Introduction to Systems Immunology

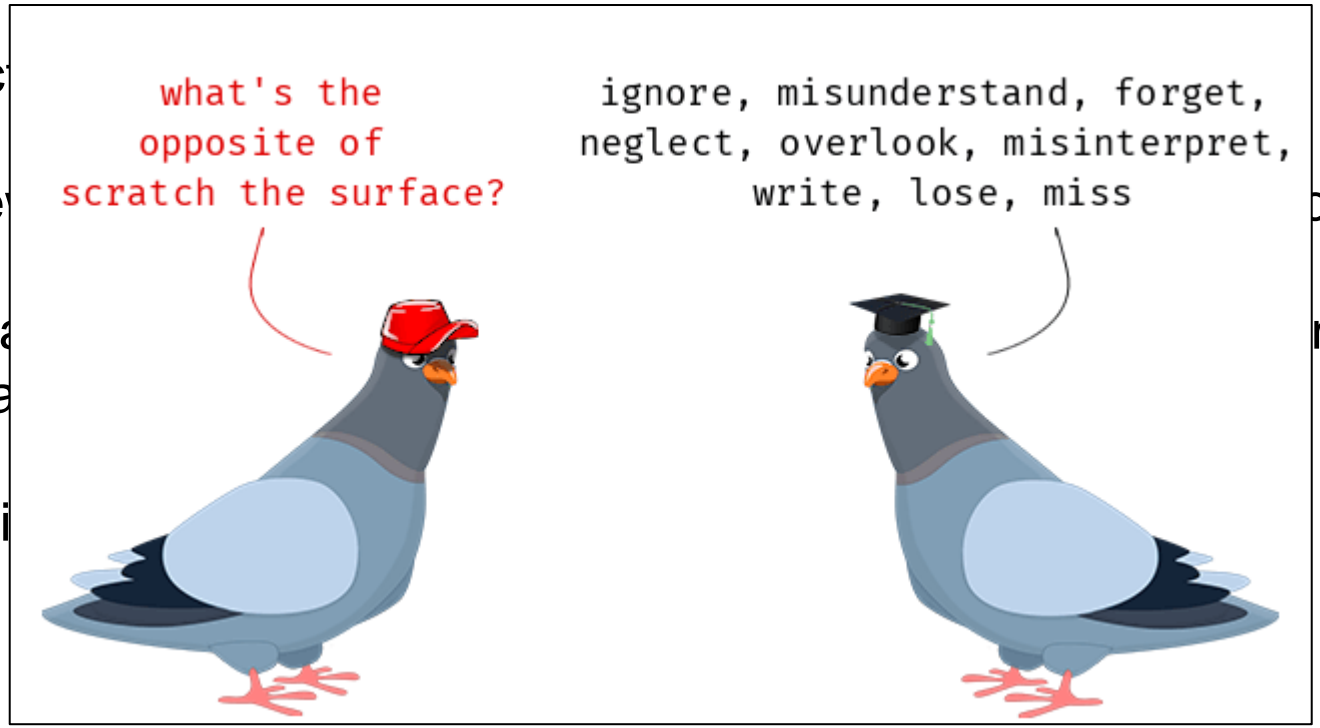
Technologies, Methods, Applications

Emanuele de Rinaldis

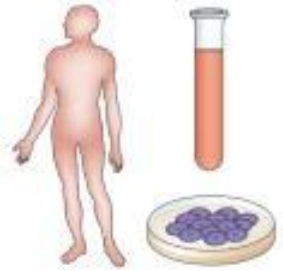
FOCIS – June 8-9 , 2026

