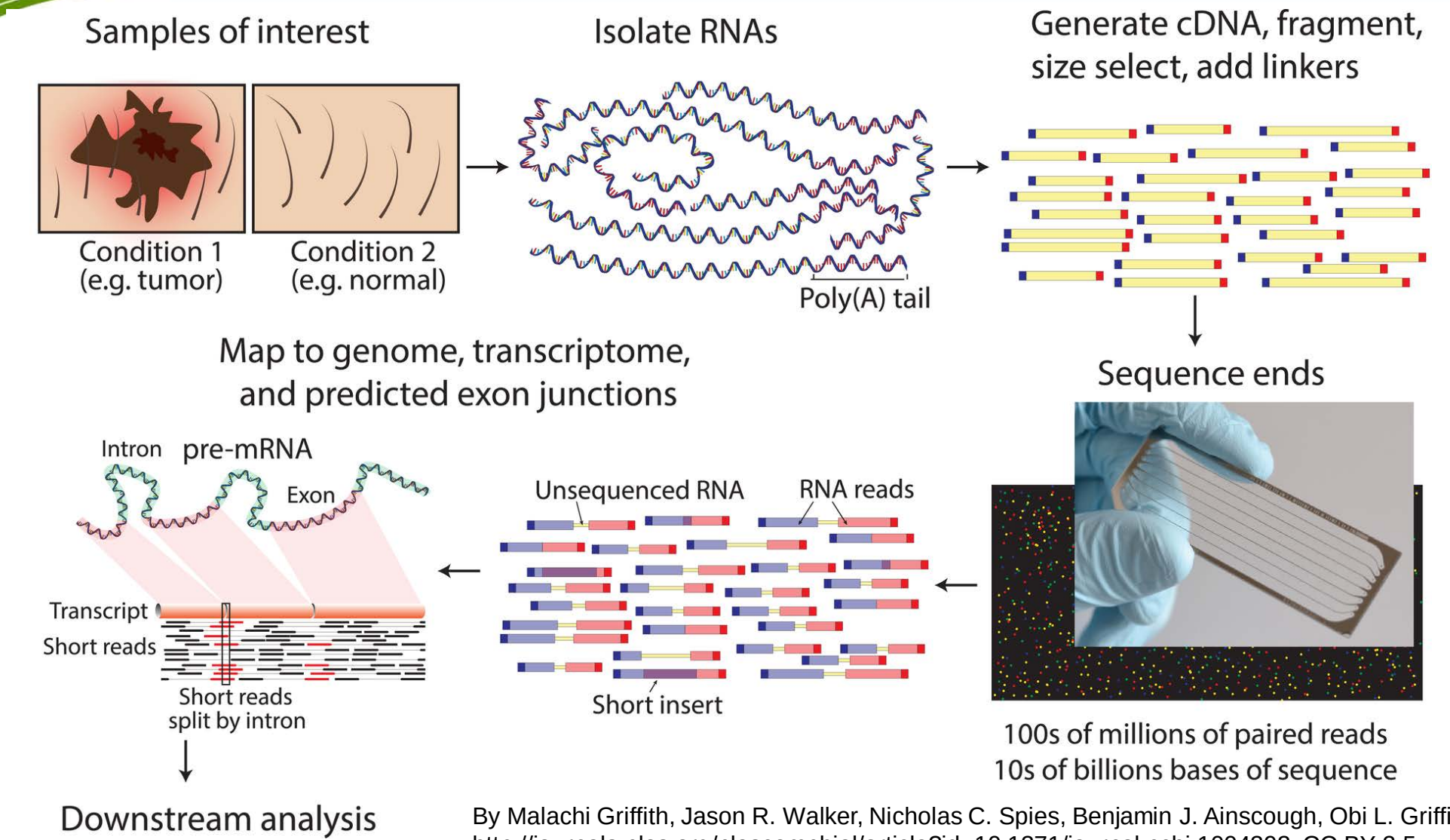
Microarrays vs RNA-seq



- While methods for analyzing microarray data are fully mature and straightforward, there is no consensus on which pipelines—or series of computational steps—to use to analyze RNA-seq data.

https://www.the-scientist.com/lab-tools/an-array-of-options-35381

Overview of RNA-seq



RNA sequencing downstream analysis

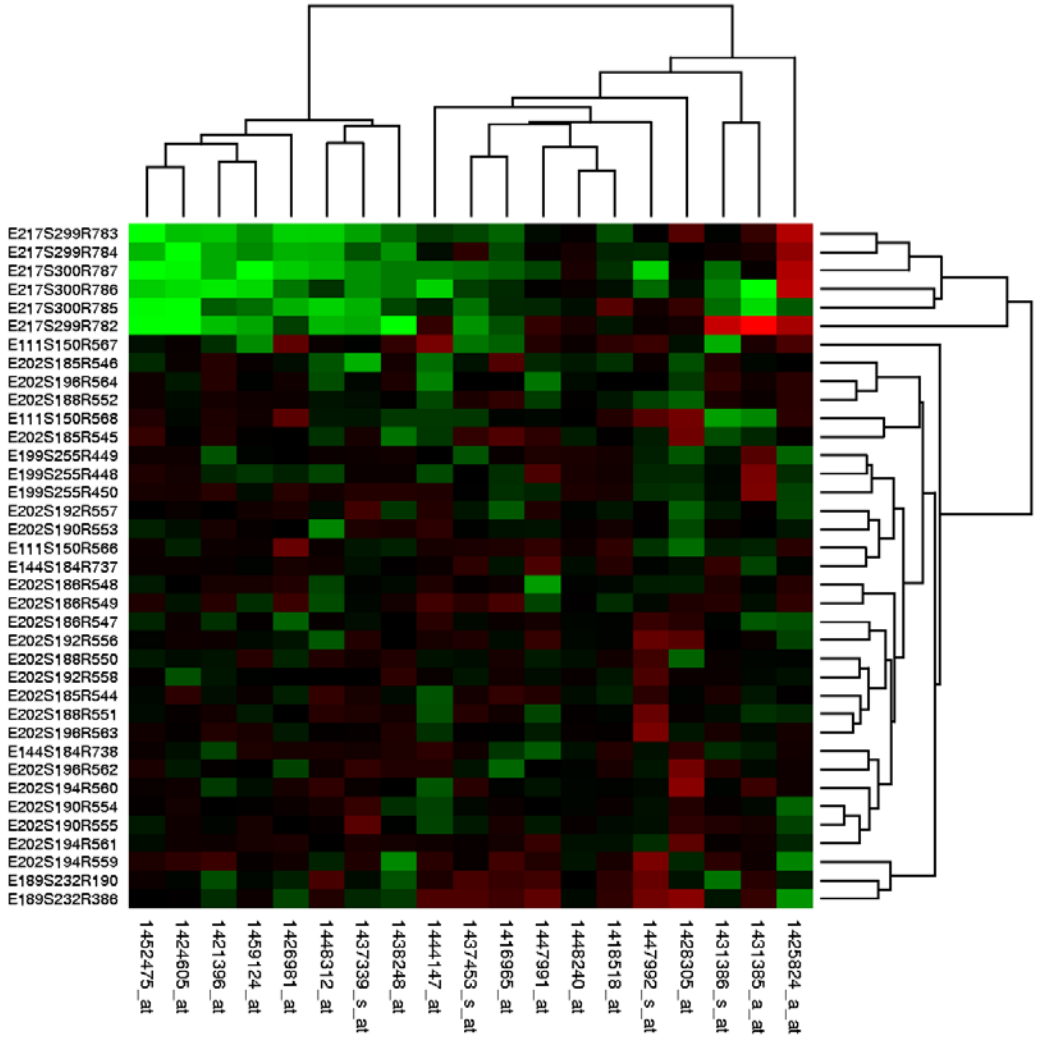
- https://www.youtube.com/watch?v=tlf6wYJrwKY (from 13:10)

- More info about microarray vs. RNA-seq at: https://www.youtube.com/watch?v=2c3t3tDEmsU

