Seeing is Believing’

Enrico Capobianco

Re-Imagining Biomedicine with Big Data.

Center for Computational Science

University of Miami

Miller School of Medicine, SCCC

http://ccs.miami.edu/

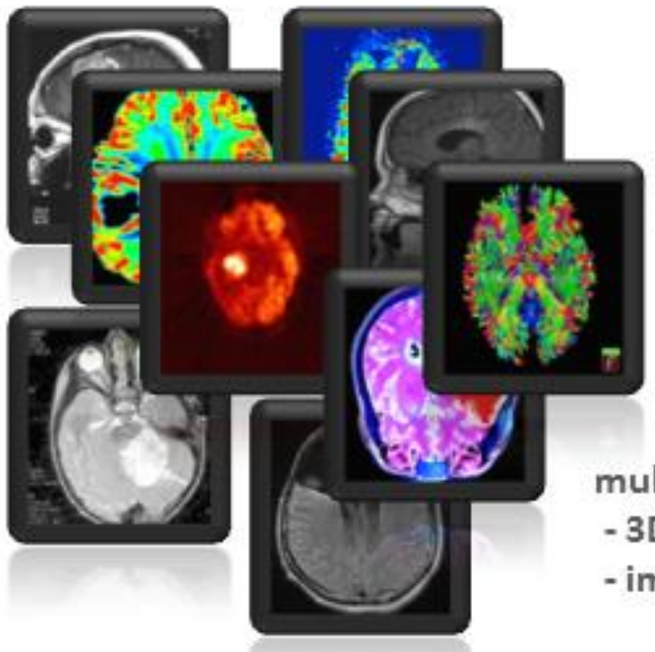
Outline

Learn = Diagnose

Predict = Prevent

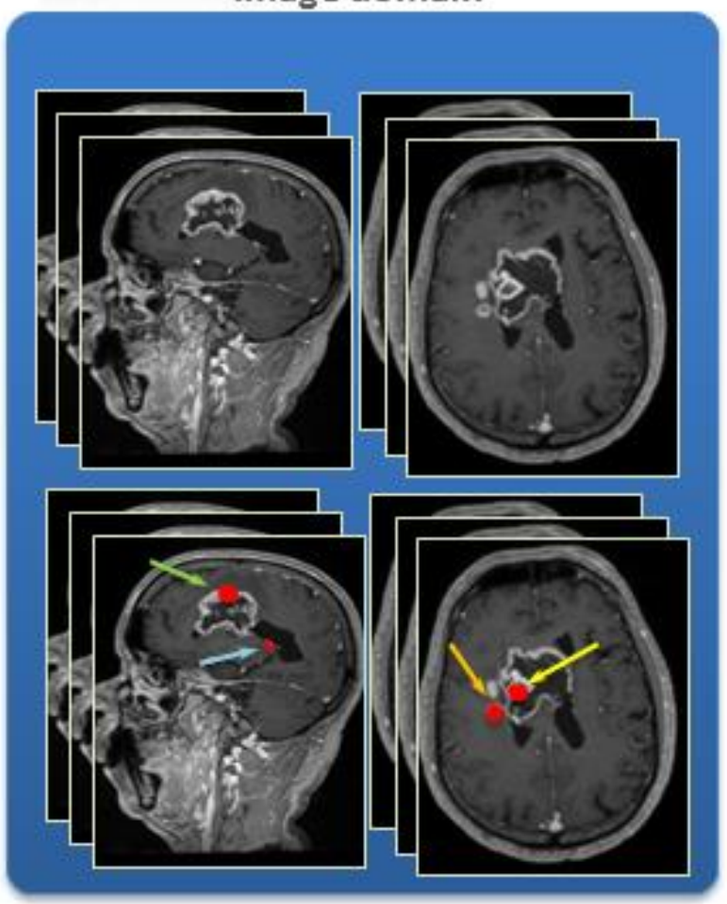
Assemble = Therapy

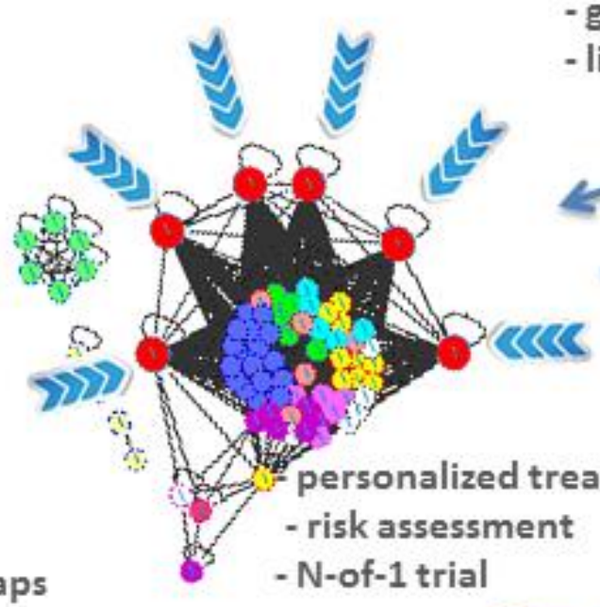
on the fly therapy optimization



multimodal imaging
- 3D Physiological maps
- imaging biomarkers

image domain





Digital Biodata
- patient EHR, drugs
- genetic profile
- lifestyle, exposome

fingerprint

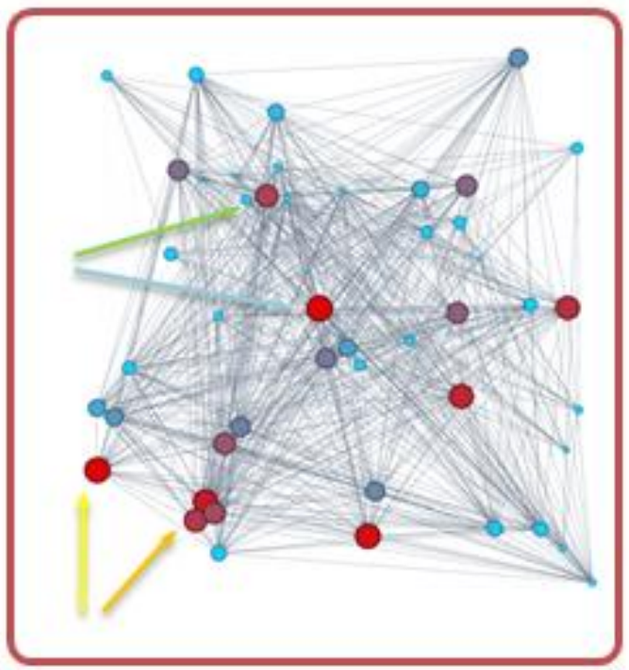
- personalized treatment
- risk assessment
- N-of-1 trial

re-phenotyping

stratifications

multicentric clinical trials

Network domain



DIGITAL BIOMARKERS

Transitions between healthy and diseased states

Expanded integrative profile

Module markers

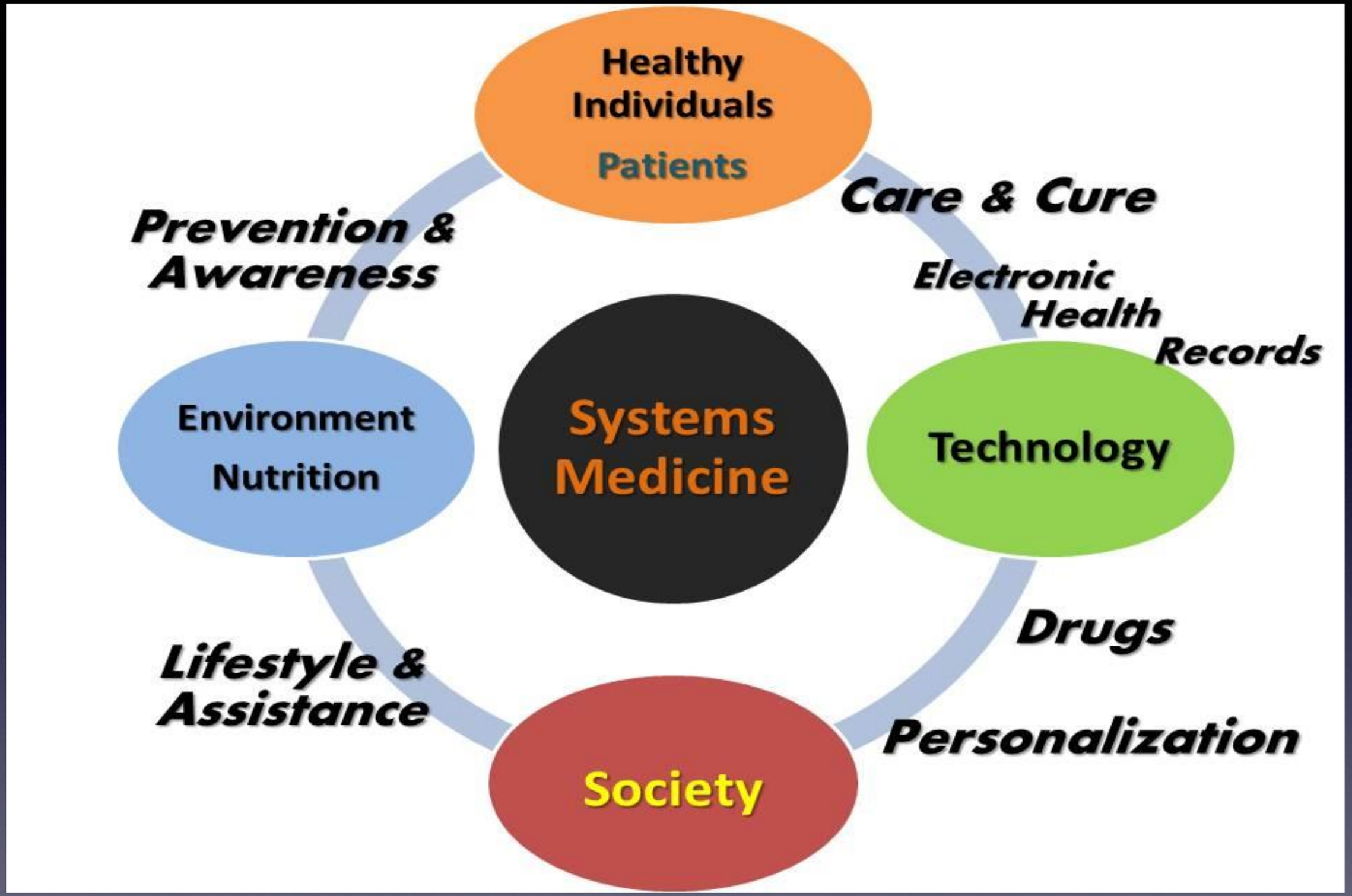
Precision Medicine Leaders Summit, San Diego, August 2017

DIGITAL HEALTH

Integrates technology in behavioral, physiological, lifestyle, social, environmental, clinical processes.

Talks about people’ needs

and business models



Health Ecosystem

Capobianco and Trivella, PAN EUROPEAN NETWORKS Science & Technology Dec, 2014.

Where were we?

human-machine communication according to natural language rules

Prototype of Watson in Yorktown Heights, NY.

27 MAGGIO 2015

Chirurgia di precisione, il robot Da Vinci "esordisce" in sala operatoria

Presentato alla clinica Santa Caterina da Siena di Torino, ha quattro braccia guidate dal medico tramite una console che consente interventi millimetrici

- **Vast amounts of spatial and temporal data**

➔ **continuous monitoring**

Active data

from clinical, behavioral, epidemiological measurements

Passive Data

data from sensors during sleep or stress

- **Influence by collective dynamics available**

Insomnia:

timing and hours of digital footprints, tweets

Health tech is a top area for venture investment in US

- Google used deep learning to detect diabetic retinopathy.
- 54 ophthalmologists rated 128,175 retinal images
- Algorithms trained on these data to detect the disease

Testing over 2 new sets of retinal images, 8 ophthalmologists

The algorithms were on average doing better than 7/8 of experts

Change in healthcare:

IoT, Sensors, Smart Devices, Big Data….