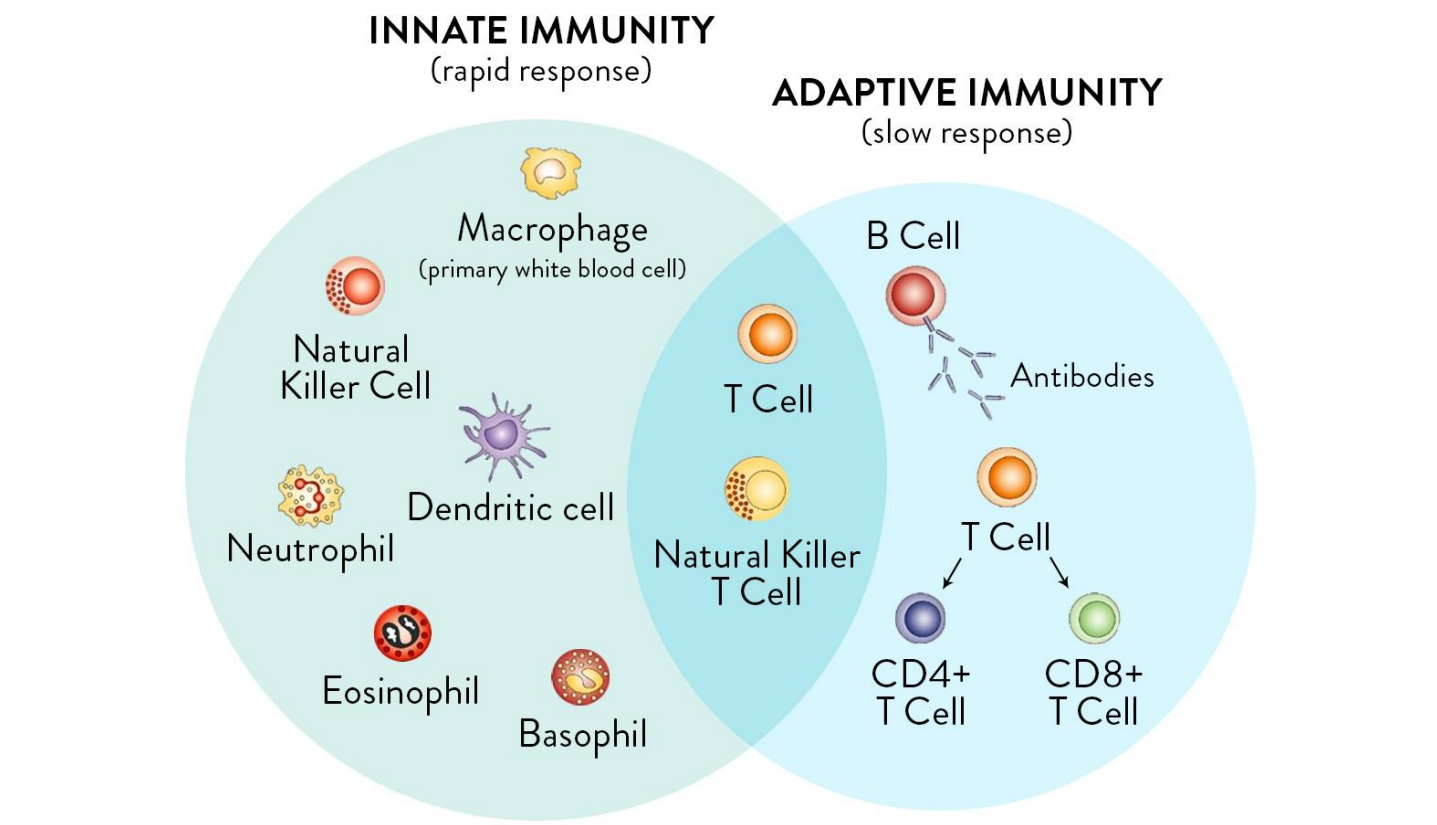
Outline

- ❑ Introduc
- ❑ Brief vie
- ❑ Data ana visualiza
- ❑ Discussi

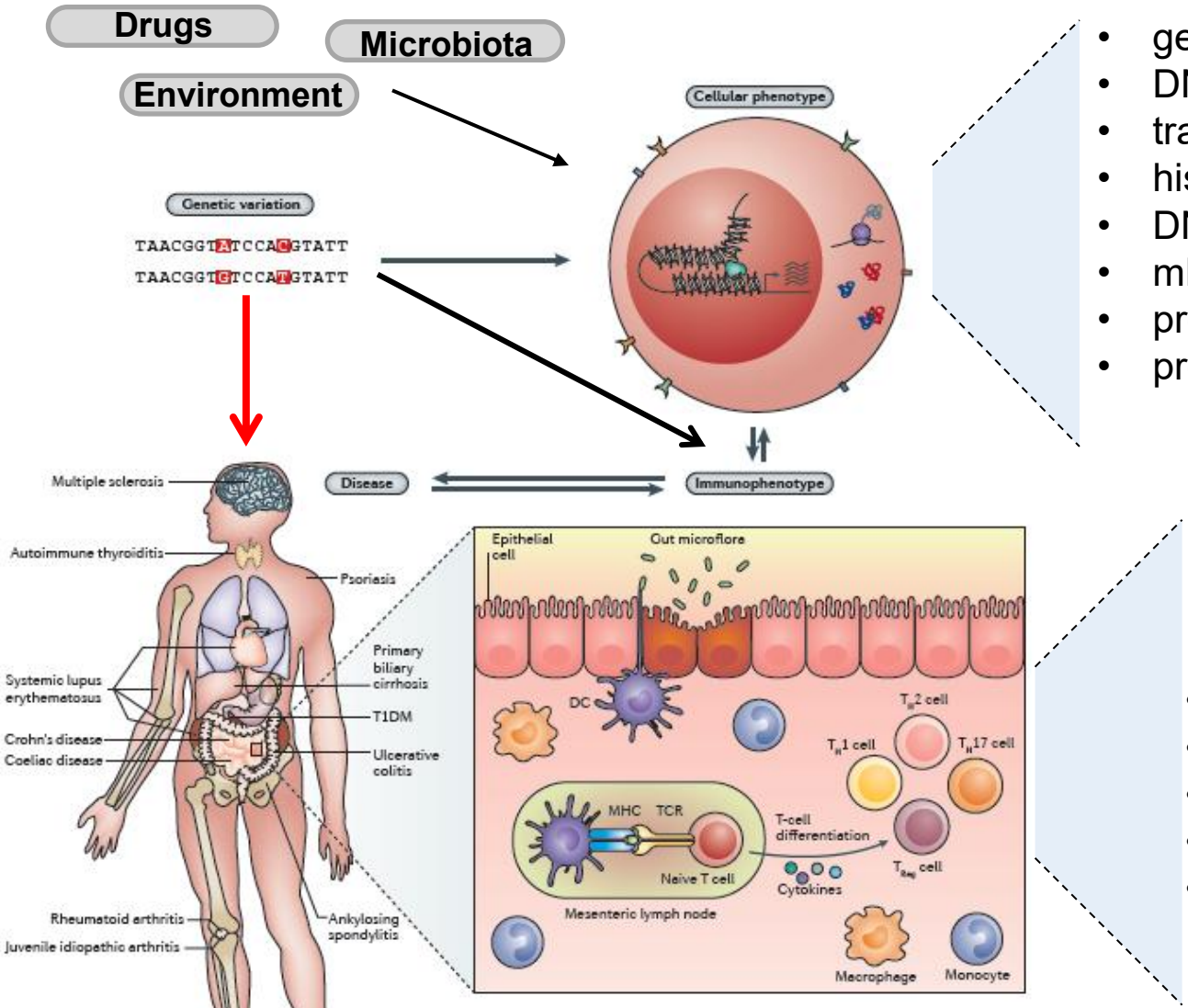


The Immune System





Systems Immunology

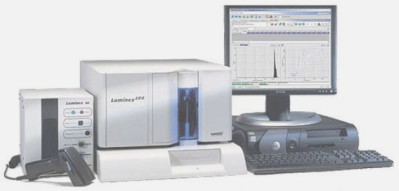


- gene transcription
- DNA–DNA interactions
- transcription factor binding
- histone modifications
- DNA methylation
- mRNA stability and translation
- protein levels
- protein–protein interactions

