SOC-B CENTRE FOR DOCTORAL TRAINING IN BIOSOCIAL RESEARCH

'OMICS & SYSTEMS BIOLOGY

Adam Stevens

Lecturer in Child Health and Growth

Division of Developmental Biology and Medicine

INTRODUCTIONS

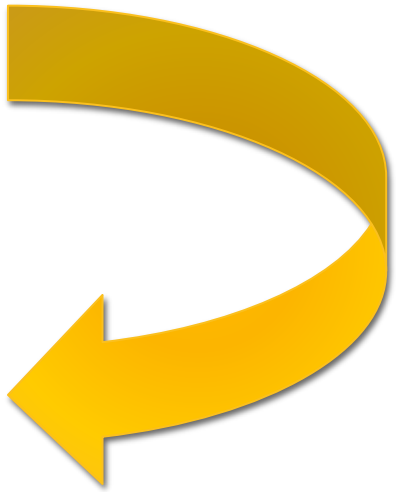
Overview

- What is biological information?
- Types of ‘omic data
- Basic analysis and tools
- Network analysis of ‘omic data

WHAT IS BIOLOGICAL INFORMATION?

Genotype *vs.* Phenotype

- **Genotype**
  - The genetic makeup of an organism
    - An organisms complete set of genes
    - Instructions for building and maintaining
      - Formation of proteins, regulation of metabolism
    - Genetic traits
    - Internally coded- not observed
    - Copied during cell division & reproduction



- **Phenotype**
  - Observable physical properties of an organism
    - Appearance, development & behavior

Phenotype is determined by an organisms genotype and ALSO environmental factors

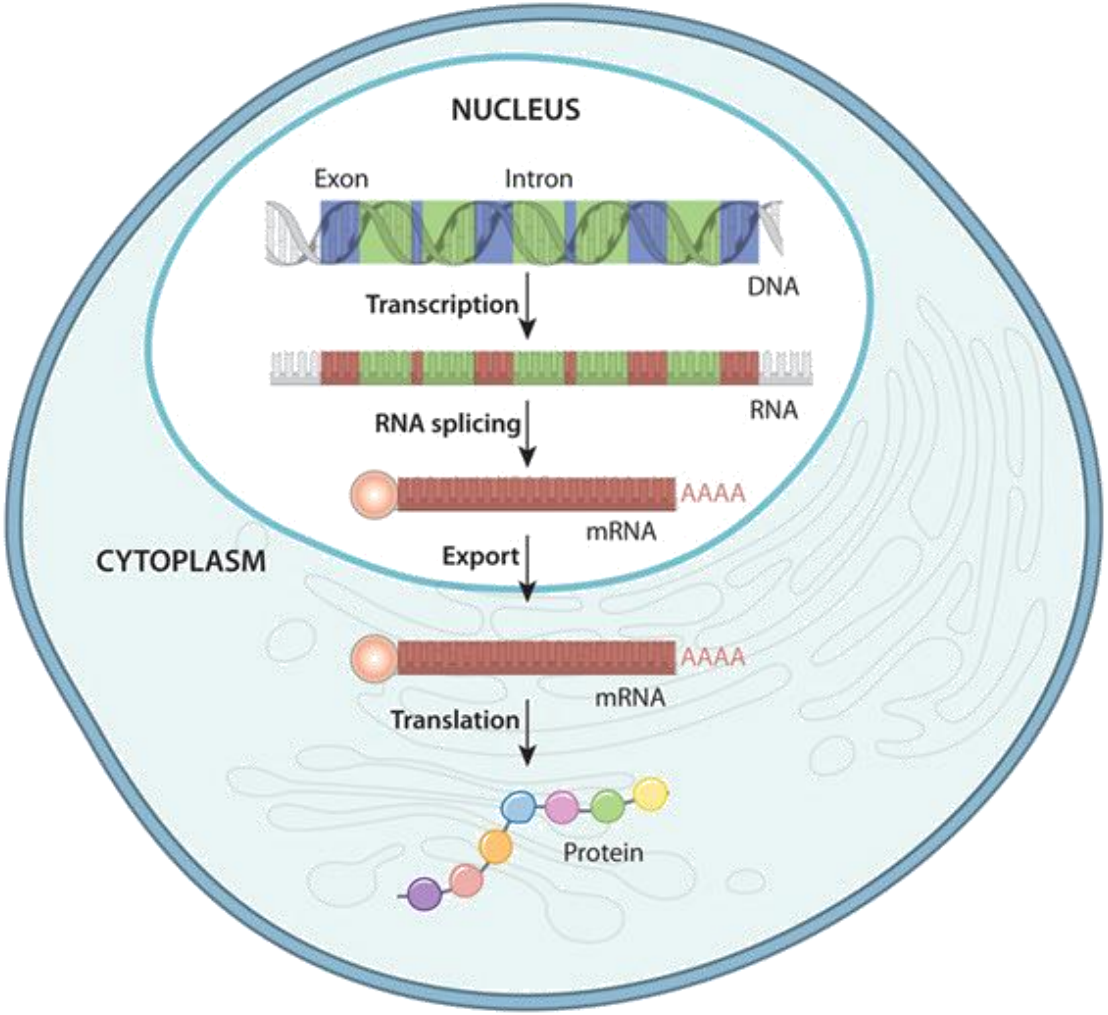
Genes + Environment = Phenotype

Central Paradigm of Molecular Biology

Genes to messenger to proteins

DNA → mRNA → Protein

Biological information flow



Replication

DNA **Information**

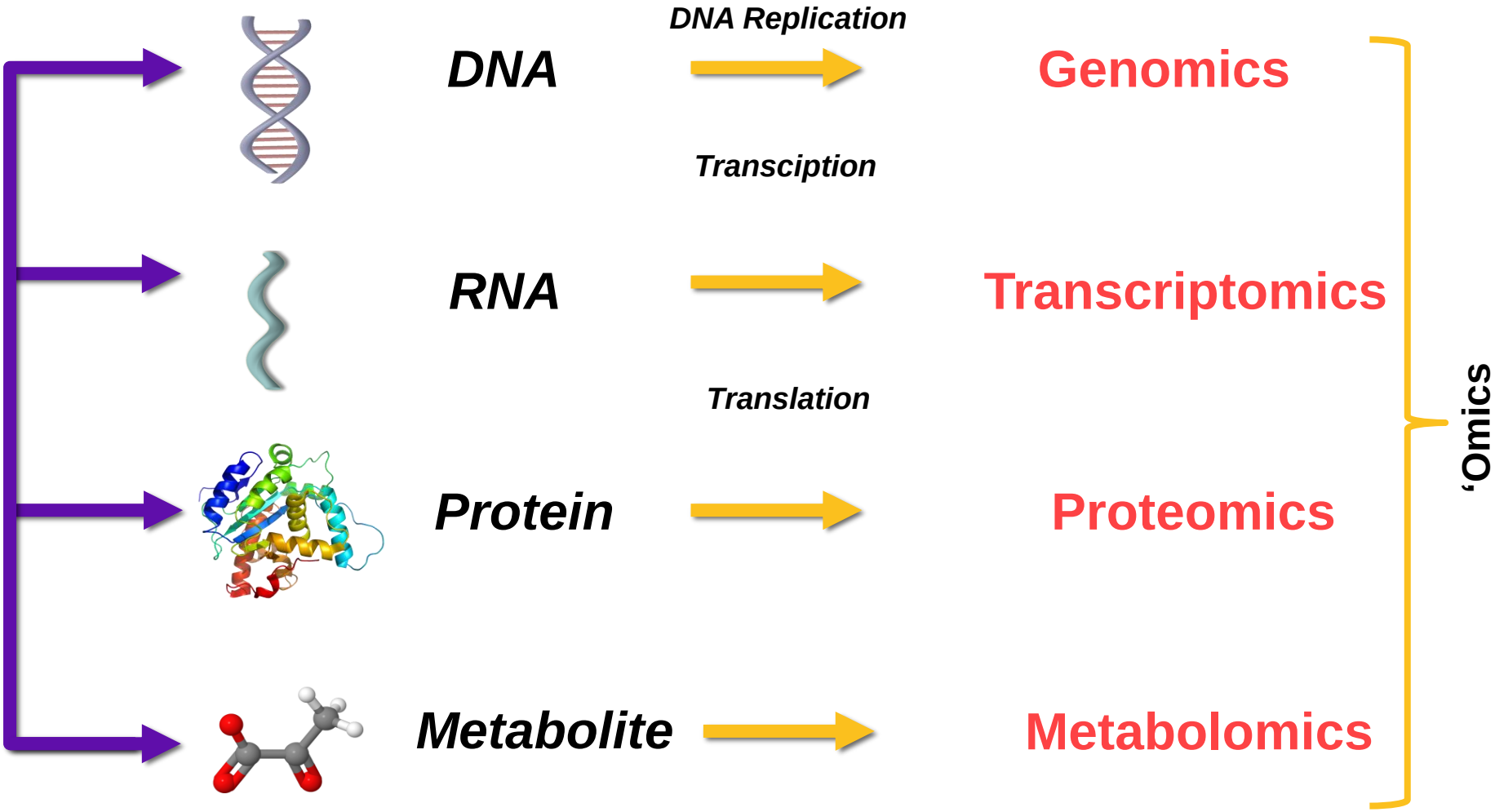
Transcription

RNA **Flow/Process**

Translation

Protein **Function**

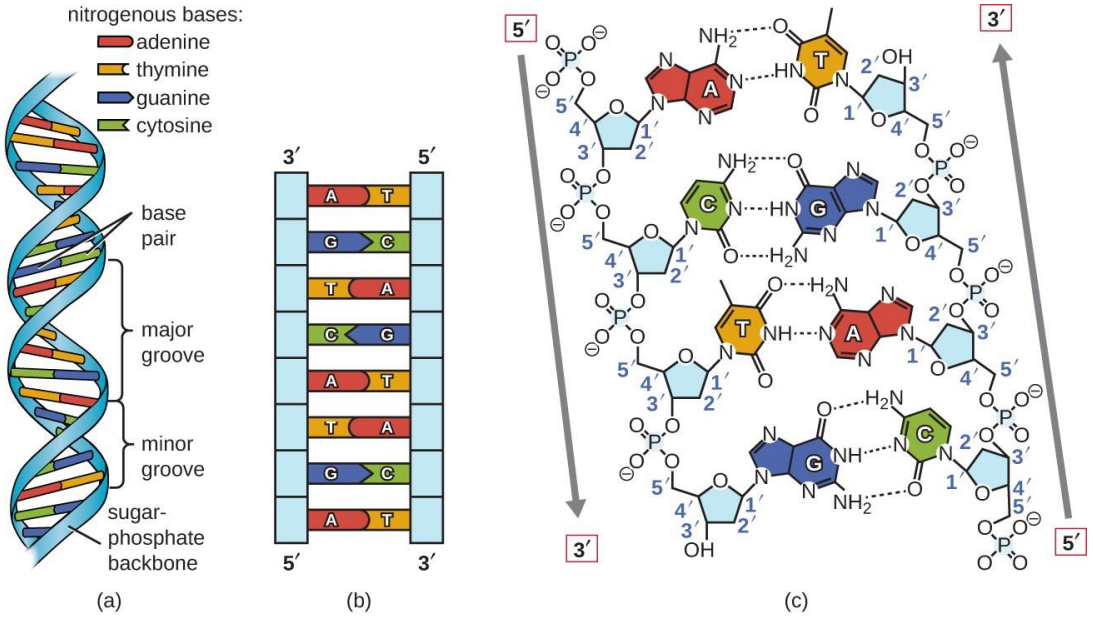
Central dogma of molecular biology



DNA & Genomics

- Deoxyribonucleic acid (DNA)
  - Polymer of nucleotides
  - Sequence of nucleotides is responsible for carrying and retaining hereditary information in a cell – Base Sequence

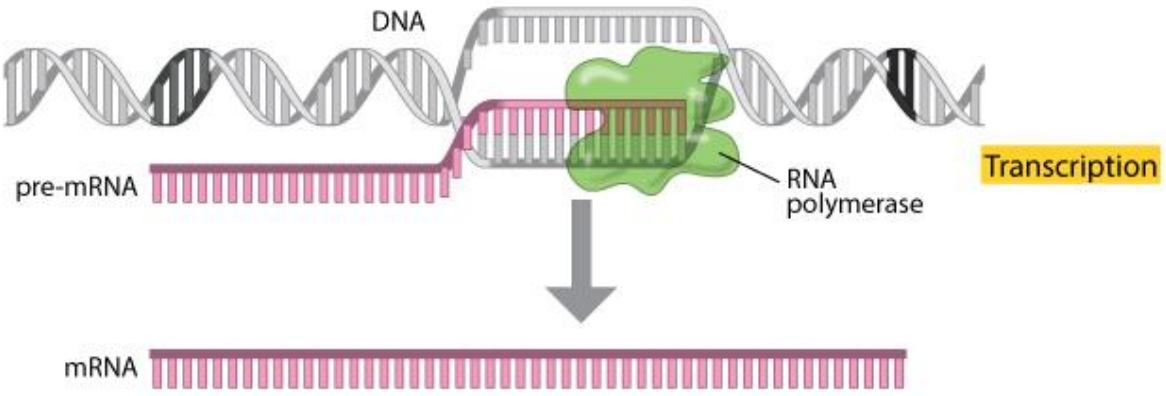
- Double helix of complementary base pairs
- Nitrogenous base
  - Adenine and Thymine
  - Cytosine and Guanine
- Phosphate group
- Deoxyribose



https://courses.lumenlearning.com/microbiology/chapter/structure-and-function-of-dna/

RNA & Transciptomics

- Ribonucleic acid (mRNA)
  - Result of TRANSCRIPTION
  - Information encoded with the DNA sequence of one or more genes is TRANSCRIBED into a strand of RNA –RNA transcript
    - Single stranded
    - A,G,C,U (T)



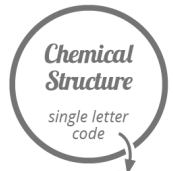
Translation: DNA to mRNA to Protein
By: Suzanne Clancy, Ph.D. & William Brown, Ph.D. (Write Science Right) © 2008 Nature Education
Citation: Clancy, S. & Brown, W. (2008) Translation: DNA to mRNA to Protein. Nature Education 1(1):101

Proteins & Proteomics

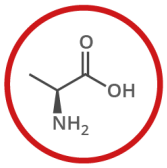
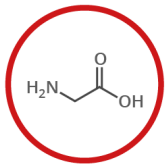
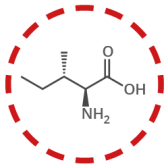
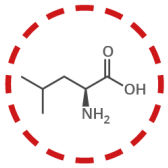
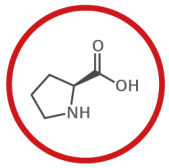
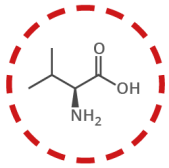
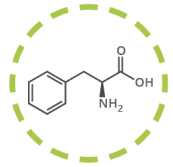
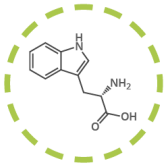
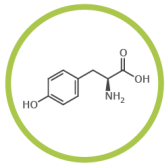
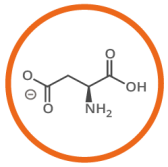
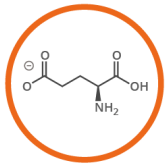
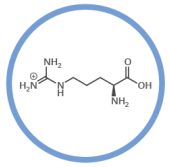
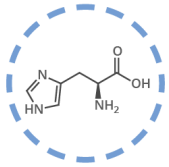
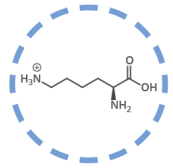
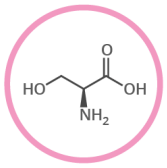
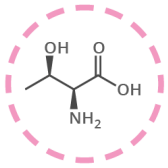
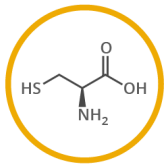
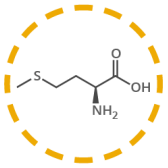
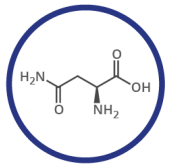
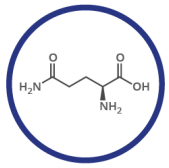
A GUIDE TO THE TWENTY COMMON AMINO ACIDS

AMINO ACIDS ARE THE BUILDING BLOCKS OF PROTEINS IN LIVING ORGANISMS. THERE ARE OVER 500 AMINO ACIDS FOUND IN NATURE - HOWEVER, THE HUMAN GENETIC CODE ONLY DIRECTLY ENCODES 20. 'ESSENTIAL' AMINO ACIDS MUST BE OBTAINED FROM THE DIET, WHILST NON-ESSENTIAL AMINO ACIDS CAN BE SYNTHESISED IN THE BODY.

Chart Key: ALIPHATIC · AROMATIC · ACIDIC · BASIC · HYDROXYLIC · SULFUR-CONTAINING · AMIDIC · NON-ESSENTIAL · ESSENTIAL



NAME (A)
three letter code
DNA codons

						
ALANINE (A)	GLYCINE (G)	ISOLEUCINE (I)	LEUCINE (L)	PROLINE (P)	VALINE (V)	
Ala	*Gly*	*Ile*	*Leu*	*Pro*	*Val*	
GCT, GCC, GCA, GCG	GGT, GGC, GGA, GGG	ATT, ATC, ATA	CTT, CTC, CTA, CTG, TTA, TTG	CCT, CCC, CCA, CCG	GTT, GTC, GTA, GTG	
						
PHENYLALANINE (F)	TRYPTOPHAN (W)	TYROSINE (Y)	ASPARTIC ACID (D)	GLUTAMIC ACID (E)	ARGININE (R)	HISTIDINE (H)
Phe	*Trp*	*Tyr*	*Asp*	*Glu*	*Arg*	*His*
TTT, TTC	TGG	TAT, TAC	GAT, GAC	GAA, GAG	CGT, CGC, CGA, CGG, AGA, AGG	CAT, CAC
						
LYSINE (K)	SERINE (S)	THREONINE (T)	CYSTEINE (C)	METHIONINE (M)	ASPARAGINE (N)	GLUTAMINE (Q)
Lys	*Ser*	*Thr*	*Cys*	*Met*	*Asn*	*Gln*
AAA, AAG	TCT, TCC, TCA, TCG, AGT, AGC	ACT, ACC, ACA, ACG	TGT, TGC	ATG	AAT, AAC	CAA, CAG

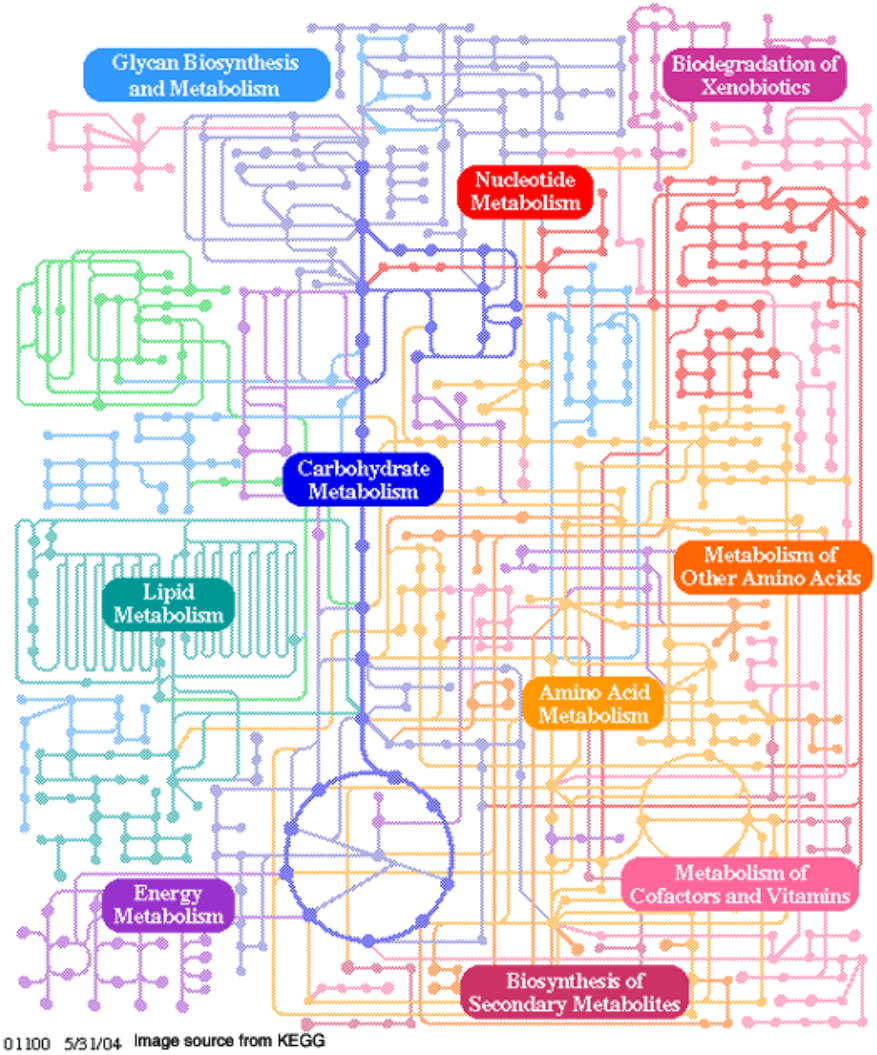
Note: This chart only shows those amino acids for which the human genetic code directly codes for. Selenocysteine is often referred to as the 21st amino acid, but is encoded in a special manner. In some cases, distinguishing between asparagine/aspartic acid and glutamine/glutamic acid is difficult. In these cases, the codes asx (B) and glx (Z) are respectively used.

© COMPOUND INTEREST 2014 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.

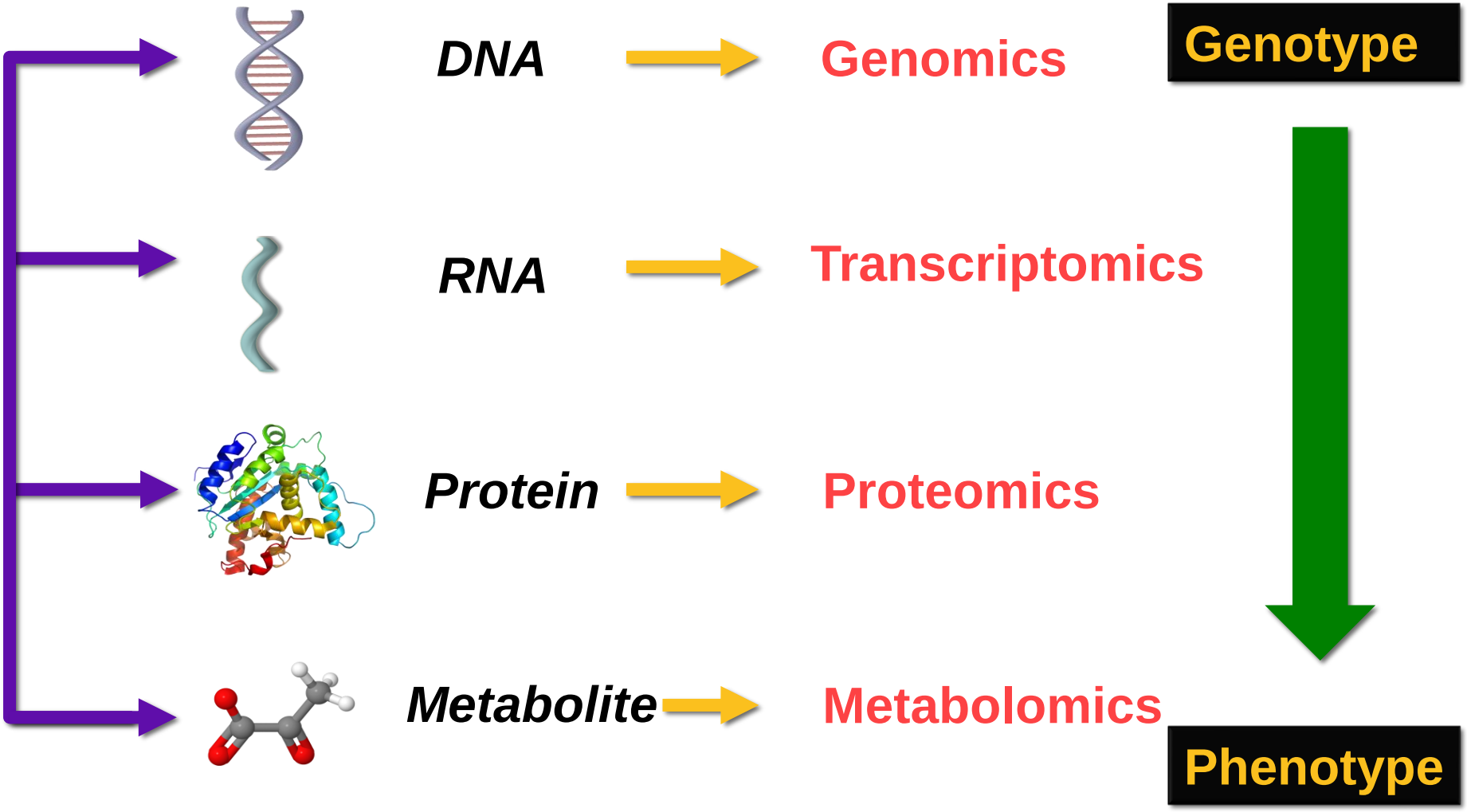
Metabolites & Metabolomics

- Low molecular weight produ
- Metabolism is a complex inte that occur within cells
- Diverse roles
  - Energy metabolism
  - Amino acid metabolism
  - Lipid metabolism

- Primary metabolism –directly development
- Secondary metabolism – en
  - Antibiotics, steroids *etc*



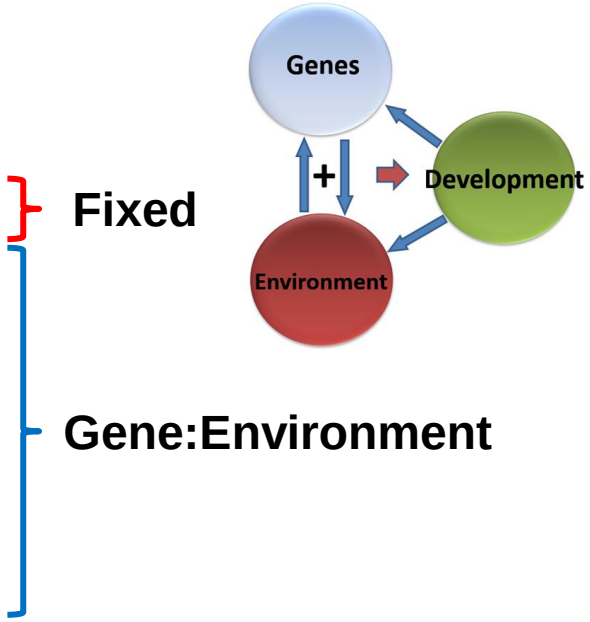
Central dogma of molecular biology



What is 'Omic Data

Omic data sets include:

- Genetics (SNPs, CNVs)
- Transcriptomics (Affymetrix, RNAseq)
- Epigenomics (DNA methylation, histone mods)
- ChIPseq
- Metabolomics
- Proteomics
- Phosphoproteomics



All have specific quality control (QC) issues and difficulties in analysis
All rely on the use of a false discovery rate correction (FDR) for analysis

frontiers in GENETICS

PERSPECTIVE ARTICLE
published: 06 January 2012
doi: 10.3389/fgene.2011.00105

When one and one gives more than two: challenges and opportunities of integrative omics

Hyungwon Choi[1] and Norman Pavelka[2]**

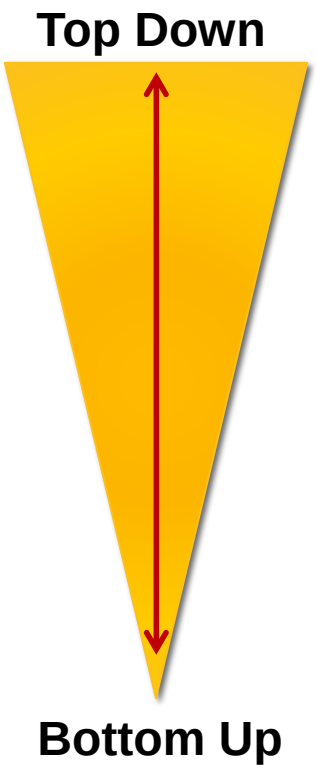
Multi-omic analysis is the integration of different omic data sets

TYPES OF 'OMICS DATA?

Why 'omics?