Gartner estimates about 5 bl connected devices,
that will become 25bl by 2020

➔ induce pervasive sensing

helping or hurting?

By 2020….how many EHR can we produce?

~ 1 terabyte (CT scan is ~ 1 giga)

and from sequencing + related analyses?

~ 6 terabytes

and accounting for life style factors?

> 1000 terabytes…

Digital Medicine

Technologies/products undergoing clinical validation

impacts:

- **diagnosis**
  - **treatment**
    - **prevention**

PatientsLikeMe.com
23AndMe.com

Why we care?

A wellness study of 108 individuals using personal, dense, dynamic data clouds.

Price et al (Hood's team), Nature Biotechnology 35, 747–756 (2017).

Through Hood ISB has pioneered the concept of

P4 Medicine

(***personalized, predictive, preventive, participatory***).

Hundred Person Wellness Project (HPWP)

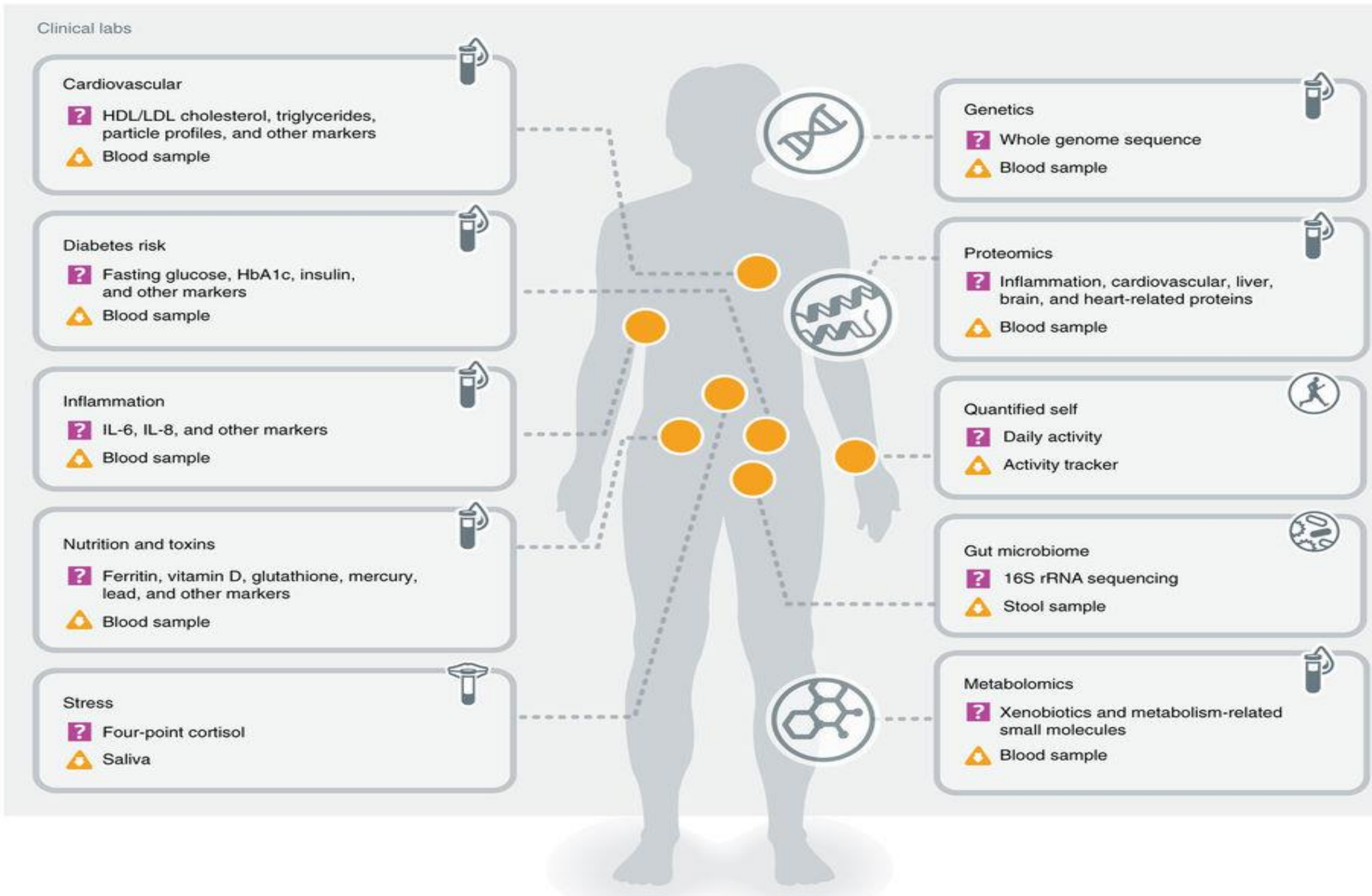
Multiple longitudinal data collected

108 individuals monitored for 9 months, 3 times

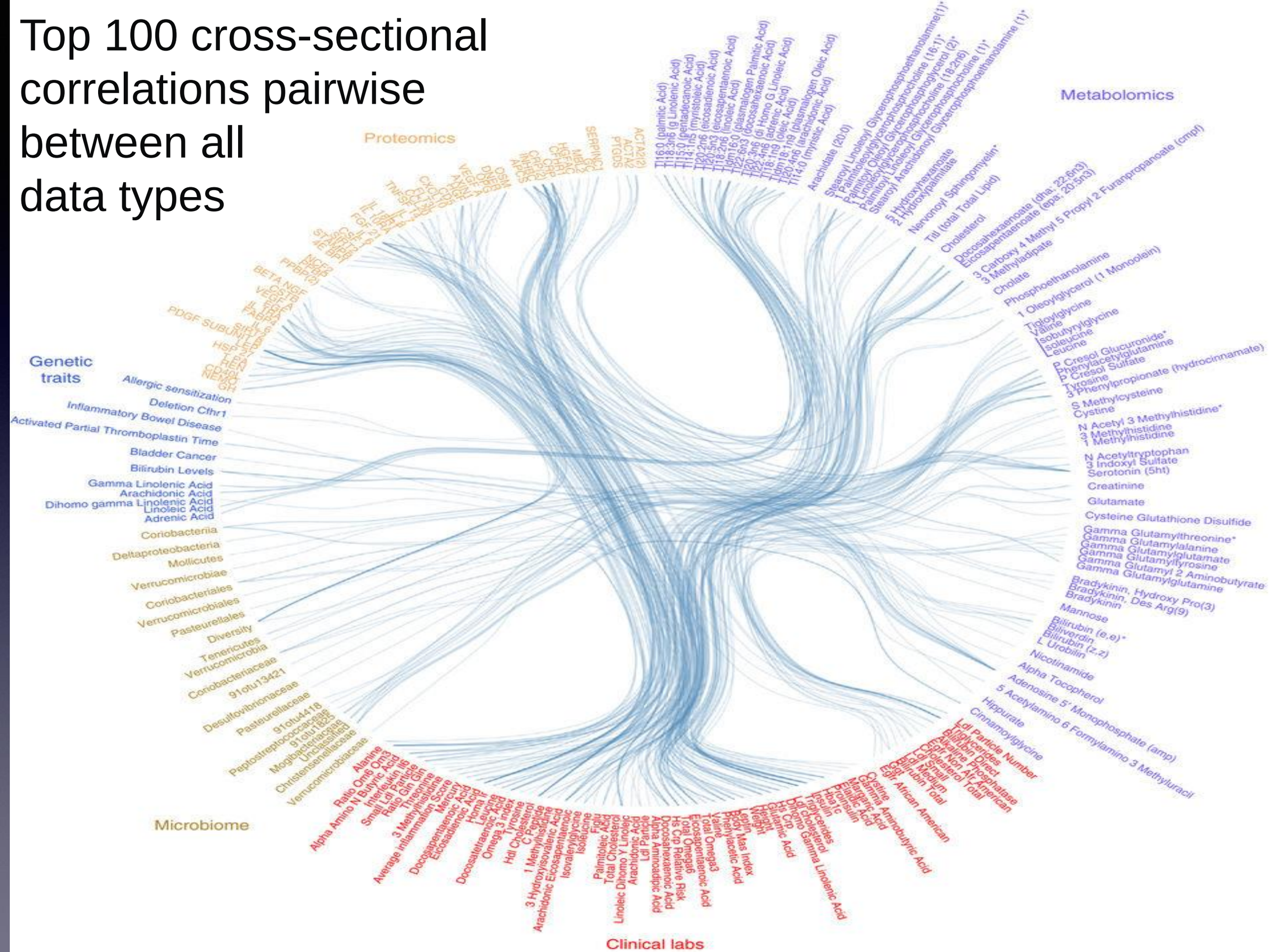
a


Month 1
Month 2
Month 3
Month 4
Month 5
Month 6
Month 7
Month 8
Month 9
Intro
Round 1
Coaching sessions
Round 2
Coaching sessions
Round 3
Coaching sessions

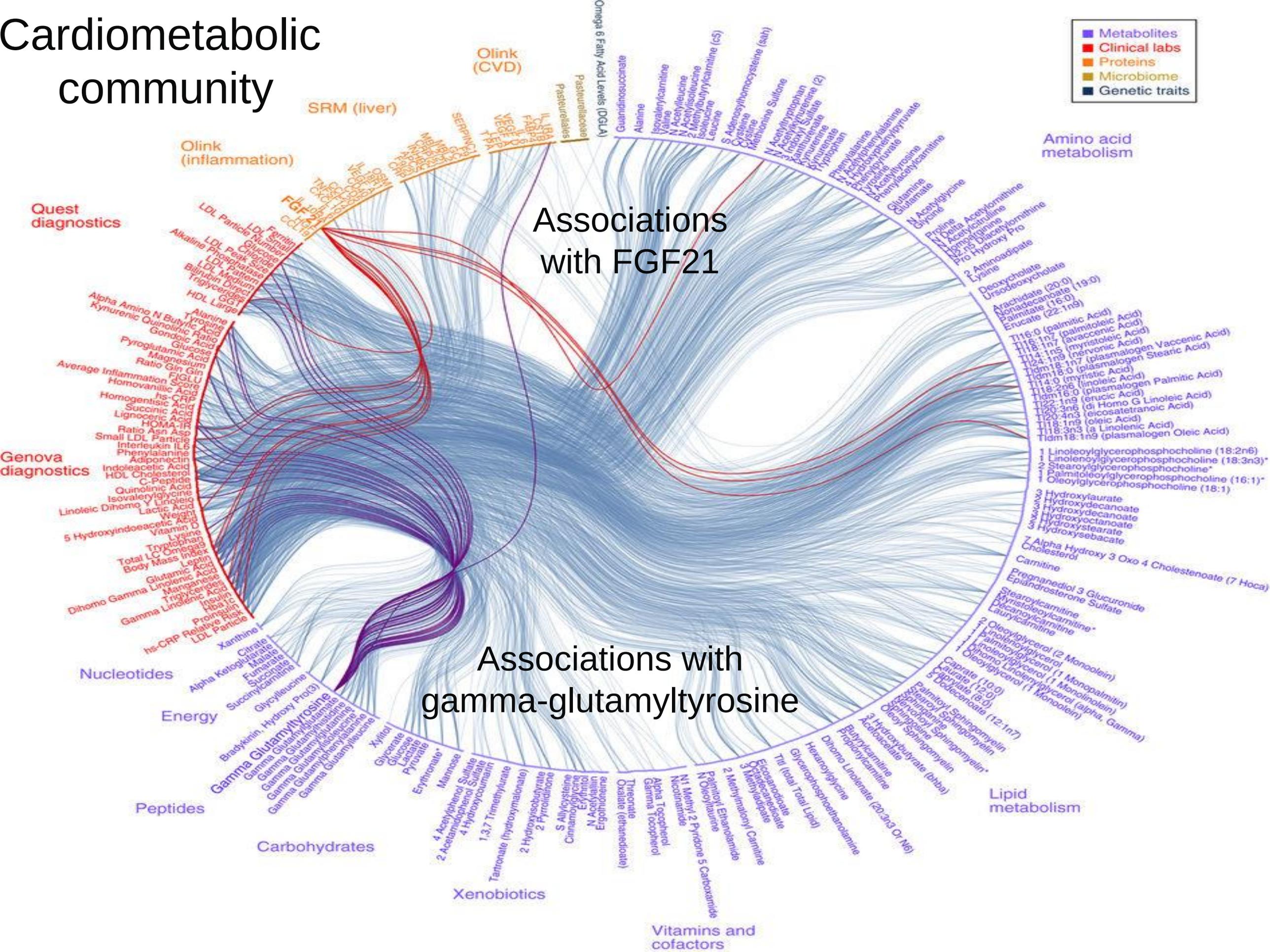
b


Clinical labs
Cardiovascular
HDL/LDL cholesterol, triglycerides, particle profiles, and other markers
Blood sample
Diabetes risk
Fasting glucose, HbA1c, insulin, and other markers
Blood sample
Inflammation
IL-6, IL-8, and other markers
Blood sample
Nutrition and toxins
Ferritin, vitamin D, glutathione, mercury, lead, and other markers
Blood sample
Stress
Four-point cortisol
Saliva
Genetics
Whole genome sequence
Blood sample
Proteomics
Inflammation, cardiovascular, liver, brain, and heart-related proteins
Blood sample
Quantified self
Daily activity
Activity tracker
Gut microbiome
16S rRNA sequencing
Stool sample
Metabolomics
Xenobiotics and metabolism-related small molecules
Blood sample