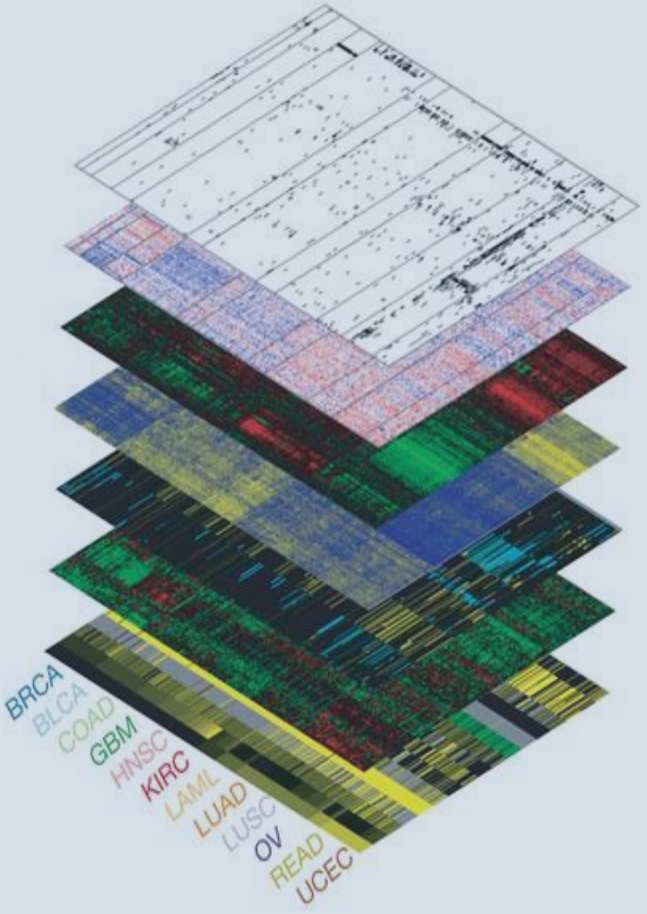
- signalling response
- cell-type abundances
- cell-cell spatial interactions
- RNA/protein levels
- cytokine production

Adapted from Nature Reviews Genetics 17, 160–174 (2016)