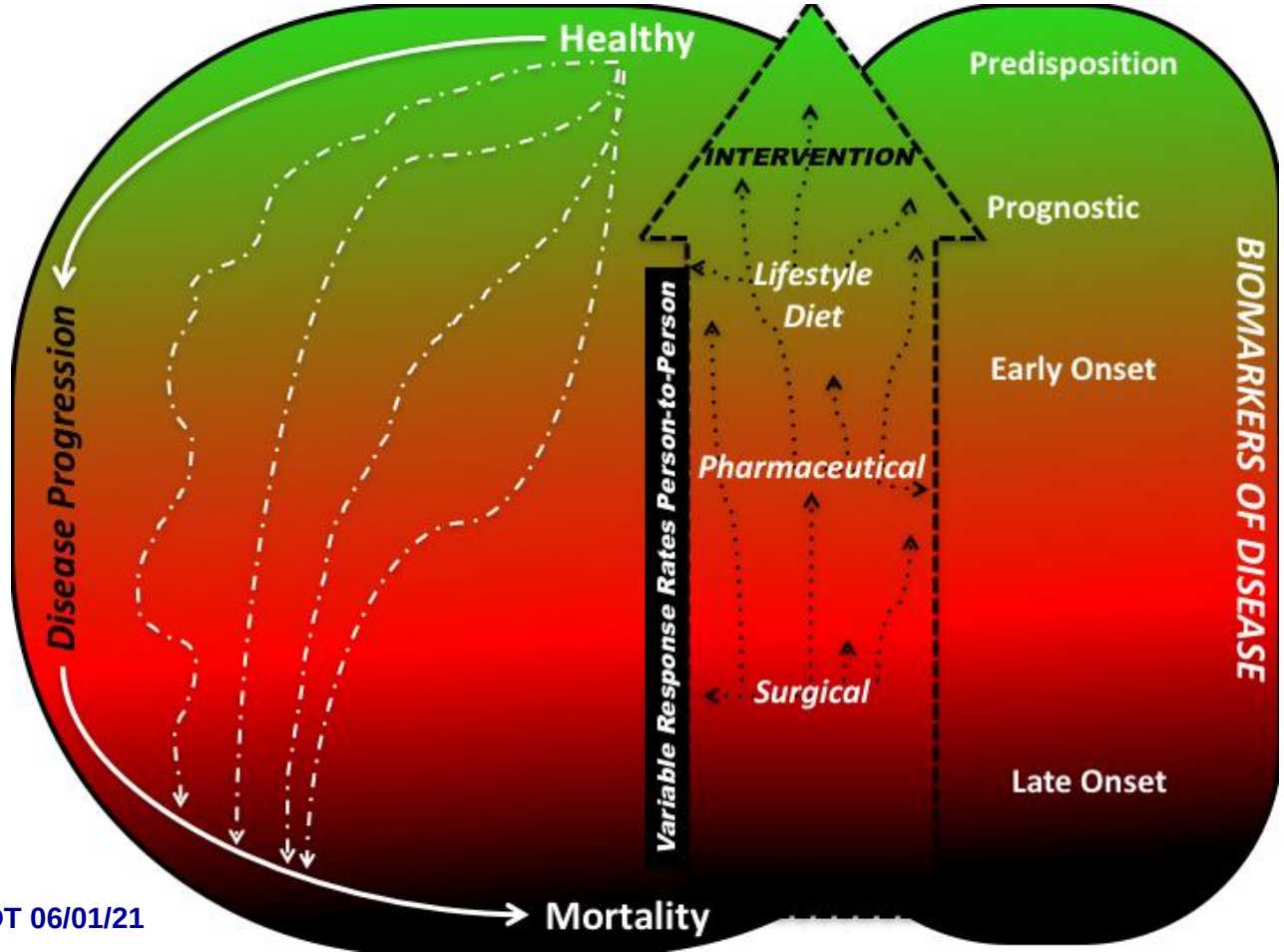
- Adopts an holistic view of all 'molecules' that make up a cell, tissue or organism
- Universal approach/Hypothesis-generating
- No analysis bias

- Many applications
  - '**BIOMARKER**' discovery
  - Early detection/population screening
  - Increasing understanding of disease aetiology
  - Drug discovery & toxicity and efficacy screens



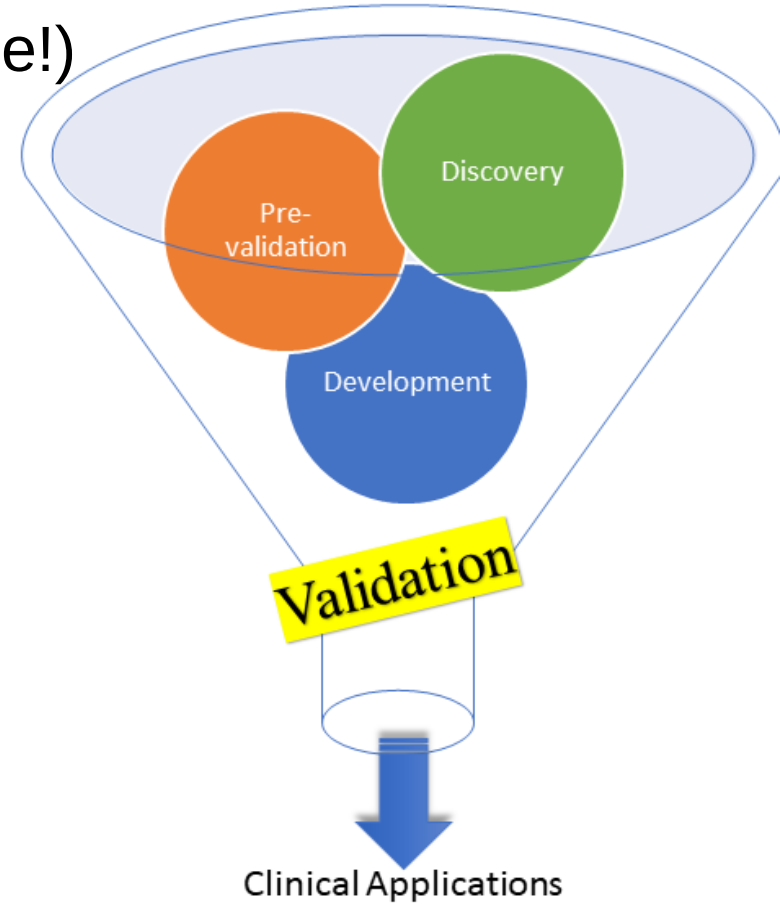
Biomarkers of disease

In medicine, a **biomarker** is a measurable indicator of the severity or presence of some disease state. More generally a **biomarker** is anything that can be used as an indicator of a particular disease state or some other physiological state of an organism.



Experimental Design

- Case *vs.* Control (an appropriate one!)
- Samples numbers?
  - How many is sufficient?
  - Test cohort + validation cohort?
- Quality of samples
  - Collection/storage/processing
- Confounding factors?
  - Age, gender, environment



Is this good experimental design?

- Metabolomics investigation of liver failure from plasma samples

Table 1. Demographic Information of the Healthy Group and Liver Failure Patient Group Investigated[a]

	healthy group ($n = 23$)	patient group ($n = 24$)
Gender (male/female)	15/8	21/3
HBsAg	Negative	Positive
Age (year)	27.39 ± 9.24	46.77 ± 13.35
ALT (U/L)	<40	172.63 ± 147.49
TB (μmol/L)	<12	457.33 ± 135.48
PT (s)	<14	26.06 ± 15.14
MELD score	/	24.68 ± 8.38

[a] Abbreviations: ALT, alanine aminotransferase; TB, total bilirubin; PT, prothrombin time; MELD, model for end-stage liver disease. The value is represented as the form of mean ± SD.

Or what about this one?

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

T. R. Golub[1,2,*,†], D. K. Slonim[1,†], P. Tamayo[1], C. Huard[1], M. Gaasenbeek[1], J. P. Mesirov[1], H. Coller[1], M. L. Loh[2], J. R. Downing…
+ See all authors and affiliations

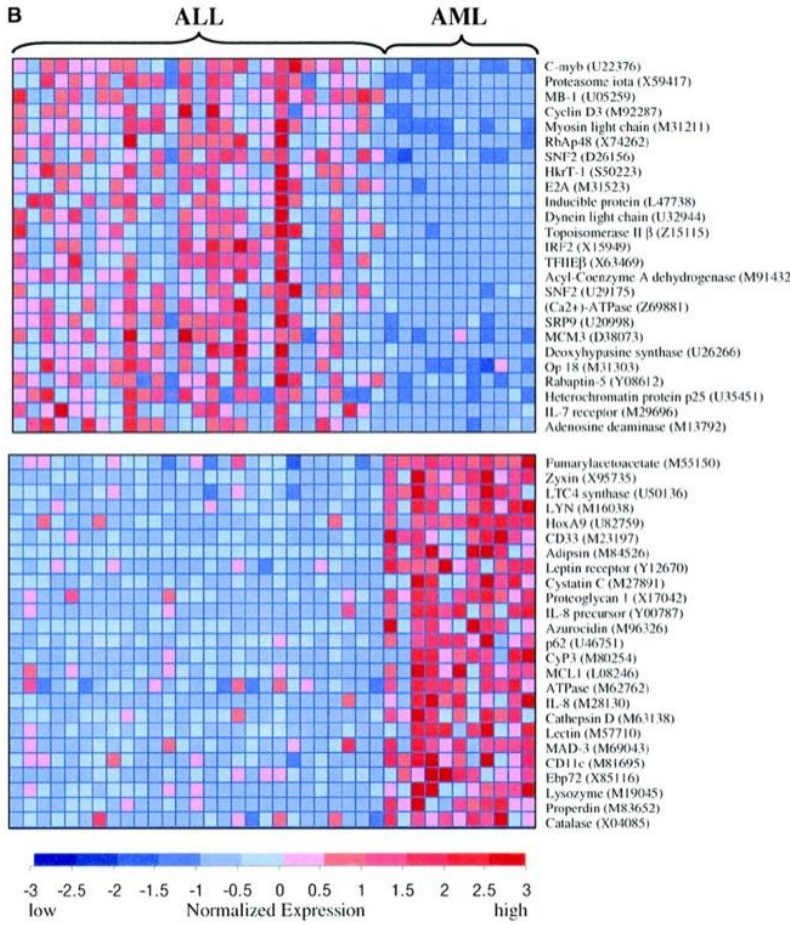
Science 15 Oct 1999:
Vol. 286, Issue 5439, pp. 531-537
DOI: 10.1126/science.286.5439.531

Generic approach to cancer classification based on gene expression monitoring by DNA microarrays applied to human acute leukemias

38 Affymetrix microarrays with 6,817 probes

27 from childhood acute lymphoblastic leukemia

11 from adult acute myeloid leukemia



Or this?

Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning

Louise C. Kenny[a,*], Warwick B. Dunn[b], David I. Ellis[b], Jenny Myers[a], Philip N. Baker[a] and the GOPEC Consortium, and Douglas B. Kell[b,*]

- Pre-eclampsia - Pregnancy-induced hypertension which may affect mother and foetus

Table 1
Demographic data for patients from whom plasma samples were taken

	Normal outcome $n=87$	Preeclampsia $n=87$
Age	30 (19–43)	31 (19–41)
Parity	0 (0–2)	0 (0–2)
BMI (weight/height^2)	25 (19–46)	26 (18–46)
Max (S) BP (mm Hg)	122 (96–147)	162 (138–220)*
Max (D) BP (mm Hg)	80 (60–93)	110 (90–140)*
Delivery gestation (weeks + days)	40 + 4 (34 + 3 to 42 + 0)	37 + 0* (26 + 3 to 41 + 1)
Birth weight (g)	3420 (2380–4420)	2410 (590–4300)*
IBR (centile)	34 (10–99)	8 (0–99)*

Median (range).
Pre-eclampsia vs normal outcome.
*$p < 0.0001$.

Genomics

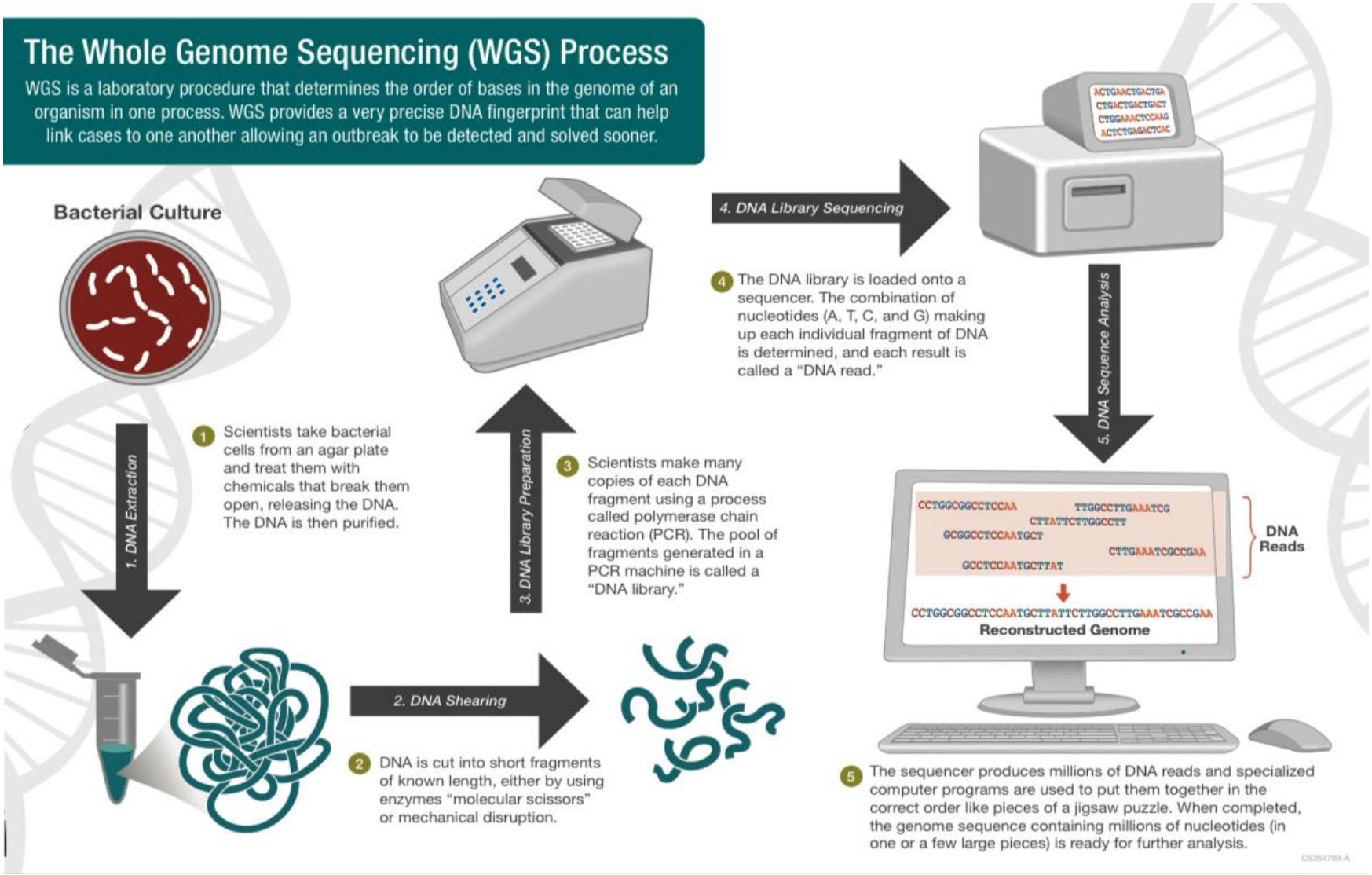
- Genome = total DNA of a cell or an organism
- Human genome = 3.2 billion bases and estimated > 30,000 protein coding genes

- Seeking mutations or alterations that may contribute towards a certain disease!
  - *i.e.* the genes BRCA1 and BRCA2 cause 60% of all cases of hereditary breast and ovarian cancers
  - BUT not a single mutation- there are >800 different mutations in BRCA1 alone

TOOLS: Whole Genome Sequencing

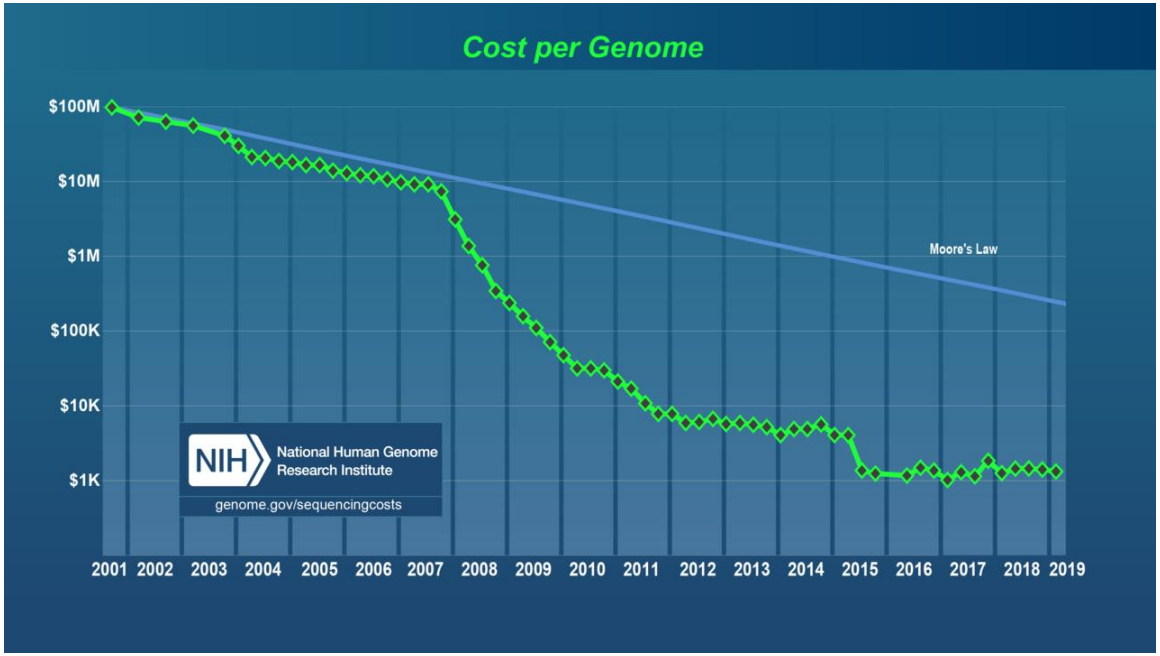
https://theanalyticalscientist.com/fields-applications/the-tools-behind-genomics

Sequencing



https://www.cdc.gov/pulsenet/pdf/Genome-Sequencing-508c.pdf

Cost of Sequencing



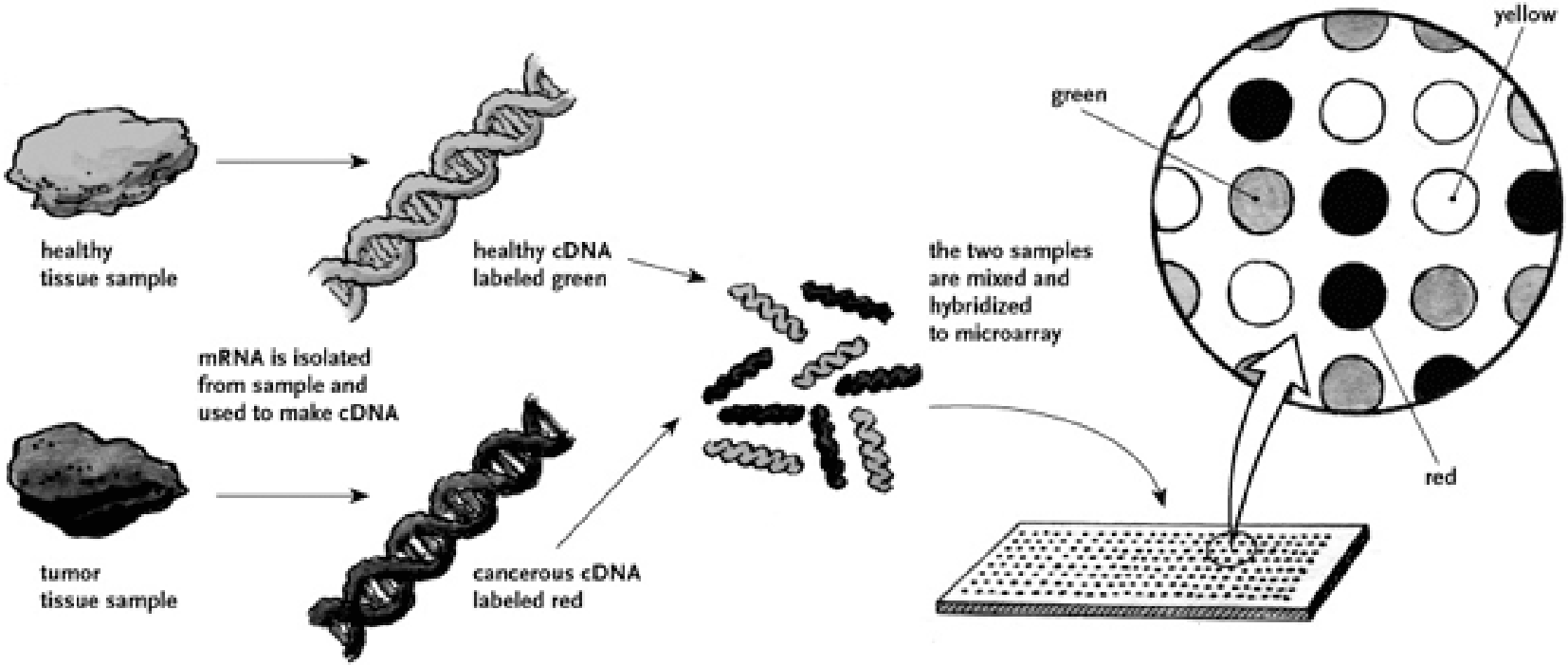
PacBio Sequencer
£300K
Single-Molecule, Real-Time (SMRT) technology
www.pacb.com

Transcriptomics

- Measuring the transcriptome
  - All the mRNA that is transcribed at a given point in a given cell or organism
- Provides direct knowledge of gene regulation and protein content information
  - Which genes are actively expressed
- Began in 1990s and is now a widespread discipline
  - Many technological advances
  - Now a routine, ‘simple’ process

- **TOOLS: Microarrays & RNA-Seq**

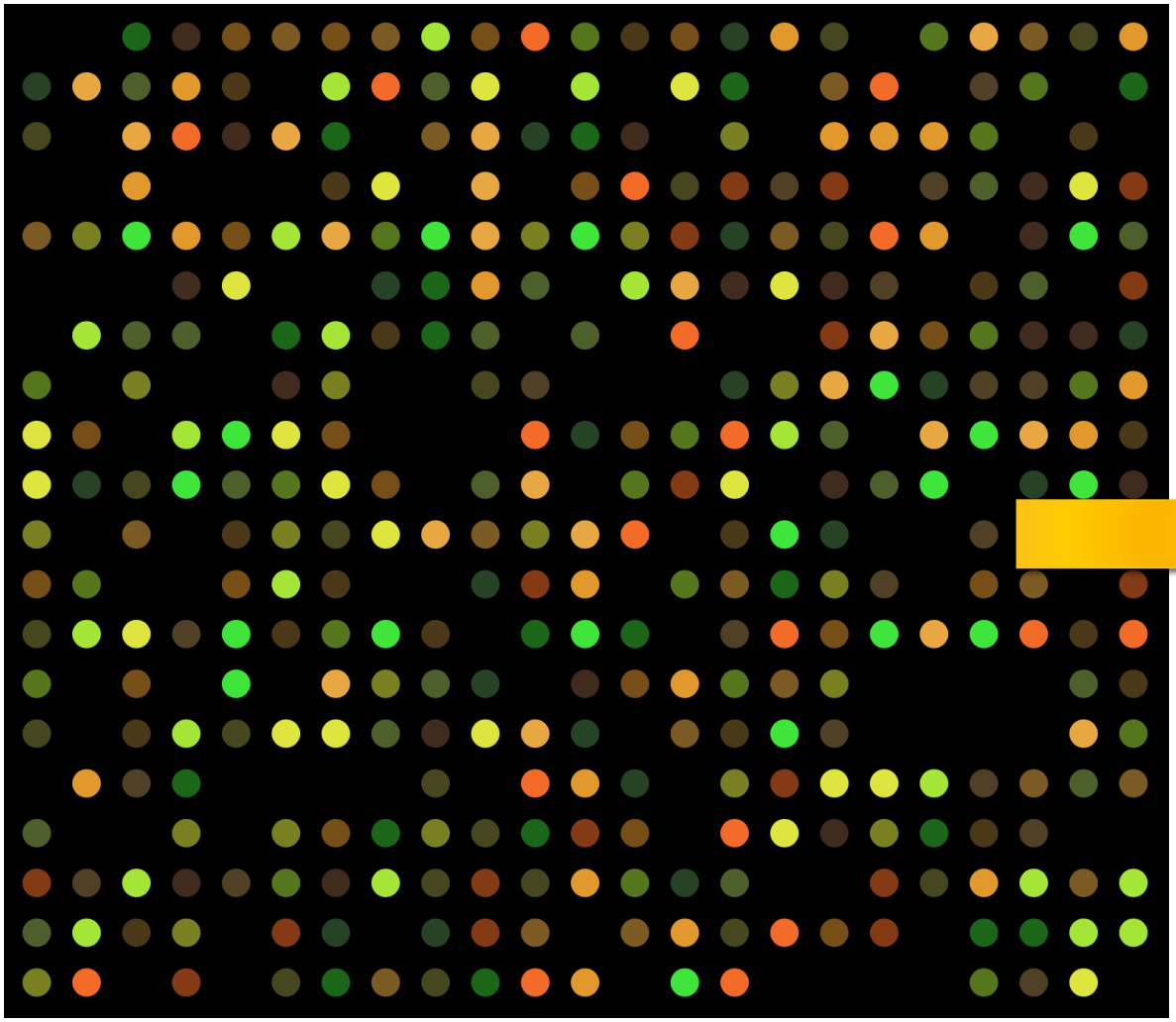
DNA Microarrays



A microarray is an orderly arrangement of rows and columns on a surface like a glass slide. Each of the spots on an array contains single-stranded DNA molecules that correspond to a single gene. An array can contain a few, or thousands, of genes.

http://www.pbs.org/wgbh/nova/education/activities/3413_genes_02.html

DNA Microarrays

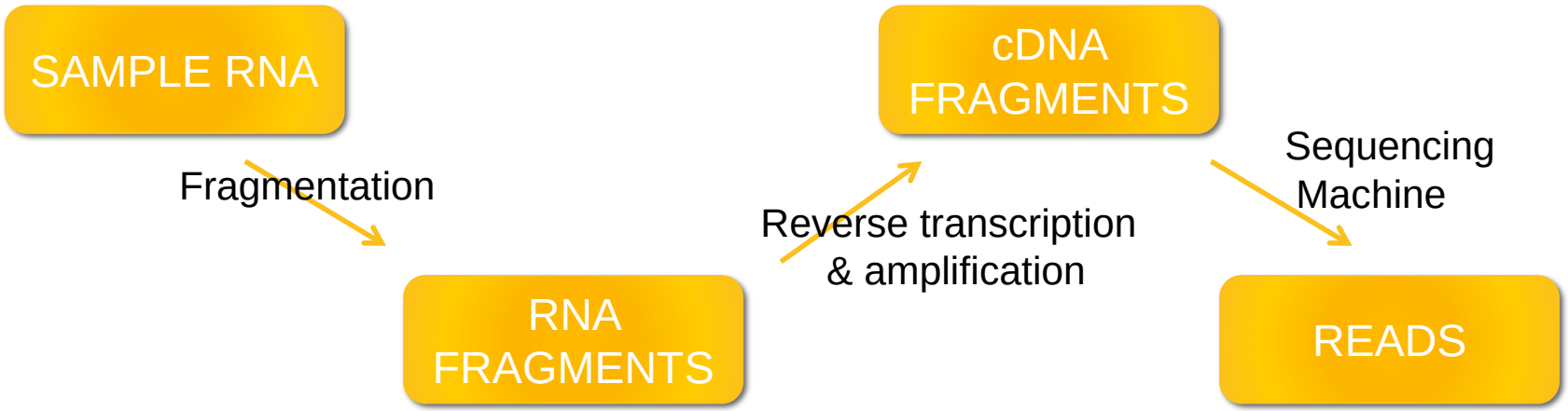


Converted to numerical values for interpretation

Limitation: Results limited to what probes you have on your chip

RNA Seq

- High throughput sequencing with computat
- No reference sequence
- Large dynamic range
- Sequences every RNA molecule and profiles the expression of a particular gene by counting the number of times its transcript has been sequenced
- Expression levels!