Top 100 cross-sectional correlations pairwise between all data types



Cardiometabolic community



Associations with FGF21

Associations with gamma-glutamyltyrosine

Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information

Plos Biology, 2017 Li et al, (M. Snyder's team)

43 individuals and 250,000 measurements,

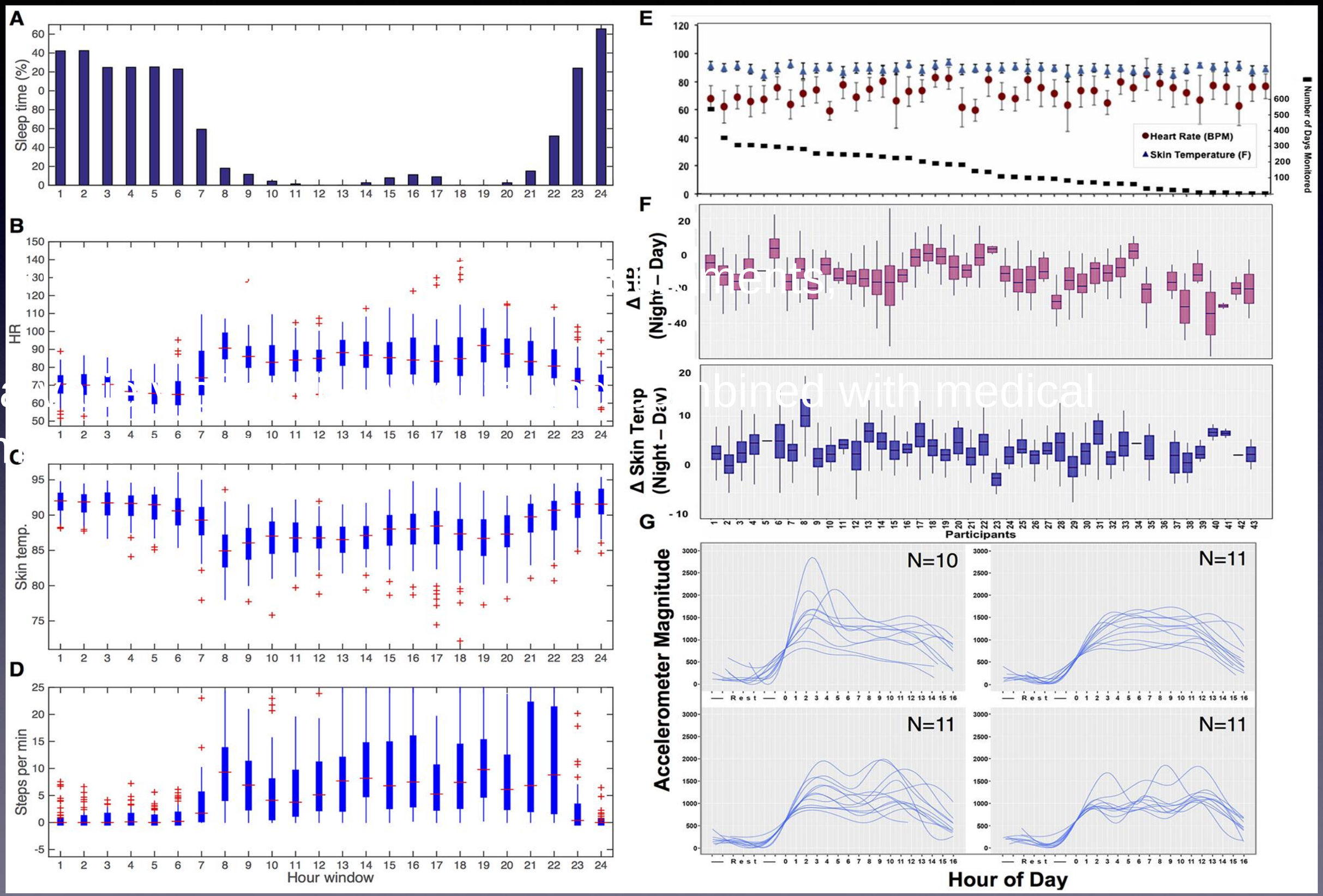
Considering a variety of environments
combined with medical tools

Circadian and diurnal patterns in physiological parameters

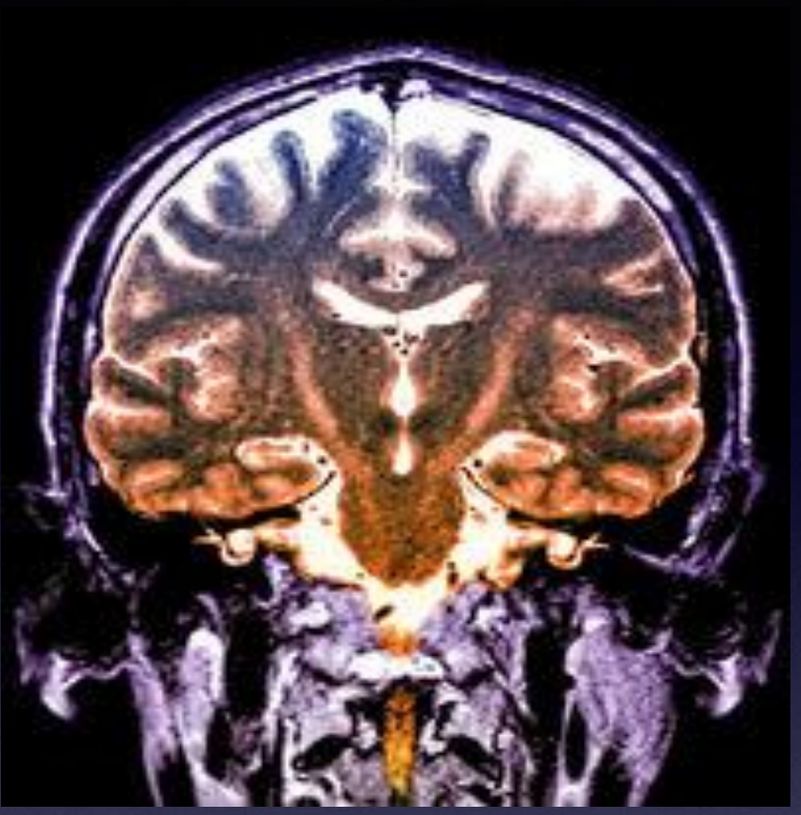
43

in a

n



More measurements: scans



Calculating Disease, Savage, Nature 2017

MRI scans help predict how fast ALS will progress

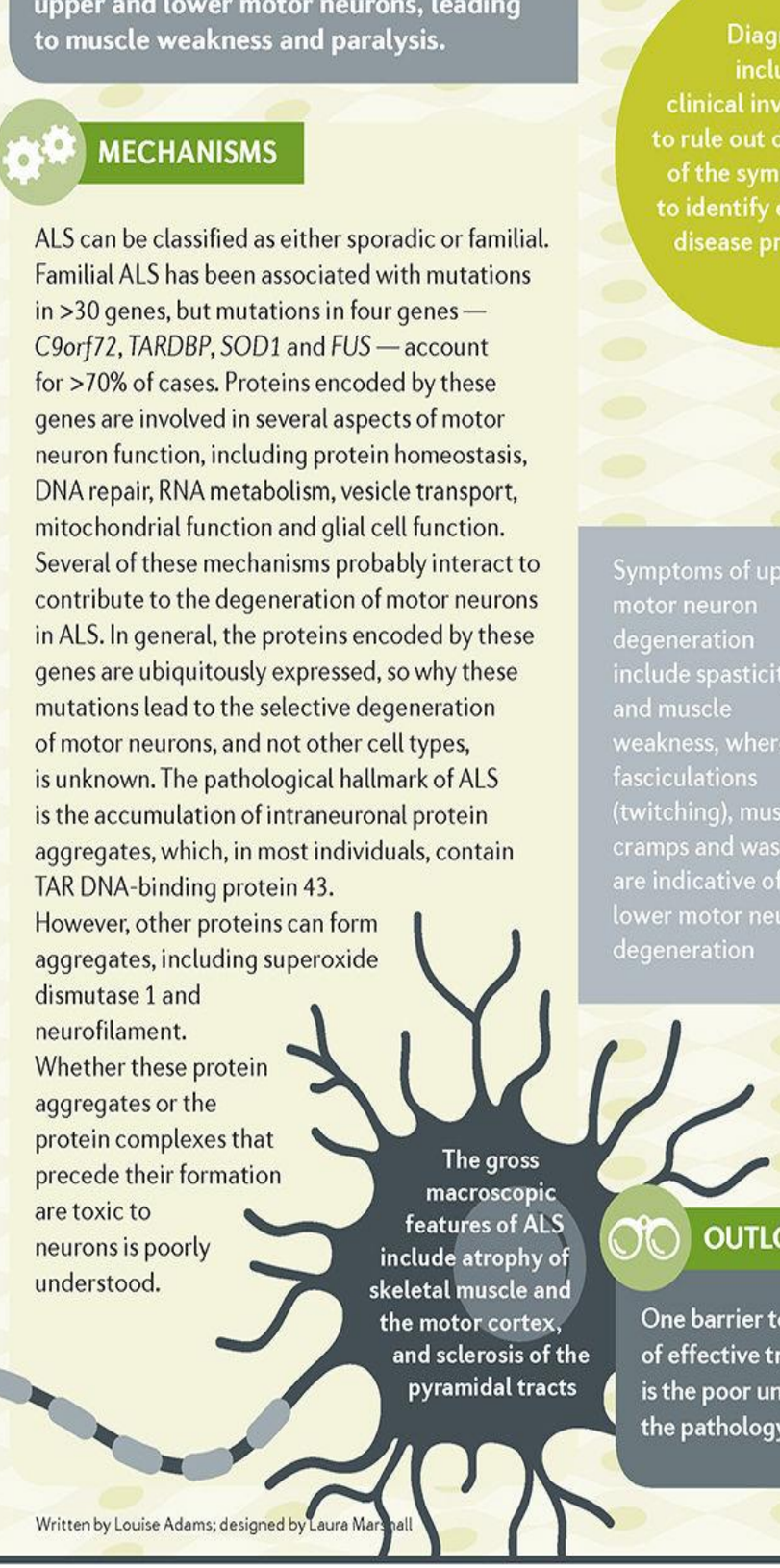
1000 US patients
6 bl data points per person

For the Primer, visit doi:10.1038/nrdp.2017.71

Amyotrophic lateral sclerosis (ALS; also known as motor neuron disease) is a rare, neurodegenerative disease that is characterized by the degeneration of upper and lower motor neurons, leading to muscle weakness and paralysis.