Technologies



Data

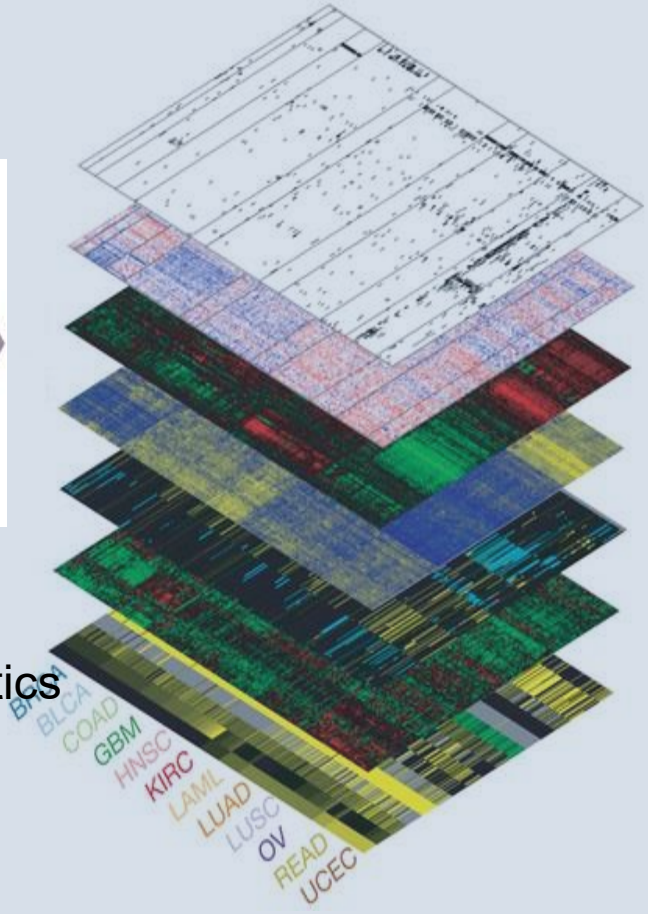


Technologies

Analysis

Math
Stats,
Bioinformatics
ML/AI

Data