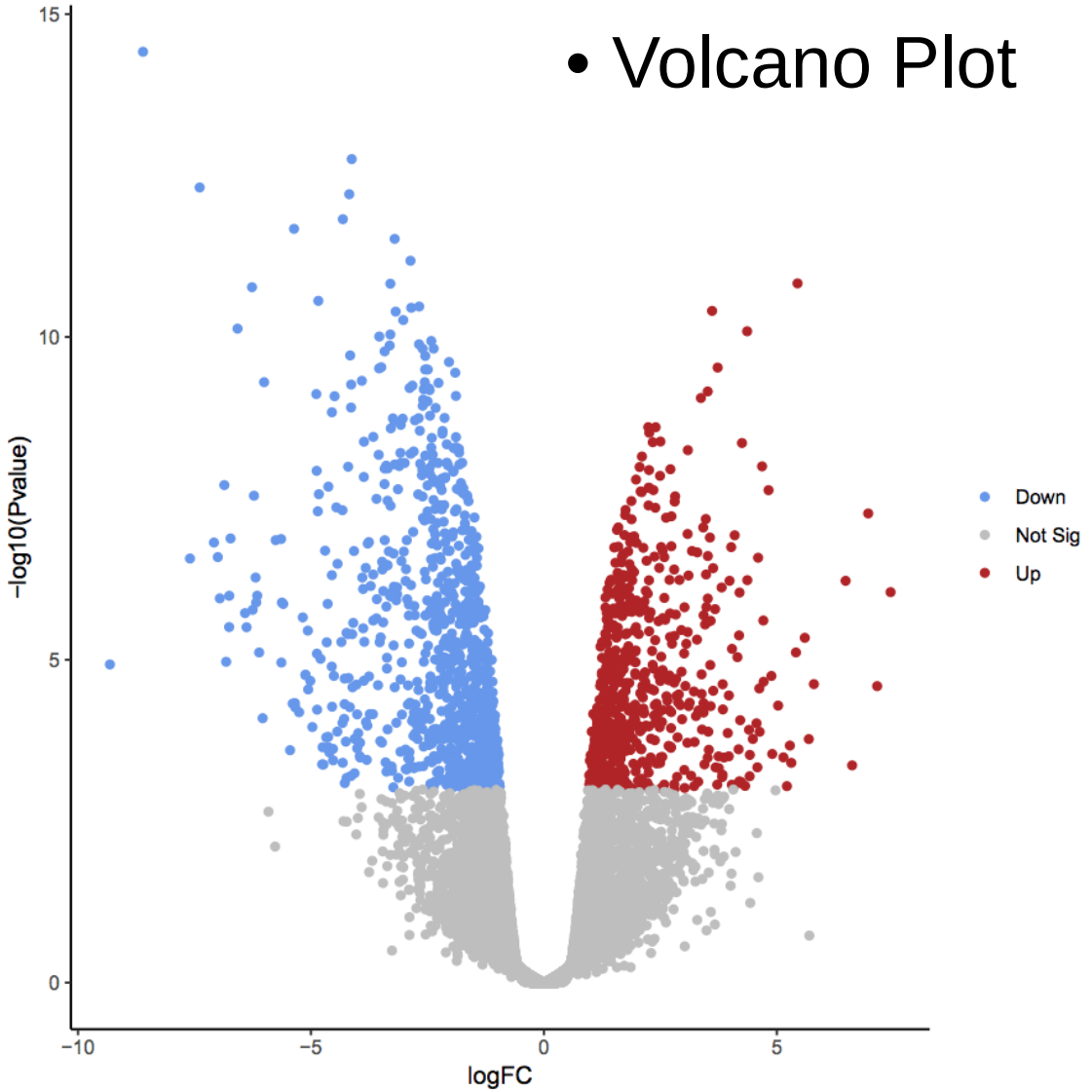
- More info RNA seq at:
- https://www.youtube.com/watch?v=MFRkwXq6v_I

- Useful detailed info about anything connected toRNA-seq
- https://www.rna-seqblog.com