MECHANISMS

ALS can be classified as either sporadic or familial. Familial ALS has been associated with mutations in >30 genes, but mutations in four genes — *C9orf72*, *TARDBP*, *SOD1* and *FUS* — account for >70% of cases. Proteins encoded by these genes are involved in several aspects of motor neuron function, including protein homeostasis, DNA repair, RNA metabolism, vesicle transport, mitochondrial function and glial cell function. Several of these mechanisms probably interact to contribute to the degeneration of motor neurons in ALS. In general, the proteins encoded by these genes are ubiquitously expressed, so why these mutations lead to the selective degeneration of motor neurons, and not other cell types, is unknown. The pathological hallmark of ALS is the accumulation of intraneuronal protein aggregates, which, in most individuals, contain TAR DNA-binding protein 43. However, other proteins can form aggregates, including superoxide dismutase 1 and neurofilament. Whether these protein aggregates or the protein complexes that precede their formation are toxic to neurons is poorly understood.

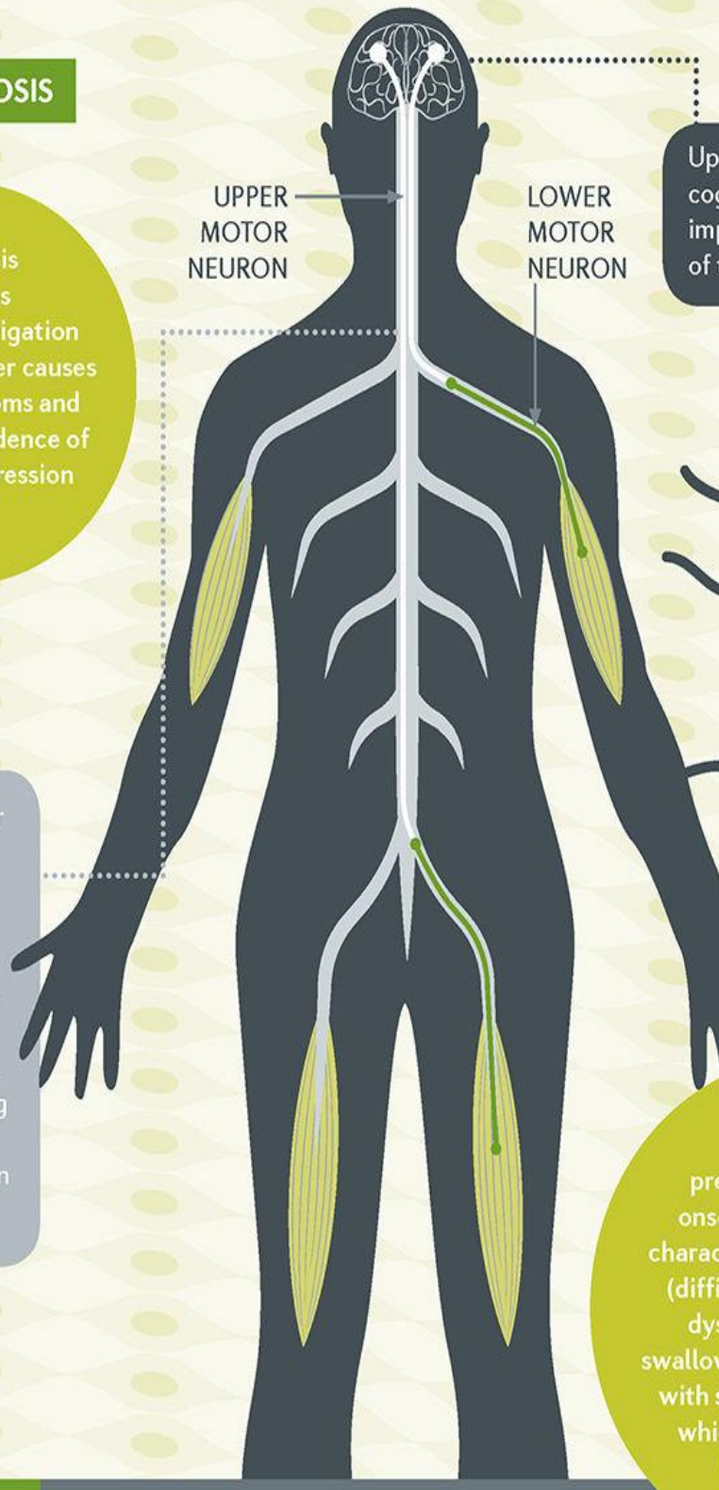


The gross macroscopic features of ALS include atrophy of skeletal muscle and the motor cortex, and sclerosis of the pyramidal tracts

DIAGNOSIS

Diagnosis includes clinical investigation to rule out other causes of the symptoms and to identify evidence of disease progression

Symptoms of upper motor neuron degeneration include spasticity and muscle weakness, whereas fasciculations (twitching), muscle cramps and wasting are indicative of lower motor neuron degeneration



Up to 50% of patients develop cognitive and/or behavioural impairments during the course of the disease

One-third of patients present with bulbar-onset disease, which is characterized by dysarthria (difficulty speaking) and dysphagia (difficulty swallowing); the rest present with spinal-onset disease, which is characterized by limb muscle weakness

EPIDEMIOLOGY

In Europe, incidence is in the range of 2–3 cases per 100,000 individuals; incidence is lower in east Asia (0.8 cases per 100,000 individuals) and south Asia (0.7 cases per 100,000 individuals). The phenotype of ALS varies between populations. For example, the age at onset of symptoms and diagnosis is higher in Europe than in Asia or South America. In addition, the proportion of individuals with bulbar-onset disease is much lower in Asia than in Europe and patient survival is lower in Europe (24 months from the onset of disease) than in central Asia (48 months).

Some forms of ALS share a genetic overlap with neuropsychiatric conditions, such as schizophrenia

MANAGEMENT

Few randomized controlled trials assessing symptomatic therapies in patients with ALS have been conducted; accordingly, many therapies are based on the management of other diseases. For example, anticholinergic drugs can be used to treat sialorrhoea (hypersalivation), baclofen can be used to treat spasticity and muscle relaxants can be used to treat cramps. The management of dysphagia includes dietary changes, swallowing manoeuvres and exercise and, if severe, use of a gastrostomy tube. Other available symptomatic treatments include speech therapy for dysarthria and noninvasive ventilation for respiratory failure.

Two disease-modifying therapies — riluzole and edaravone — have been approved by the US FDA

OUTLOOK

One barrier to the development of effective treatments for ALS is the poor understanding of how the pathology of disease affects the overall integrity and function of brain networks. Improvements in model systems to study ALS and better ways to study the disease in humans will enhance our understanding of ALS and enable us to target therapies to specific aspects of the pathophysiology.

Written by Louise Adams; designed by Laura Marshall

Article number: 17072; doi:10.1038/nrdp.2017.72; published online 5 Oct 2017

What we gain?

- Improved understanding of disease progression
- Earlier Diagnosis
- Personalization of treatment
- Novel Patient stratifications

- Comparison of records from patients worldwide

Translational Benefits

- **Decision support systems for clinicians**
  (early warnings, disease trajectories,…)

- **Clinical trials**
  (better design, assessment of feasibility,…)

- **Drug safety management**
  (adverse reactions, off-target, occurrence of comorbidities,…)

Computational environments hosting huge projects, UK with 100,000 Genomes Project (same in US and China)

but

Medical relevance of much genetic variation revealed by next generation sequencing

is currently unknown

Large-Scale Initiatives

The Global Alliance for Genomics and Health (2013)

➔ Data underlying genomic medicine must be federated, distributed across many db, data centers

and

➔ virtually connected/interfaced for global access.

Knowledge Bank approach
to allow decision-support predictions

Algorithmic Personalized Care
to predict interventions with long run impacts

***N of 1* medicine**

specific to patient's needs
(customized medical options)

Patient-empowered health

How valuable the real-world data?

measuring transition from wellness to disease

Yet, applications are still used sparingly, especially in care delivery.

Why?

LIMITATIONS

EHR not aligned with care or treatment goals!

➔ *don’t capture outcomes & reflect different scopes.*

Need to include/redefine endpoints:

- to train algorithms
- to explain variability

Problems with data-driven medicine
(*NEJM survey*)

- Functionality within digital systems
- Sophisticated analytics needed
- Difficulty in collecting data
- Difficulty in interpreting results

Lack of interoperability
(> 70% respondents)

Need scientific thinking
…… and more

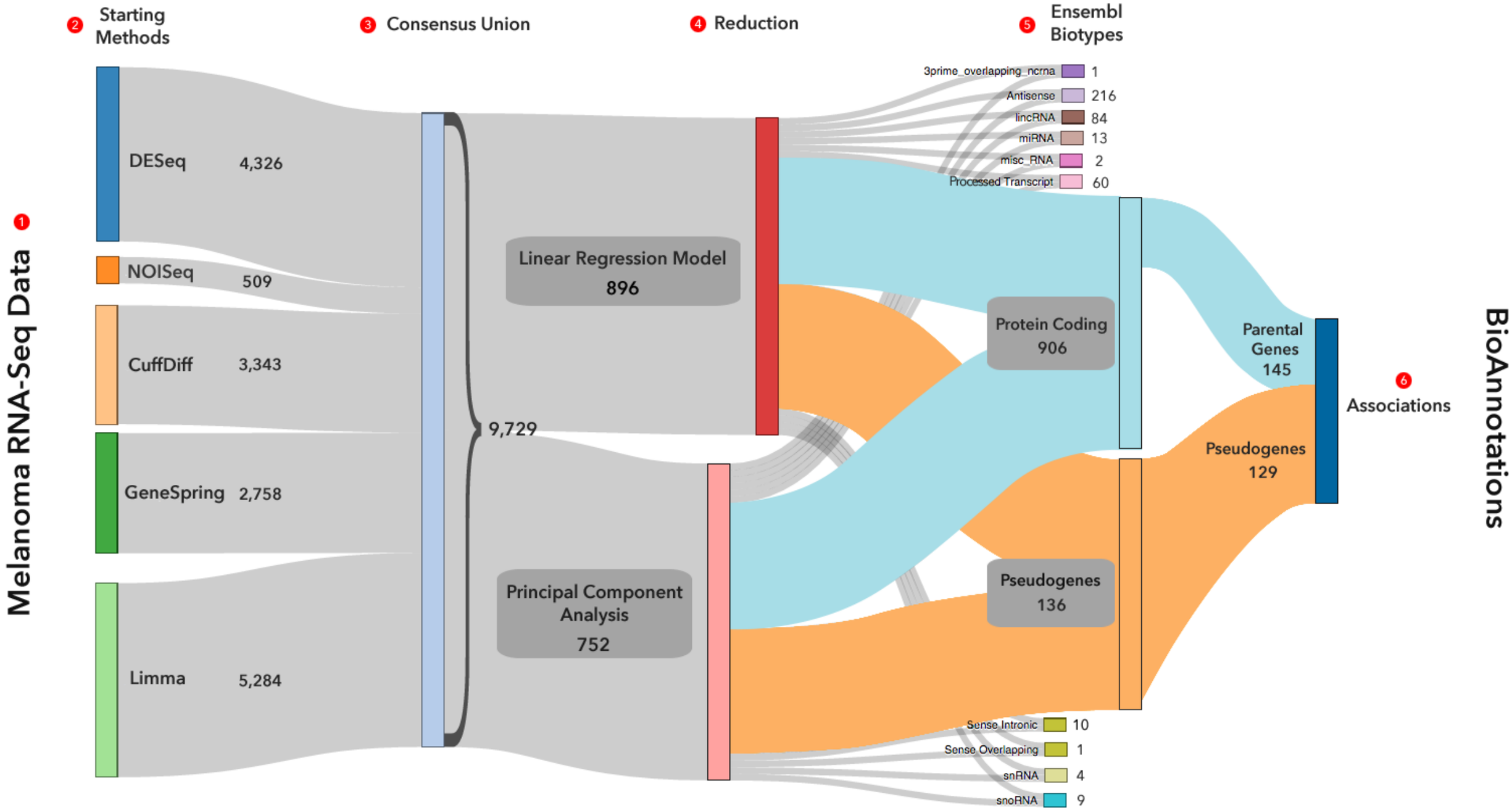
New research paradigm:

end of theory?