Immune Cell Profiling: Flow Cytometry and Gene Expression

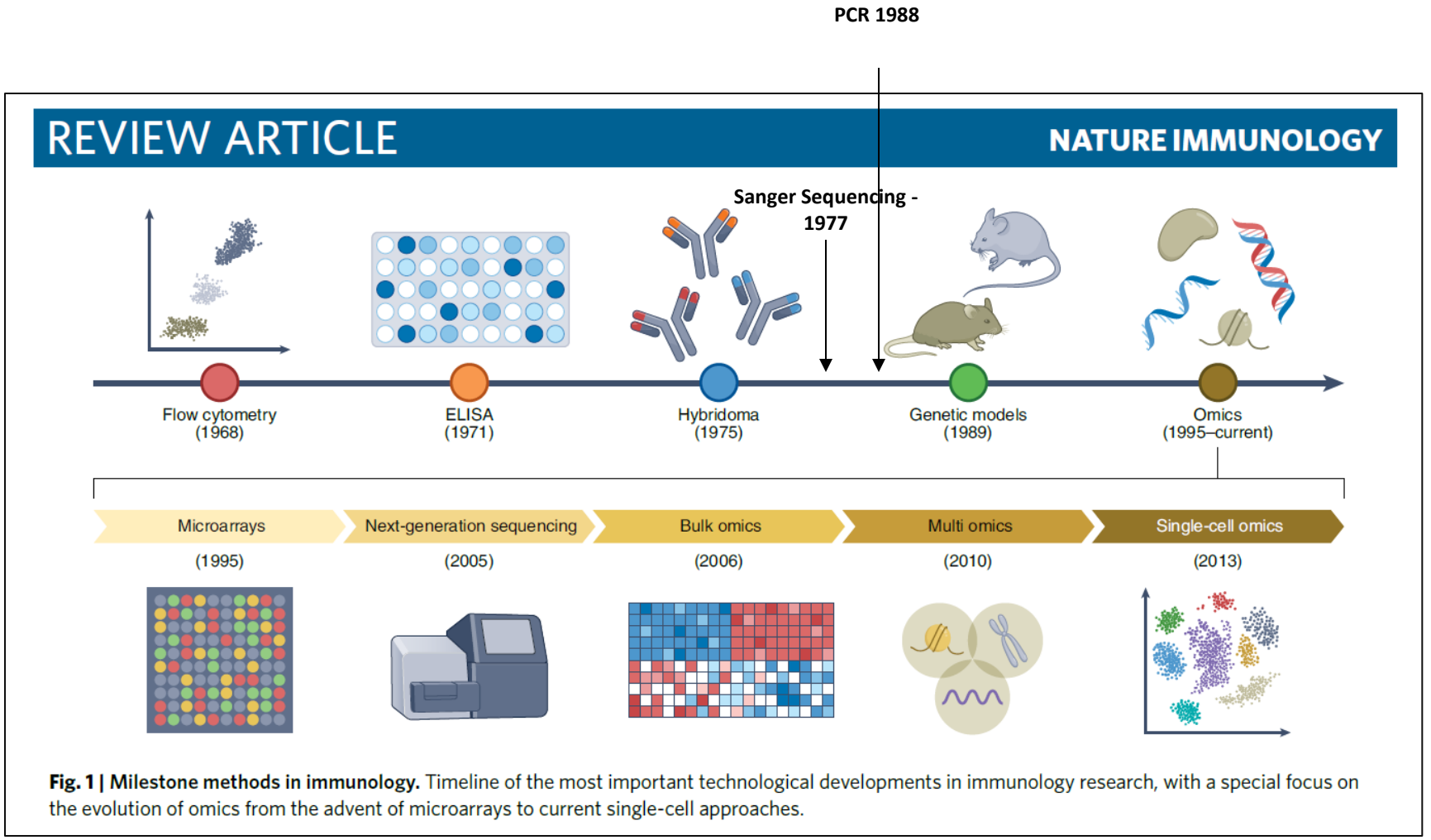
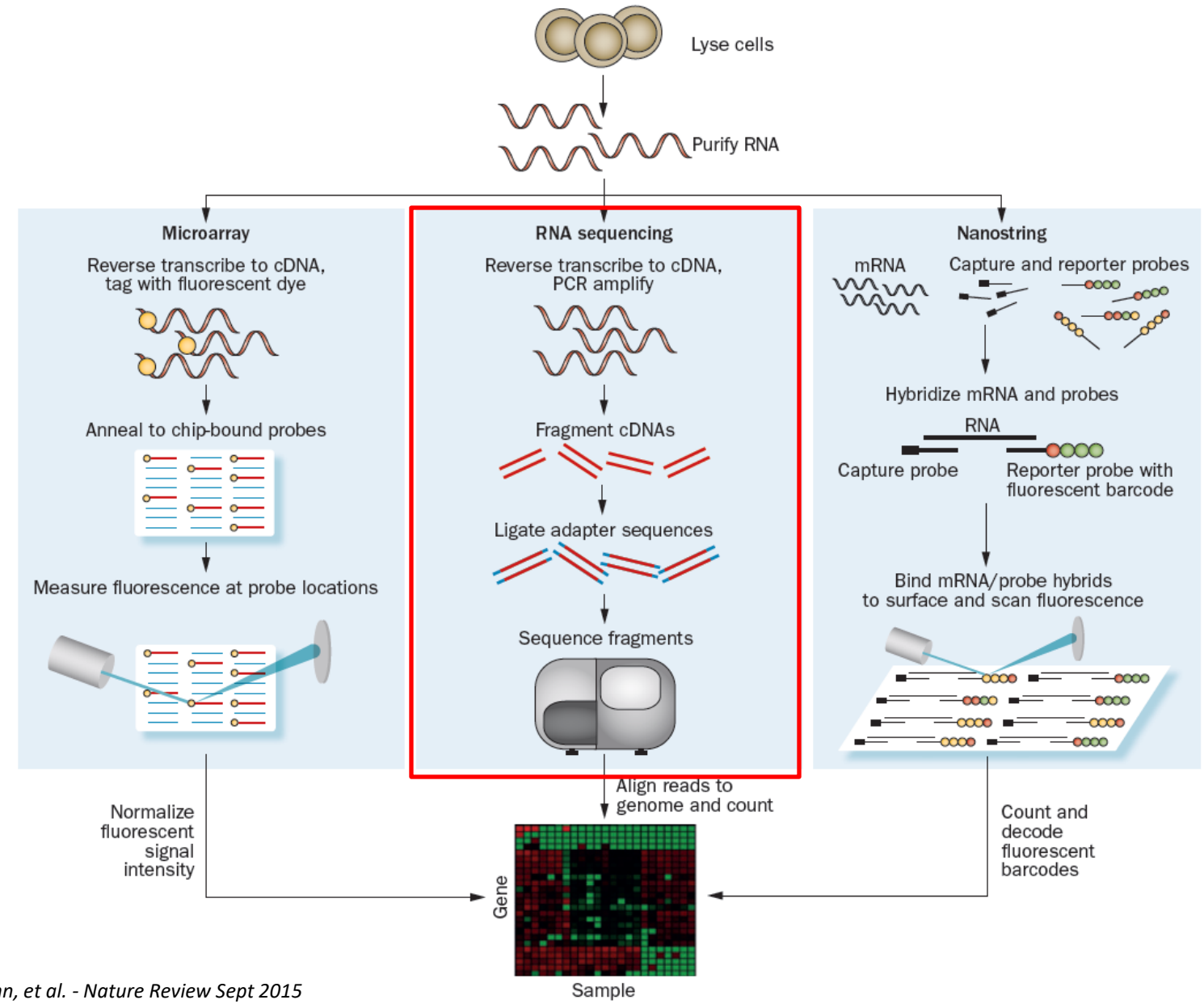