Examples of transcriptomics data outputs



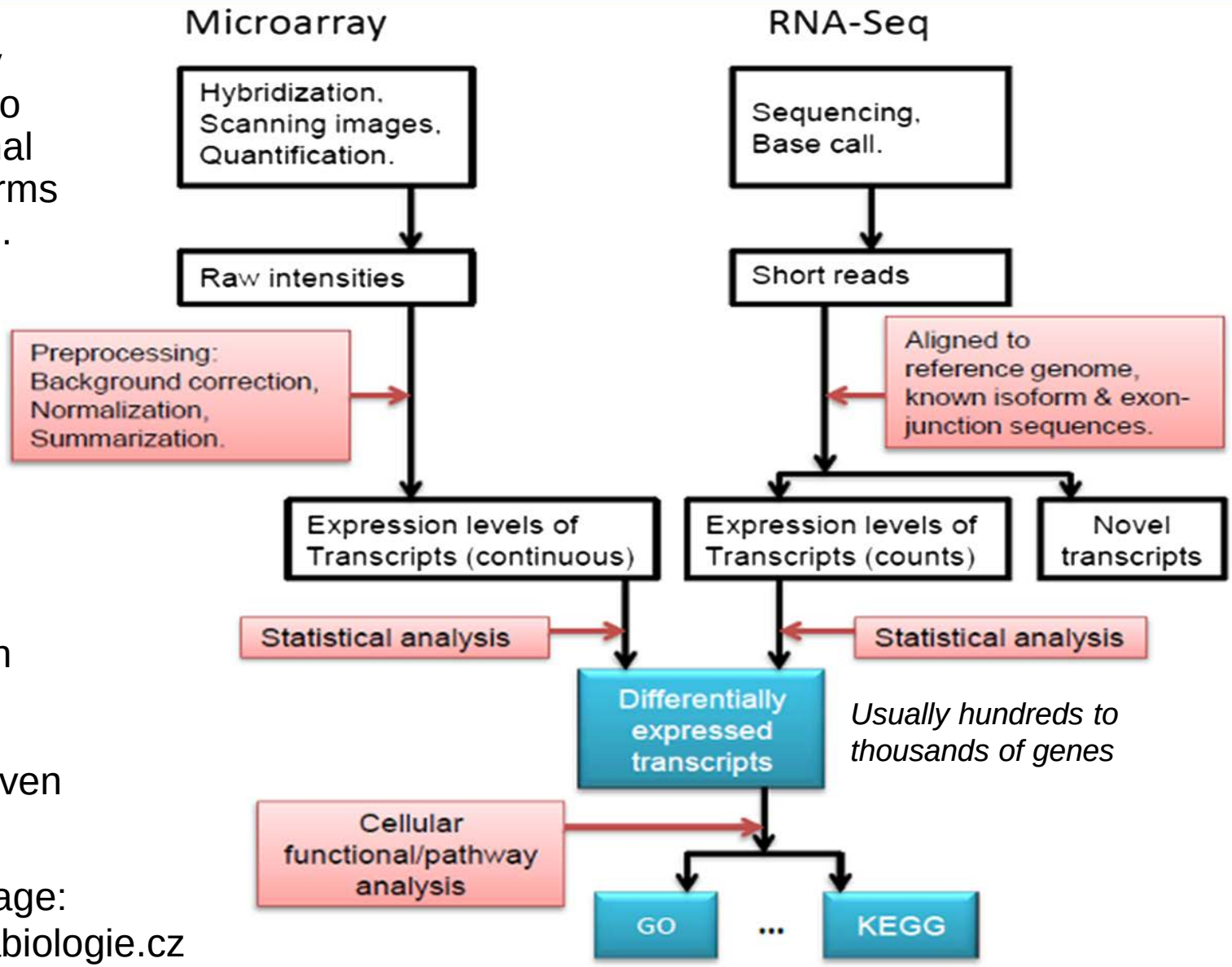
- Heat map



- Volcano Plot

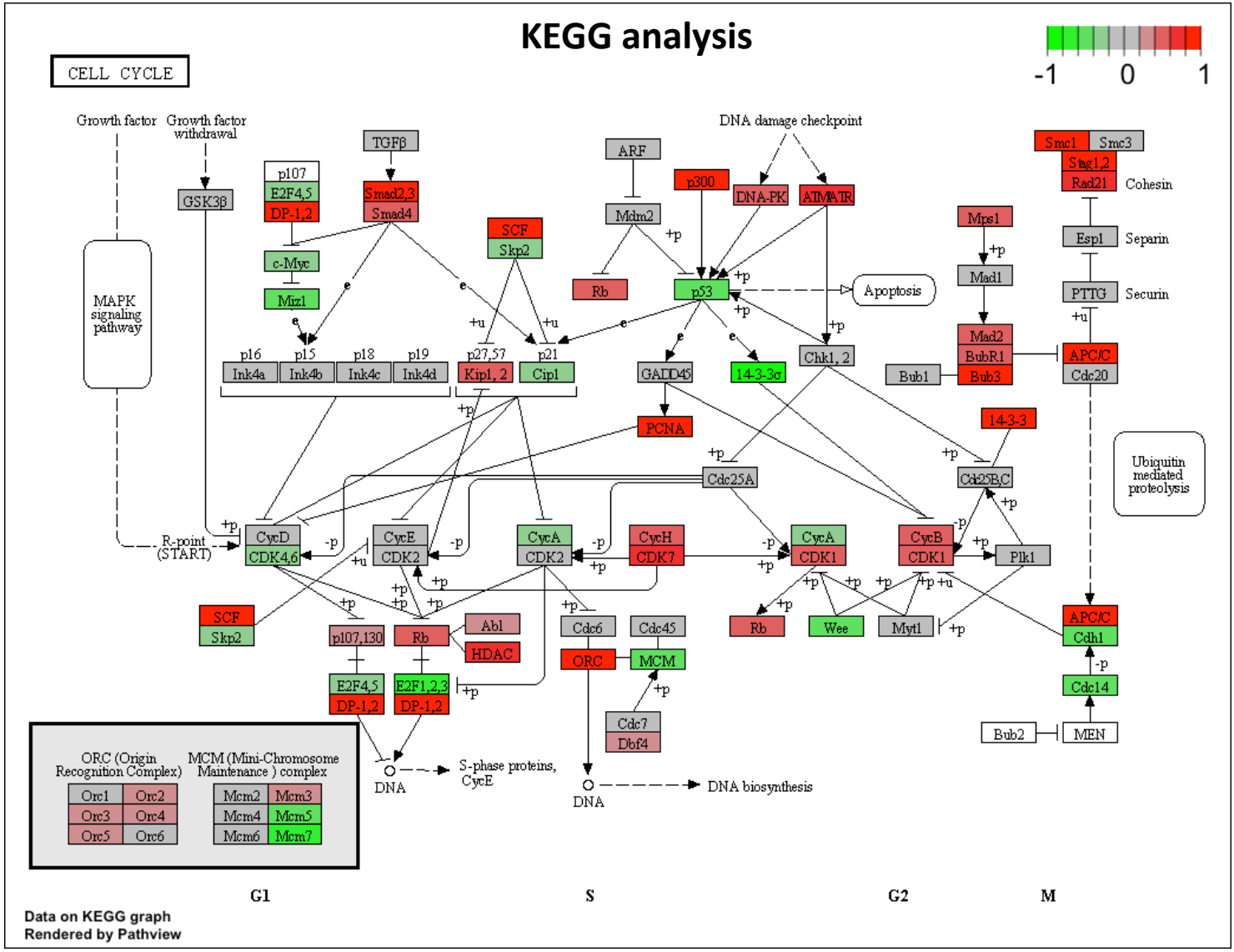
Cellular/functional/pathway analysis

- Cellular/functional/pathway analysis is a valuable tool to summarize high-dimensional gene expression data in terms of biologically relevant sets.
- Genes are aggregated into gene sets on the basis of shared biological or functional properties as defined by a reference knowledge base.
- Knowledge bases are database collections of molecular knowledge which may include molecular interactions, regulation, molecular product(s) and even phenotype associations.
- Useful info in Czech language:
  https://portal.matematickabiologie.cz



Database resources for understanding high-level functions and utilities of the biological system

- Database tools:
  - **KEGG** (Kyoto Encyclopedia of Genes and Genomes)
  - (https://www.genome.jp/kegg/)
  - Disadvantage – does not provide statistical significance of particular pathways
  - And many others available online

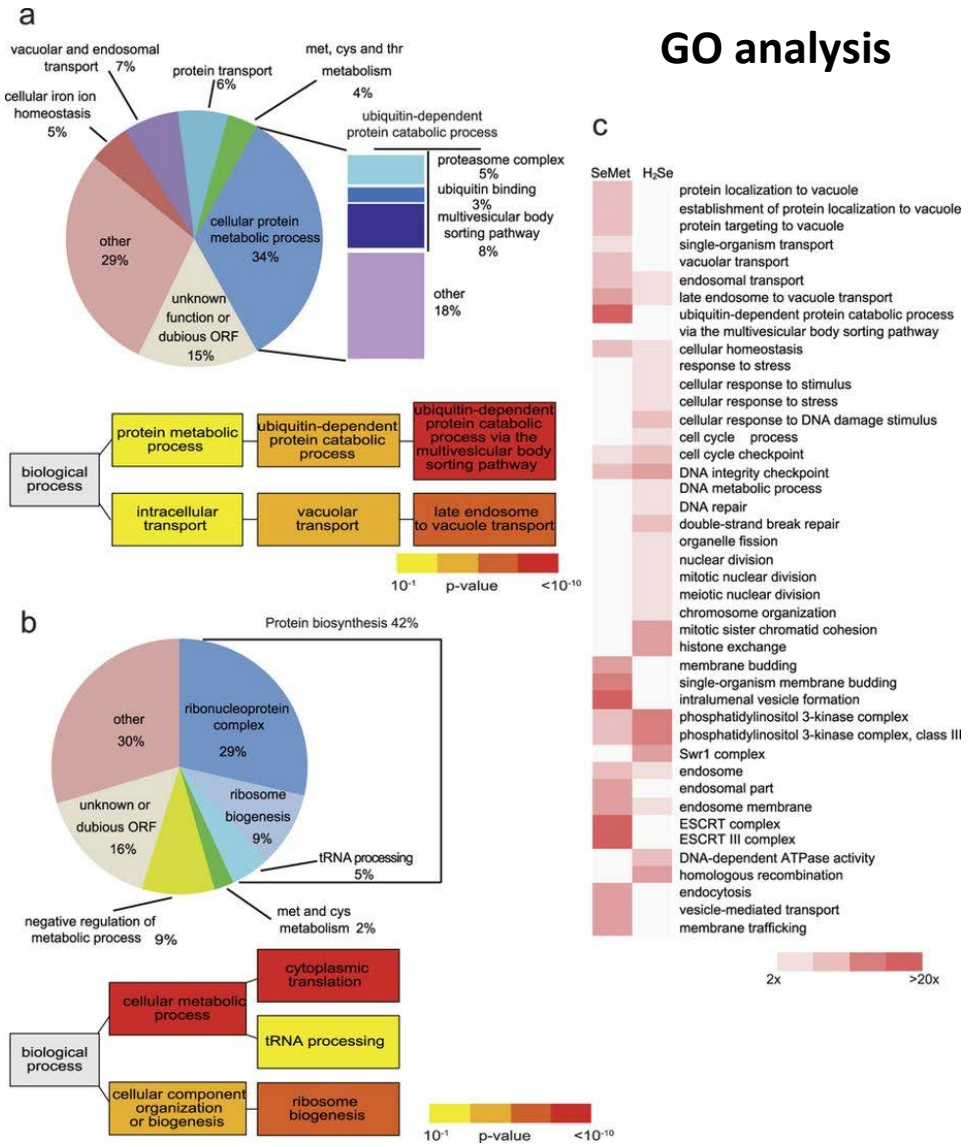


Gene-set analysis (GSA)/Pathway analysis

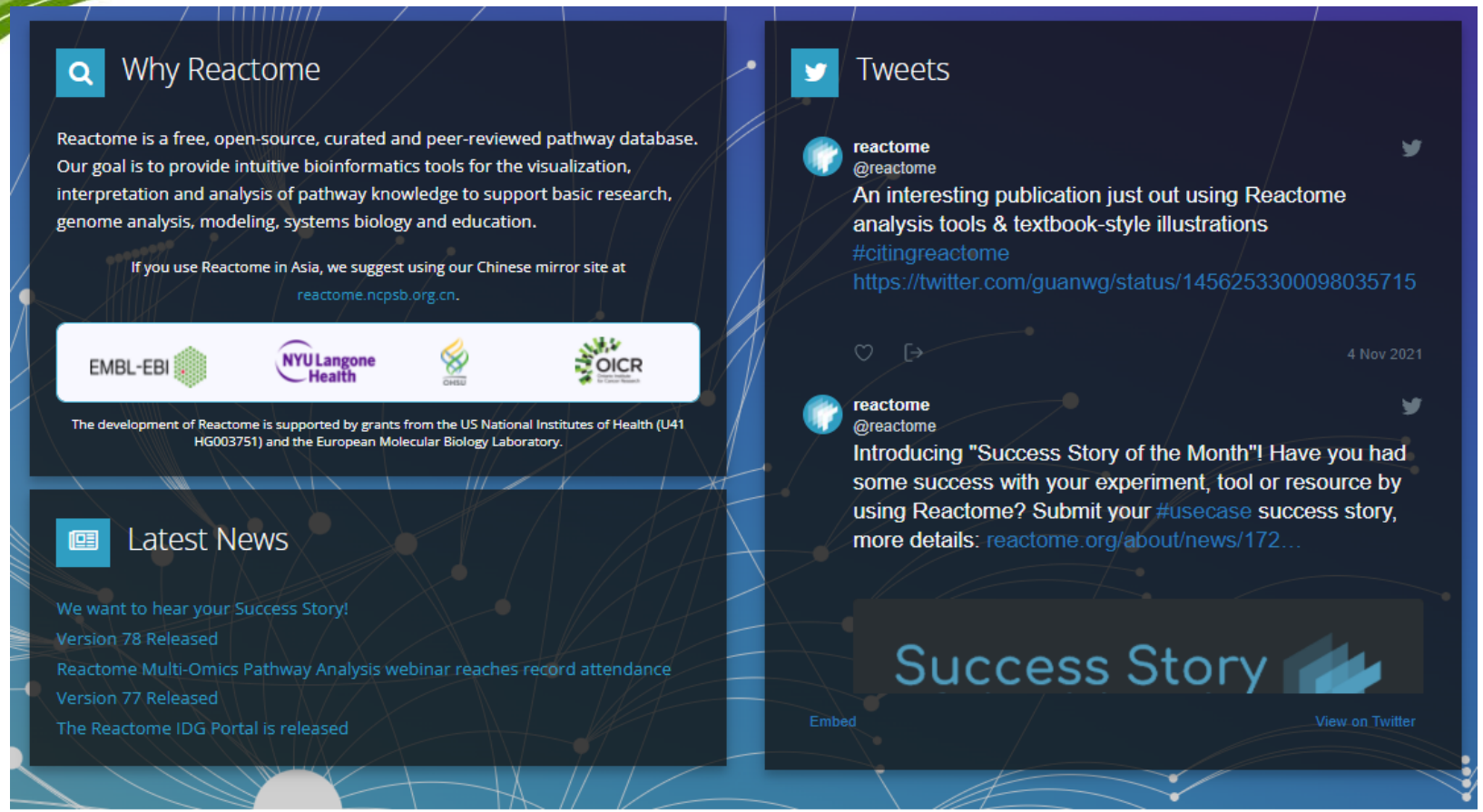
- Gene Ontology (GO) analysis (http://geneontology.org/)