‘**mine the data**’

Dark Matter or Dark Genome (98%)

- ✓ **Dark matter** in genomics, especially non-coding RNA (ENCODE).
- ✓ Uncharacterized functional associations (genes/proteins) emerge massively, but experimental validations not routine.
- ✓ Inaccurate, speculative associations only putatively hypothesized by NGS findings



Notes

The flow refers to differentially expressed biotypes measured by different methods.

1. 103 Melanoma samples from The Cancer Genome Atlas (TCGA), and 2 Melanocyte control samples from the Gene Expression Omnibus (GEO)
2. Algorithms used for detecting differential expression
3. Consensus of differentially expressed bioentities
4. Reduction in size by linear regression model (LRM) and principal component analysis (PCA)
5. Final decomposition of biotypes
6. Particular focus goes to the association of pseudogenes and parental protein coding genes

Ensemble Modeling Approach Targeting Heterogeneous RNA-Seq data: Application to Melanoma Pseudogenes.

Capobianco E, Valdes C, Sarti S, Jiang Z, Poliseno L, Tsinoremas NF. ***Scientific Reports*****, 2017**

…however

- Opportunity to **formulate new paradigms** that govern biological systems.
- **Better diagnosis** from a richer characterization of tumorigenesis.
- **New drugs** designed with increased RNA specificity will result in targeted therapeutic outcomes

Toward Omnigenomics

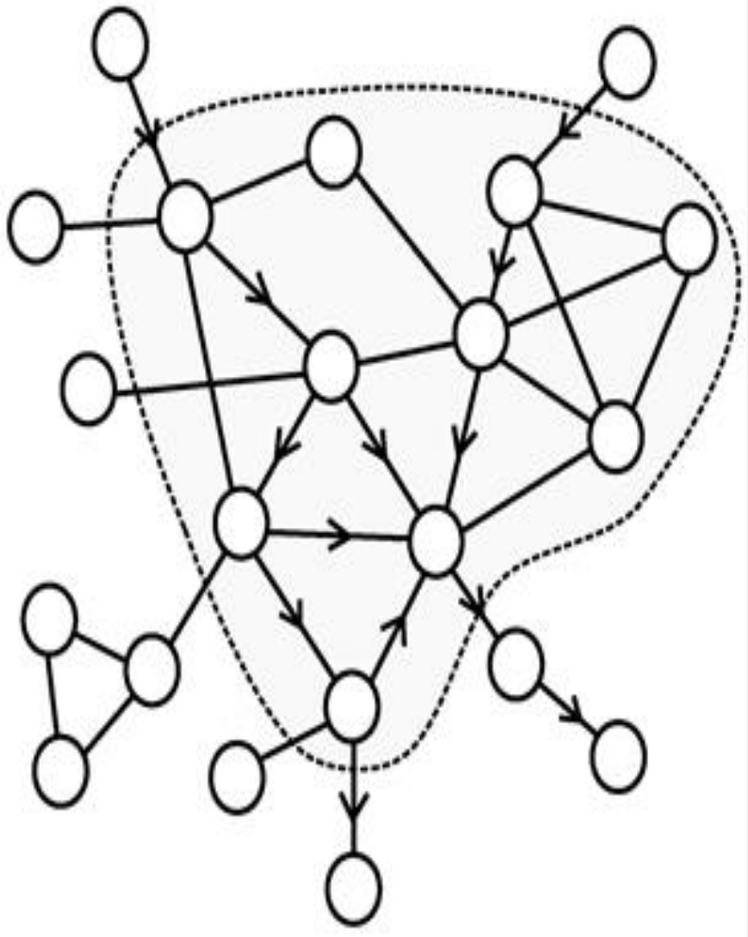
➔ Gene regulatory networks represent omnigenic models

Network medicine….
Barabasi et al,
Nat Rev Gen 2011

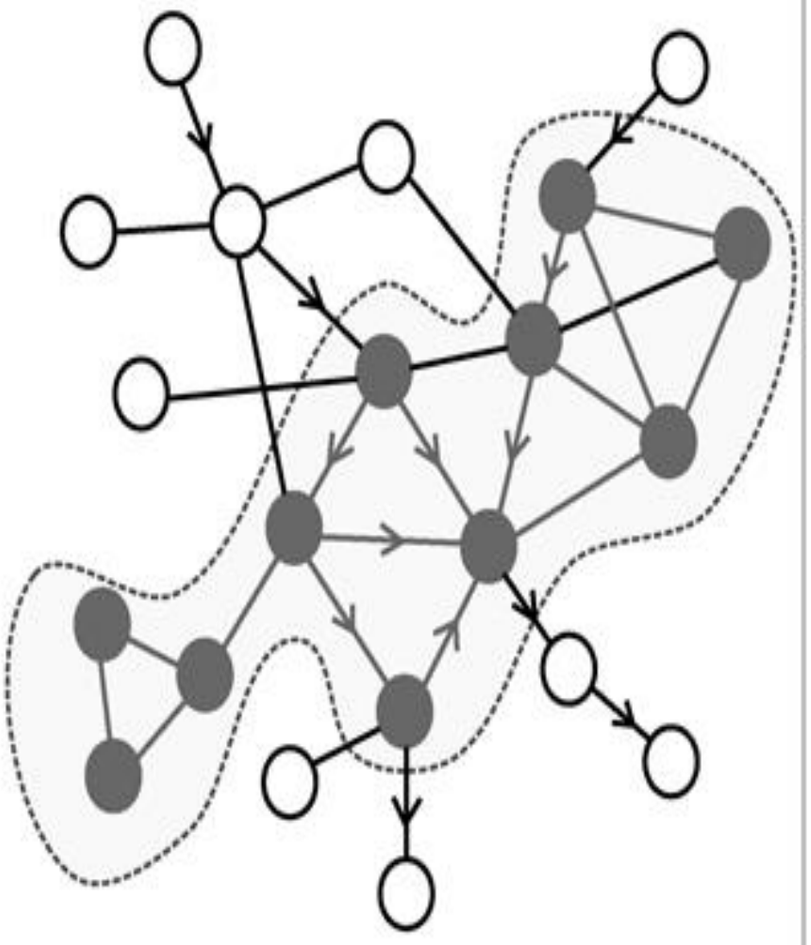
Networks

Identification of **influence** through modularity and controllability of markers/drivers/targets

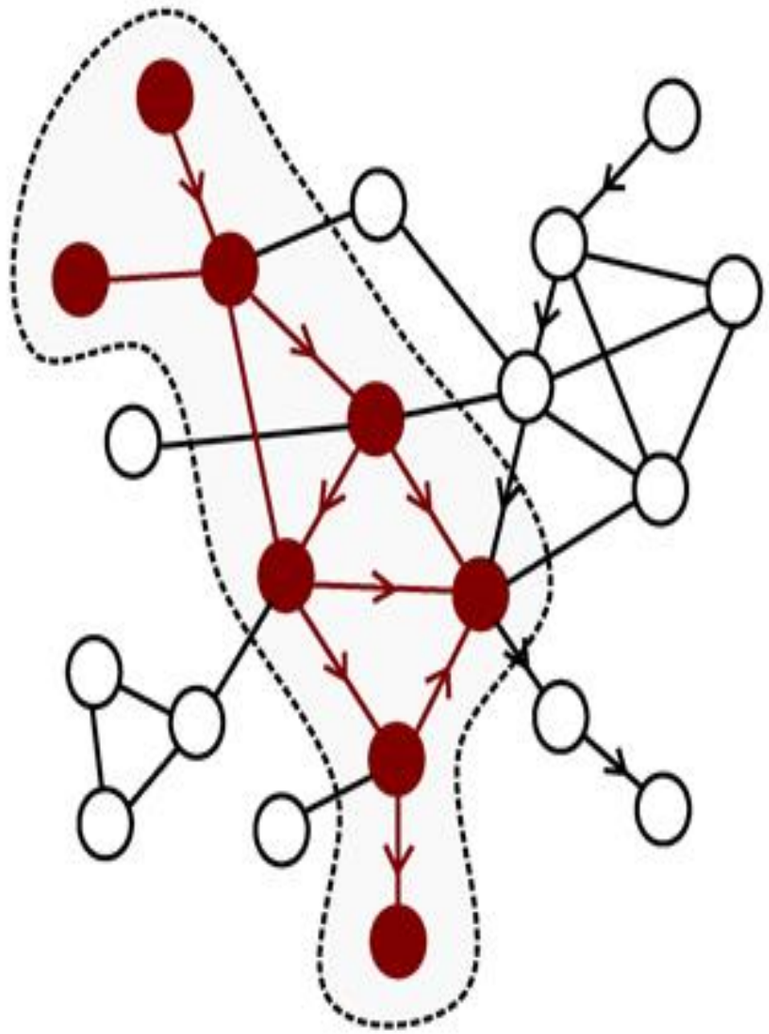
to go fast, go alone
to go far, go together



Topological module



Functional module



Disease module

- Topologically close genes (or products)
- Functionally similar genes (or products)
- Disease genes (or products)
- Undirectional interactions
- Directional interactions

Network medicine….
Barabasi et al,
Nat Rev Gen 2011

influence through modularity

Multitype Network-Guided Target Controllability in Phenotypically Characterized Osteosarcoma: Role of Tumor Microenvironment

A. Sharma, C. Cinti, E. Capobianco

Frontiers Immunology, 2017

PPI-miRNA network: Invasive phenotype

A

influence through target controllability

Color mapping: Gene

Edge: Regulation

Max(-) log(FC) Max(+)