Proteomics

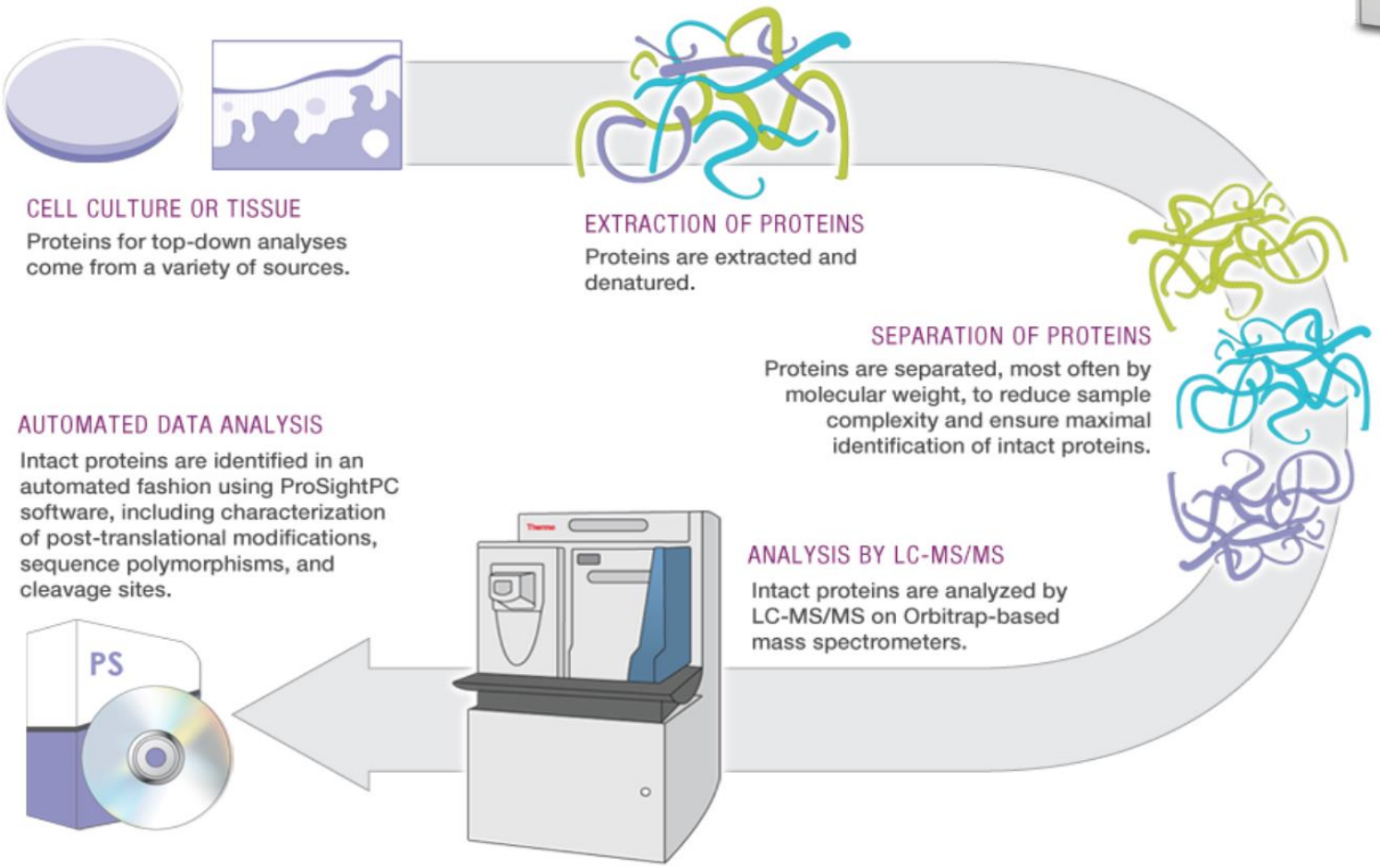
- Investigating the entire complement of proteins within a cell, tissue or organism.
- >100,000 proteins
- Large dynamic range

- Some questions
  - When & where proteins are expressed
  - The involvement of proteins with particular phenotypes
  - How proteins are modified or how they interact with each other

- Blood, urine, tissues

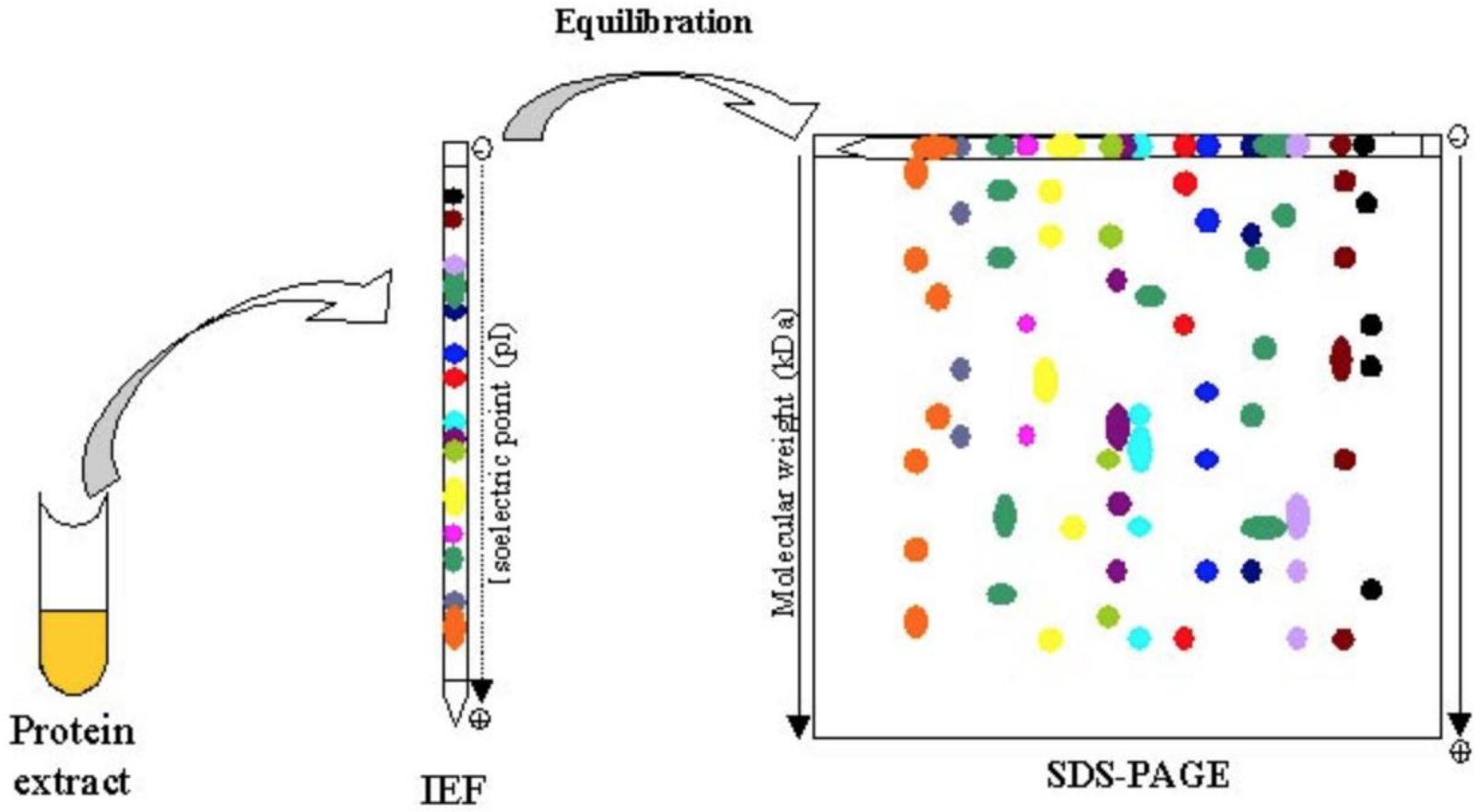
Proteomic Tools

- Mass spectrometry



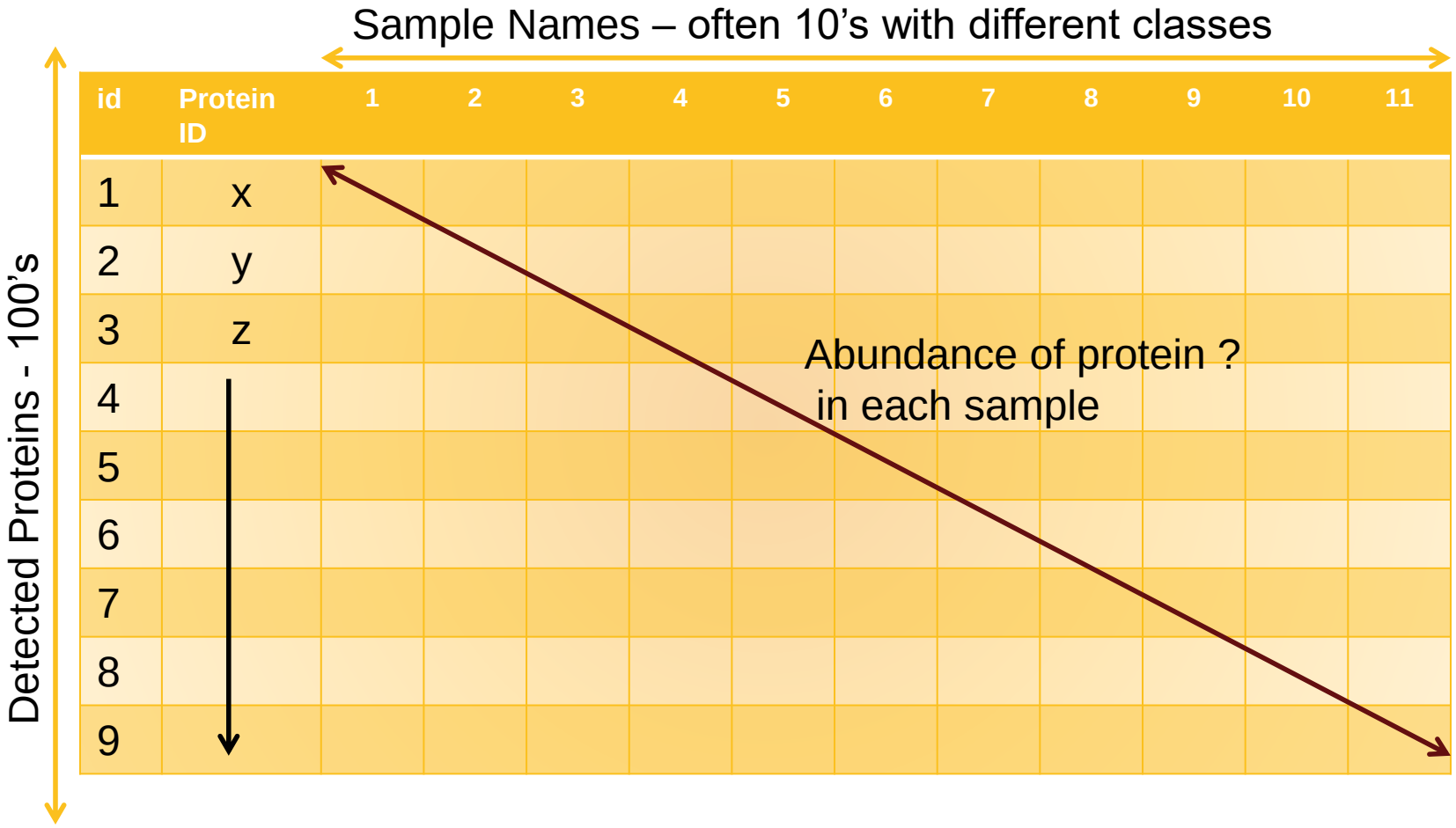
http://planetorbitrap.com/data/fe/image/Workflows_TopDownProt(1).png

Proteomic Tools



https://www.creative-proteomics.com/services/2d-electrophoresis.htm

Proteomics Data



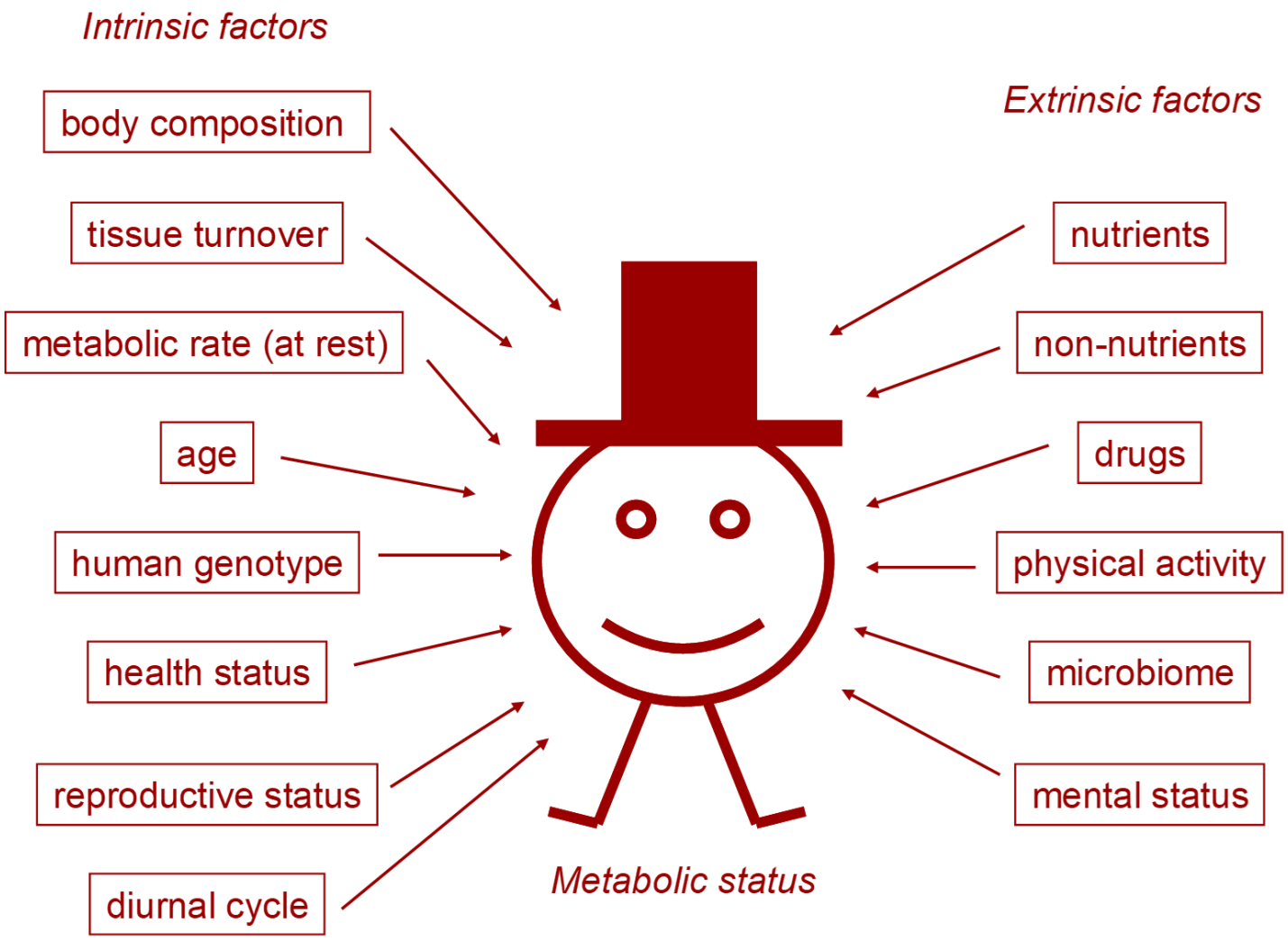
Do statistical analysis to decide which features change between classes

Metabolomics

- Investigating the entire complement of low molecular weight molecules in an sample of interest.

- ~ 5000 metabolites in a biological sample from a human
  - GREATLY impacted by environment!
  - GREATLY impacted by time!

- Experimental design & sample collection is VERY important!



Goodacre, R. (2007) Metabolomics of a superorganism. *Journal of Nutrition* **137**, 259S-266S.

Biological Matrices

Primary Sources

Serum
Plasma
CSF
BAL
Saliva
Semen
Urine
Faeces

Sweat
Breath

Secondary Sources

Tissue Biopsies

Brain
Nerve
Lung
Pancreas
Liver
Heart
Gut
Skin

Additional Sources

Animal Models

Mammalian Cell Culture

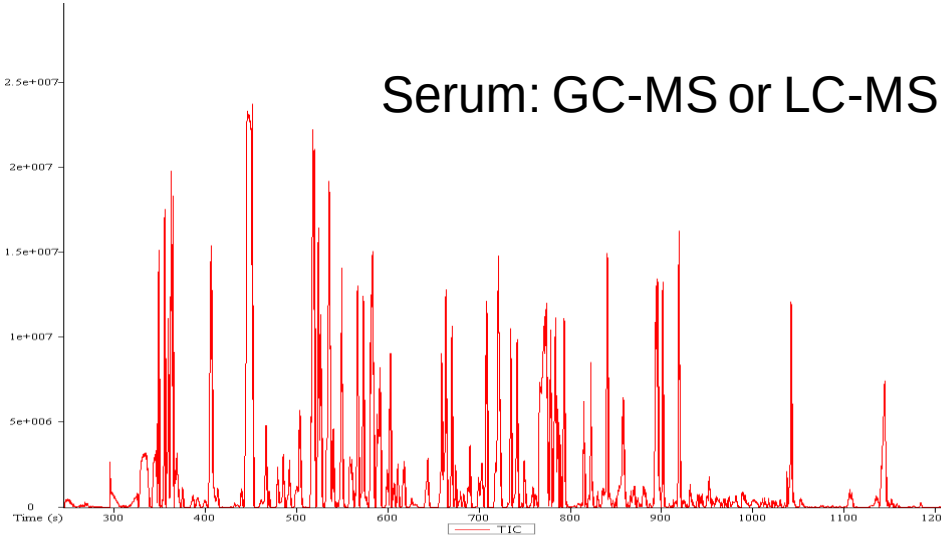
IVF culture medium

Chromatography linked to Mass Spectrometry

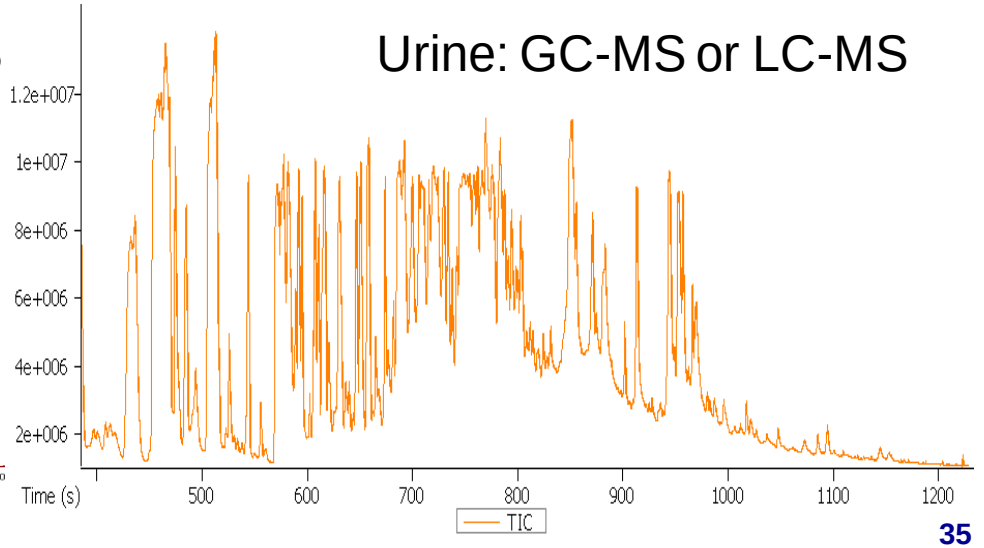
Higher affinity to stationary phase

Lower affinity to stationary phase

Serum: GC-MS or LC-MS



Urine: GC-MS or LC-MS

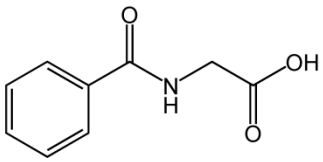


Chromatography linked to Mass Spectrometry

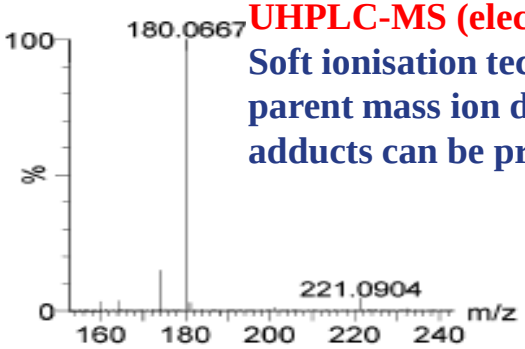
Higher affinity to stationary phase

Lower affinity to stationary phase

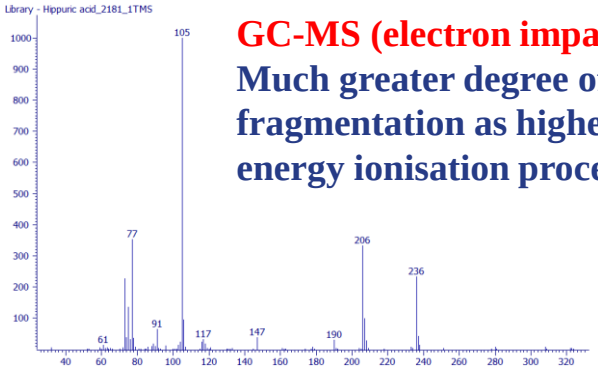
Hippuric acid



***m/z*=179.17**



UHPLC-MS (electrospray (ESI))
Soft ionisation technique, intact parent mass ion detected, but many adducts can be produced



GC-MS (electron impact)
Much greater degree of fragmentation as higher energy ionisation process

Matching of the chromatographic retention time and fragmentation mass spectra between a sample analyte and a reference standard is required for definitive id.
We have *ca*. 1600 analytes in our GC-MS library

Sumner L.W. *et al*. (2007) *Metabolomics* **3**, 211-221

Metabolomics Data

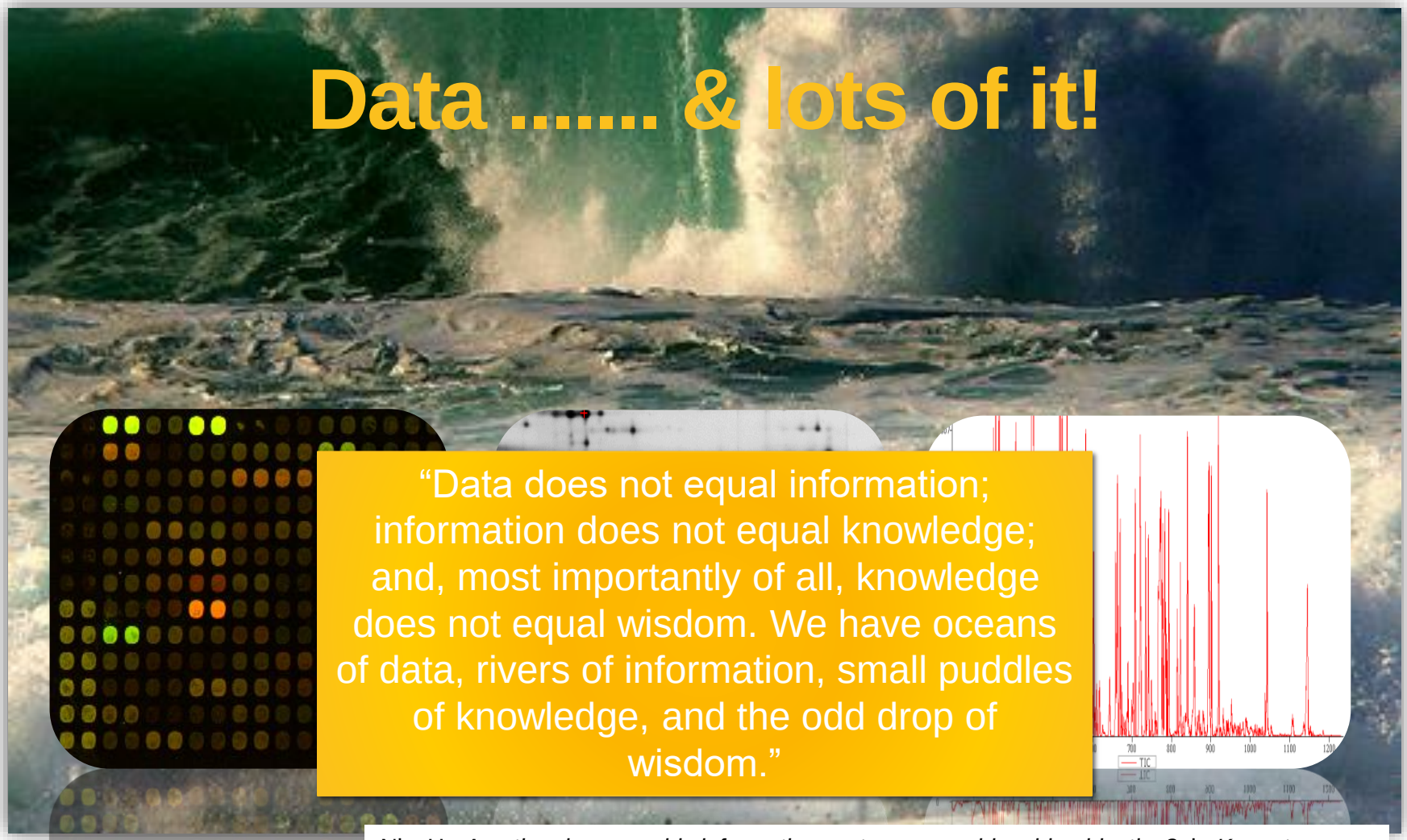
Sample Names – often 100's with different classes

Detected Features- 1000's

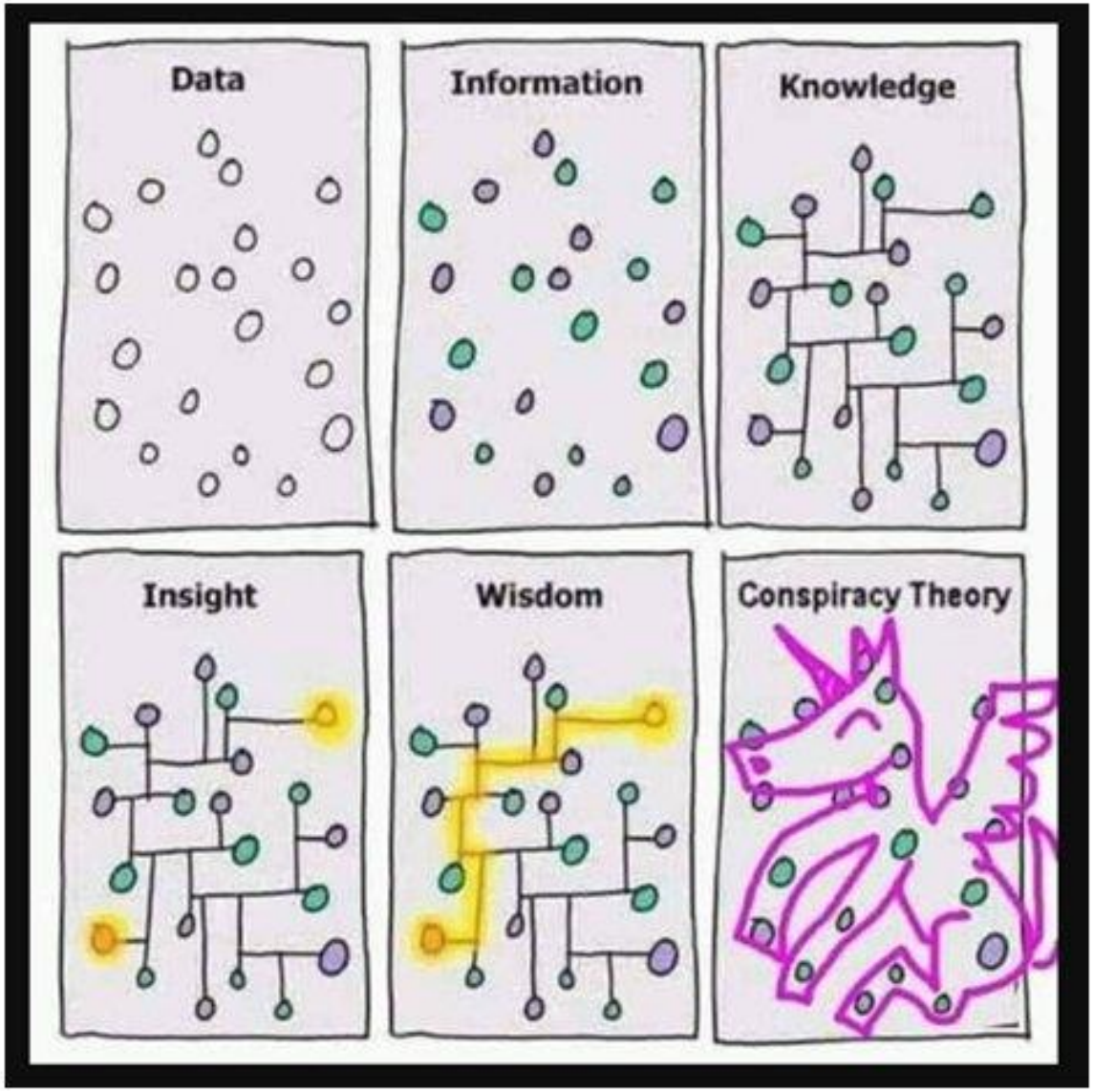
id	RT (min)	m/z	ID	1	2	3	4	5	6	7	8	9	10	11
1														
2														
3														
4														
5														
6														
7														
8														
9														

Use RT & m/z to identify the metabolite – MAJOR CHALLENGE!
Do statistical analysis to decide which features change between classes

Common Theme



Nix, H., *A national geographic information system - an achievable objective?*, in *Keynote address*. 1990: AURISA.