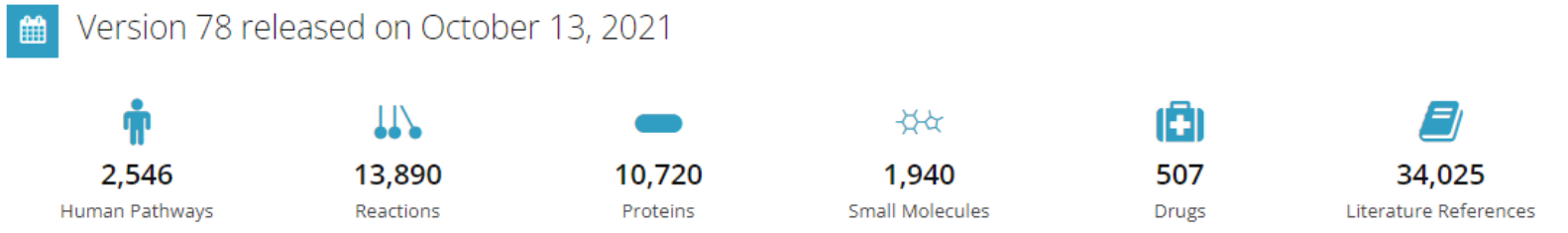
Example data of GO enrichment analysis

- **GO enrichment analysis**
- One of the main uses of the GO is to perform enrichment analysis on gene sets. For example, given a set of genes that are up-regulated under certain conditions, an enrichment analysis will find which GO terms are over-represented (or under-represented) using annotations for that gene set.
- 3 main GO aspects (molecular function, biological process, cellular component)
- http://geneontology.org/docs/go-enrichment-analysis/



Reactome Knowledgebase





- More info at:
- https://www.youtube.com/user/Reactome/videos

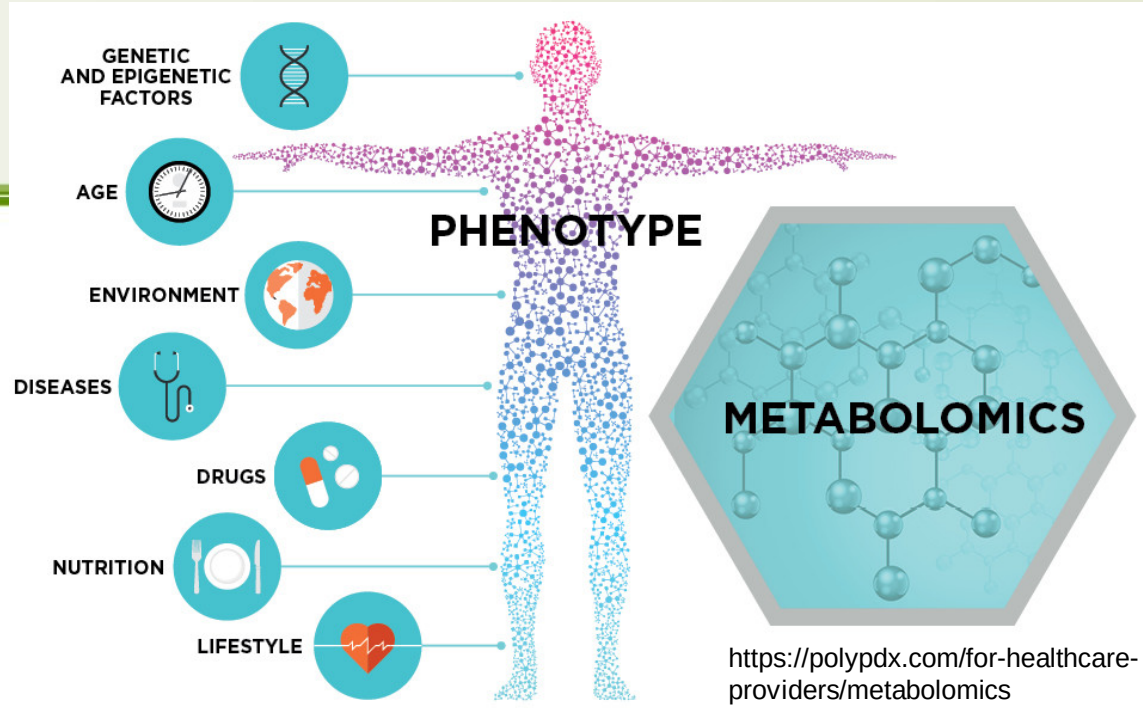
Metabolomics

- Metabolomics – large-scale systematic study of the metabolome

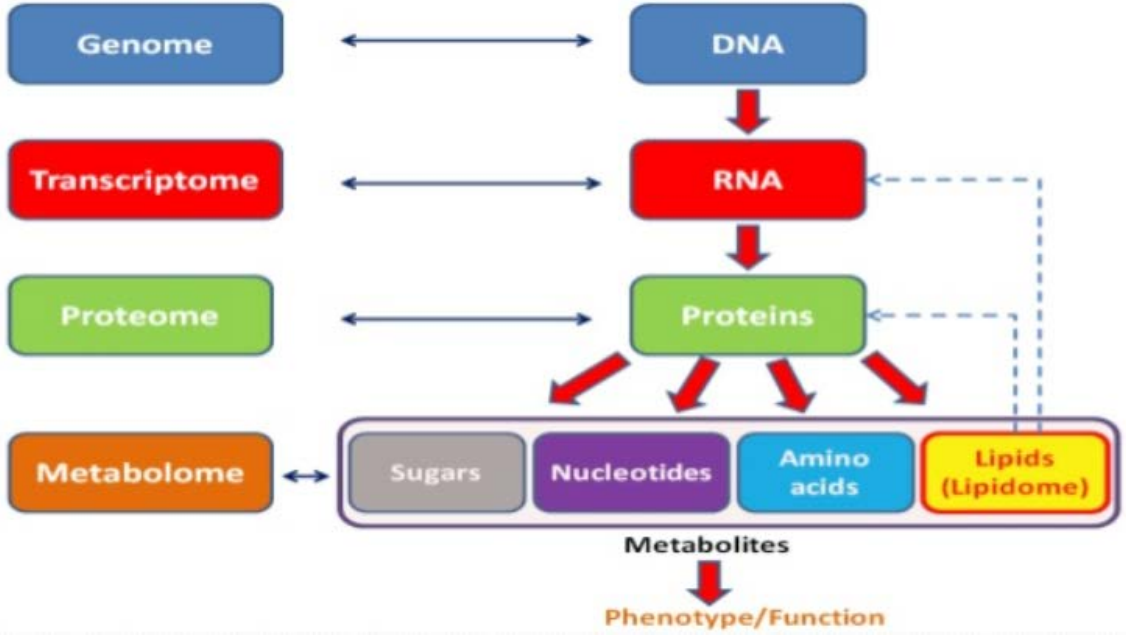
- Metabolome - total complement of metabolites present in a biological sample under given genetic, nutritional or environmental conditions
  - the unique biochemical fingerprint of all cellular processes

- Metabolite - low molecular (usually 50 – 1,500 Da) weight organic compound, typically involved in a biological process as a substrate or product.

- Metabolomics yield many insights into basic biological research in areas such as systems biology, metabolic modelling, pharmaceutical research, nutrition and toxicology



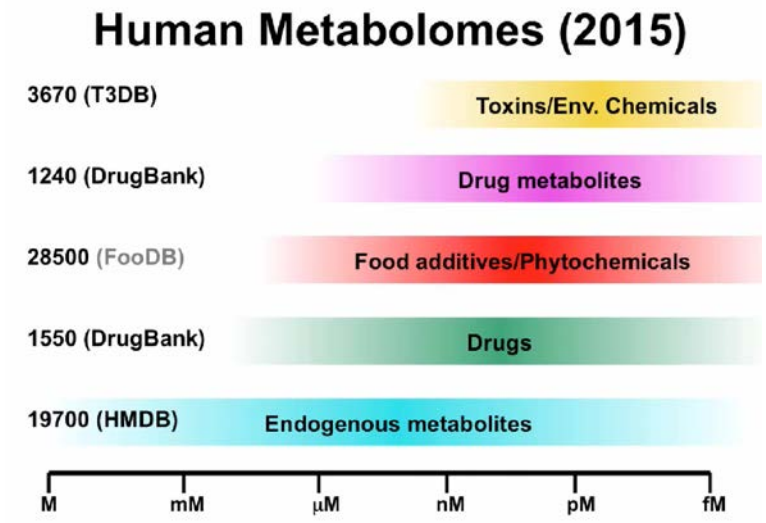
https://polypdx.com/for-healthcare-providers/metabolomics