-ve Regulation +ve

high low

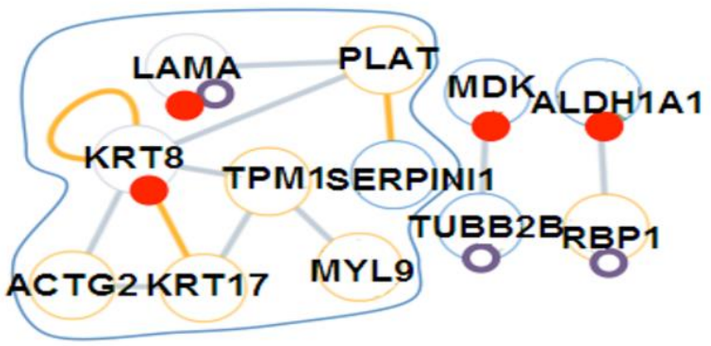
Node size

low Degree high

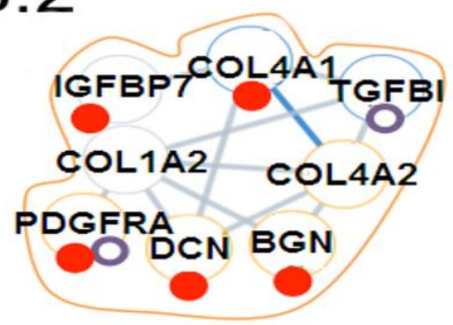
Critical node

Participating in protein complex

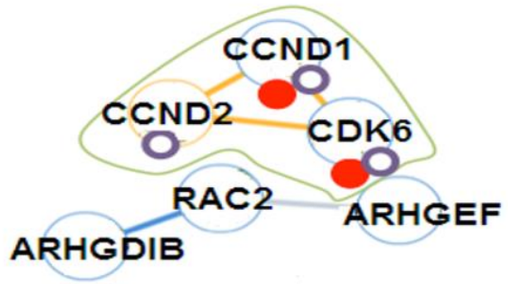
B.1



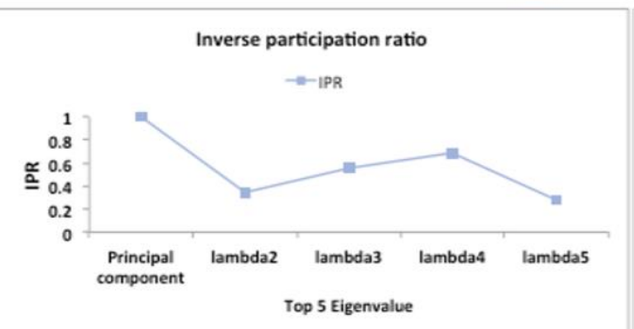
B.2

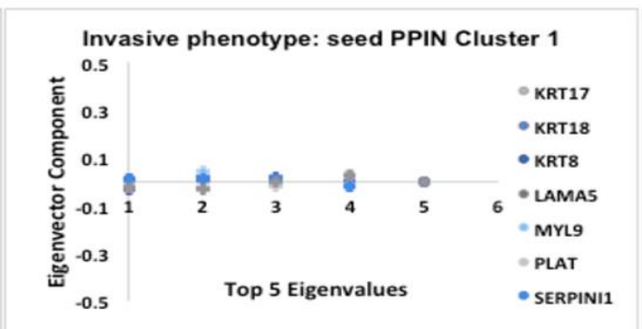


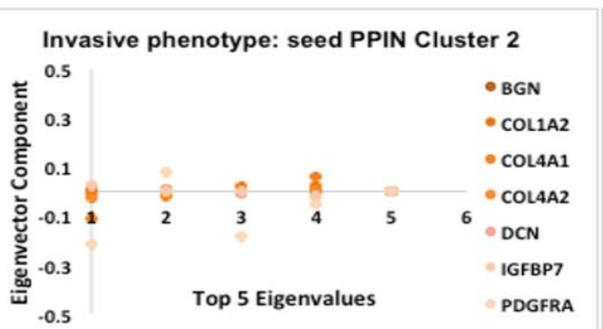
B.3

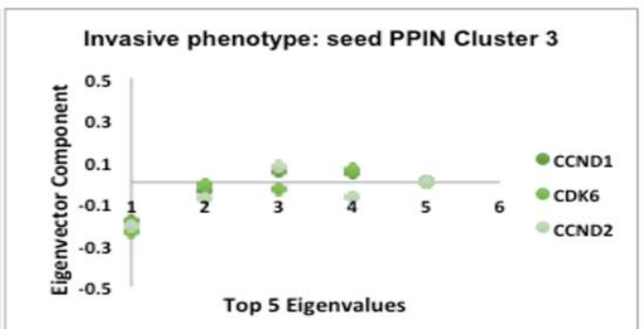


C





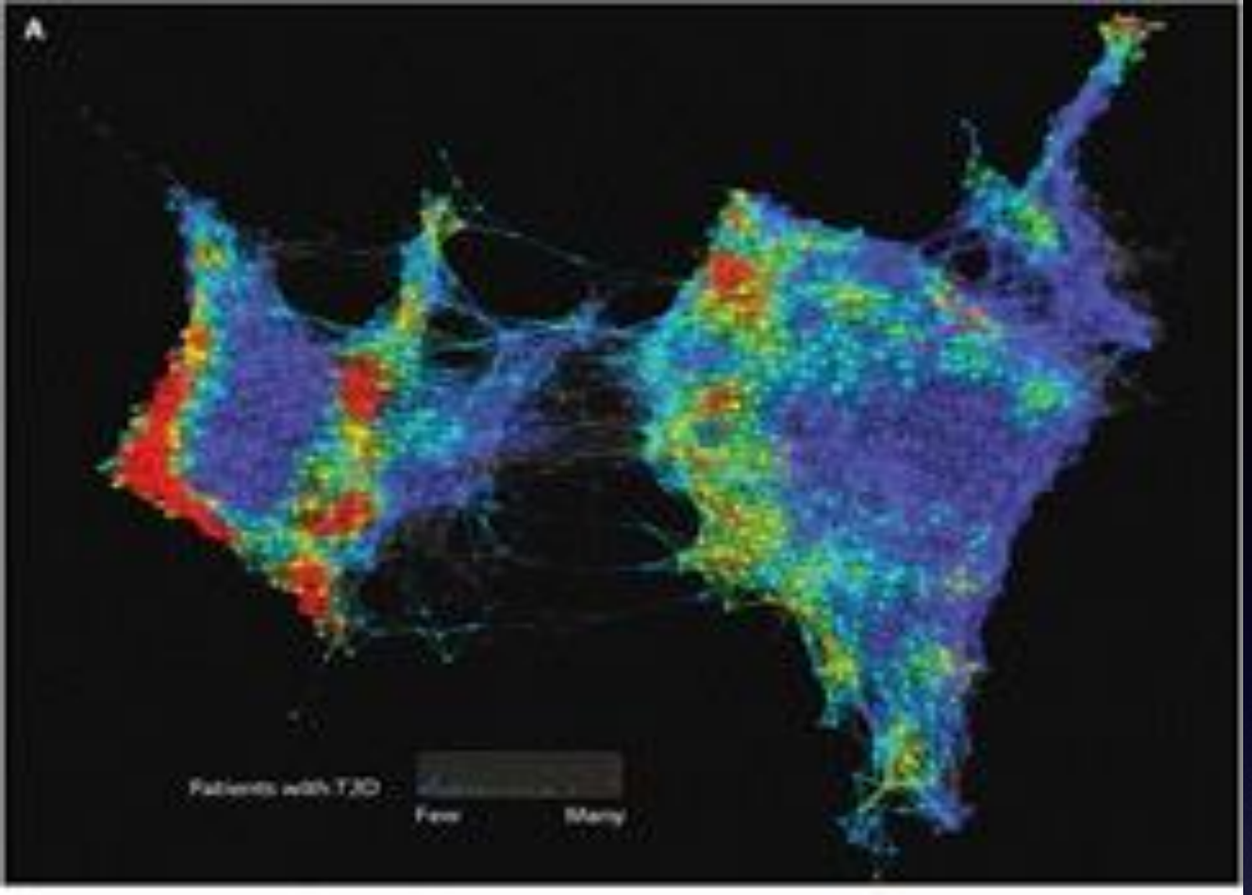




Identification of type 2 diabetes subgroups throuth topological analysis of patient similarity.
Li et al, Sci Transl Med, 2015.

Complexity of diabetic population (11,210 individual EHR) and 3 subgroups found from topological patient networks.

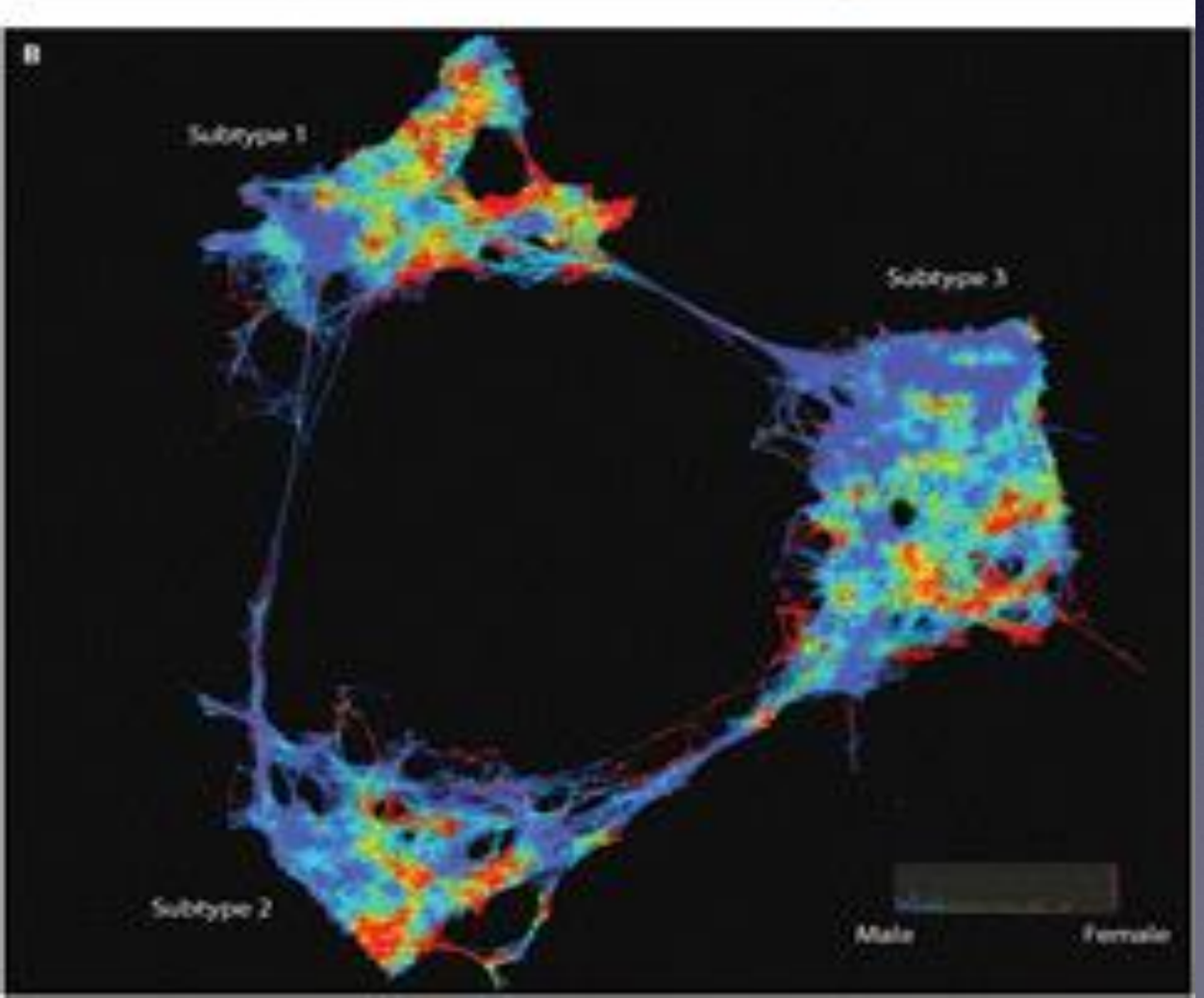
1. **Diabetic complications (nephropathy and retinopathy)**
2. **Co-occurrence of cancer and cardiovascular diseases**
3. **Association with neurological diseases, allergies, infections.**



A) Patient-patient network for topology patterns

Nodes = groups of patients with significant similarity of clinical features. Edges indicate shared patients.

Red color = high enrichment for T2D
Blue color = low enrichment for T2D



B) Patient-patient network for topology patterns on 2551 T2D patients.