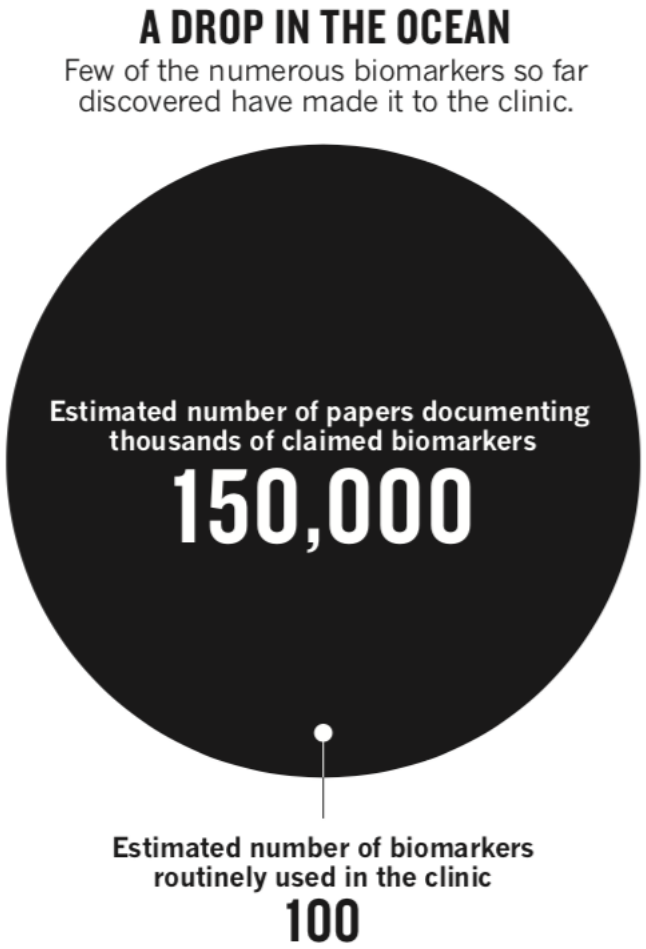


https://www.quora.com/What-are-the-similarities-and-differences-between-data-information-knowledge-and-wisdom

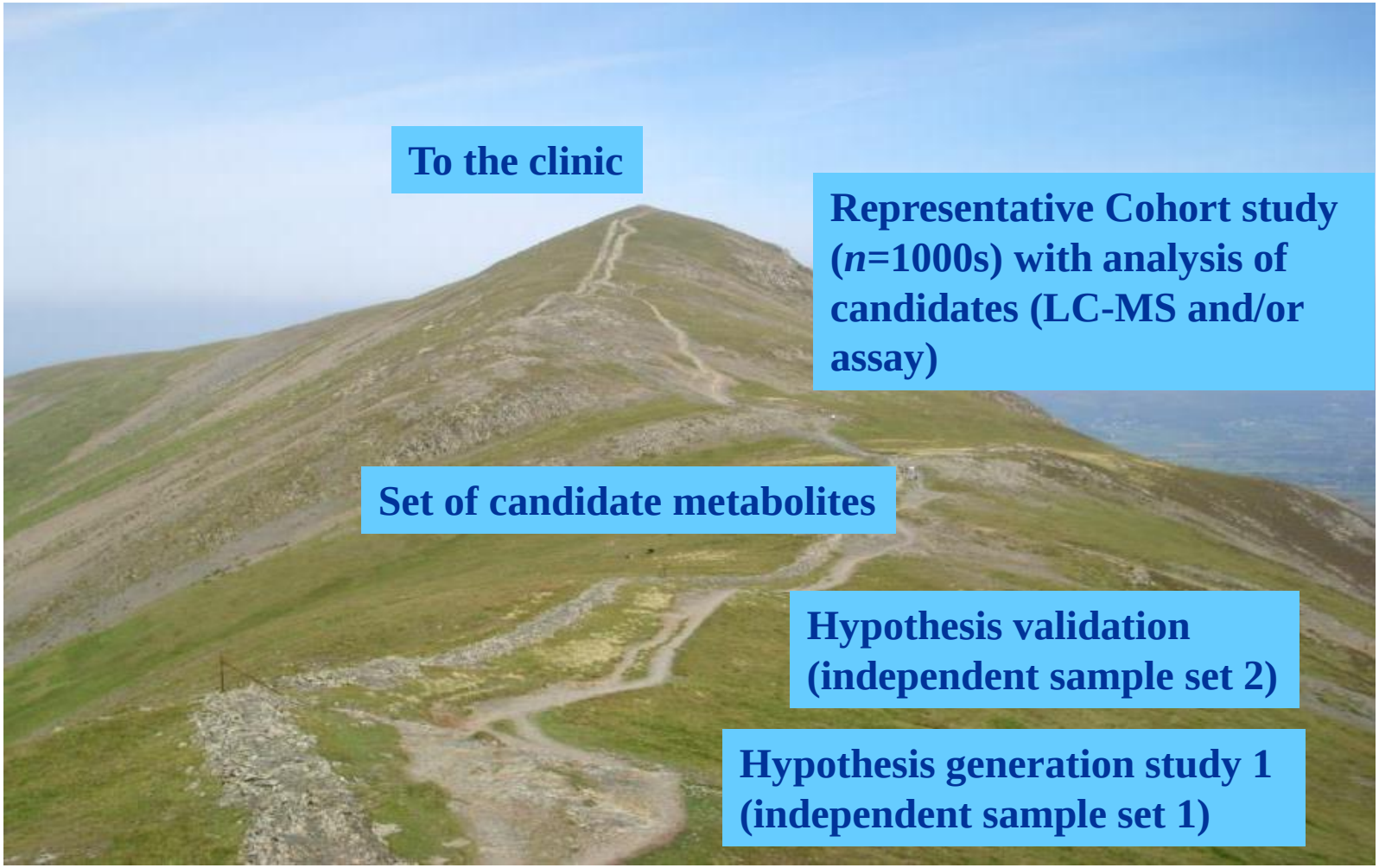
The challenge of biomarkers?

- Poste, G., *Bring on the biomarkers.* Nature, 2011. **469**(7329): p. 156-157.

- Multi-Centre, large cohort studies



Biomarker Discover: From lab to bedside



Increasing sample numbers and throughput

Conception (objectives, collaborations, design of experiment)

SOME EXAMPLES.....

Personalised/Precision Medicine

- Is this the future of healthcare?
  - 2015 State of the Union address President Obama announced that he was launching the Precision Medicine Initiative
- Assess the genotype (SNPs) and phenotype (metabolome) of a patient before they undergo any treatment
  - Population monitoring & data collation
  - Seeking cures & preventative screening
- Offering a well-designed screening program at a reasonable cost may not always be possible due to the numerous associated challenges
  - monetary limitations (labour and consumable costs) as well as ethical, legal and social considerations for an opt-in test

(Metabol)'omics for the masses?

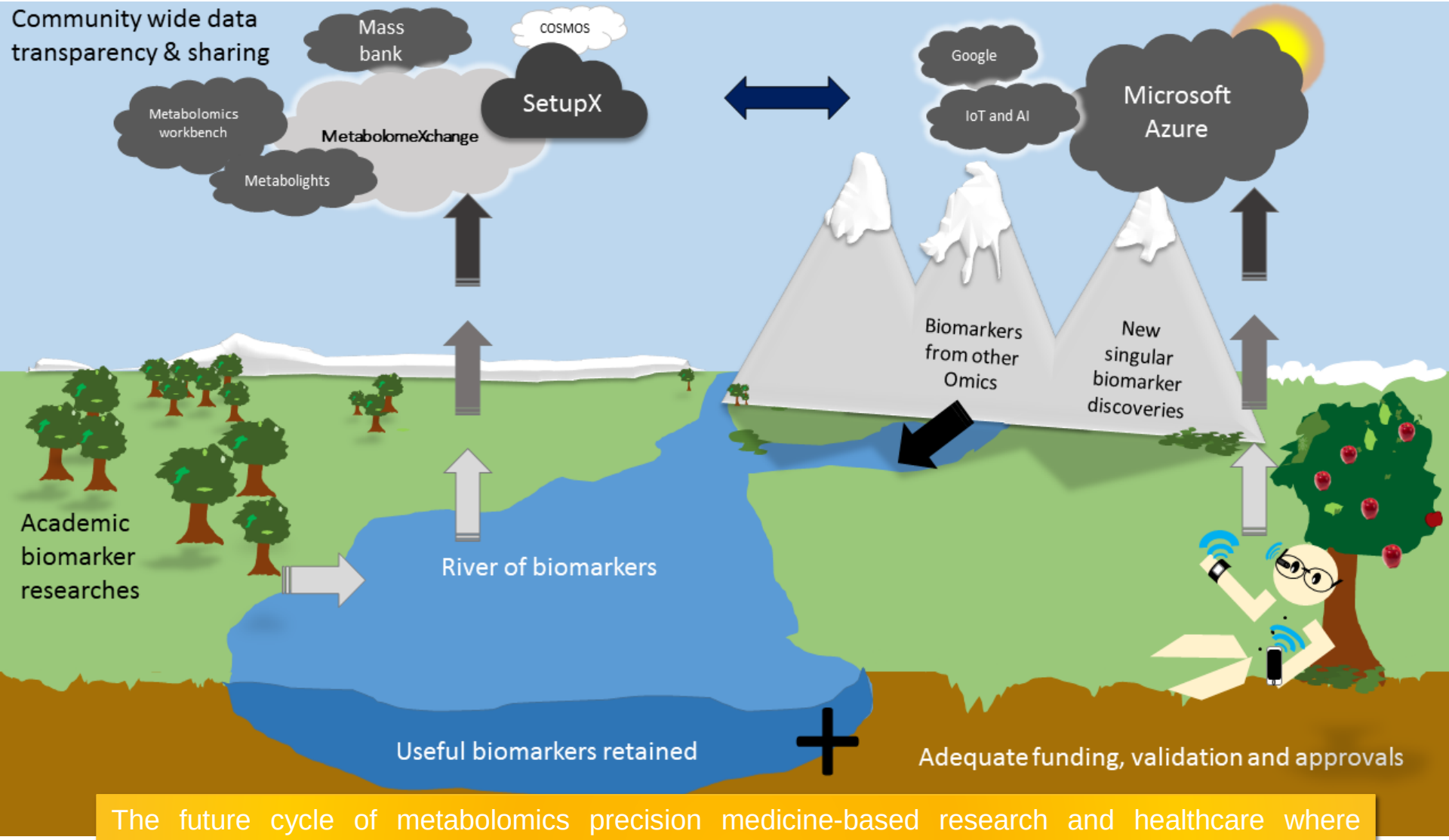
- Wearable technologies – smartphones, smart-watches, health bands, necklaces, glucose monitoring contact lenses
  - Innovations for collecting personal information

- mPower - mobile Parkinson's Disease study that attempts to research the occurrence, presentation and management of PD symptoms *via* survey telemetry data using a smartphone app
  - Bot, B.M., et al., *The mPower study, Parkinson disease mobile data collected using ResearchKit.* Scientific Data, 2016. **3**: p. 160011.

- Smart-phone app to monitor the association between pain and the weather for people suffering from rheumatoid arthritis
  - Reade, S., et al., *Cloudy with a Chance of Pain: Engagement and Subsequent Attrition of Daily Data Entry in a Smartphone Pilot Study Tracking Weather, Disease Severity, and Physical Activity in Patients With Rheumatoid Arthritis.* JMIR Mhealth Uhealth, 2017. **5**(3): p. e37.
  - Dixon, W.G., et al. How the weather affects the pain of citizen scientists using a smartphone app. Npj Digital Medicine (2019) 2:105

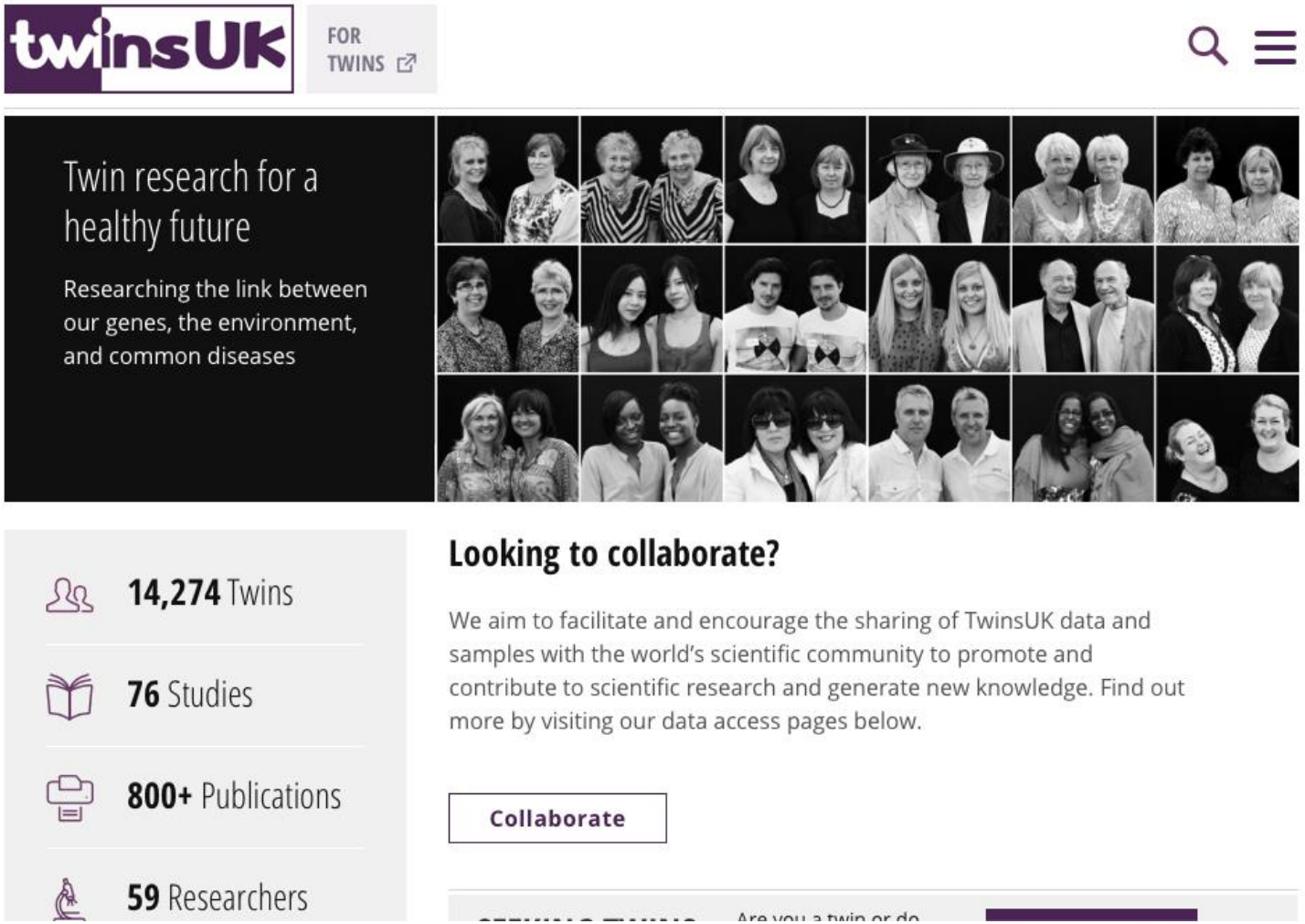
CLOUDY
W TH A CHANCE OF
P A I N

Patients with chronic pain commonly believe their pain is related to the weather. Scientific evidence to support their beliefs is inconclusive, in part due to difficulties in getting a large dataset of patients frequently recording their pain symptoms during a variety of weather conditions. Smartphones allow the opportunity to collect data to overcome these difficulties. Our study *Cloudy with a Chance of Pain* analysed daily data from 2658 patients collected over a 15-month period. The analysis demonstrated significant yet modest relationships between pain and relative humidity, pressure and wind speed, with correlations remaining even when accounting for mood and physical activity. This research highlights how citizen-science experiments can collect large datasets on real-world populations to address long-standing health questions. These results will act as a starting point for a future system for patients to better manage their health through pain forecasts.



The future cycle of metabolomics precision medicine-based research and healthcare where academia, industrial partners, corporate data analytics work with patients' wearable data collection devices to provide health monitoring solutions.

TWINS UK


twinsUK
FOR TWINS
Twin research for a healthy future
Researching the link between our genes, the environment, and common diseases
14,274 Twins
76 Studies
800+ Publications
59 Researchers
Looking to collaborate?
We aim to facilitate and encourage the sharing of TwinsUK data and samples with the world's scientific community to promote and contribute to scientific research and generate new knowledge. Find out more by visiting our data access pages below.
Collaborate

Research Areas

Omics

- Epigenomics
- Expression-omics
- Genetics
- Glycomics
- Human Microbiome
- Metabolomics
- Nutriomics

Disease & Ageing

- Allergy
- Bone
- Brain
- Cardiovascular
- Diet
- Hearing
- Pain
- Skin
- Vision

TWINS UK

- Comprehensive study
- Unique design with internal controls
- Well documented & controlled
- Wealth of scientific publications

The super smeller

INDY/PULSE

THIS WOMAN CAN SMELL PARKINSON'S DISEASE

The woman who can smell Parkinson's disease

By Elizabeth Quigley
BBC Scotland news

22 October 2015 Share

Lifestyle › Health and Fitness › Body

Meet the woman who can smell Parkinson's disease

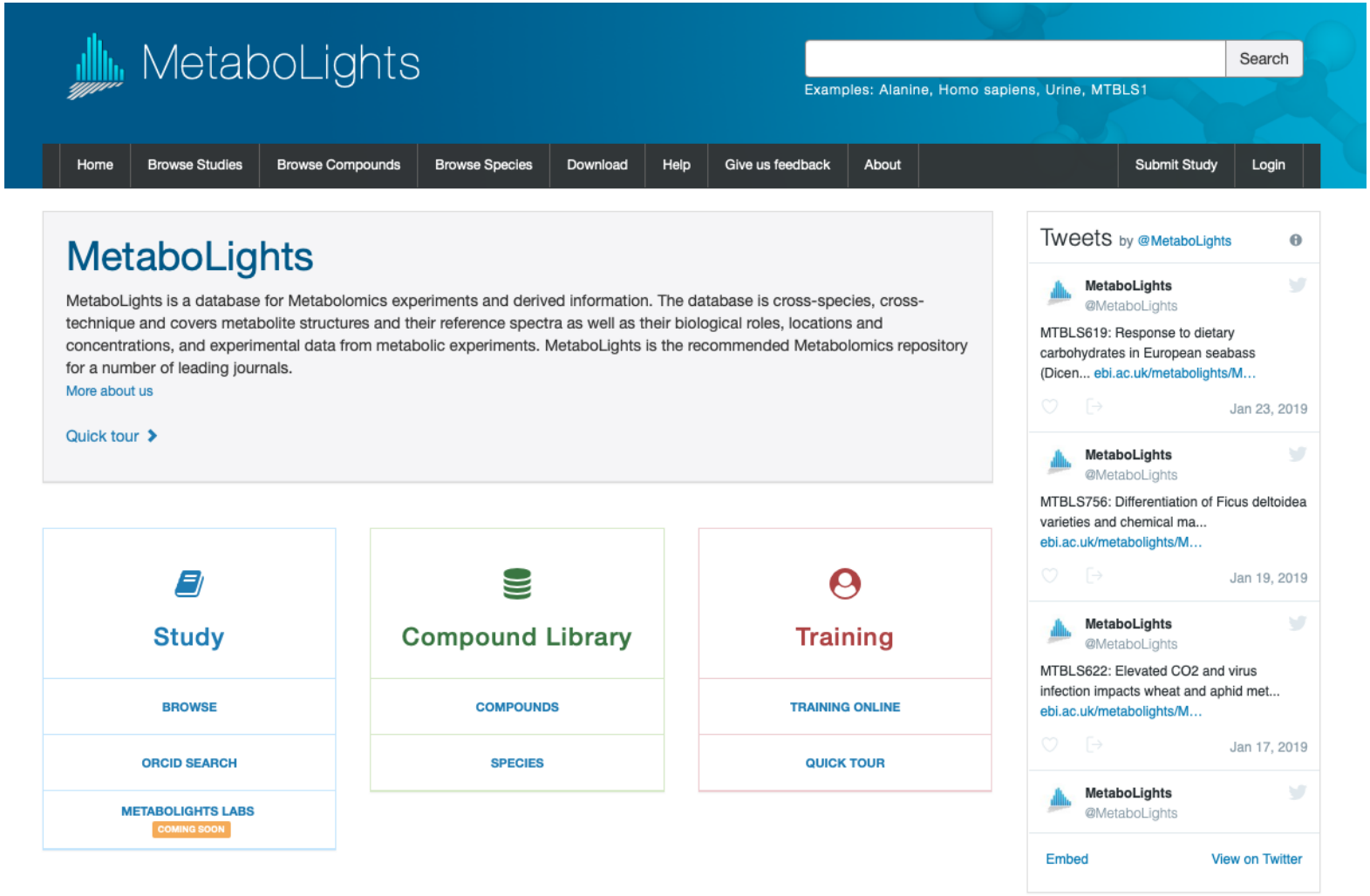
share 7

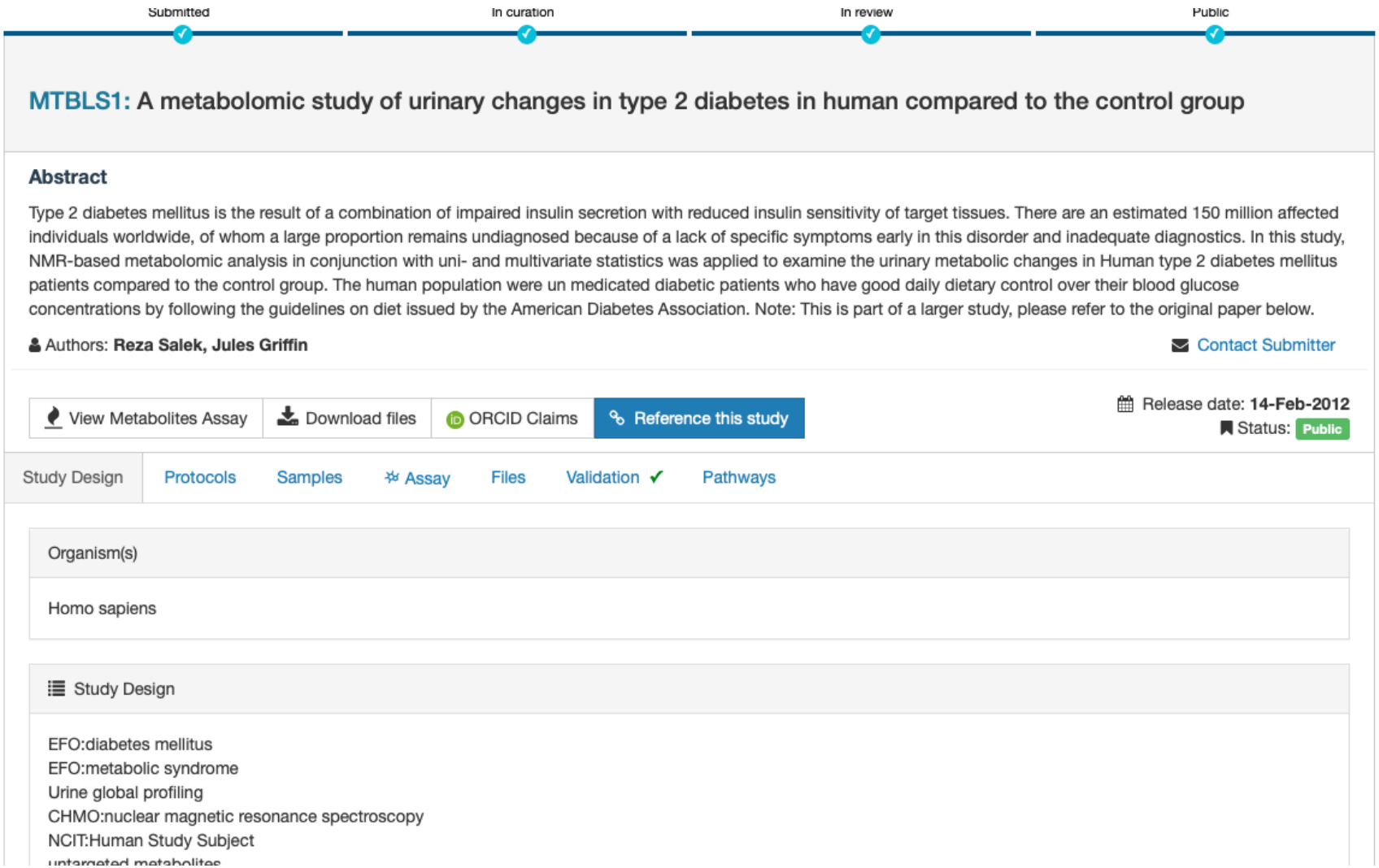
Joy Milne, right, gives the author a sniff CREDIT: CHRIS WATT

Biomarker detection

Gaining distance on biomarker discovery.....

Data Repositories




Submitted
In curation
In review
Public
MTBLS1: A metabolomic study of urinary changes in type 2 diabetes in human compared to the control group
Abstract
Type 2 diabetes mellitus is the result of a combination of impaired insulin secretion with reduced insulin sensitivity of target tissues. There are an estimated 150 million affected individuals worldwide, of whom a large proportion remains undiagnosed because of a lack of specific symptoms early in this disorder and inadequate diagnostics. In this study, NMR-based metabolomic analysis in conjunction with uni- and multivariate statistics was applied to examine the urinary metabolic changes in Human type 2 diabetes mellitus patients compared to the control group. The human population were un medicated diabetic patients who have good daily dietary control over their blood glucose concentrations by following the guidelines on diet issued by the American Diabetes Association. Note: This is part of a larger study, please refer to the original paper below.
Authors: Reza Salek, Jules Griffin
Contact Submitter
View Metabolites Assay
Download files
ORCID Claims
Reference this study
Release date: 14-Feb-2012
Status: Public
Study Design
Protocols
Samples
Assay
Files
Validation
Pathways
Organism(s)
Homo sapiens
Study Design
EFO:diabetes mellitus
EFO:metabolic syndrome
Urine global profiling
CHMO:nuclear magnetic resonance spectroscopy
NCIT:Human Study Subject
untargeted metabolites

EMBL-EBI

Services | Research | Training | About us

PRIDE Archive

Search

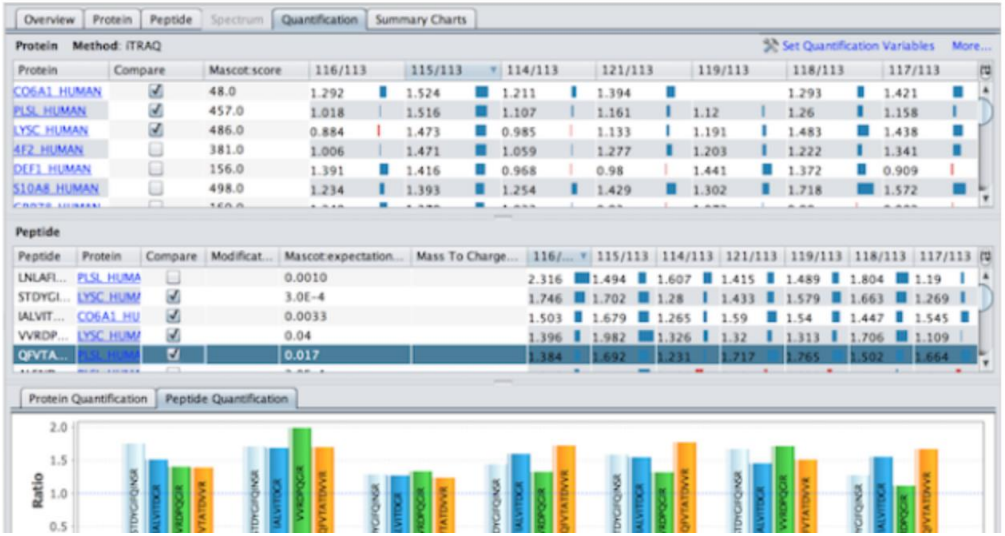
Examples: stress, human, blood, P02768, MDPNTIIEALR*

Home | Submit data | Browse data | Help | Publications | About PRIDE Archive | Register | Login | Feedback

PRIDE > Archive

PRIDE Archive - proteomics data repository

The PRIDE PRoteomics IDEntifications (PRIDE) database is a centralized, standards compliant, public data repository for proteomics data, including protein and peptide identifications, post-translational modifications and supporting spectral evidence. PRIDE is a core member in the ProteomeXchange (PX) consortium, which provides a single point for submitting mass spectrometry based proteomics data to public-domain repositories. Datasets are submitted to PRIDE via ProteomeXchange and are handled by expert biocurators.



Datasets

- 8159 projects
- 76426 assays

News

JAN 16 @CellCellPress "Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons" link Dataset #PXD008136

JAN 06 (2/2) … corresponding dataset #PXD004736 available @proteomexchange @pride_ebi link

JAN 06 (1/2) @naturemethods "Biotinylation by antibody recognition—a method for proximity labeling" link

ISB Home

PeptideAtlas

PeptideAtlas Home
Seattle Proteome Center

PeptideAtlas:
Overview
Contacts
Data Contributors
Publications
Software
Database Schema
Feedback
FAQ

Atlas Data:
Data Repository
Human Plasma (Farrah, et al.)
HPPP Data Central
PeptideAtlas Builds
Search Database

Contribute Data
Genome Browser Setup

Related:
SRMAtlas
PASSEL
Phosphopep
Unipep

PeptideAtlas

Search PeptideAtlas: GO

Expanded Search

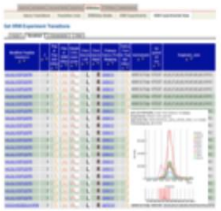
PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. More...