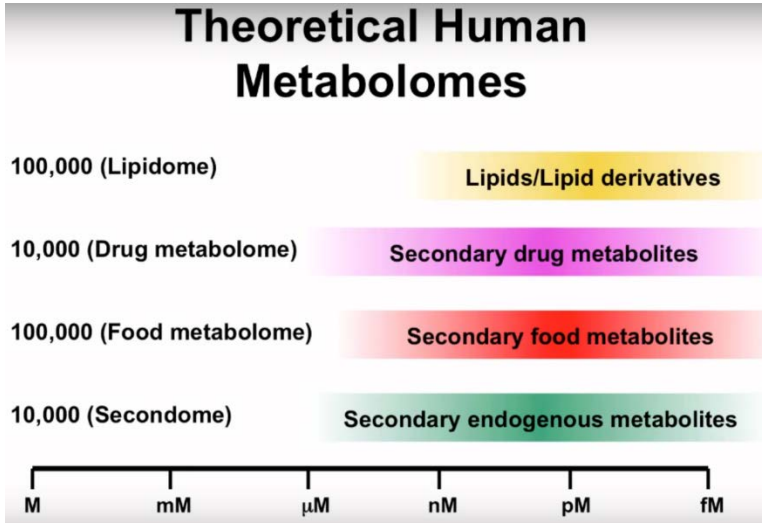


Metabolites are important

- >95% of all diagnostic clinical assays test for small molecules
- 89% of all known drugs are small molecules
- 50% of all drugs are derived from pre-existing metabolites
- 30% of identified genetic disorders involve diseases of small molecule metabolism
- Small molecules serve as cofactors and signaling molecules to 1000's of proteins

Metabolomics can therefore be seen as bridging the gap between genotype and phenotype



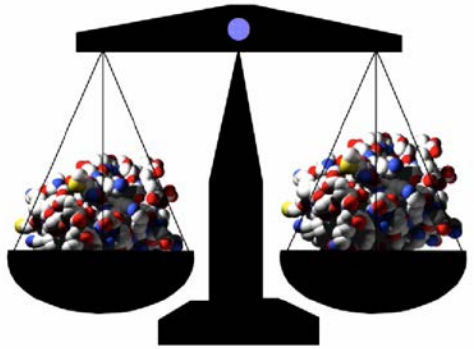


Metabolomics technologies

- UPLC, HPLC
- CE/microfluidics
- LC-MS
- FT-MS
- QqQ-MS
- NMR spectroscopy
- X-ray crystallography
- GC-MS
- FTIR

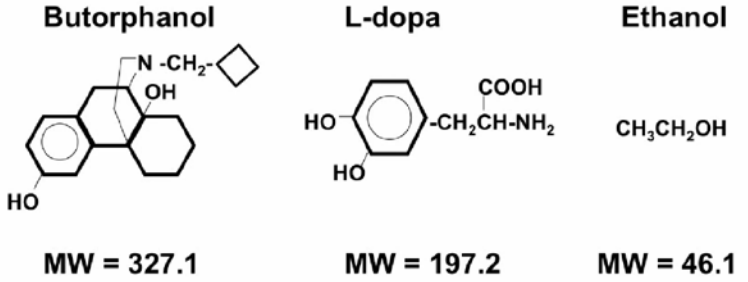
Mass Spectrometry

Analytical method to measure the molecular or atomic weight of samples



MS Principles

- Different compounds can be uniquely identified by their mass

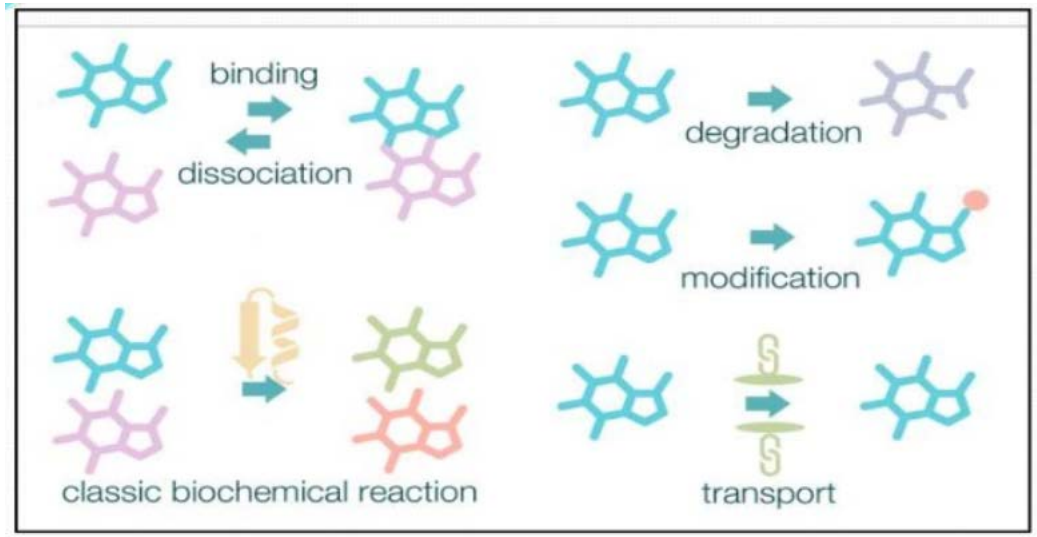


Metabolomics – ‚a snapshot' in time

Conceptual approaches in metabolomics:

• Target analysis: has been applied for many decades and includes the determination and quantification of a small set of known metabolites (targets) using one particular analytical technique of best performance for the compounds of interest.

• Metabolite profiling: aims at the analysis of a larger set of compounds, both identified and unknown with respect to their chemical nature. This approach has been applied for many different biological systems using GC-MS, including plants, microbes, urine, and plasma samples.



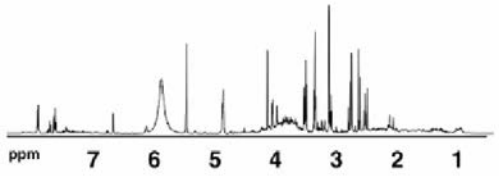
A diagram showing the main different types of metabolic reactions that take place in a cell. These are shown as they are represented in the database *Reactome*.

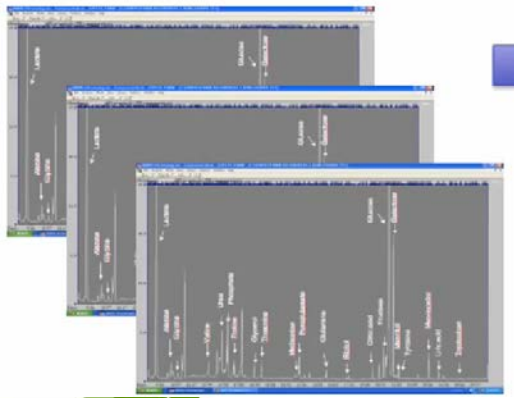
• Metabolomics: employs complementary analytical methodologies, for example, LC-MS/MS, GC-MS, and/or NMR, in order to determine and quantify as many metabolites as possible, either identified or unknown compounds.

• Metabolic fingerprinting: a metabolic "signature" or mass profile of the sample of interest is generated and then compared in a large sample population to screen for differences between the samples. When signals that can significantly discriminate between samples are detected, the metabolites are identified and the biological relevance of that compound can be elucidated, greatly reducing the analysis time.