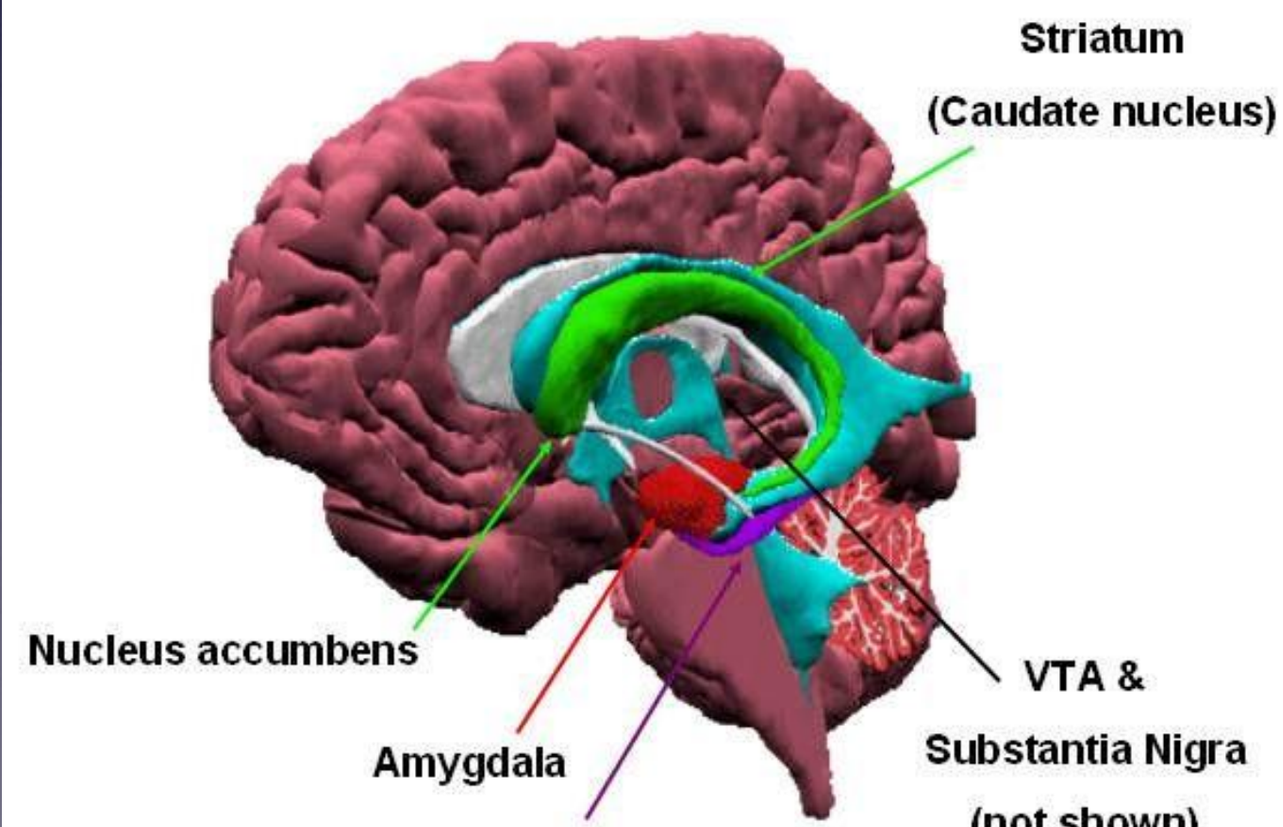
Red color = enrichment for females

Blue color = enrichment for males.

Aequam memento servare mentem.
(Ricordati di mantenere la mente serena, Orazio)

Cocaine Addiction Project

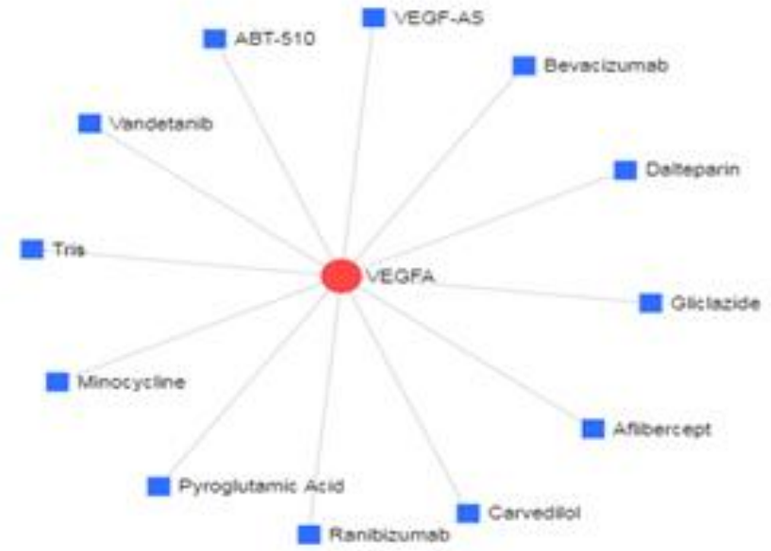
Un of Miami, Neurology, CCS,
Brain Endowment Bank,
Broad Inst., Mc Gill Univ