PeptideAtlas Chromosome Explorer (Human only)

SRMAtlas

SRMAtlas interface for selection of best available SRM transitions

PeptideAtlas Raw Data Repository



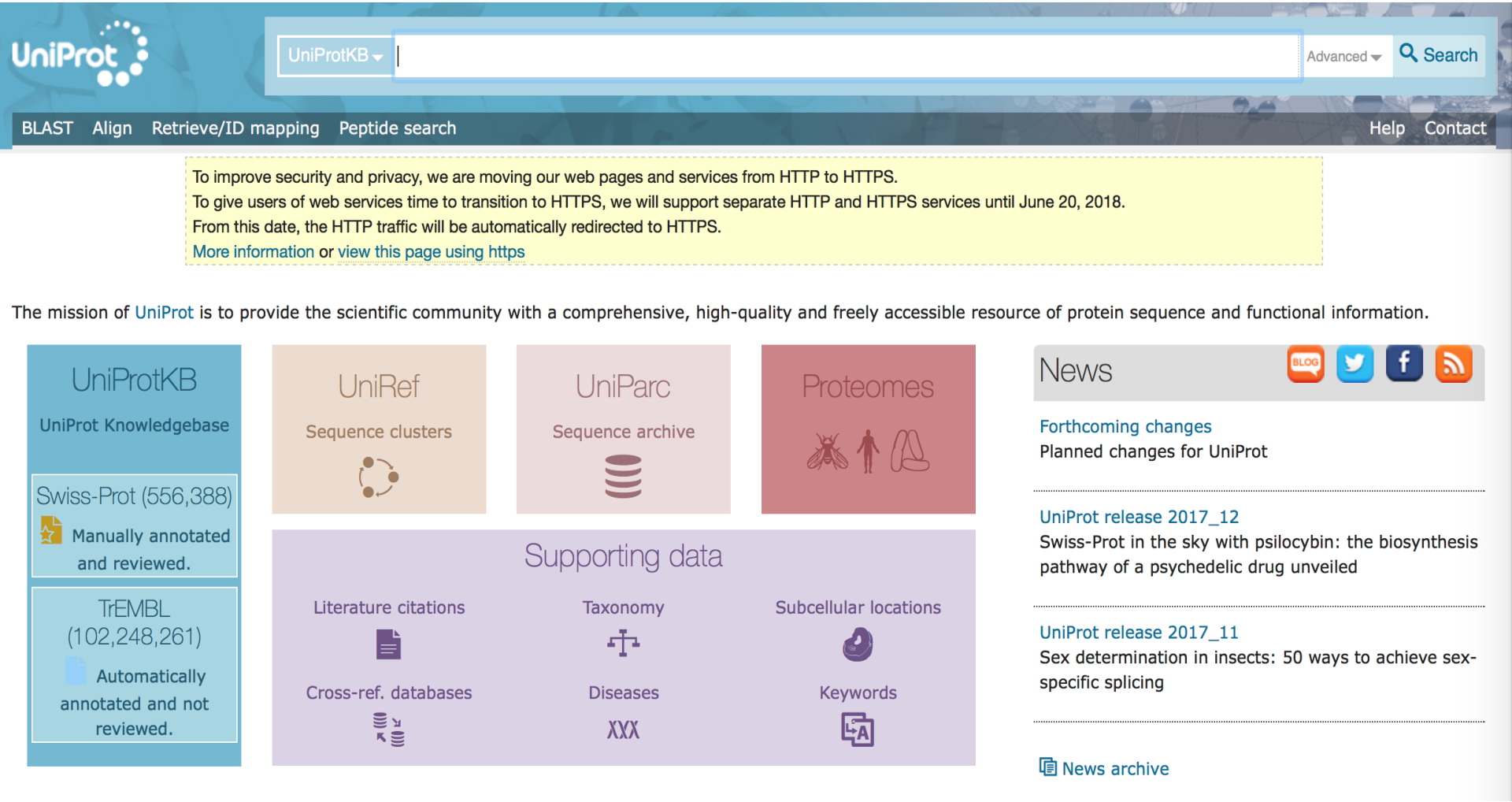
PeptideAtlas SRM Experiment Library (PASSEL)

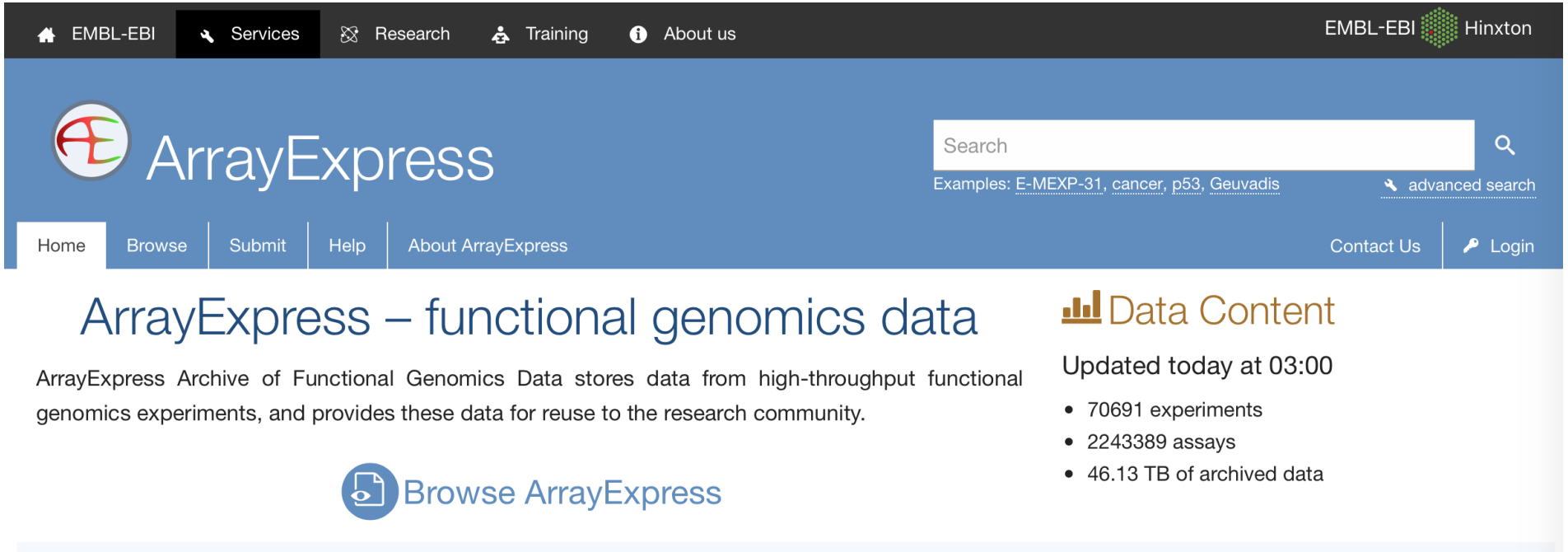
CHPP

PeptideAtlas and the Chromosome-Centric Human Proteome Project

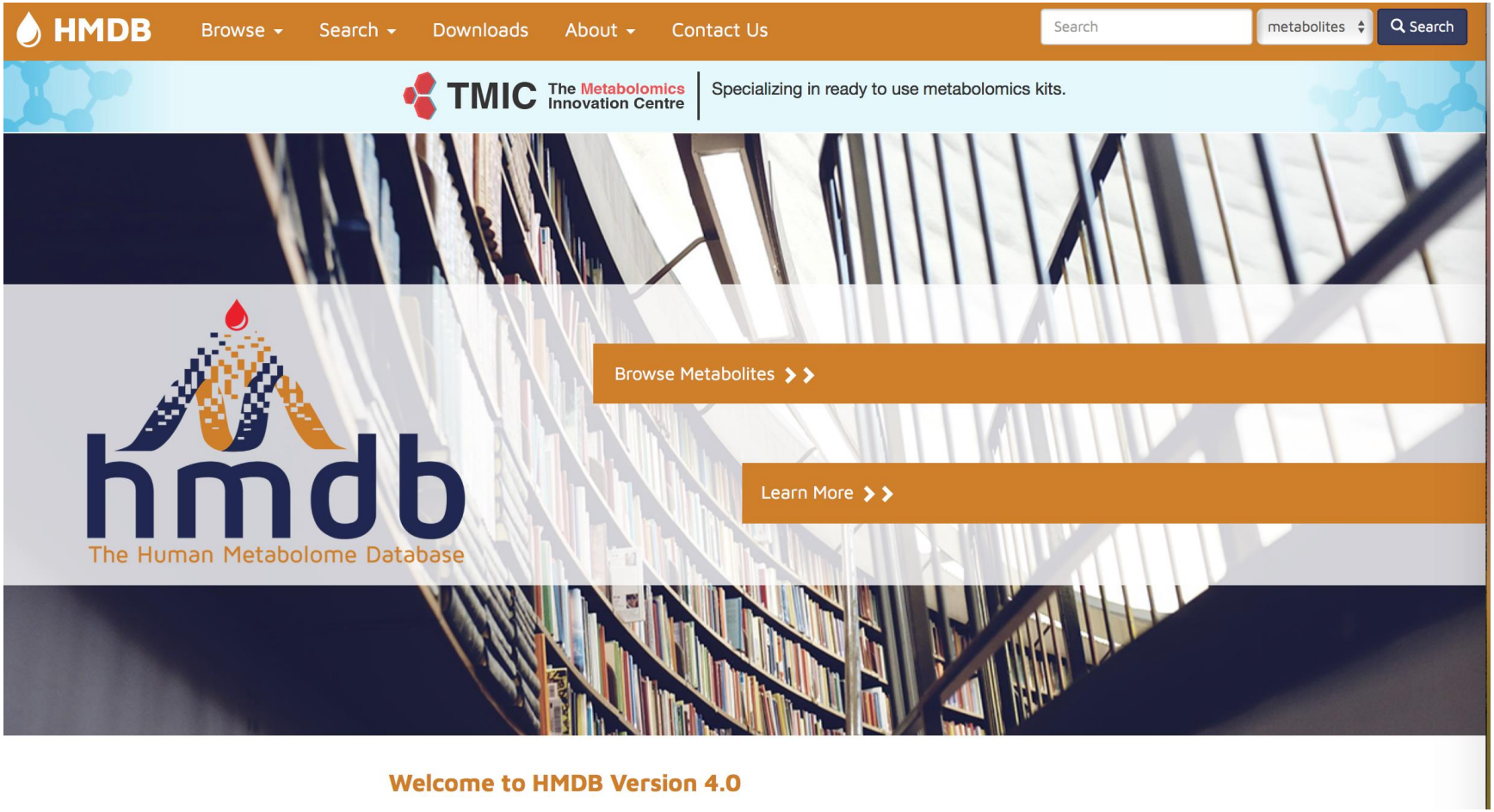
Submit

your raw data to PeptideAtlas

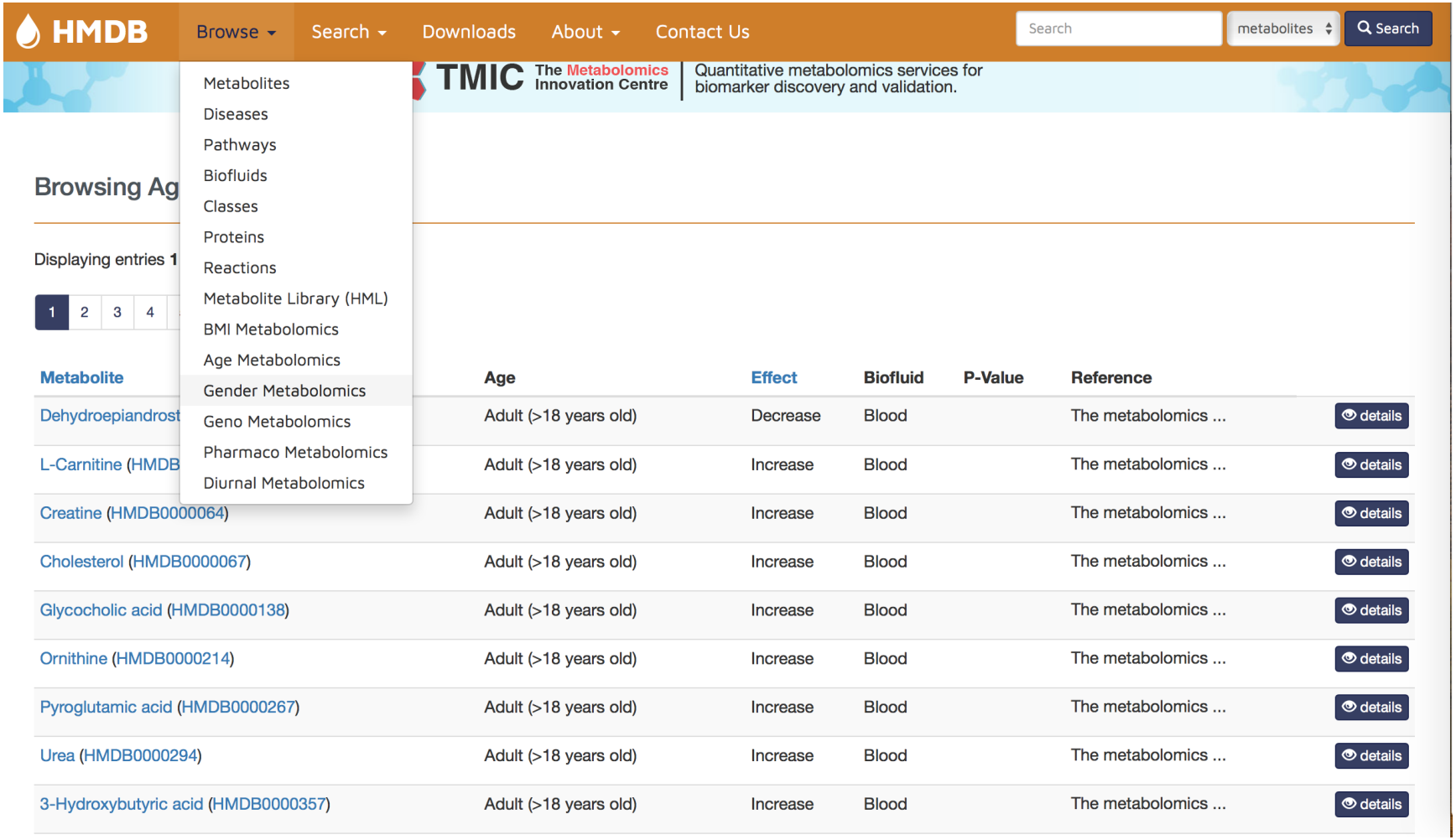

UniProt
UniProtKB
Advanced
Search
BLAST
Align
Retrieve/ID mapping
Peptide search
Help
Contact
To improve security and privacy, we are moving our web pages and services from HTTP to HTTPS.
To give users of web services time to transition to HTTPS, we will support separate HTTP and HTTPS services until June 20, 2018.
From this date, the HTTP traffic will be automatically redirected to HTTPS.
More information or view this page using https
The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
UniProtKB
UniProt Knowledgebase
Swiss-Prot (556,388)
Manually annotated and reviewed.
TrEMBL (102,248,261)
Automatically annotated and not reviewed.
UniRef
Sequence clusters
UniParc
Sequence archive
Proteomes
Supporting data
Literature citations
Taxonomy
Subcellular locations
Cross-ref. databases
Diseases
Keywords
News
Forthcoming changes
Planned changes for UniProt
UniProt release 2017_12
Swiss-Prot in the sky with psilocybin: the biosynthesis pathway of a psychedelic drug unveiled
UniProt release 2017_11
Sex determination in insects: 50 ways to achieve sex-specific splicing
News archive


EMBL-EBI
Services
Research
Training
About us
EMBL-EBI Hinxton
ArrayExpress
Search
Examples: E-MEXP-31, cancer, p53, Geuvadis
advanced search
Home
Browse
Submit
Help
About ArrayExpress
Contact Us
Login
ArrayExpress – functional genomics data
ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.
Browse ArrayExpress
Data Content
Updated today at 03:00
70691 experiments
2243389 assays
46.13 TB of archived data

Human Metabolome Database



Search Parameters



Standardisation

- MSI formed in 2005 to unify and to engage with the growing metabolomics community so that experiments can be reproduced by others and are based on solid sample collection, analysis and data processing. – Complete Transparency!
- Working group now working on how to perform experimental design better.
- Pre-requisite for publication in Metabolomics

Metabolomics (2007) 3:175–178
DOI 10.1007/s11306-007-0070-6

BRIEF REPORT

The metabolomics standards initiative (MSI)

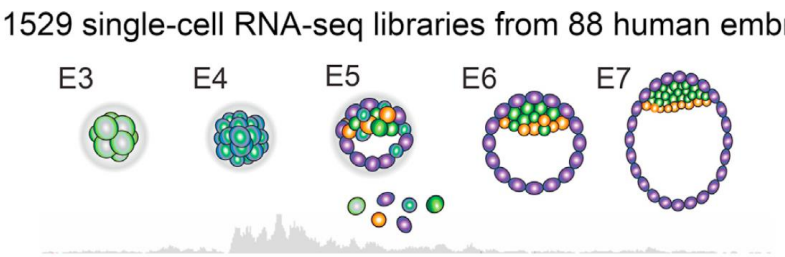
Oliver Fiehn · Don Robertson · Jules Griffin · Mariet van der Werf · Basil Nikolau · Norman Morrison · Lloyd W. Sumner · Roy Goodacre · Nigel W. Hardy · Chris Taylor · Jennifer Fostel · Bruce Kristal · Rima Kaddurah-Daouk · Pedro Mendes · Ben van Ommen · John C. Lindon · Susanna-Assunta Sansone

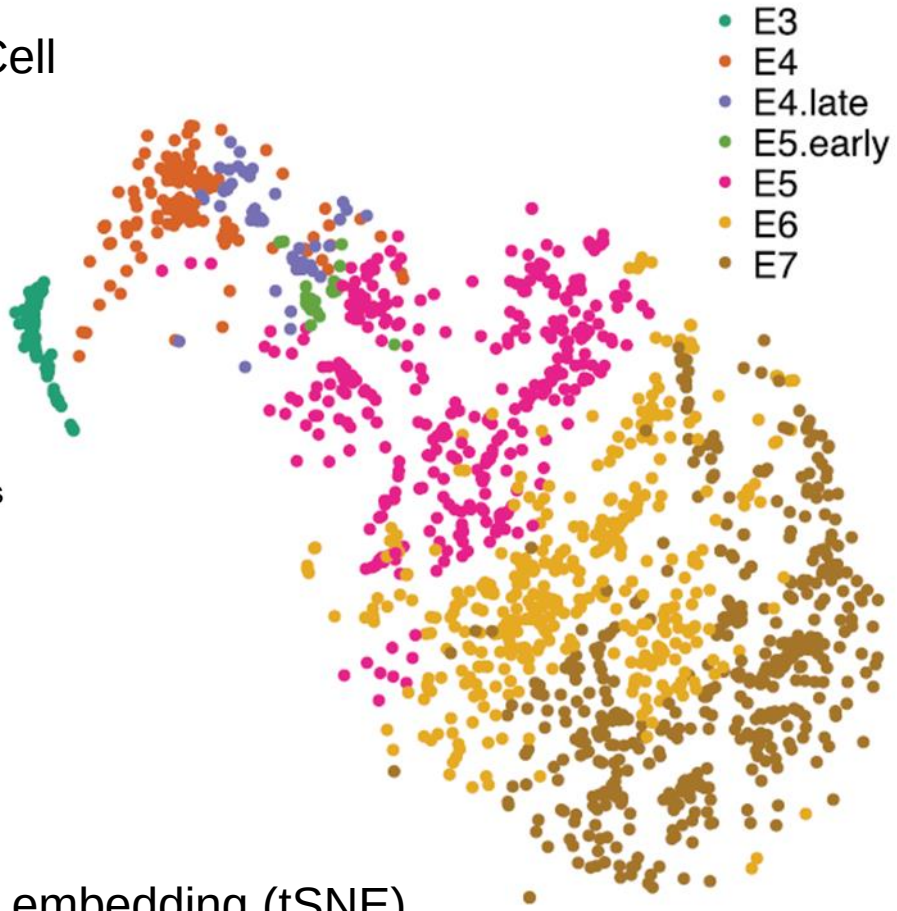
Metabolite Identification
MSI Leve1 1 – Definitive
MSI Level 2 - Putative

NETWORK ANALYSIS

Single Cell Transcriptomic Data

Petropoulos, S., et al. (2016). "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos." Cell 165(4): 1012-1026.

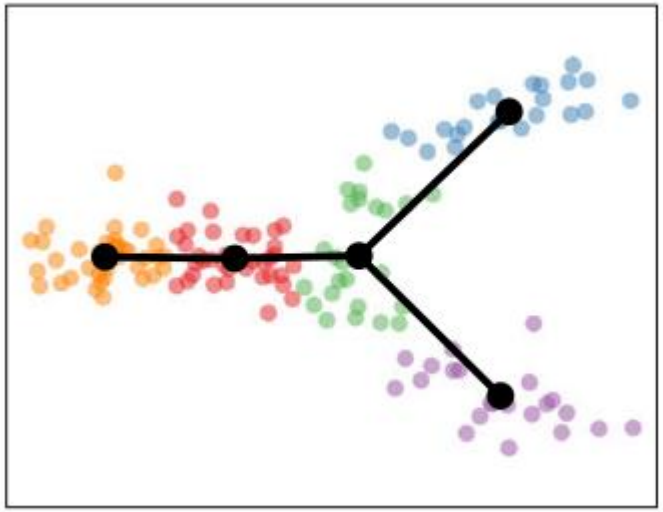




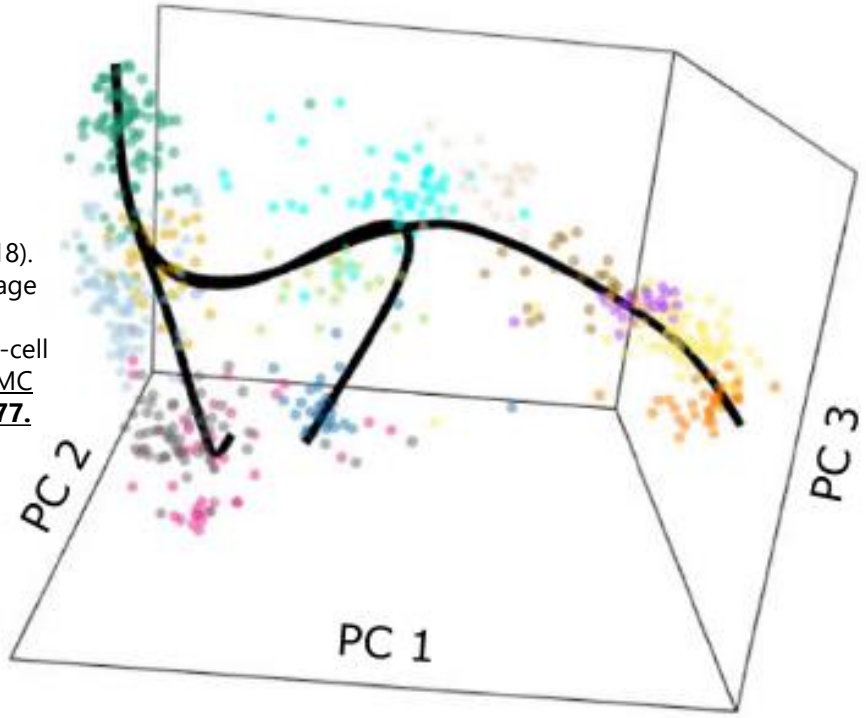
T-distributed stochastic neighbourhood embedding (tSNE)

Developmental trajectories

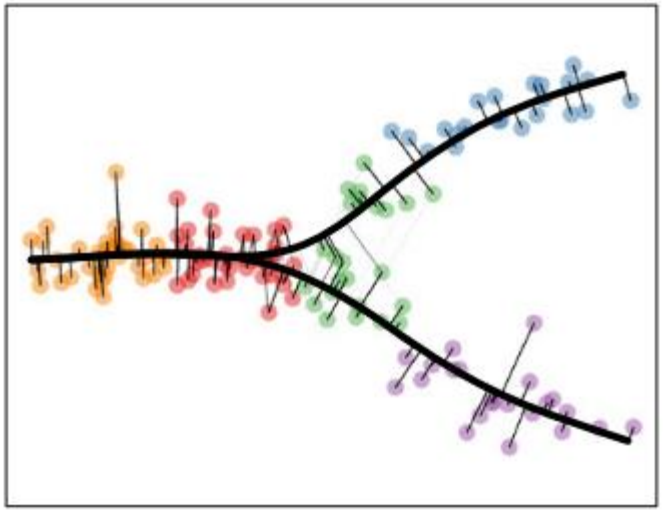
Minimal Spanning Tree (MST)

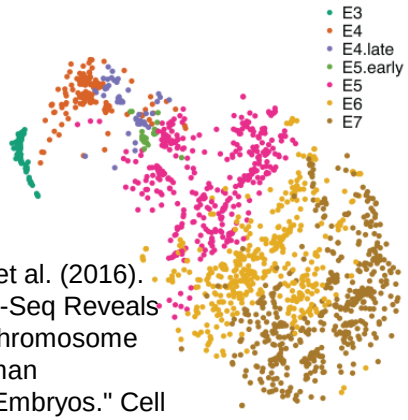


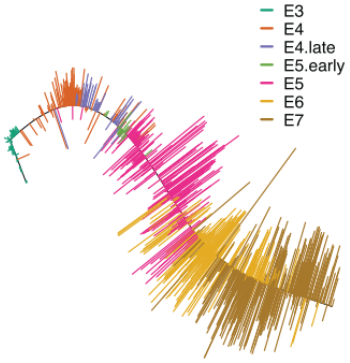
Street, K., et al. (2018). "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics." BMC Genomics **19(1): 477.**



Principal Curve (PC)

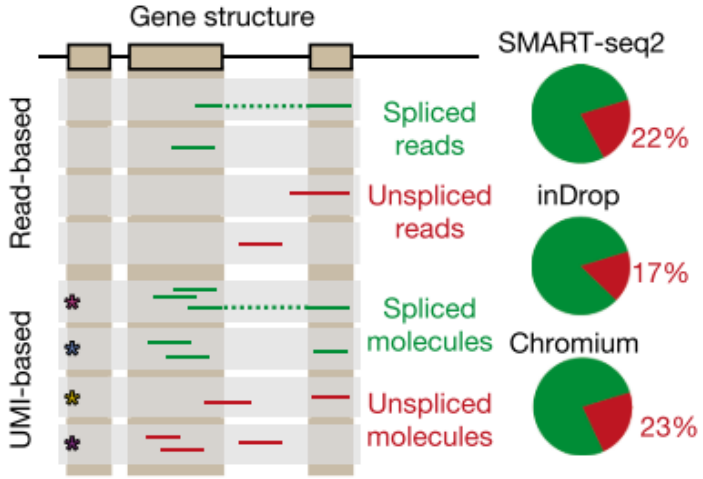


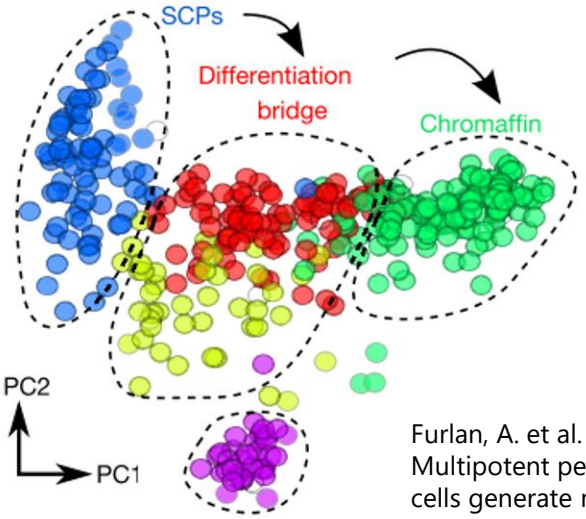




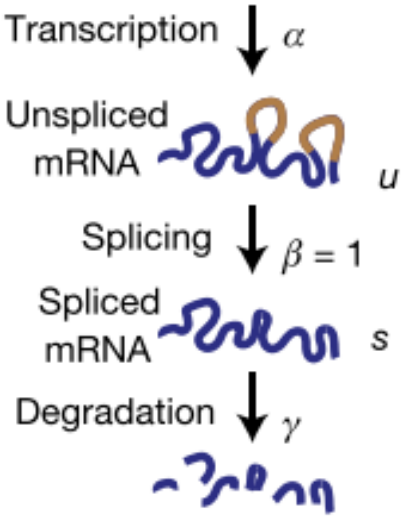
Petropoulos, S., et al. (2016). "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos." Cell 165(4): 1012-1026.