Metabolomics data analysis

From Spectra to Lists

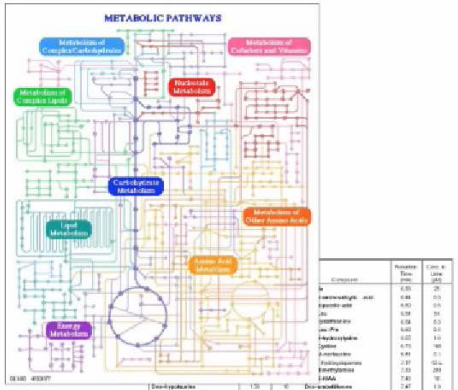


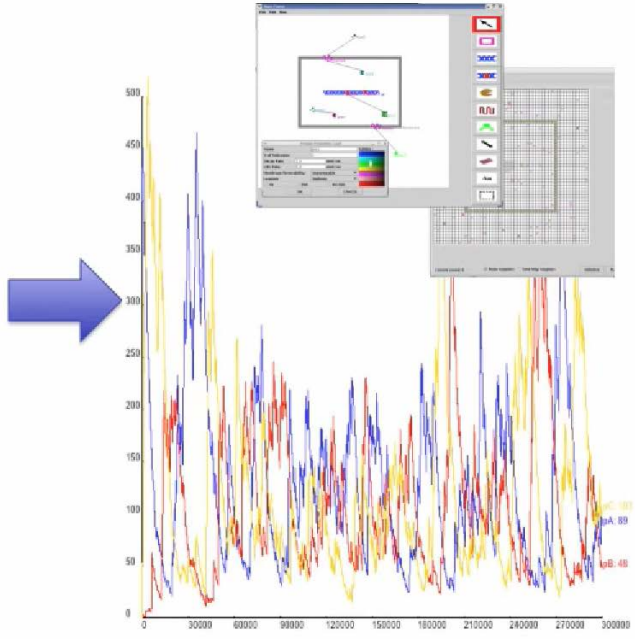


From Lists to Pathways

Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L.	Dns-Ile	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	0.99	<D.L.	Dns-pipecolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathionine	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	Dns-Leu-Pro	6.60	0.4
Dns-6-dimet hylamine purine	1.20	<D.L.	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl -histidine	1.22	80	Dns-Cystine	6.73	160
Dns-taurine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferone	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3-diaminopropionic acid	7.53	<D.L.
Dns-guanidine	1.62	<D.L.	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetyamidophenol	7.73	51
Dns-allantoin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-citrulline	1.87	2.9	Dns-homocystine	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylguanidine	2.20	<D.L.	Dns-5-methyo- xysalicylic acid	8.04	2.1
Dns-Ser	2.24	511	Dns-Lys	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	Dns-Leu-Phe	8.22	0.3
Dns-Glu	2.57	21	Dns-His	8.35	1560
Dns-Asp	2.60	90	Dns-4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<D.L.
Dns-epinephrine	3.05	<D.L.	Dns-1-ephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.53	0.4
Dns-aminoadipic acid	3.17	80	Dns-pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.66	533	Dns-5-hydroxytryptophan	9.25	0.22
Dns-4-aminolevulinic acid	3.97	30	Dns-1,3-diaminopropane	9.44	0.23
Dns-4-amino-butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns-p-amino-hippuric acid	3.98	2.9	Dns-1,2-diaminopropane	9.66	0.1
Dns-5-hydro xymethyluricil	4.08	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	<D.L.	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-methoxy -tyramine	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cystamine	10.44	<D.L.

From Pathways & Lists to Models & Biomarkers





Where to look for metabolomics data

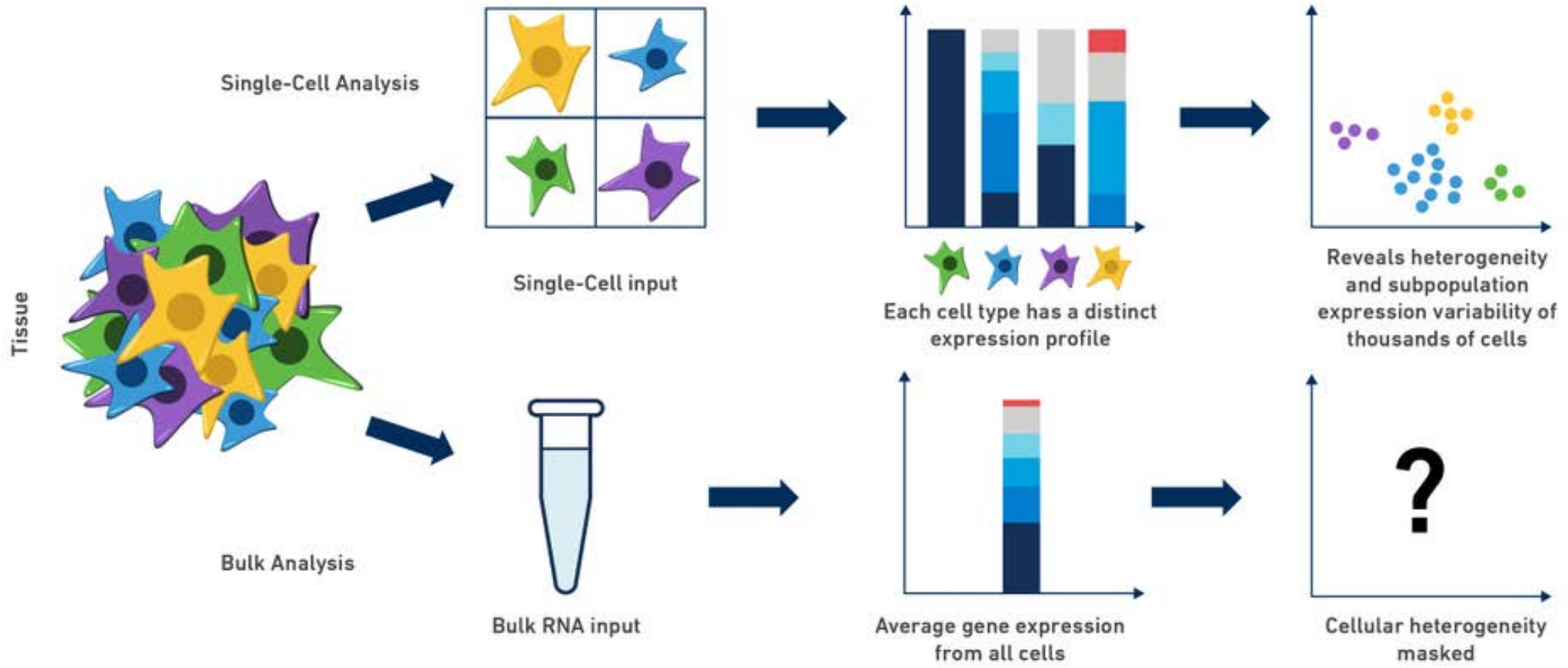
Metabolic pathway databases

- Pathway viewers KEGG (http://www.genome.ad.jp/kegg/),
- Atomic Reconstruction of Metabolism database (http://www.metabolome.jp/),
- BioCyc (http://biocyc.org) (Paleyand Karp 2006),
- MetaCyc (http://metacyc.org/) (Caspiet al. 2006),
- AraCyc (http://www.Arabidopsis.org/tools/ aracyc/) (Zhang et al. 2005), MapMan (http://gabi.rzpd.de/projects/MapMan/)
- (Thimm et al. 2004), KaPPA-View (http://kpv.kazusa.or.jp/kappa-view/) (Tokimatsu et al.2005) and
- BioPathAT (http://www.ibc.wsu.edu/research/lange/public%5Ffolder/) (Lange and Ghassemian 2005),
- the data model for plant metabolomics experiments ArMet (http://www.armet.org/)

https://www.slideshare.net/TNAUgenomics/metabolomics-13725538?next_slideshow=1

Cutting edge: Single-cell -omics

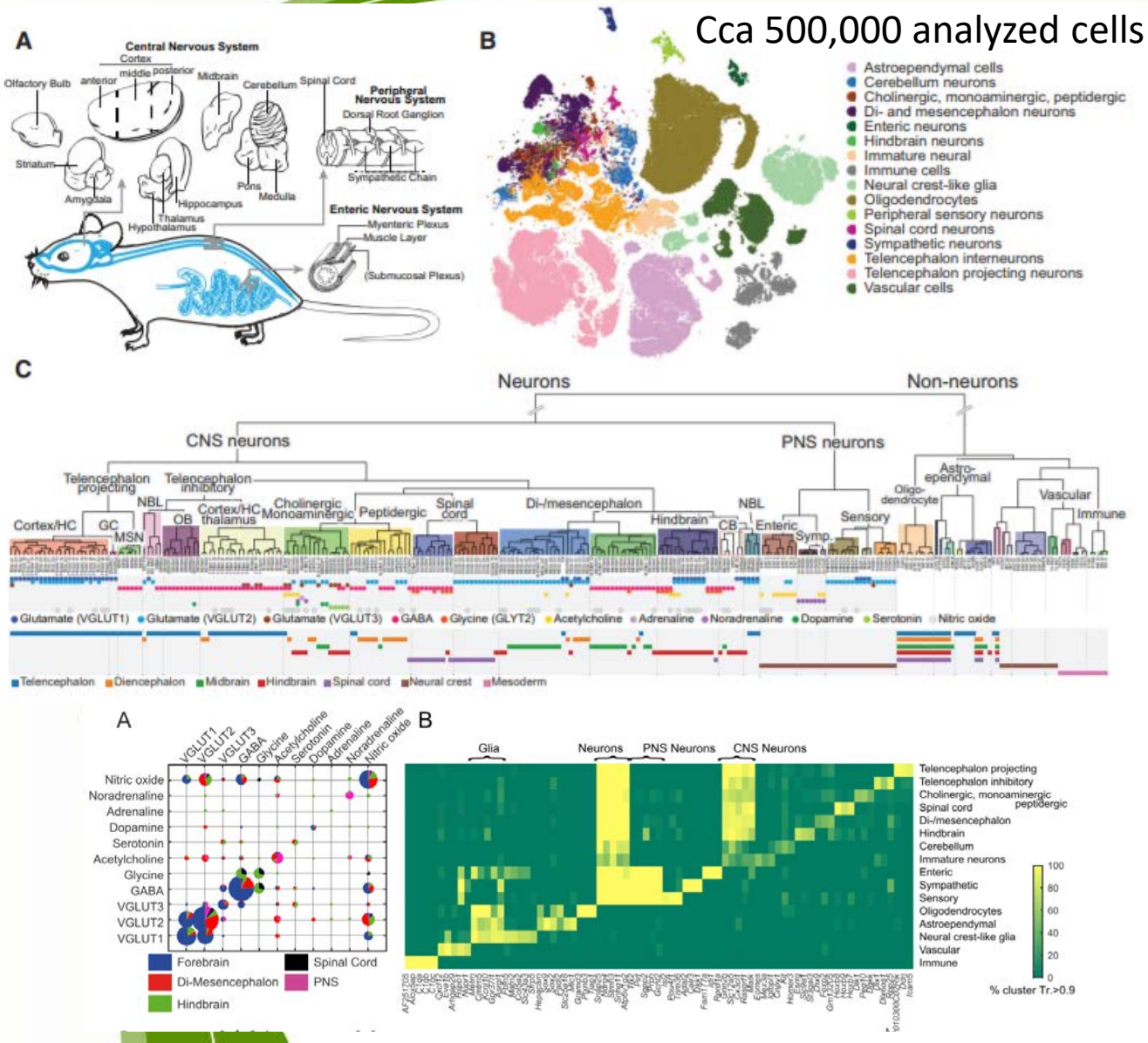
- Application of whole genome, whole transcriptome sequencing and other –omics methods to single cells, scRNA-seq is now the top method



https://community.10xgenomics.com/t5/10x-Blog/Single-Cell-RNA-Seq-An-Introductory-Overview-and-Tools-for/ba-p/547