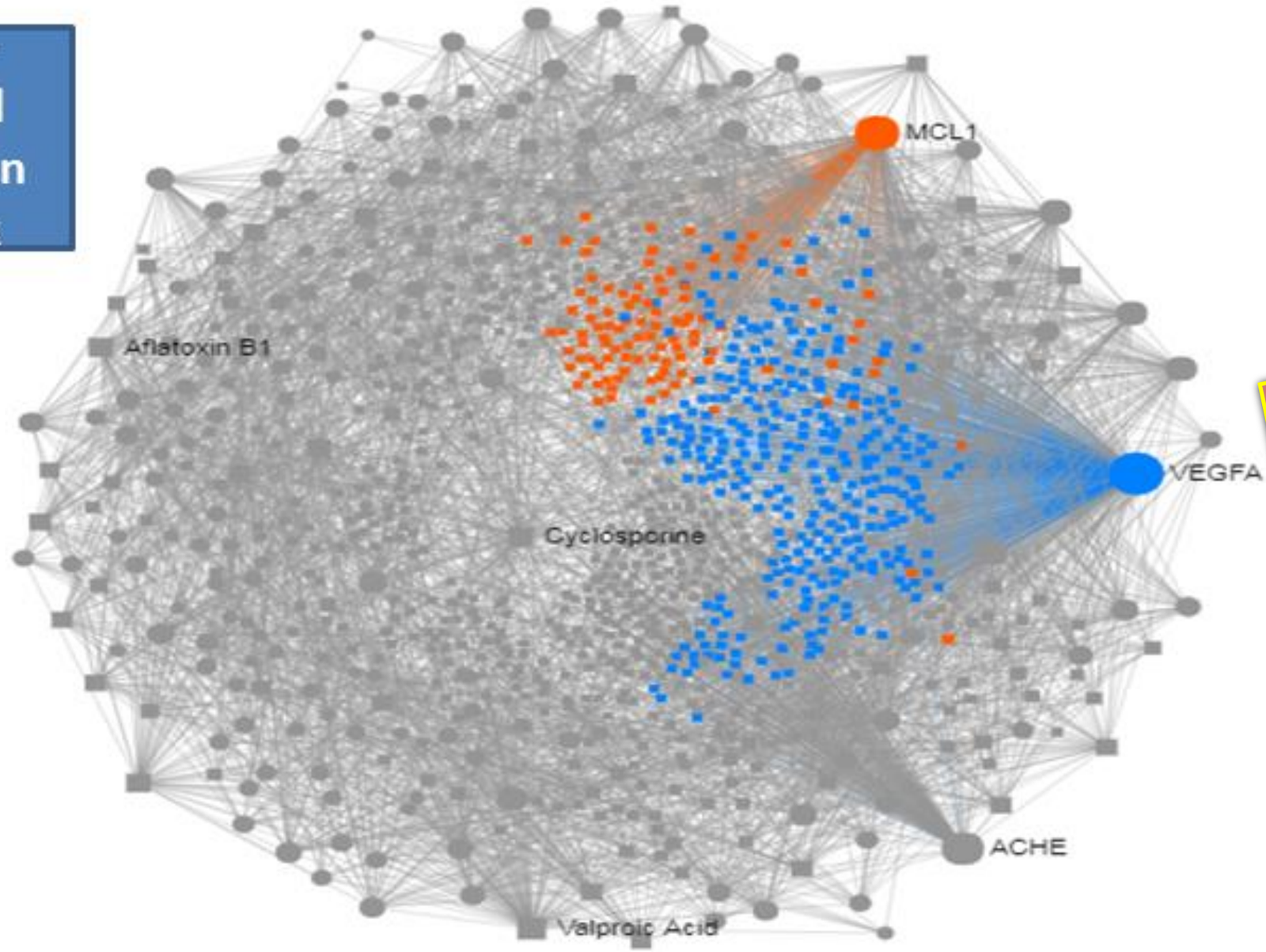
RNA-Seq and Methylation experiments

Brain Tissues:
Nucleus Accumbens region
Caudate region

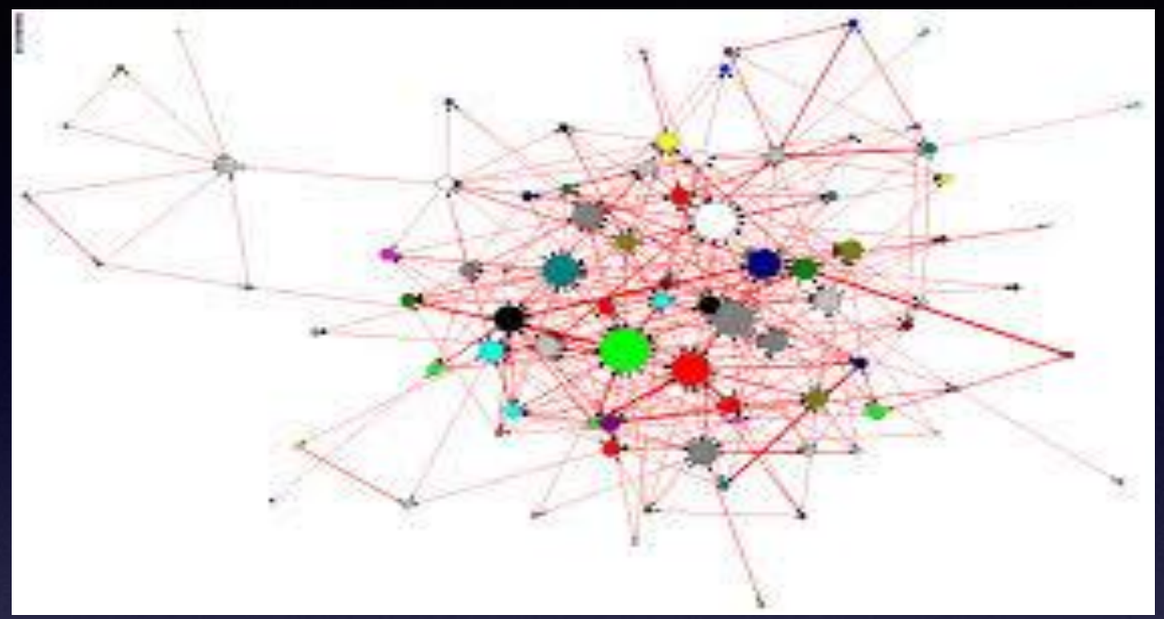
drug – protein interaction network



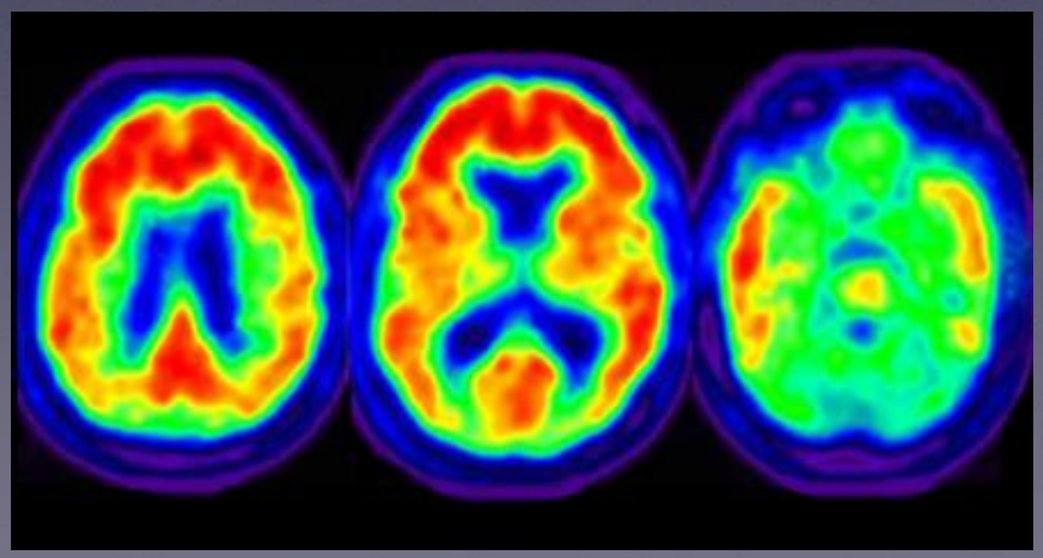
protein-chemical interaction network

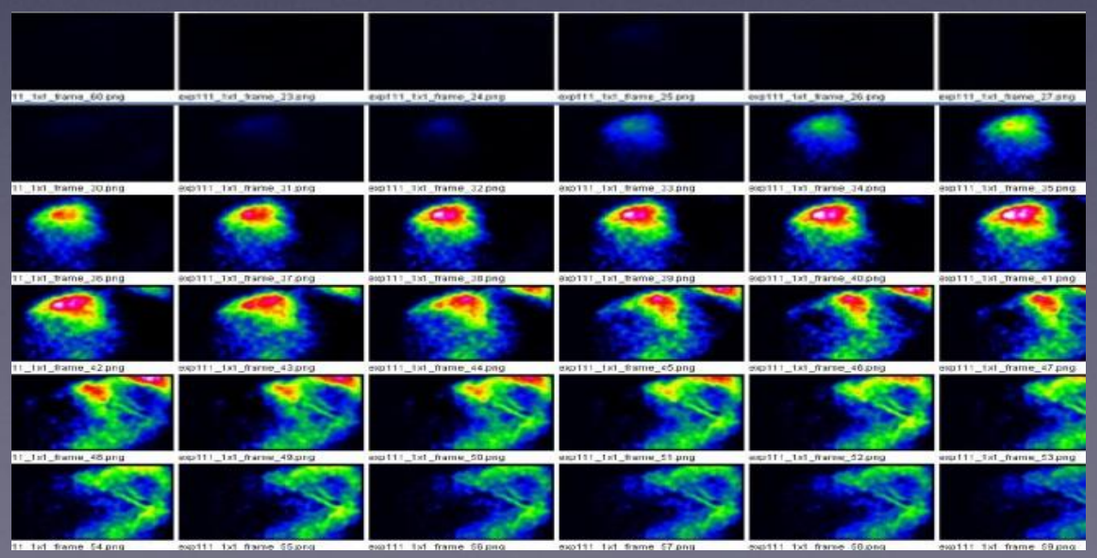


COMPLEX NETWORKS



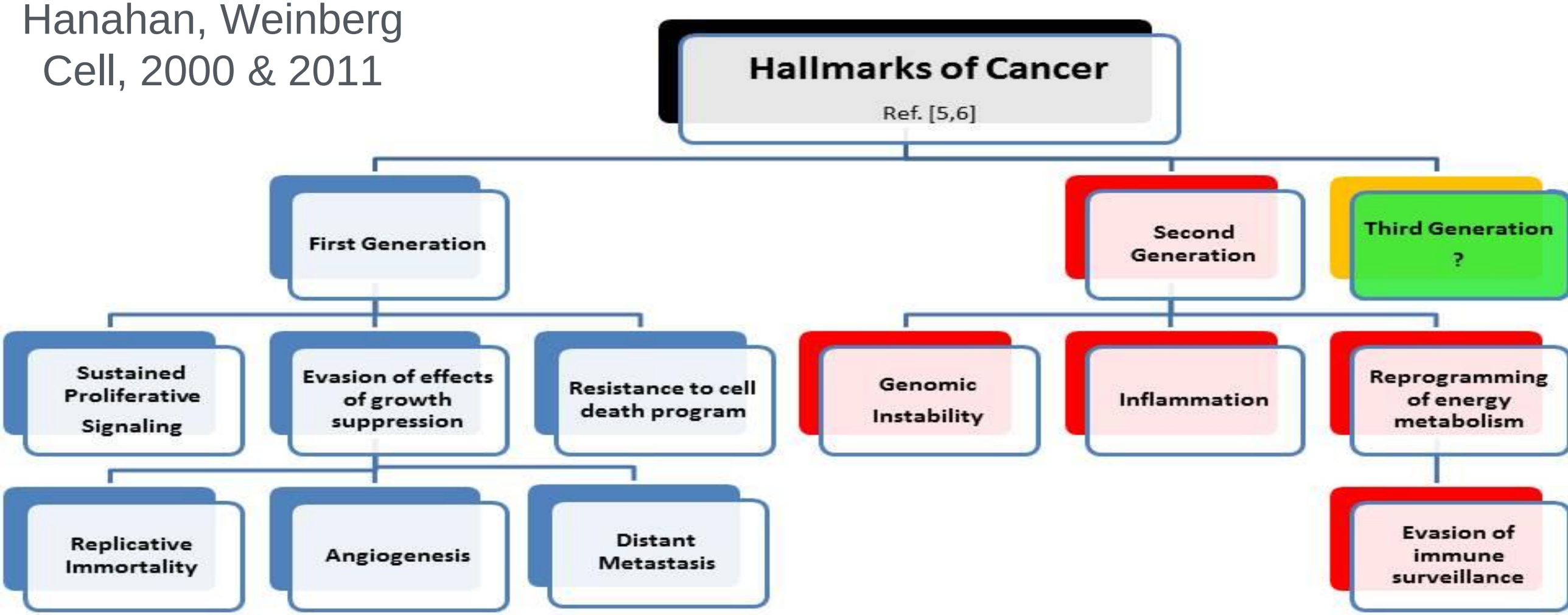
Molecular Imaging





M.Dominietto, N. Tsinoremas and E. Capobianco (2015) Integrative Analysis of Cancer Imaging Readouts by Networks. *Molecular Oncology*, 9(1), 1-16

Hanahan, Weinberg
Cell, 2000 & 2011



Development of Predictive Biomarkers for Immunotherapy (Uveal Melanoma)

Benign lesion

Network N_B

Network configuration

Malignant tumor

Network N_M

Network configuration

Differential Analysis

Tissue differentiation

Network configuration

Tumor Fingerprint

Hallmarks scoring

H 1

H 2

H 10

S1

S2

S10

before

Treatment

after

Network adaptation

Network configuration ex -ante

Network remodularization

Network configuration ex-post

Differential Network Assessment

Evaluation of treatment response

Computational and Statistical trade-off

data growth !

➔ deeper complexity

➔ increased accuracy ?

Algorithm weakening:

as data accumulate, consider simpler algorithmic strategies to achieve the desired risk/accuracy.

Design approximation algorithms for semi-intractable problems (combinatorial optimization or controllability)

How messy data reflect the complexity of nature?

Theoretically

- Unknown distributional laws and associations
- Unclear links between data and computational complexity (structural vs algorithmic one).

Empirically

- Need domain-oriented data methods
- New ways to summarize essential data characteristics.

The Gaussian world

$N_d(\mu,\sigma)$

Works under Central Limit th.

(*same experimental conditions for all data)*

$\sum_j w_j N_d(\mu_j,\sigma_j)$

Concatenated data sets collected under different experimental conditions

Statistical summaries hard to obtain

Quantum biology

A quantum biological system subject to perturbations (environment, stressors, drugs…) fluctuates in and out of equilibrium state

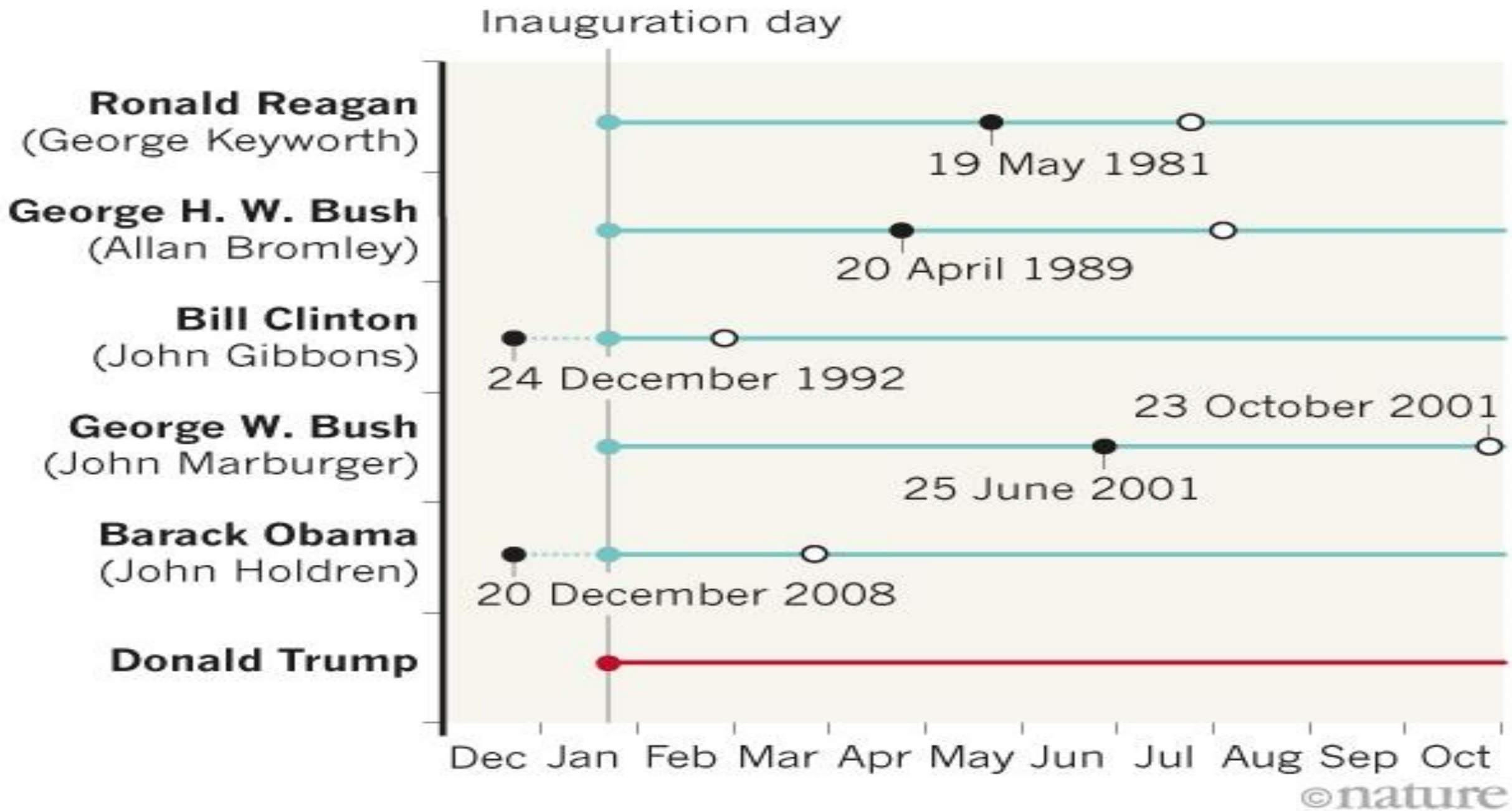
~

evolve into a state of lower entropy or increased entanglement.

HELP WANTED

Donald Trump has now gone longer without a science adviser than any first-term US president since at least 1976. By contrast, his predecessor Barack Obama took the least amount of time in announcing his science adviser pick (in brackets), naming John Holdren one month before Inauguration Day.

● Science adviser announced ○ Confirmed by Senate



Ex nihilo, nihil

(Dal nulla, nulla)

Remember **Ithaca** (greek poet CP Cavafy, 1911):
the true prize of a research career is the journey

expect the unexpected!


#AlwaysRunWithYourBike

Thanks!