RNA velocity

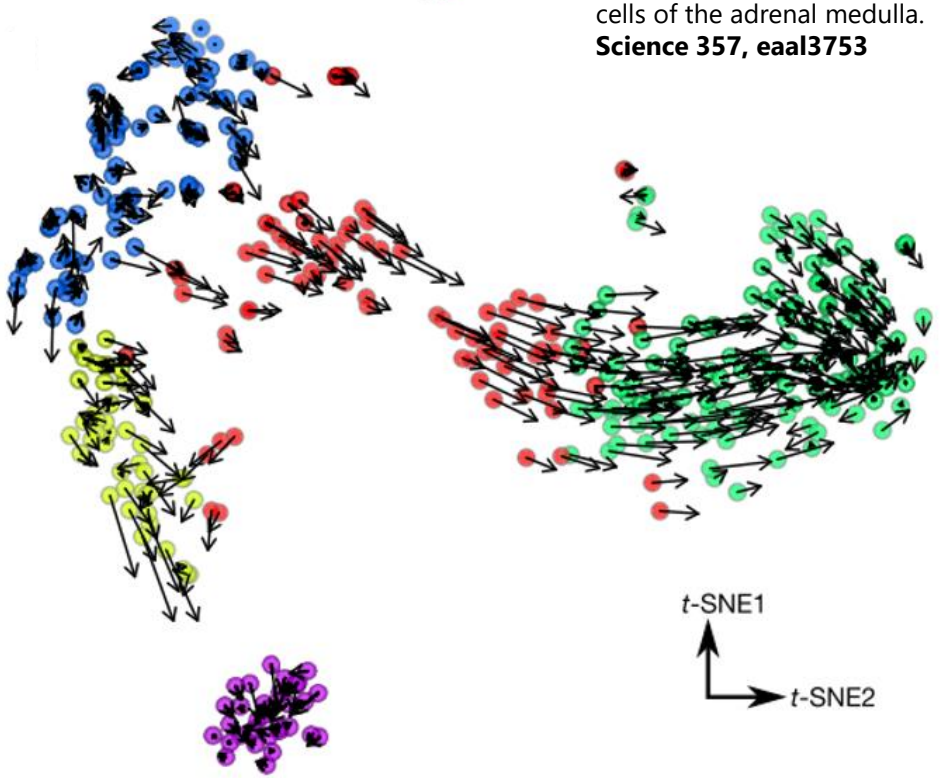




Furlan, A. et al. (2017). Multipotent peripheral glial cells generate neuroendocrine cells of the adrenal medulla. **Science 357, eaal3753**



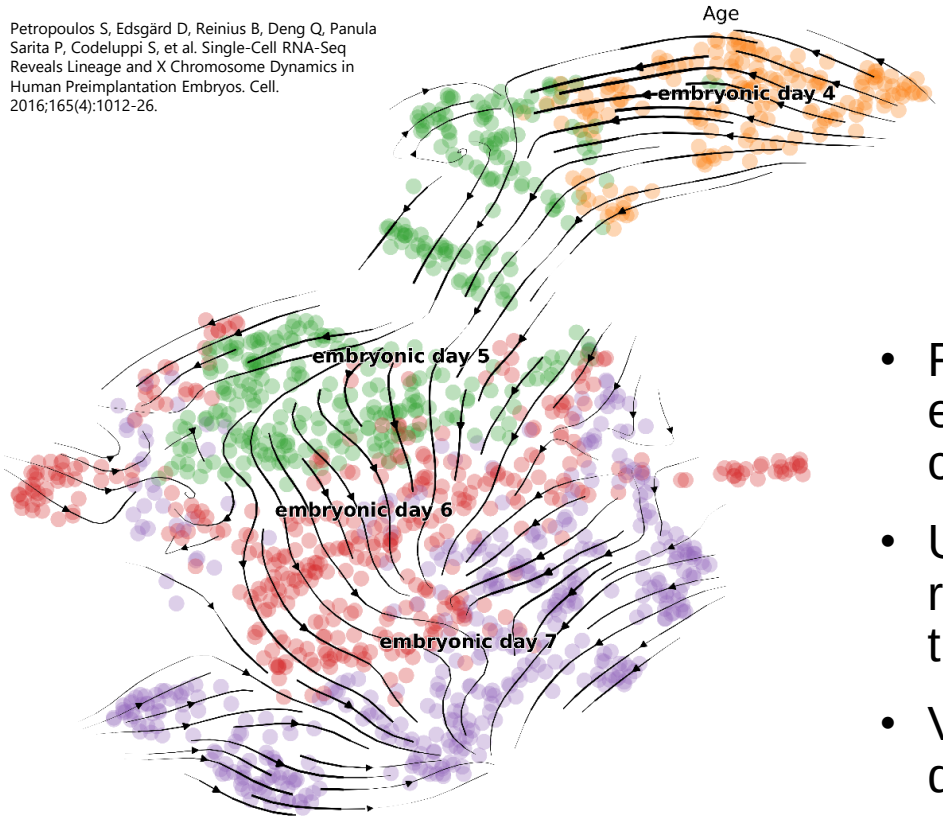
$$\frac{ds}{dt} = u - \gamma s$$

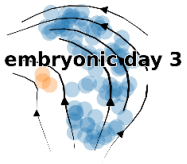


La Manno, G., et al. (2018). "RNA velocity of single cells." Nature **560(7719): 494-498.**

RNA Velocity of Early Human Embryo

Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula Sarita P, Codeluppi S, et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 2016;165(4):1012-26.

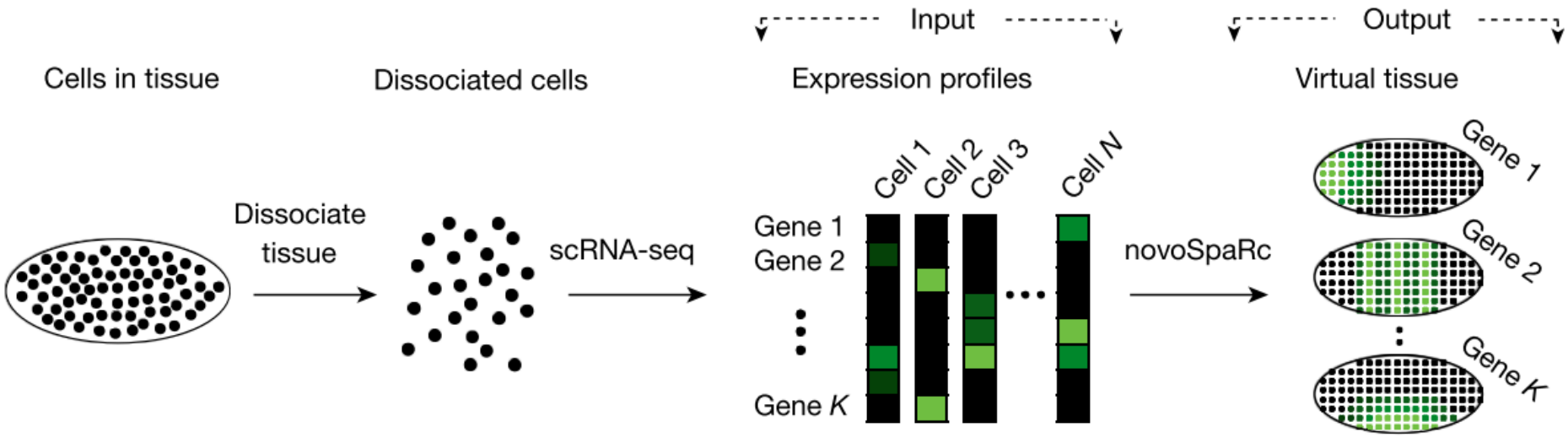


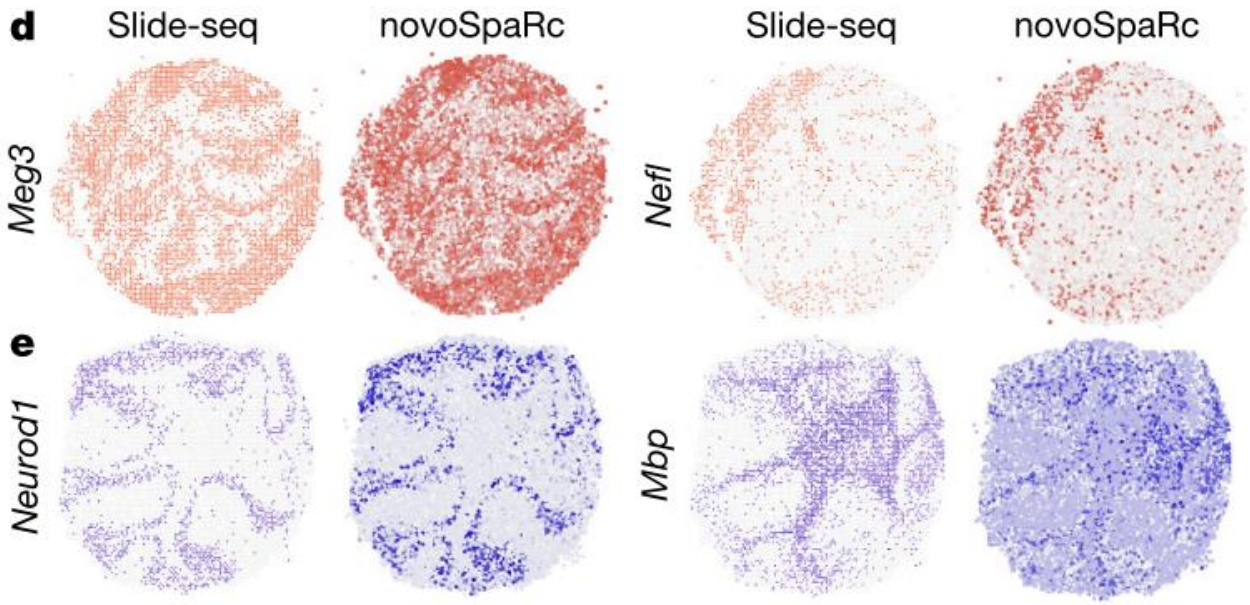


- Reanalysis of >1500 cells from Petropoulos et al. (2016) represented as points, coloured based on age
- UMAP visualisation – dimensionality reduced – cells positioned based on transcriptome similarity
- Velocity streams overlaid on top – captures developmental direction

MANCHESTER 1824
The University of Manchester

RNA Cartography

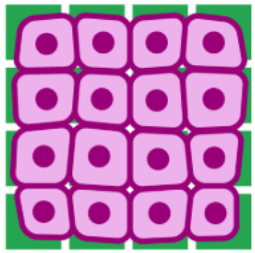


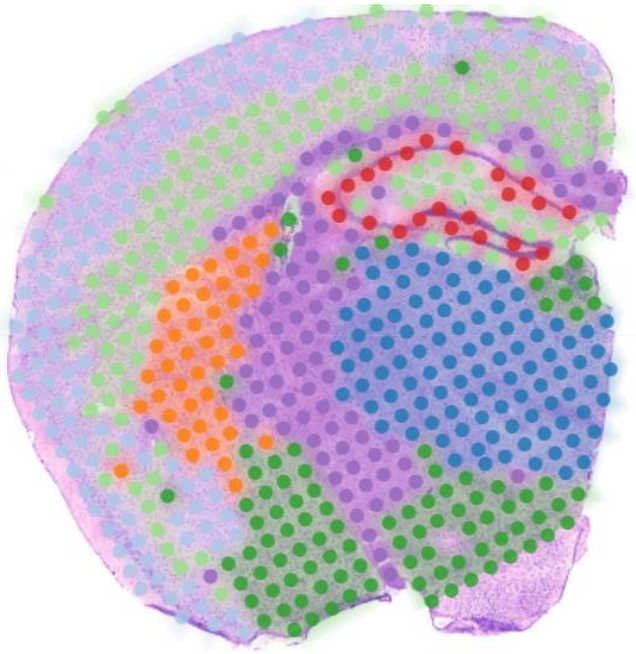


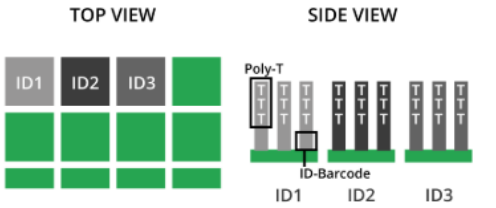
Nitzan, M., et al. (2019). "Gene expression cartography." Nature **576(7785): 132-137.**

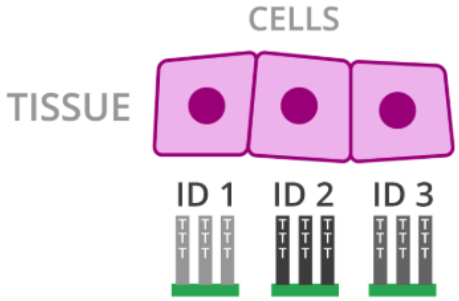
Spatial RNAseq

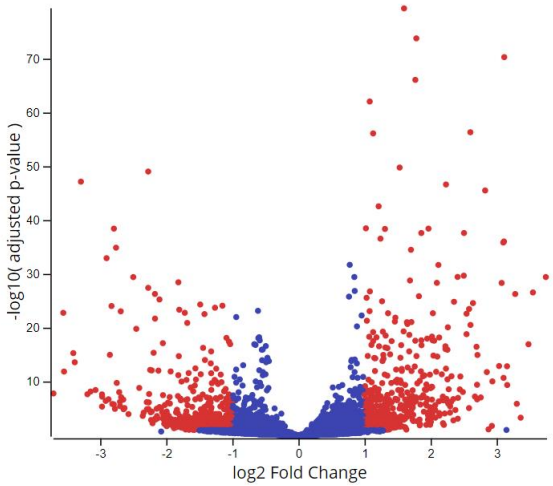
https://spatialtranscriptomics.com

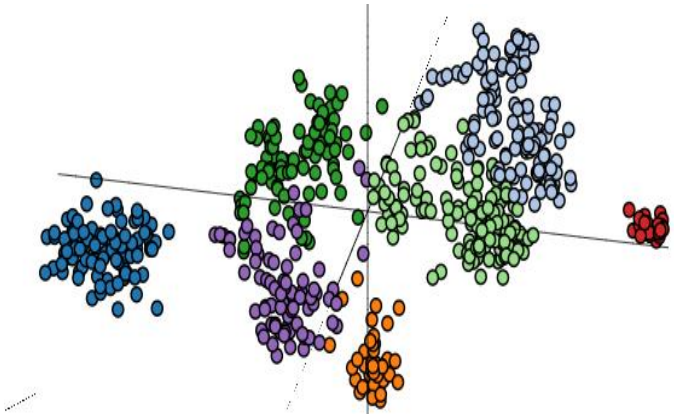












Individual Omic Data Set Analysis

	varID	Gene Symbol	q-Value	Fold change
1	203820_s_at	IGF2BP3	1.19574e-16	3.03628
2	203819_s_at	IGF2BP3	9.37733e-11	2.62628
3	240143_at	---	2.07997e-08	1.76953
4	206659_at	---	3.21346e-07	1.32887
5	228988_at	ZNF711	1.25012e-05	1.56114
6	201417_at	SOX4	1.33199e-05	1.71911
7	207996_s_at	C18orf1	1.33199e-05	1.38872
8	203038_at	PTPRK	1.33317e-05	1.65887
9	201310_s_at	C5orf13	1.38711e-05	1.42943
10	222344_at	---	1.67269e-05	1.40778
11	213808_at	ADAM23	1.88826e-05	1.61604
12	214744_s_at	RPL23	2.49258e-05	1.29967

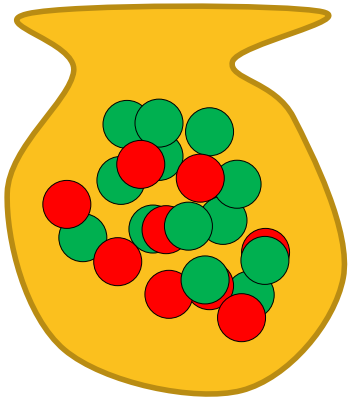
Age group 2 Vs Rest, ANOVA,
q<0.1 = 1524 probe-sets

What does a p-value threshold mean?

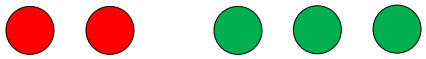
What does a Fold change cut off mean?

Lists and Cut-offs!!

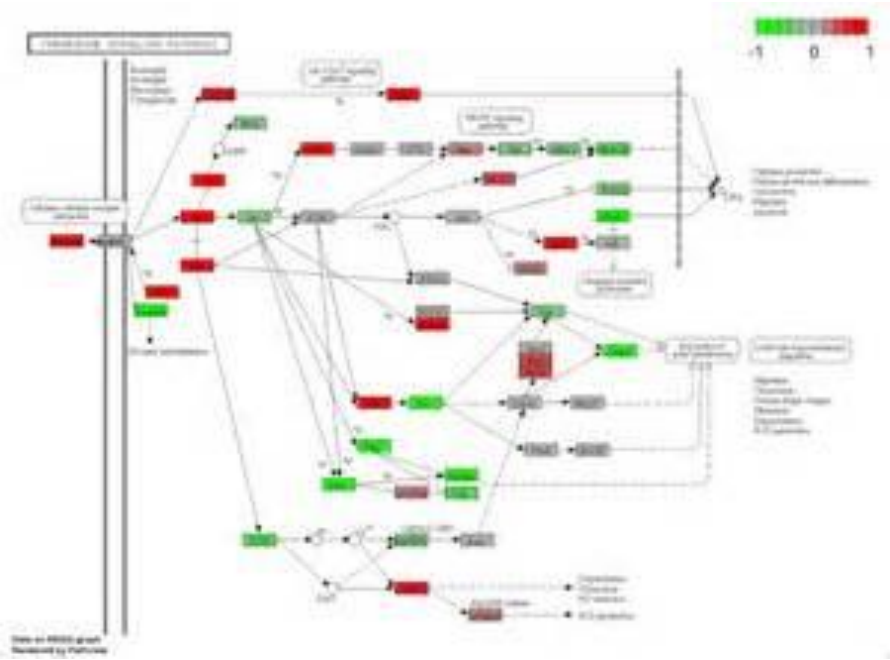
Fisher's Exact Test / Hypergeometric Test



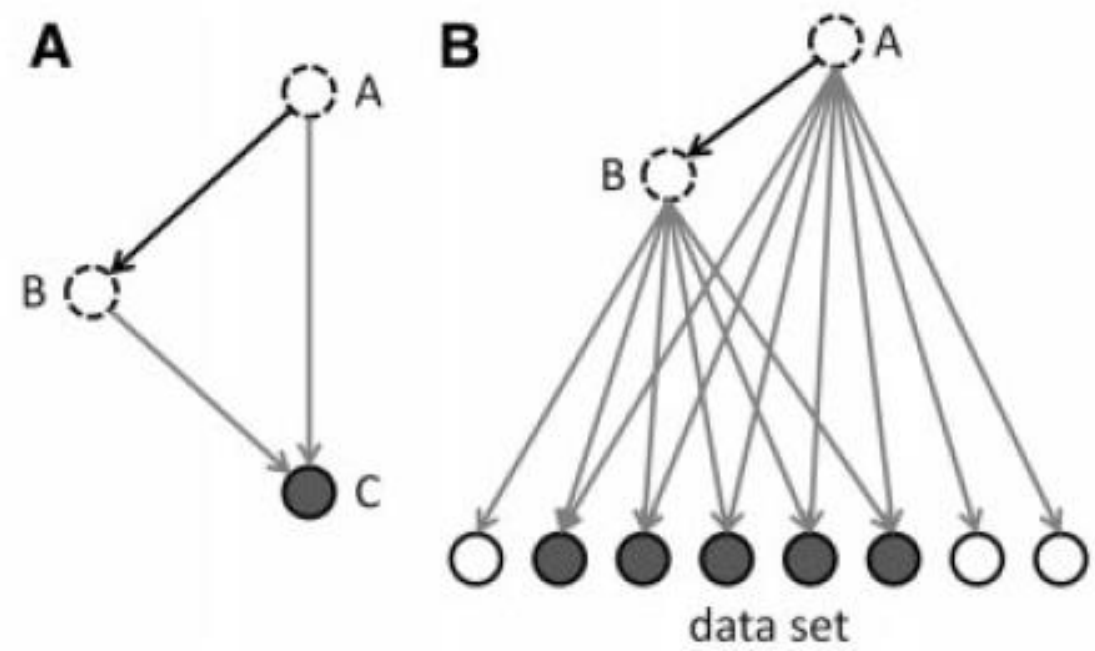
Pick 5 marbles



P(2 are red)



Upstream Regulator Analysis



Enrichment of ‘causal transitive triangles’

Literature Driven analysis

IPA® Interactive pathway analysis of complex ‘omics data

BIOINFORMATICS ORIGINAL PAPER
2014, pages 1–8
doi:10.1093/bioinformatics/btt703
Systems biology
Advance Access publication December 13, 2013

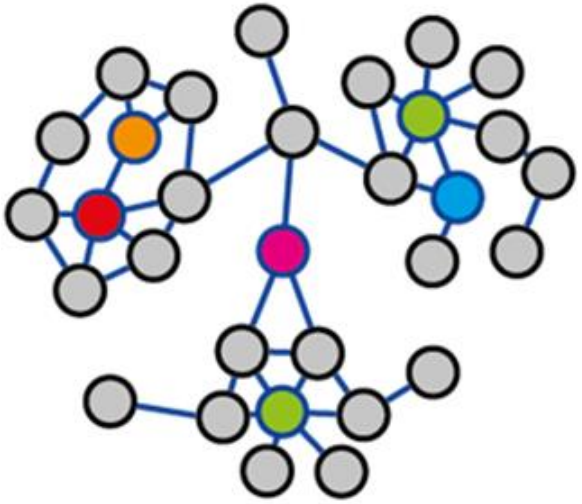
Causal analysis approaches in Ingenuity Pathway Analysis

Andreas Krämer[1,*], Jeff Green[1], Jack Pollard, Jr[2] and Stuart Tugendreich[1]
[1]Ingenuity Systems, 1700 Seaport Boulevard, Redwood City, CA and [2]Translational and Experimental Medicine—Bioinformatics, Sanofi-Aventis, 270 Albany Street, Cambridge, MA, USA
Associate Editor: Jonathan D. Wren

What is Network Biology

Biological networks are:

- **"Scale free"**
- **Resistant to random error**
- **Exhibit "small world" properties**



Nature **393**, 440-442 (4 June 1998) | doi:10.1038/30918; 1998

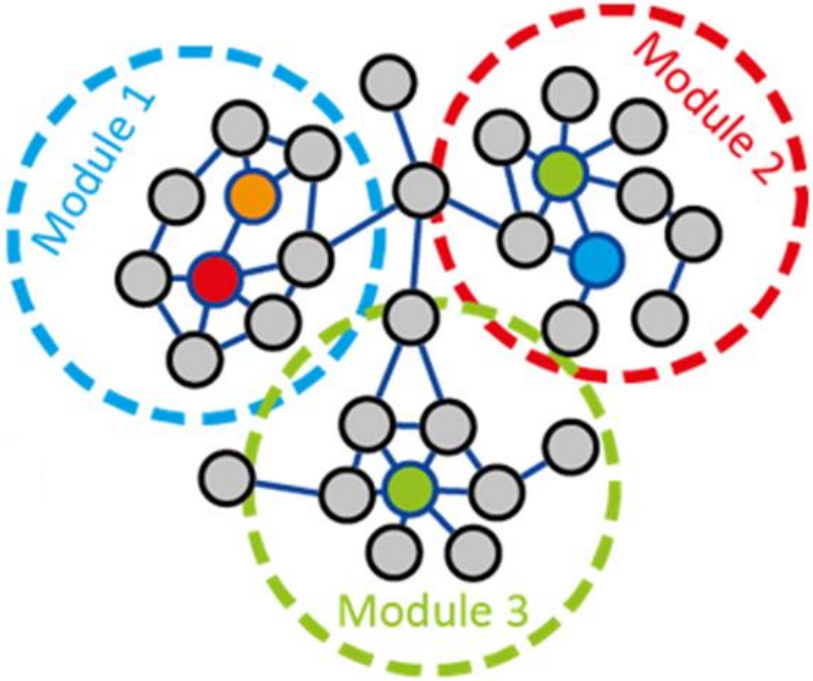
Collective dynamics of 'small-world' networks

Duncan J. Watts[2] & Steven H. Strogatz[1]

$$L \propto \log N$$

typical distance L between two randomly chosen nodes (the number of steps required) grows proportionally to the logarithm of the number of nodes N in the network

small-world network - most nodes are not neighbours of one another, but most nodes can be reached from every other by a small number of steps.



< Previous Article

Volume 159, Issue 5, p1212–1226, 20 November 2014

Resource

Switch to Standard View

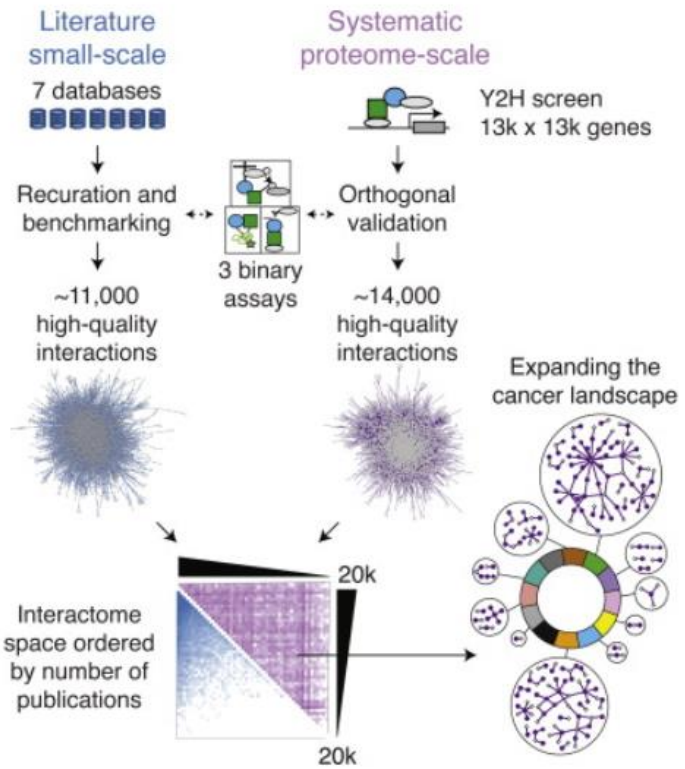
A Proteome-Scale Map of the Human Interactome Network

Thomas Rolland[19], Murat Taşan[19], Benoit Charloteaux[19], Samuel J. Pevzner[19], Quan Zhong[19], Nidhi Sahni[19], Song Yi[19], Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D. Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E. Begg, Pascal Braun, Marc Brehme, Martin P. Broly, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Jasmin Coulombe-Huntington, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J. Gutierrez, Madeleine F. Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan Ning Lin, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R. Murray, Alexandre Palagi, Matthew M. Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviana Romero, Elien Ruyssinck, Julie M. Sahalie, Annemarie Scholz, Akash A. Shah, Amitabh Sharma, Yun Shen, Kerstin Spirohn, Stanley Tam, Alexander O. Tejeda, Shelly A. Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E. Cusick, Yu Xia, Albert-László Barabási, Lilia M. Iakoucheva, Patrick Aloy, Javier De Las Rivas, Jan Tavernier, Michael A. Calderwood[20], David E. Hill[20], Tong Hao[20], Frederick P. Roth, Marc Vidal

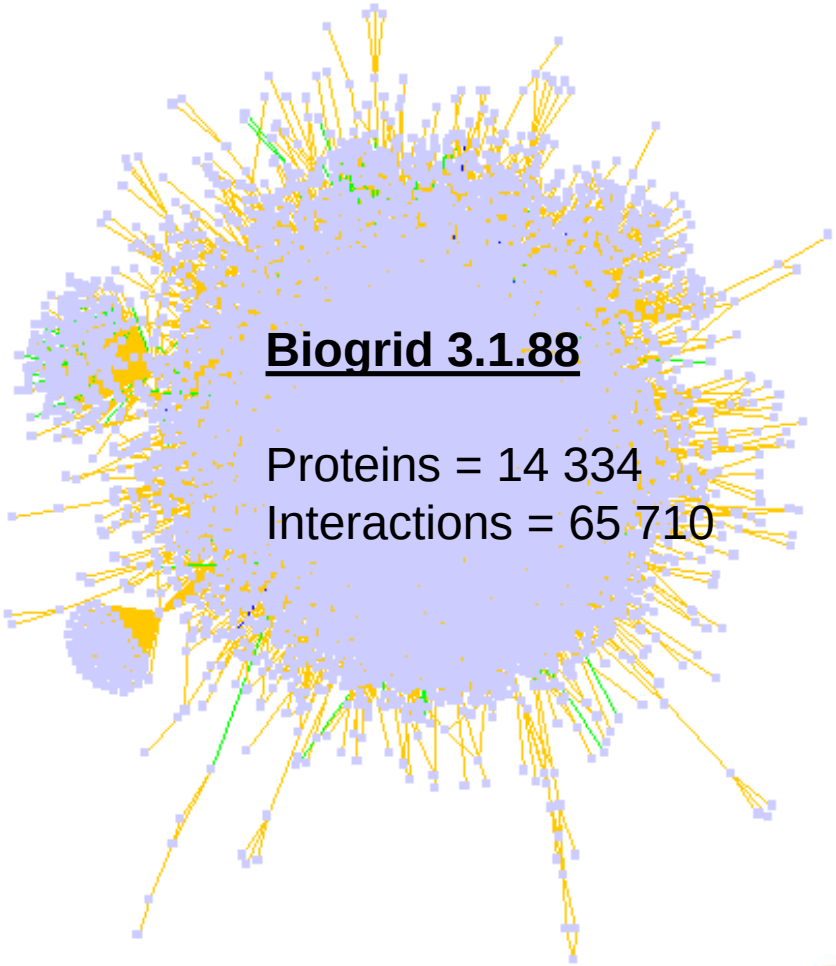
[19] Co-first author

[20] Co-senior author

DOI: http://dx.doi.org/10.1016/j.cell.2014.10.050

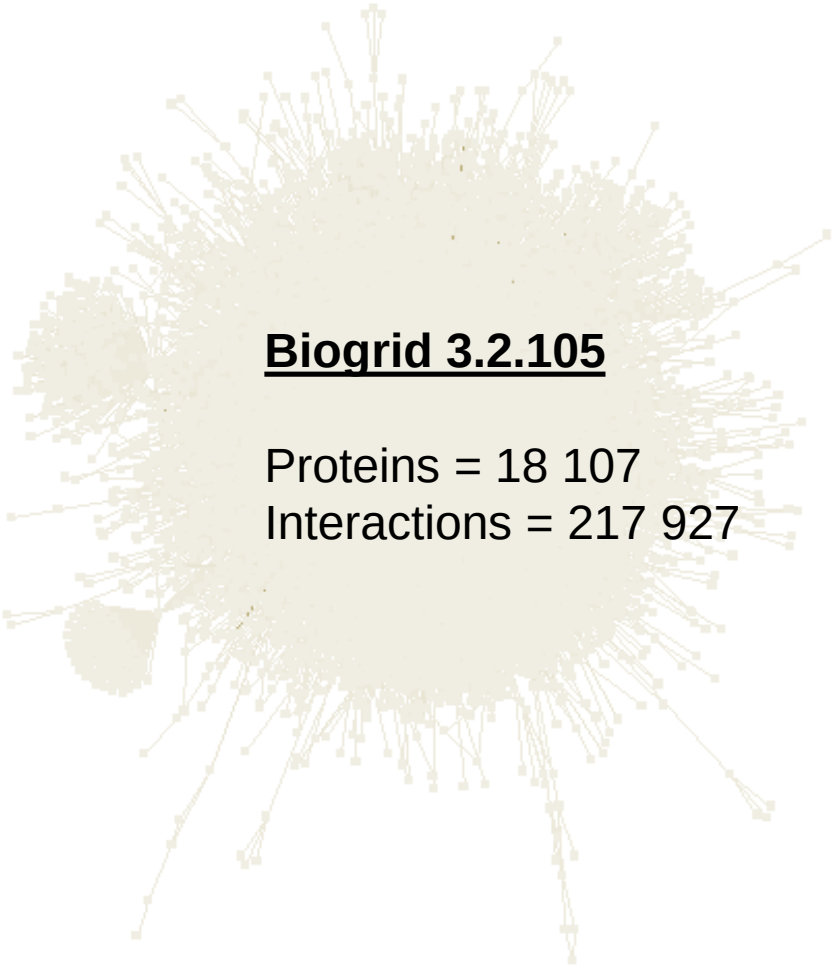


Human Interactome.