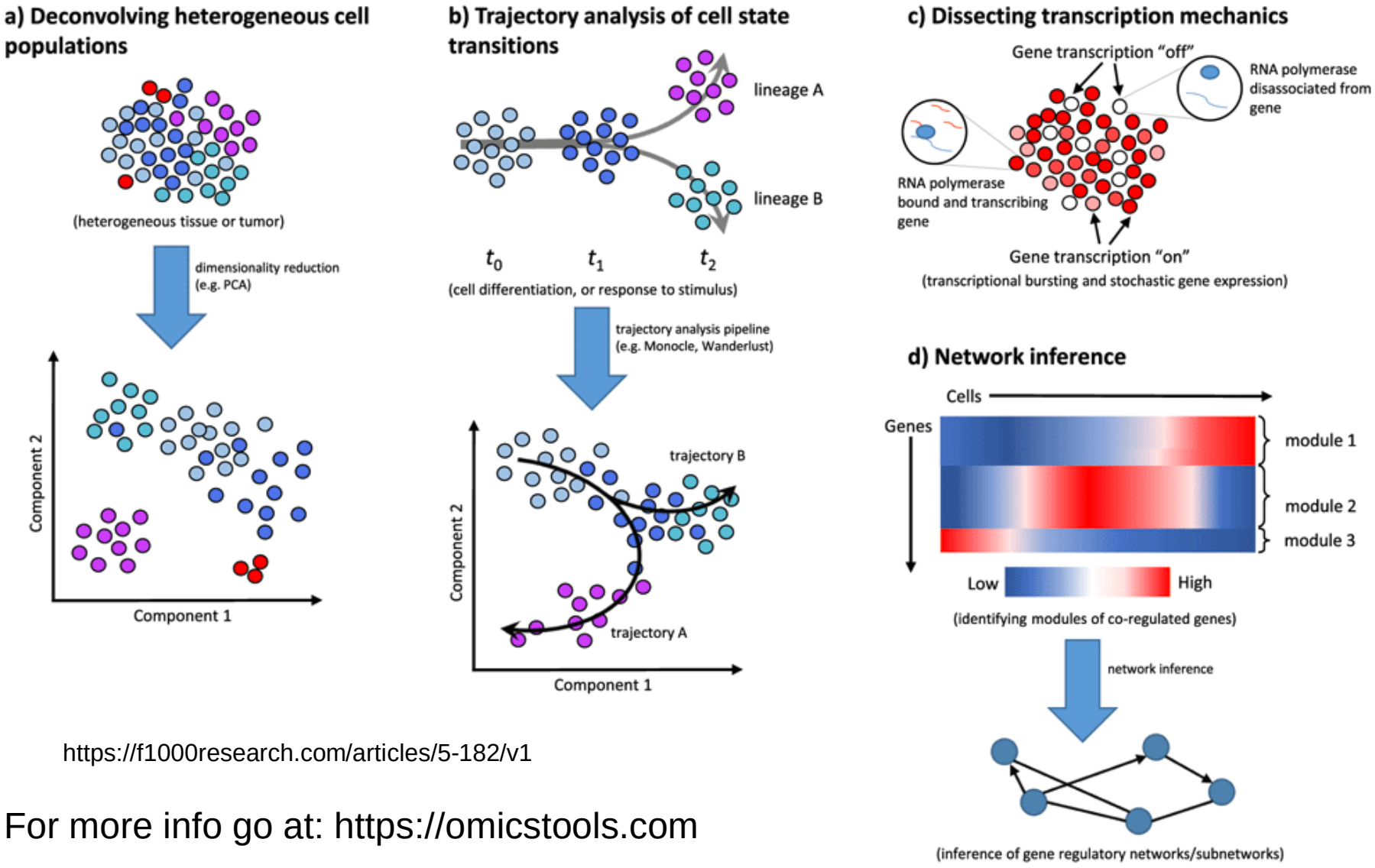
scRNA-seq data visualization

Cca 500,000 analyzed cells



Molecular Architecture of the Mouse Nervous System
DOI:https://doi.org/10.1016/j.cell.2018.06.021