Biogrid 3.1.88

Proteins = 14 334
Interactions = 65 710



Biogrid 3.2.105

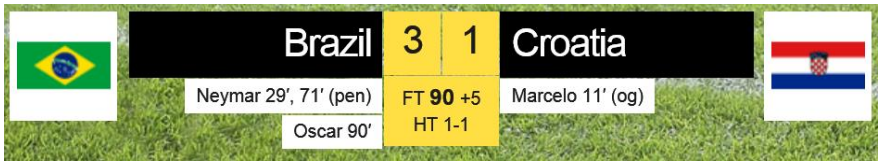
Proteins = 18 107
Interactions = 217 927

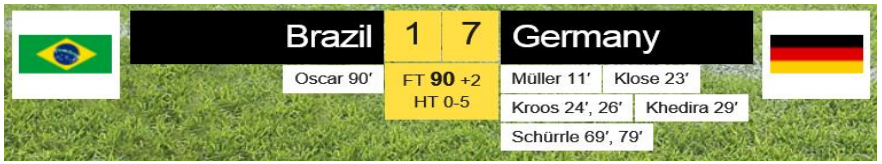
BioGRID 3.2

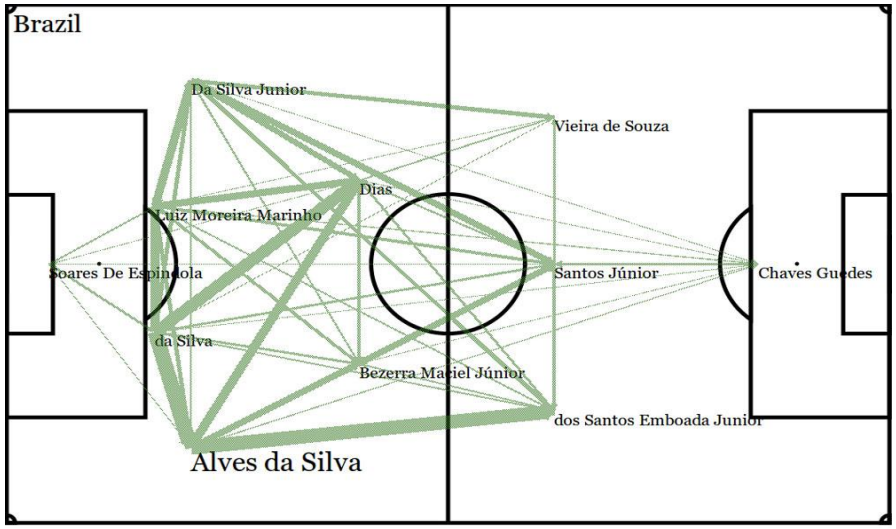
Cytoscape

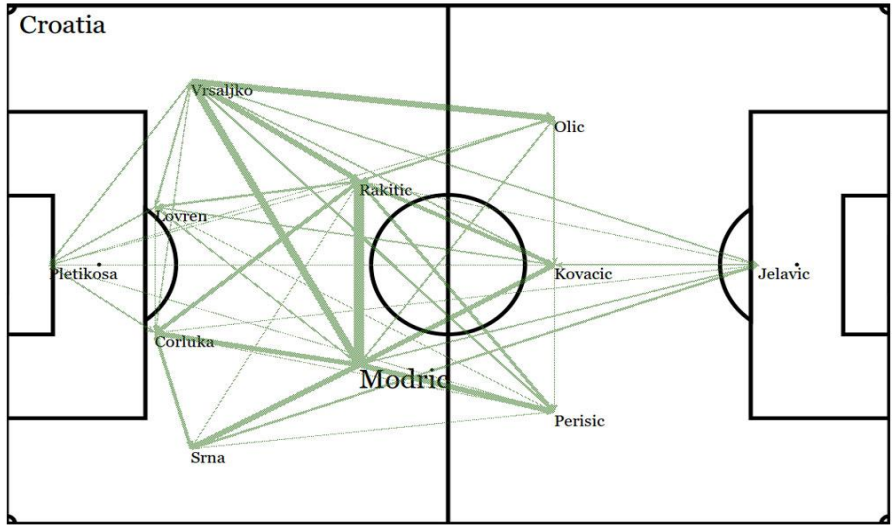
What is a network model

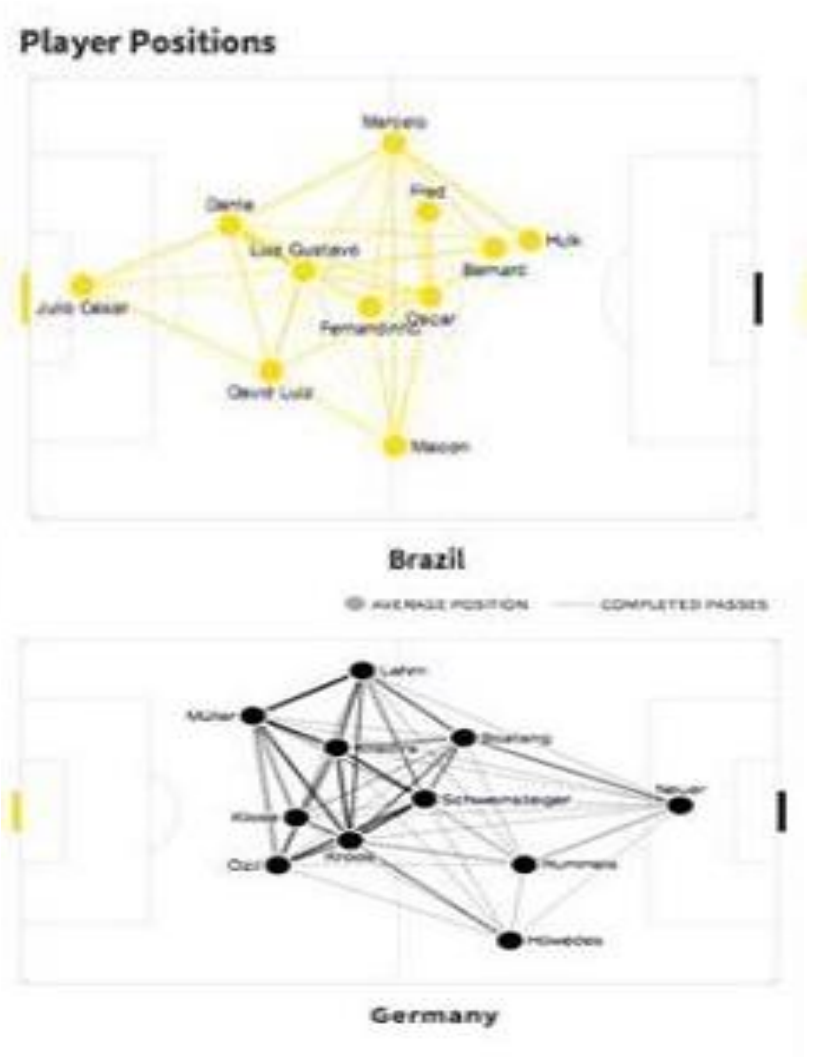
FIFA World Cup 2014







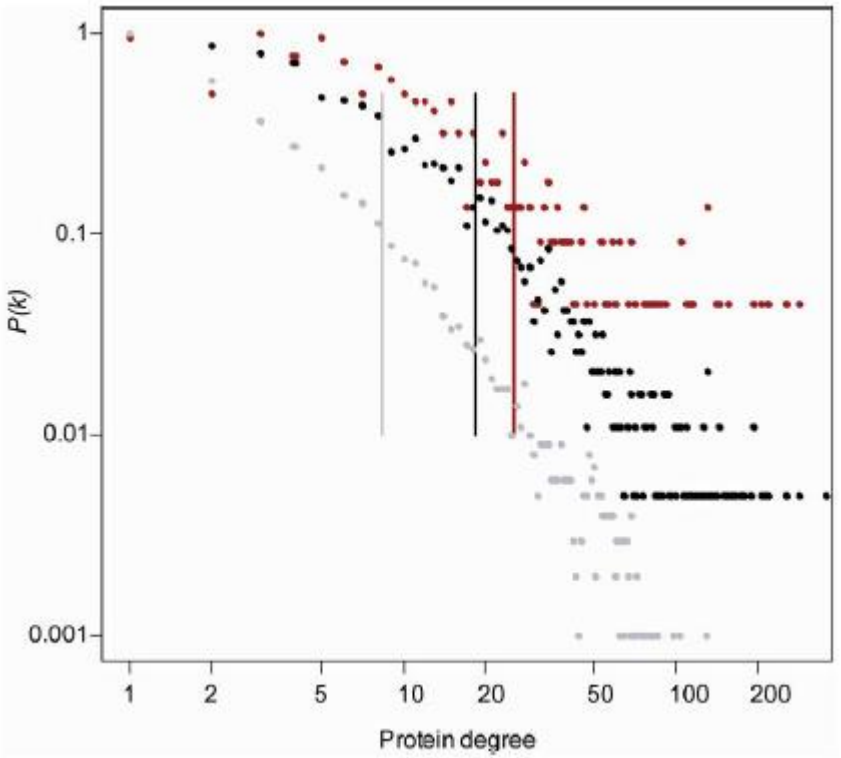


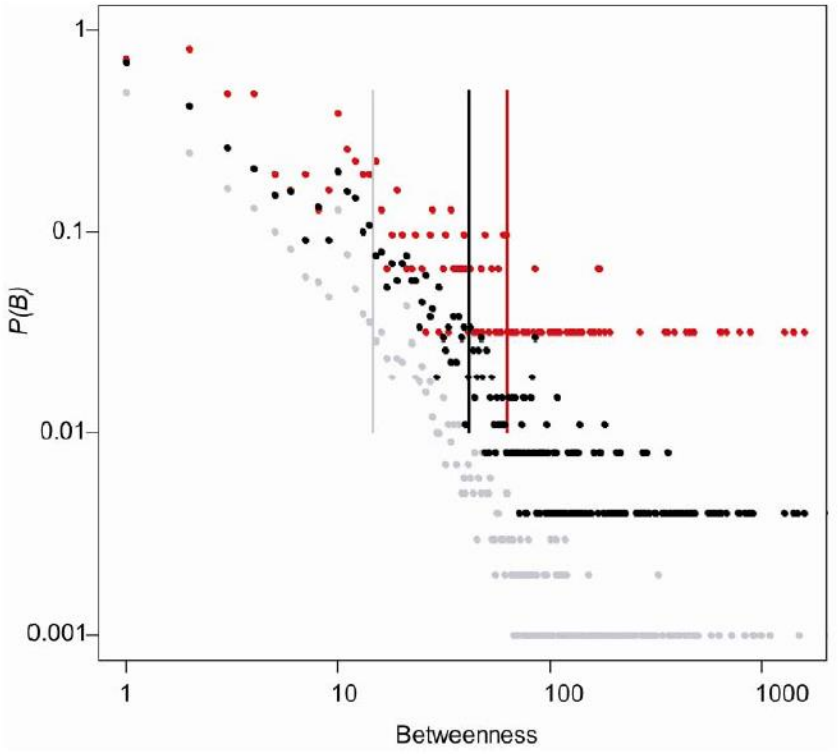


http://blog.physicsworld.com/2014/06/19/a-network-analysis-of-the-fifa-world-cup/

Network Topology Is Associated with Essential Function

Sun *et al* (2010), BMC Genomics **11** S5





Red = Cancer Genes, **Black = Essential Genes,** Grey = Control Genes

Sun and Zhao *BMC Genomics* 2010, **11**(Suppl 3):S5
http://www.biomedcentral.com/1471-2164/11/S3/S5

BMC Genomics

RESEARCH Open Access

A comparative study of cancer proteins in the human protein-protein interaction network

Jingchun Sun[1,2], Zhongming Zhao[1,2,3*]

Chavali *et al. BMC Systems Biology* 2010, **4**:78
http://www.biomedcentral.com/1752-0509/4/78

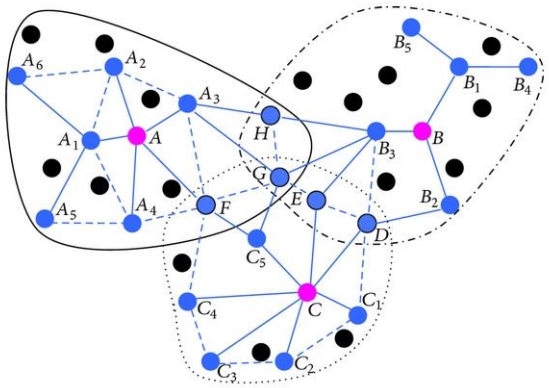
BMC Systems Biology

RESEARCH ARTICLE Open Access

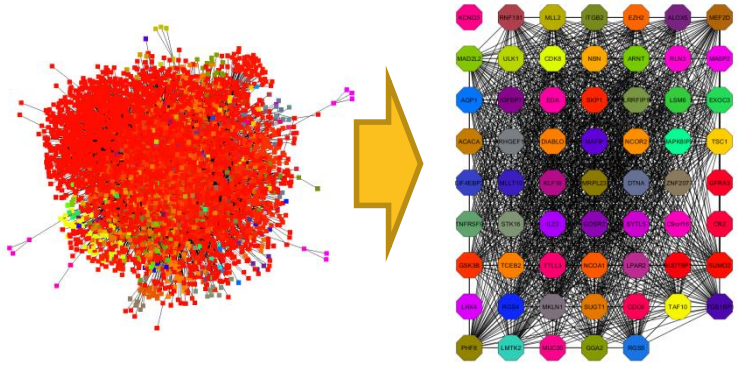
Network properties of human disease genes with pleiotropic effects

Sreenivas Chavali*†, Fredrik Barrenas†, Kartiek Kanduri and Mikael Benson

Network Clusters are associated with hierarchy of biological function



"overlapping modules"



"Overlapping Community Clustering"

Hierarchy of functions and pathways ranked by centrality of cluster

Cluster	Hierarchy of Functions and Pathways	Core nodes of cluster
SUMO2	Cellular responses to stress, Apoptosis, Circadian clock, DNArepair, Metabolism of Proteins ($q<1x10^{-3}$). G1/S DNA damage checkpoint ($q<6.2x10^{-8}$)	**SUMO2**, UBC, MDM2, TP53, EEF1A1, PRKDC, RPS4X, SRRM2, RPS13, RPS20
SKP1	Cell Cycle, Circadian Clock ($q<1x10^{-5}$). Wnt & Prolactin signalling ($q<1.0x10^{-5}$).	**SKP1**, BTRC, CUL1,GSK3B, CTNNB1, SKP2, NFKBIA, CLSPN, FBXW11, FBXO6
GSK3B	Apoptosis, Signal transduction ($q<0.01$). Wnt, PI3/AKT signalling ($q<0.01$).	**GSK3B**, AXIN1, APP, AKT1, MAPT, CTNNB1, ELAVL1, KIF5B, AXIN2, HIPK2
NCOA1	Transcriptional regulation of white adipocyte differentiation, Regulation of lipid metabolism by PPARalpha, Circadian Clock, Mitochondrial biogenesis ($q<0.01$).	**NCOA1**, NCOA6, ESR1, PPARG, MLL3, RXRA, ESR2, MLL4, NCOR2, VDR
EZH2	Cellular Senescence, Epigenetic regulation of gene expression ($q<1.0x10^{-4}$)	**EZH2,** EED, SUZ12, EZH1, JARID2, SON, SRSF7, NRF1, FBL, HDAC1
TCEB2	Cellular response to hypoxia ($q<1x10^{-5}$). TGF-beta Signalling ($q<0.05$)	**TCEB2,** VHL, CUL5, TCEB1, CUL2, TCEB3, ASB9, STK16, NEDD8, COPS6
NCOR2	Transcriptional regulation of white adipocyte differentiation, Regulation of lipid metabolism by PPARalpha, Circadian Clock, Mitochondrial biogenesis ($q<0.01$).	**NCOR2**, NCOA6, HDAC3, BCL6, RARA, AR, ANKRD11, THRB, HDAC1, KDM5B
DIABLO	Immune system, Apoptosis ($q<0.05$) Toll-Like Receptors Cascades ($q<1.0x10^{-4}$)	**DIABLO**, XIAP, BIRC2, BIRC6, UBE2D4, BIRC3, BIRC7, TRAF2, BIRC5, ELAVL1
NBN	DNA Repair, Cell Cycle ($q<1.0x10^{-4}$) G2/M DNA damage checkpoint ($q<6.7x10^{-4}$)	**NBN**, MRE11A, MDC1, RAD50, BRCA1, H2AFX, FANCD2, ATM, TP53BP1, ATR
TSC1	Cellular response to stress ($q<0.01$). Insulin, IGF1, mTOR, PI3K signalling ($q<1.0x10^{-5}$).	**TSC1**, RHEB, TSC2, YWHAE, MTOR, YWHAB, RAF1, RPTOR, BECN1, MAST3

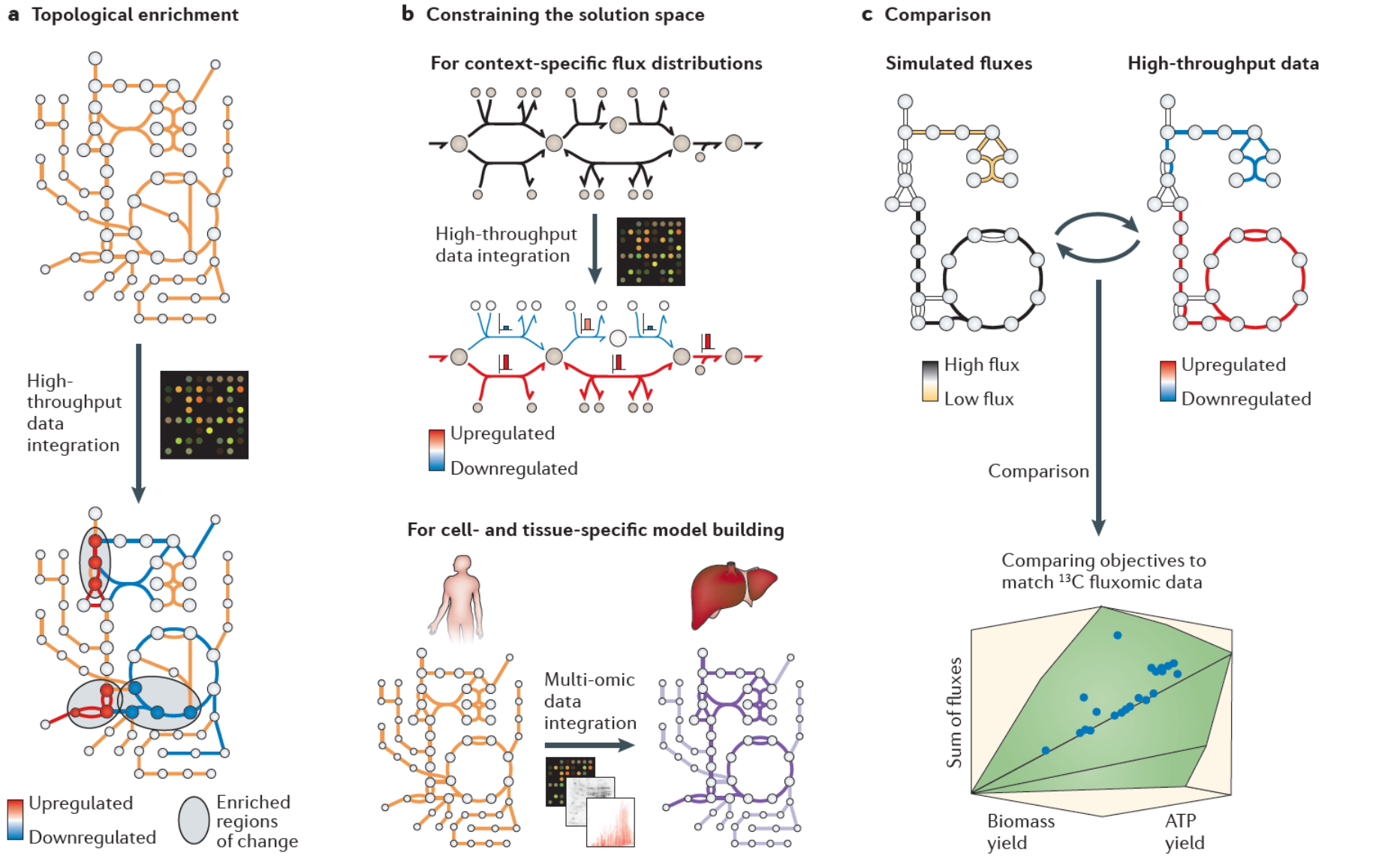
Bioinformatics (2012) 28: 2202-2204

Data and text mining

ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality

Máté Szalay-Bekő[1,‡] and Robin Palotai[1,‡], Balázs Szappanos[2], István A. Kovács[1,3], Balázs Papp[2] and Péter Csermely[1,*]

Constraint Based Modelling



Why are methods of prioritising Omic data important?

Answer: Lists don't necessarily deliver the solution!

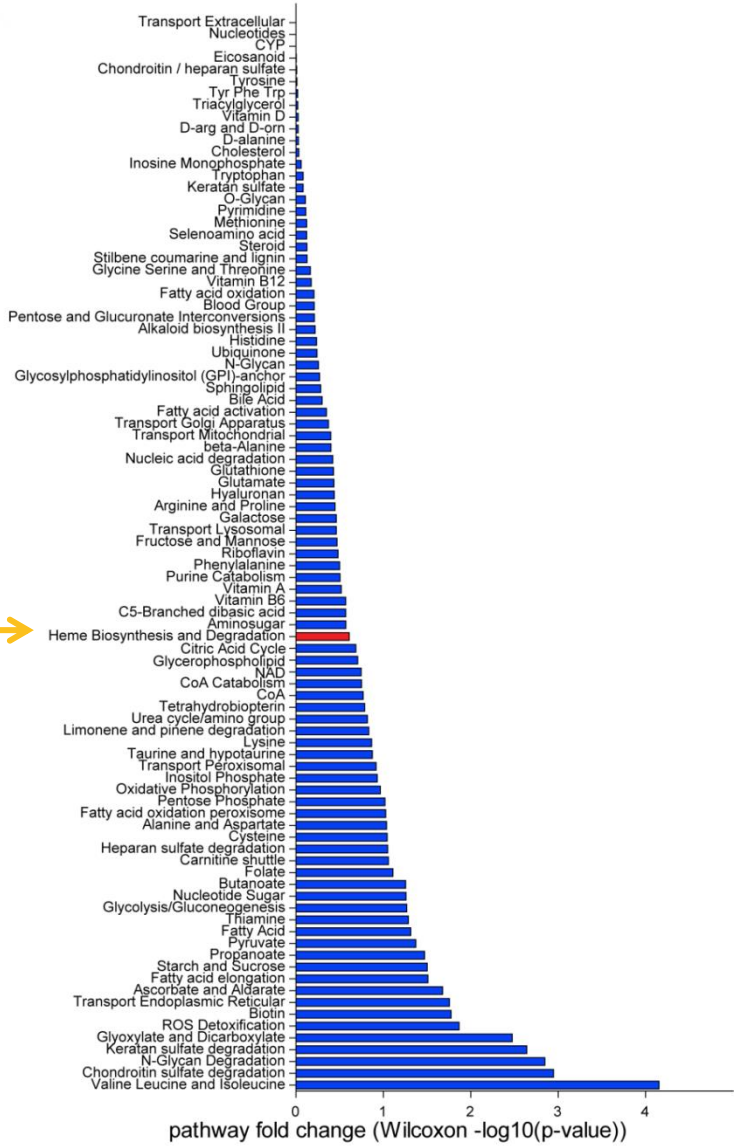
Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase

Christian Frezza, Liang Zheng, Ori Folger, Kartik N. Rajagopalan, Elaine D. MacKenzie, Livnat Jerby, Massimo Micaroni, Barbara Chaneton, Julie Adam, Ann Hedley, Gabriela Kalna, Ian P. M. Tomlinson, Patrick J. Pollard, Dave G. Watson, Ralph J. Deberardinis, Tomer Shlomi, Eytan Ruppin & Eyal Gottlieb

Affiliations | Contributions | Corresponding author

Nature **477**, 225–228 (08 September 2011) | doi:10.1038/nature10363
Received 13 July 2010 | Accepted 11 July 2011 | Published online 17 August 2011

Gene set analysis of Transcriptomic data



Systems Analysis by constraint based modelling predicts correctly

Exponential Random Graphs – Network statistics

Many structural elements can be used to model networks

E.G.

Tie

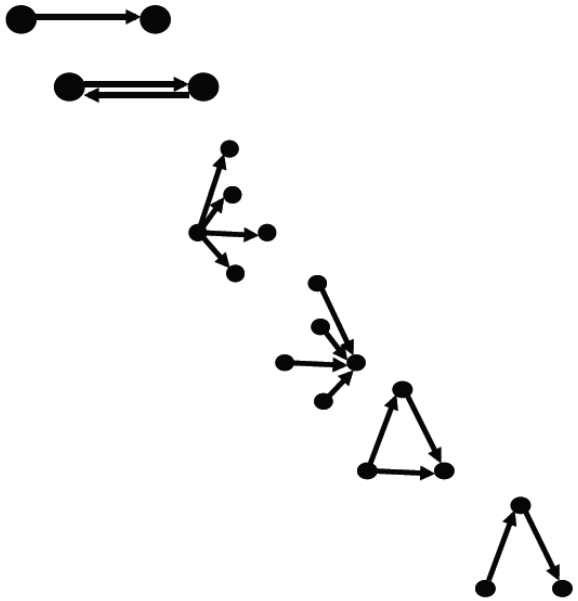
Reciprocity

Activity

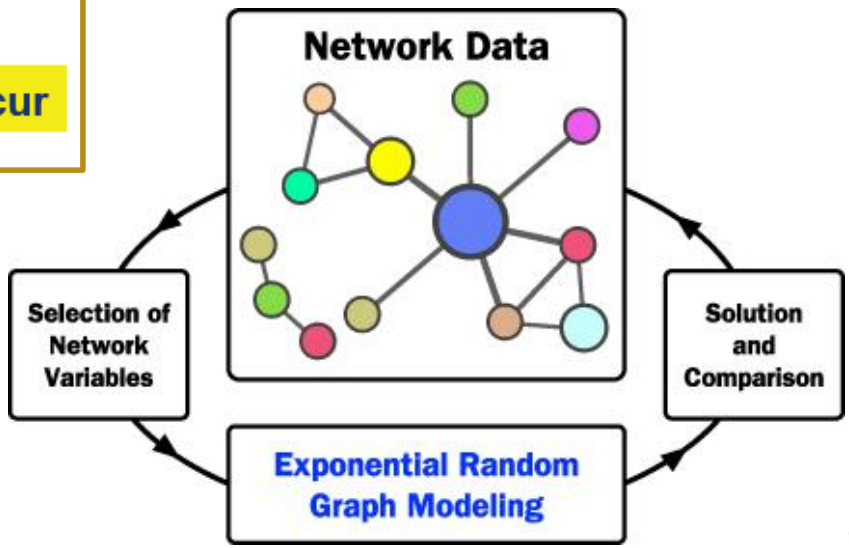
Popularity

Triads

Brokerage



No single rule explains why ALL network ties occur



Integration of Multi-Omic Data

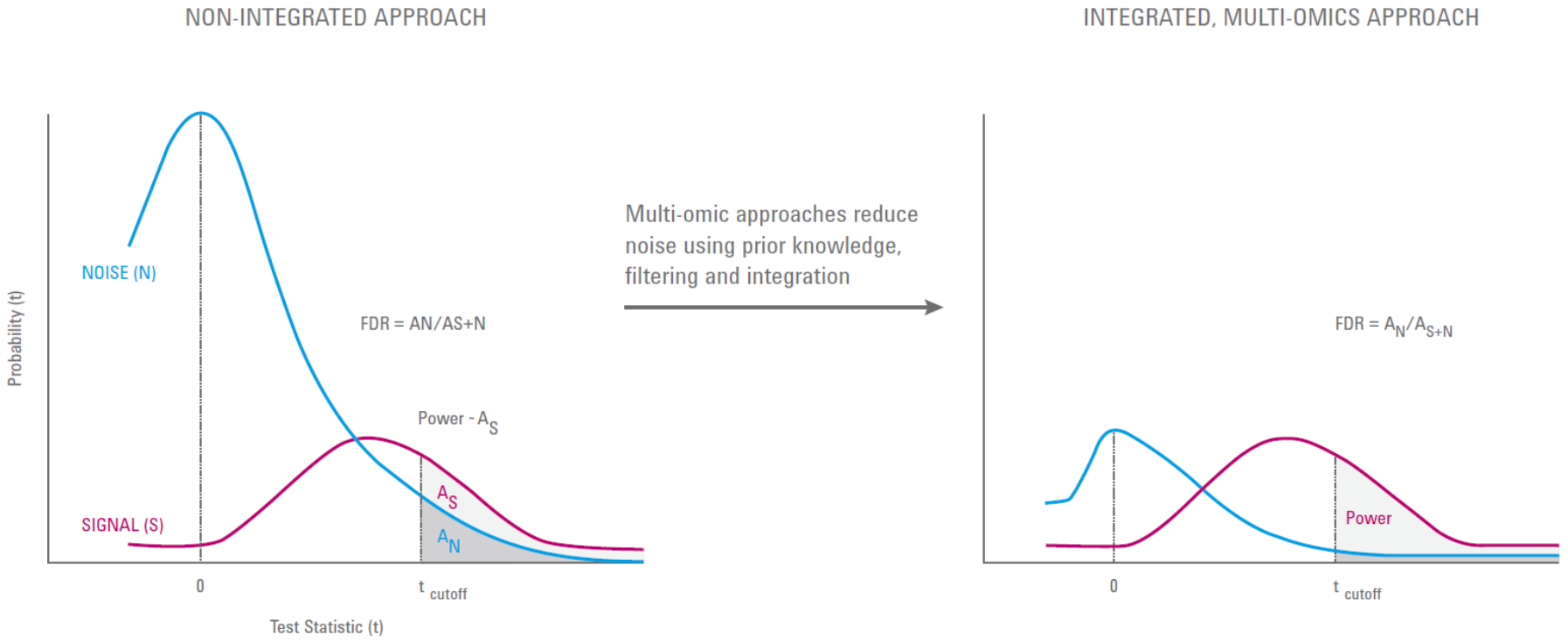
Cell Volume 144, Issue 6, p860–863, 18 March 2011