Common applications of scRNA-seq

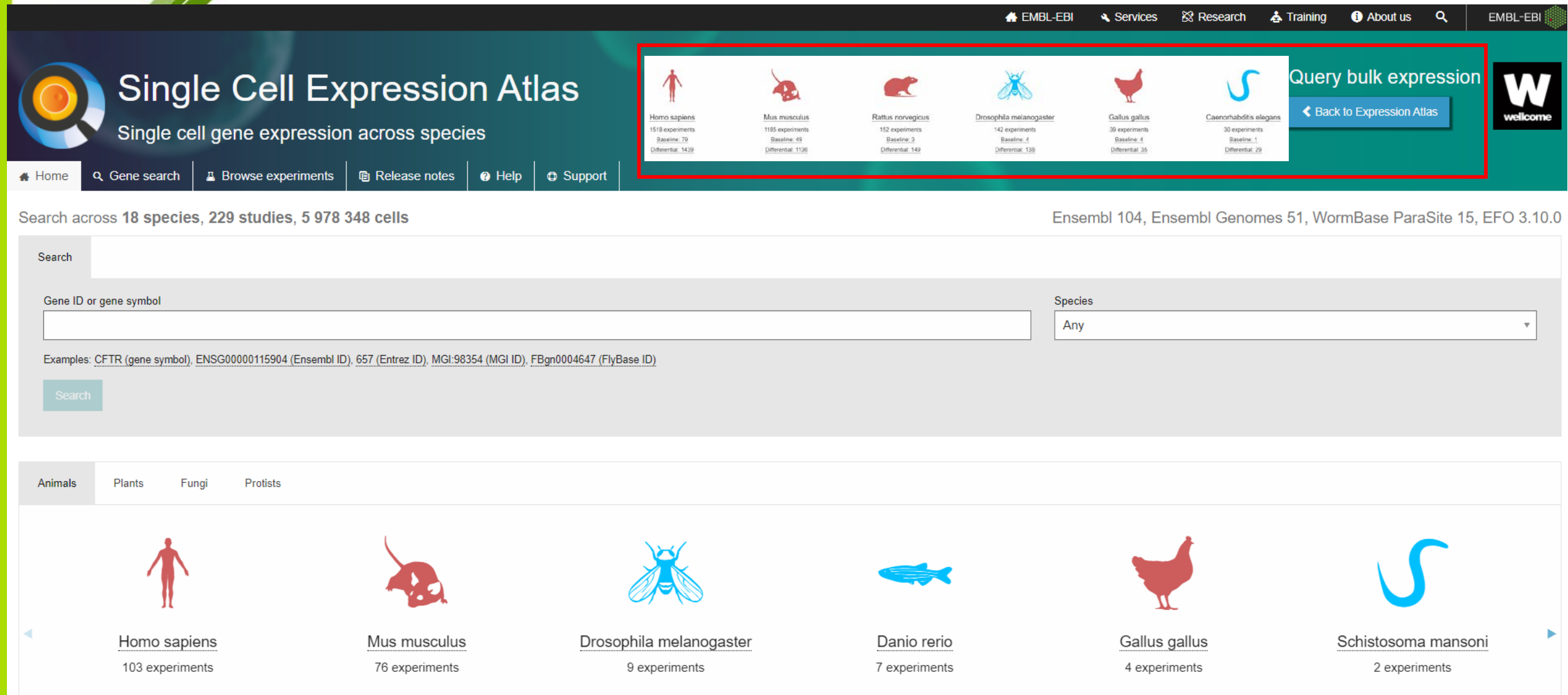


https://f1000research.com/articles/5-182/v1

For more info go at: https://omicstools.com

ScRNA-seq databases

https://www.ebi.ac.uk/gxa/sc/home



Single-cell multi-omics

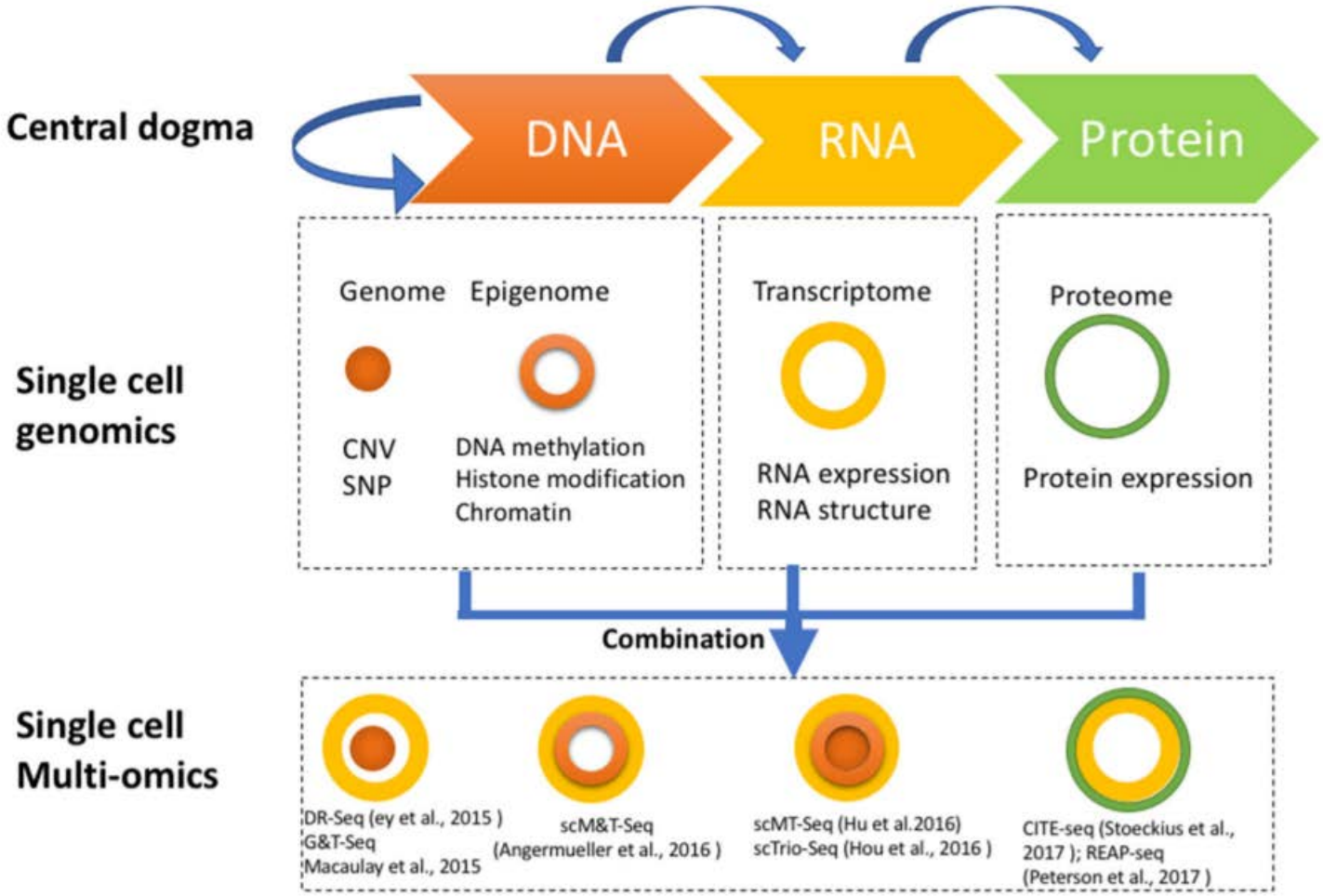


FIGURE 2 | Strategies for multi-omics profiling of single cells. Three major types of molecules relating to biological central dogma **(Top)**. Single cell genomics methods profiling the genome, epigenome, transcriptome, and proteome are shown by different shapes with variable colors **(Middle)**. Single cell multi-omics methods are built by combining different single cell sequencing methods to simultaneously profile multiple types of molecules of a single cell genome wide **(Bottom)**. For example, G&T-seq was built by combining genome (orange) and transcriptome (yellow) to simultaneously detect DNA and RNA of the same cell genome wide.

https://www.frontiersin.org/articles/10.3389/fcell.2018.00028/full

Challenges:

- There are no commercial kits available yet for any single-cell multi-omics techniques, and many are technically challenging.
- Researchers must modify existing single-cell protocols so that they're compatible with multiple types of molecules and take great care to minimize the loss or contamination of samples

https://www.the-scientist.com/lab-tools/integrating-multiple--omics-in-individual-cells-64829

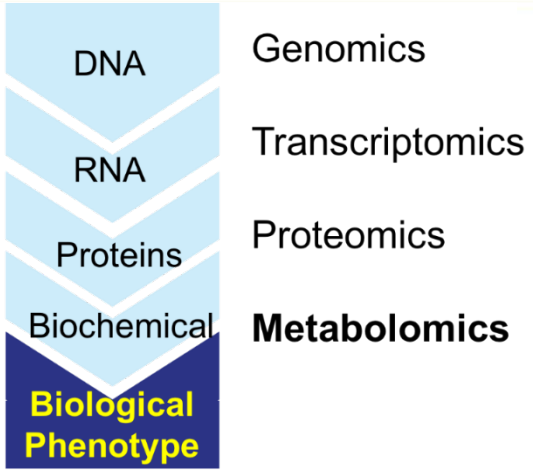
Difficulty squared

Combining modalities only multiplies the difficulty. All the weaknesses, all the noise, all the challenges from each technology, it just gets exacerbated by combining them into a multimodal assay.

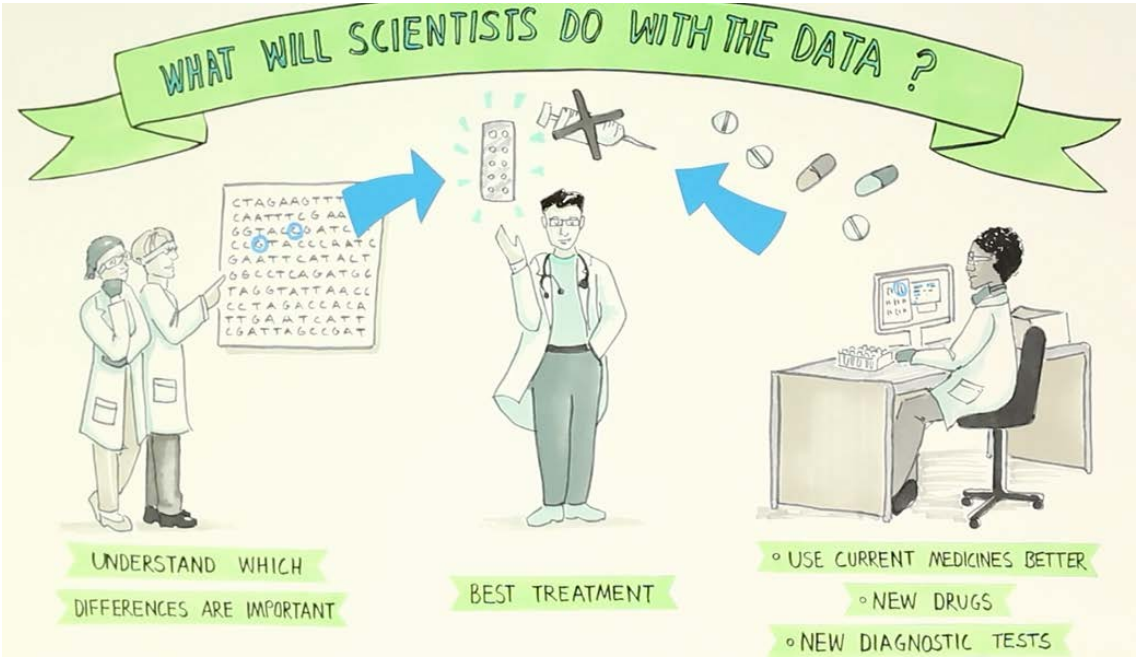
Single-cell analysis enters the multiomics age
https://www.nature.com/articles/d41586-021-01994-w#correction-1

Summary

- Omics technologies - „the data deluge“
- Genomics and Transcriptomics rely on two main approaches: microarrays (hybridization) and NGS (sequencing by synthesis)
- Proteomics and Metabolomics rely heavily on mass spectrometry



- Omics technologies are revolutionizing science and medicine
- From data to actionable knowledge - Integrated Omics data
- Precision medicine is the ultimate goal of many –omics efforts
- Despite the progress made we have still a long way to go …



Take home messages

- We have been generating Big data, but we hardly understand it ☹
- Big data is publicly available, go through the databases before you even start planing your experiment – it can save you enourmous time and money
- Databases contain huge datasets of patients you would never be able to gather by yourself, test your hypothesis in silico before the „wet-lab“ work
- If you cannot find the „yes/no“ or „a few genes“ answer, use the Cellular/functional/pathway analyses to help you out ☺
- Learning bioinformatics skills (e.g. programing in R) is a good investment plan for your future (scientific) career

Thank you for your attention

Any Questions?

- Jay Flatley, Executive Chairman of Illumina:
- *„Everyone is going to get sequenced, it is gonna be part of their health record and it will be used to manage their health care throughout their lifetime“.*