Leading Edge
Minireview

Boosting Signal-to-Noise in Complex Biology: Prior Knowledge Is Power

Trey Ideker,[1,2,*] Janusz Dutkowski,[1] and Leroy Hood[3]

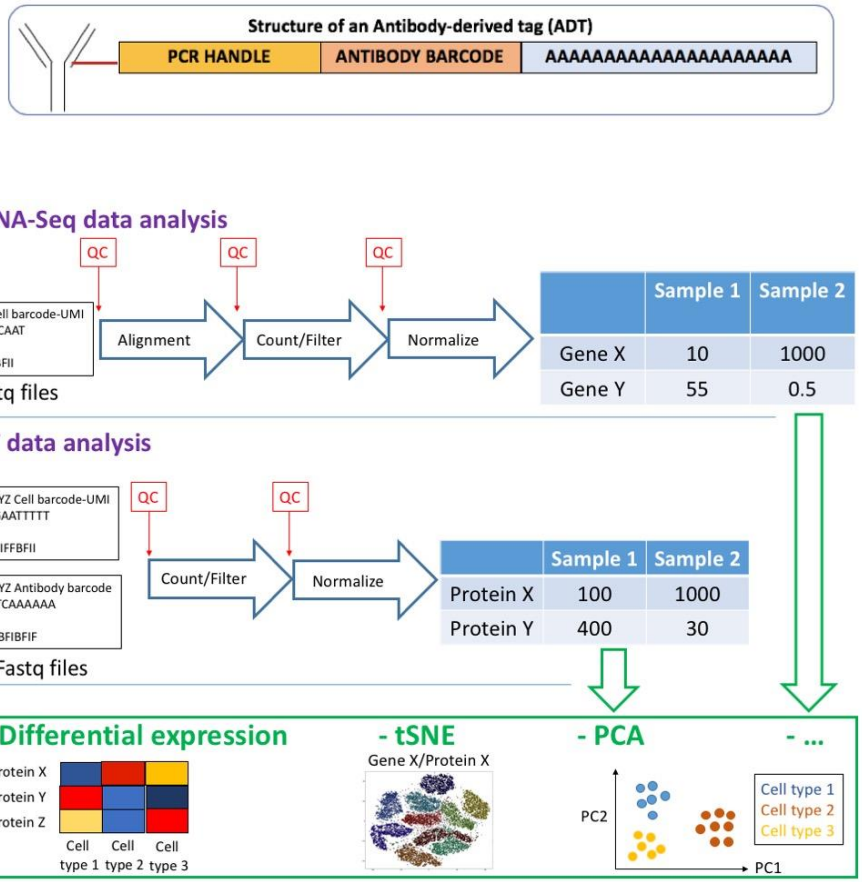
"Network Biology is a primary tool"

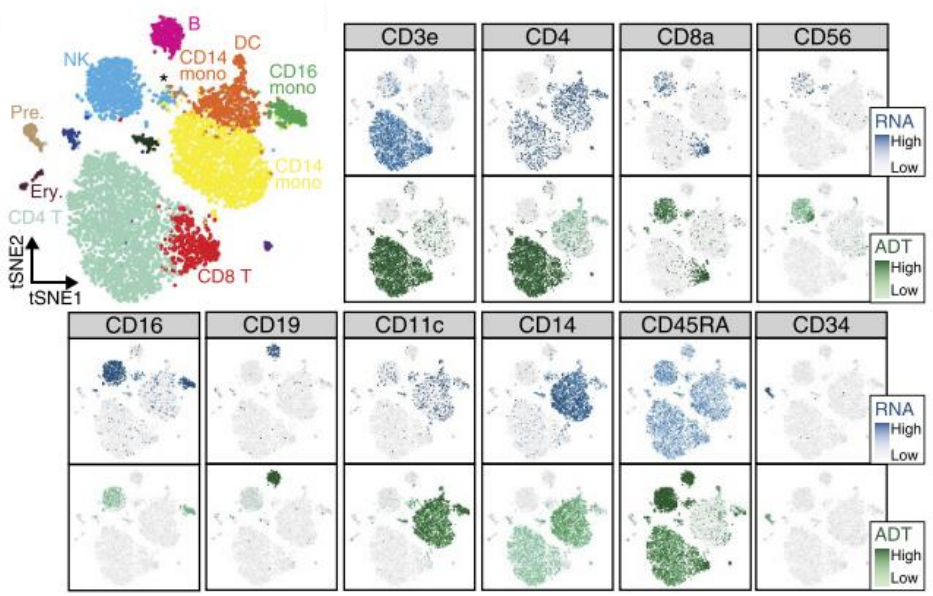


The statistical power of omics experiments can be enhanced through bioinformatics methods that decrease noise through the use of (1) complementary datasets, and (2) incorporation of prior knowledge about the system (e.g., aggregating measurements from entities that belong to the same pathway). This results in an effective decrease in the False Discovery Rate (FDR), at a given t statistic cutoff, or in the ability to relax such cutoff while maintaining the same FDR.

RNAseq + Proteomics

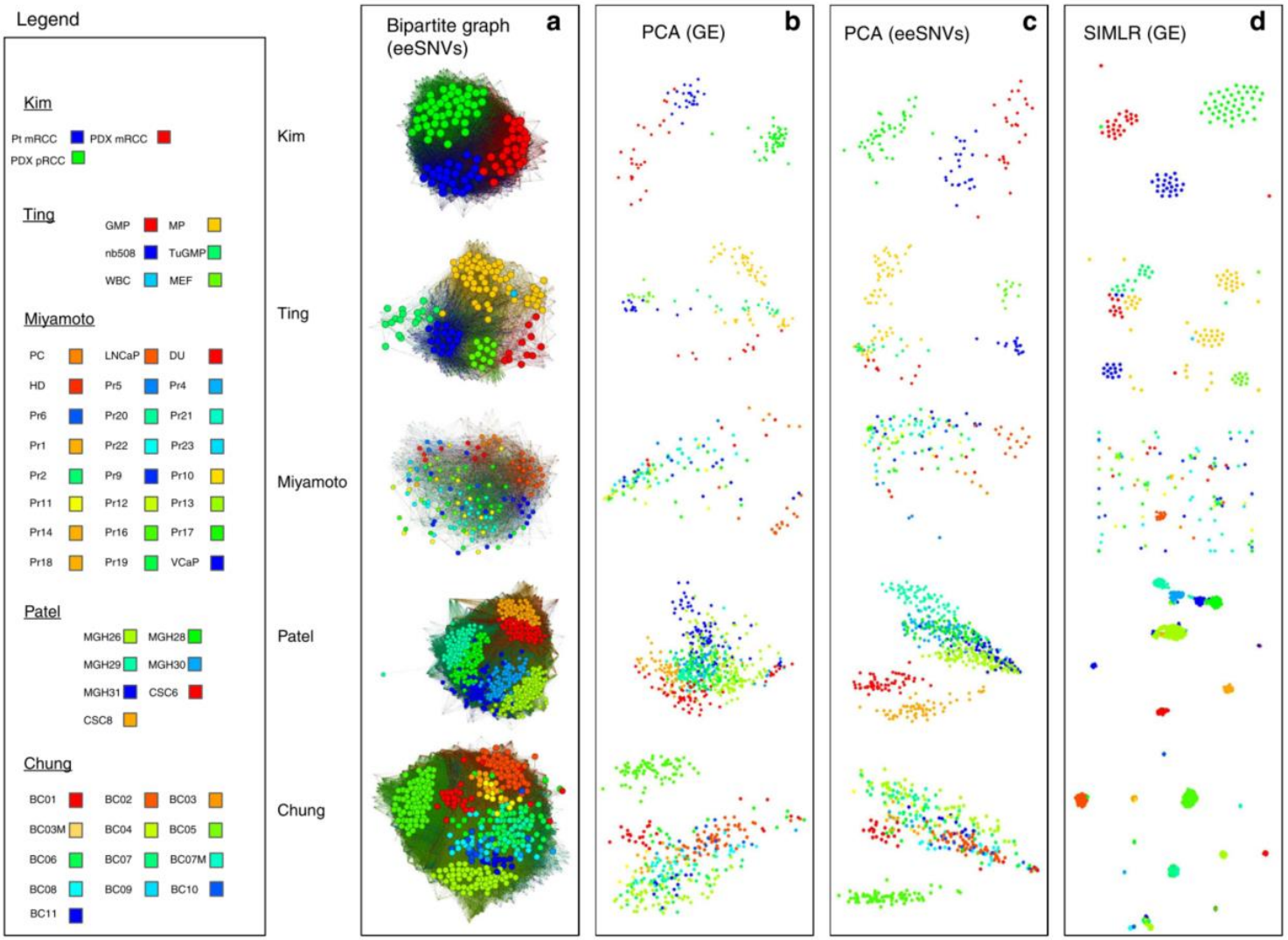
CITE-Seq (**C**ellular **I**ndexing of **T**ranscriptomes and **E**pitopes by **Seq**uencing)





Stoeckius, M., et al. (2017). "Simultaneous epitope and transcriptome measurement in single cells." Nature Methods **14(9): 865-868.**

Using single nucleotide variations in single-cell RNA-seq to identify subpopulations and genotype-phenotype linkage



expressed nucleotide variations = eeSNVs

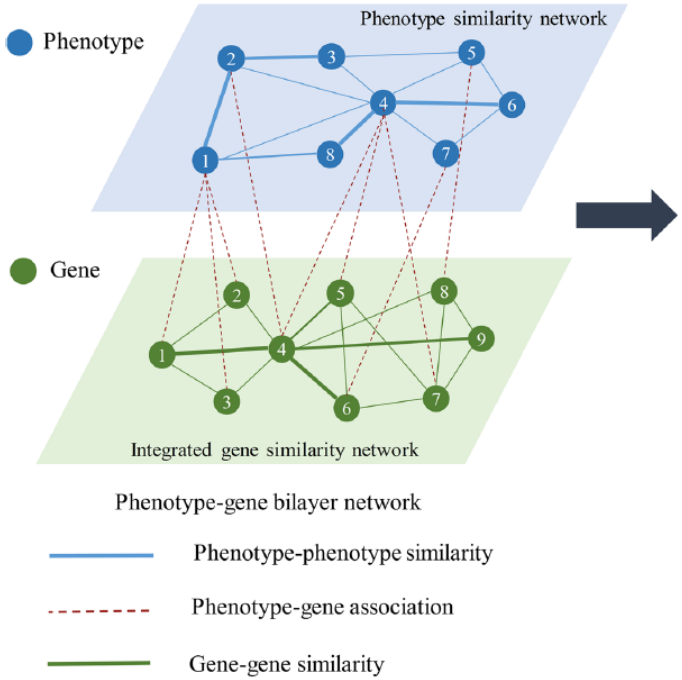
Poirion, O., et al. (2018). "Using single nucleotide variations in single-cell RNA-seq to identify subpopulations and genotype-phenotype linkage." Nat Commun 9(1): 4892.

Integration of Networks to Enhance Clinical Prediction

NATURE METHODS | VOL.11 NO.3 | MARCH 2014

Similarity network fusion for aggregating data types on a genomic scale

Bo Wang[1,5], Aziz M Mezlini[1,2], Feyyaz Demir[1,2], Marc Fiume[2], Zhuowen Tu[3], Michael Brudno[1,2], Benjamin Haibe-Kains[4,5] & Anna Goldenberg[1,2]



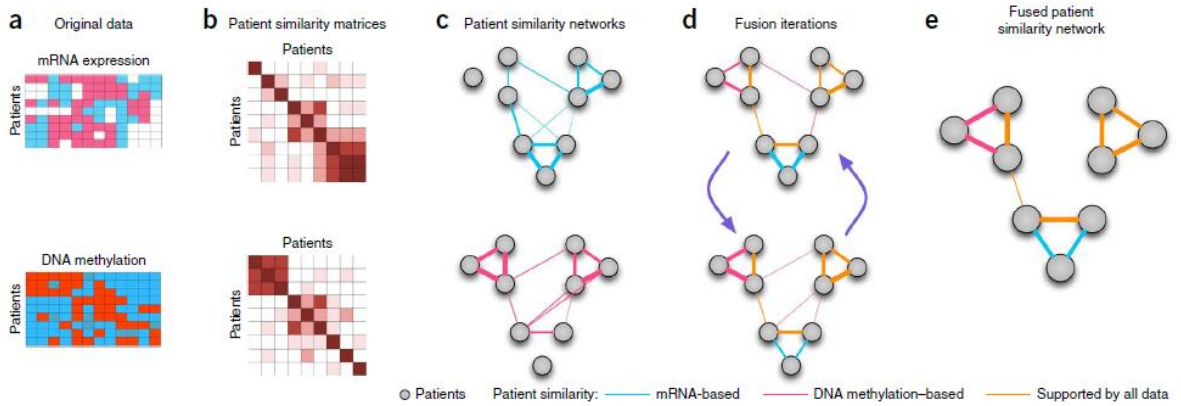
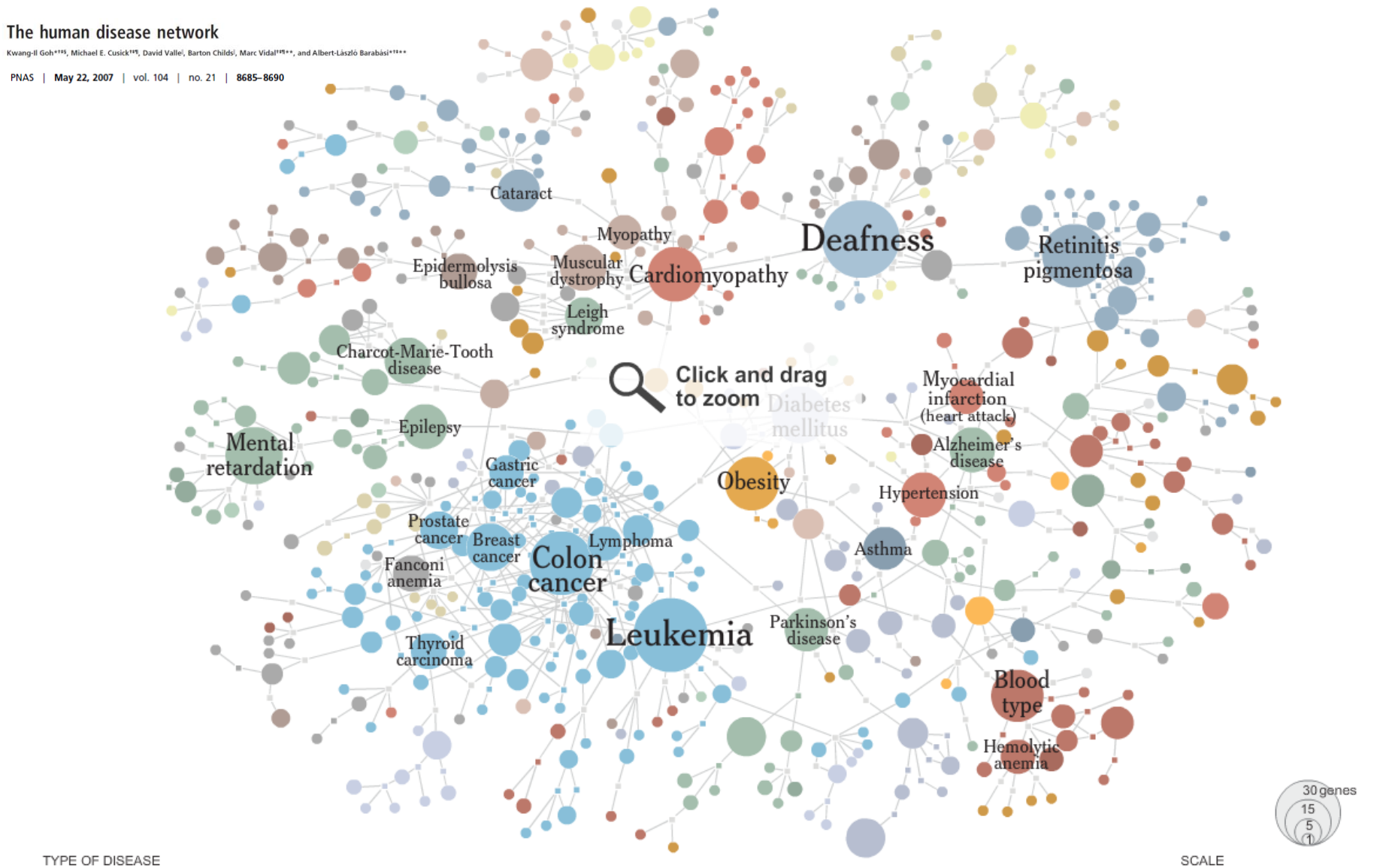


Figure 1 | Illustrative example of SNF steps. (**a**) Example representation of mRNA expression and DNA methylation data sets for the same cohort of patients. (**b**) Patient-by-patient similarity matrices for each data type. (**c**) Patient-by-patient similarity networks, equivalent to the patient-by-patient data. Patients are represented by nodes and patients' pairwise similarities are represented by edges. (**d**) Network fusion by SNF iteratively updates each of the networks with information from the other networks, making them more similar with each step. (**e**) The iterative network fusion results in convergence to the final fused network. Edge color indicates which data type has contributed to the given similarity.

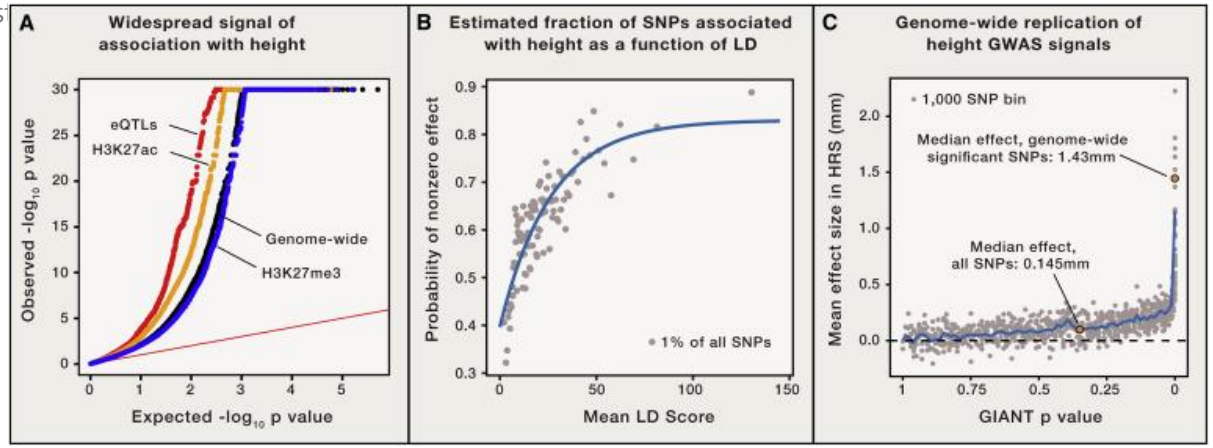
Mapping the Human 'Diseasome'

Researchers created a map linking different diseases, represented by circles, to the genes they have in common, represented by squares.

Related Article: Redefining Disease, Genes and All



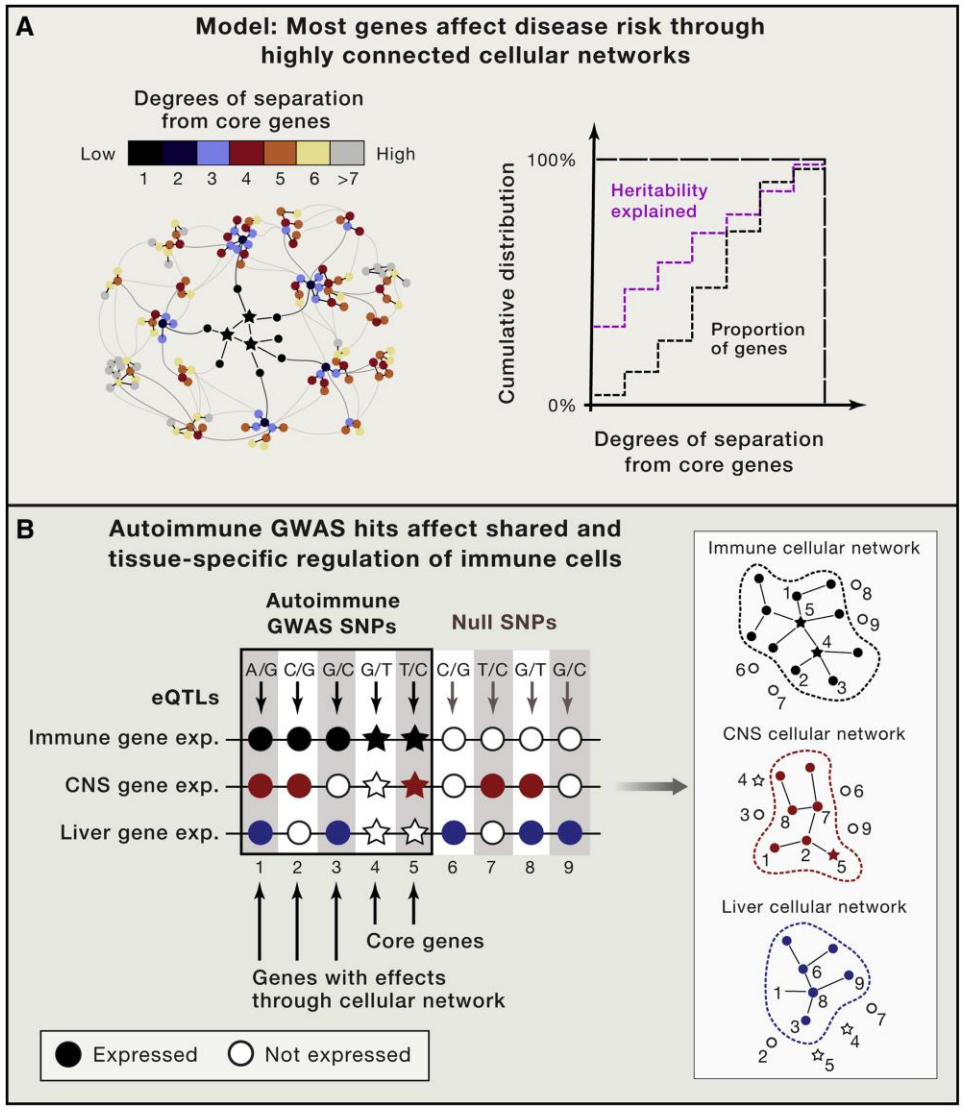
An Expanded View of Complex Traits: From Polygenic to Omnigenic



- A central goal of genetics is to understand the links between genetic variation and disease.
- Intuitively, one might expect disease-causing variants to cluster into key pathways that drive disease etiology.
- But for complex traits, association signals tend to be spread across most of the genome—including near many genes without an obvious connection to disease.
- We propose that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that **most heritability can be explained by effects on genes outside core pathways**.
- We refer to this hypothesis as an “omnigenic” model.

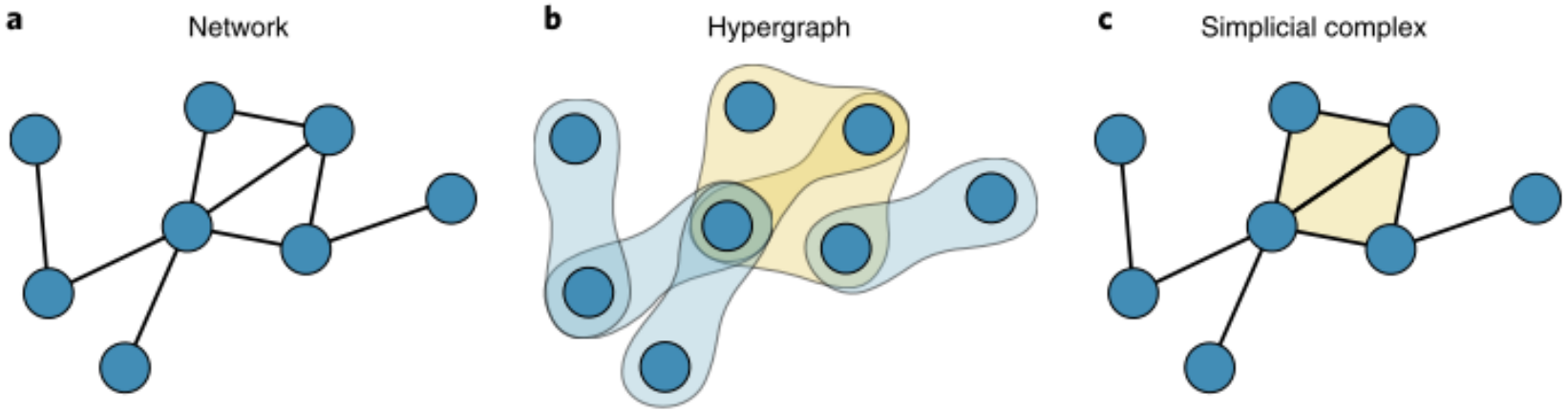
Cell - Volume 169, Issue 7, p1177–1186, 15 June 2017

An Expanded View of Complex Traits: From Polygenic to Omnigenic



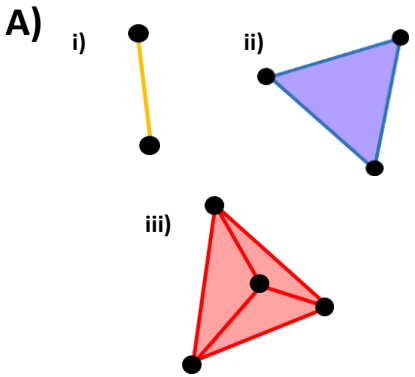
Cell - Volume 169, Issue 7, p1177–1186, 15 June 2017

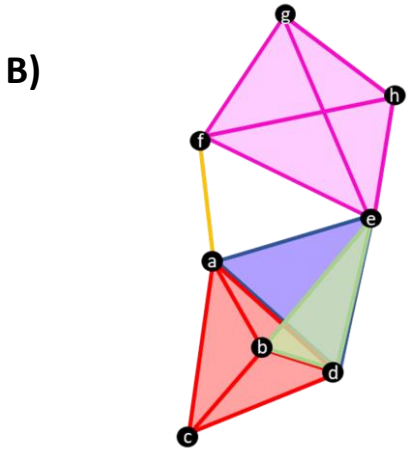
Hypernetworks



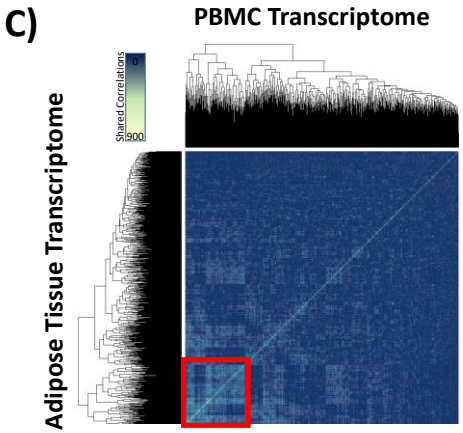
Battiston, F., et al. (2021). "The physics of higher-order interactions in complex systems." Nature Physics **17(10): 1093-1098.**

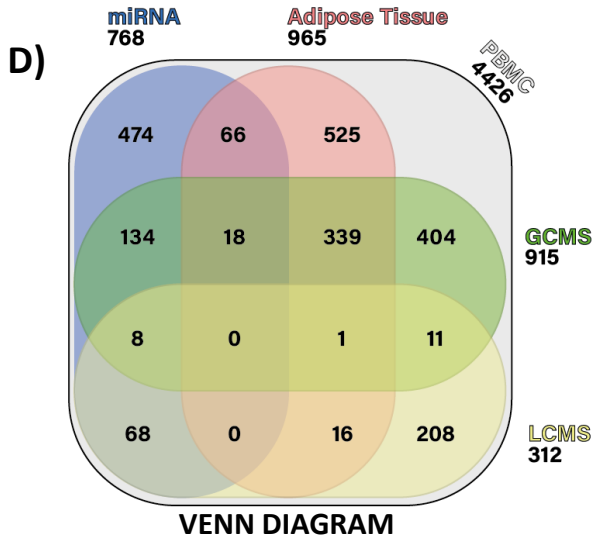
Hypernetworks





Chantzichristos, D., et al. (2021). "Identification of human glucocorticoid response markers using integrated multi-omic analysis from a randomized crossover trial." Elife **10: e62236.**





Thanks

Kat Hollywood

Senior Experimental Officer