Fig. 1 | Milestone methods in immunology. Timeline of the most important technological developments in immunology research, with a special focus on the evolution of omics from the advent of microarrays to current single-cell approaches.

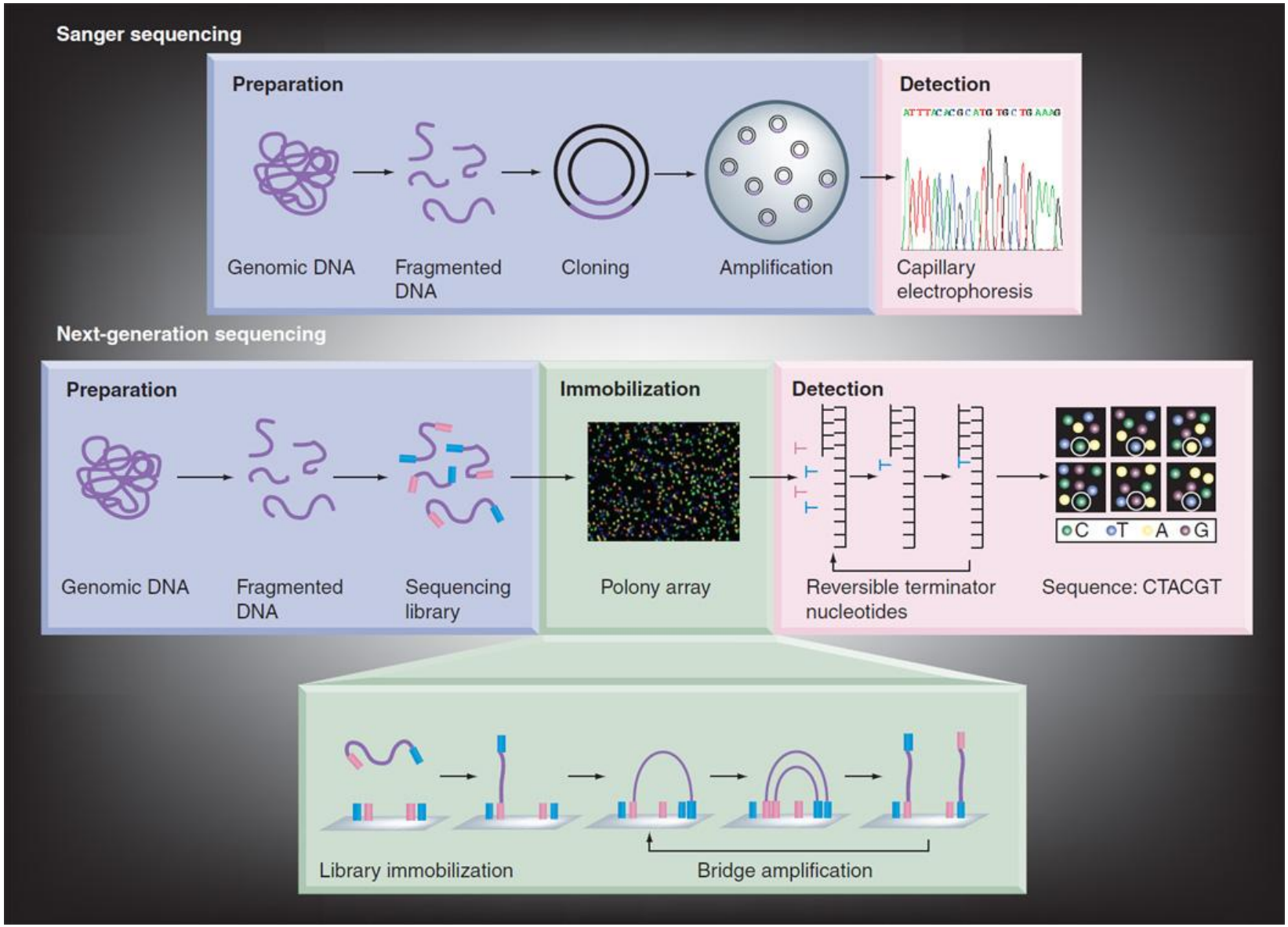
Nature Immunology volume 23, pages1412–1423 (2022)

Gene Expression in “Bulk” Samples



Joerg Ermann, et al. - Nature Review Sept 2015

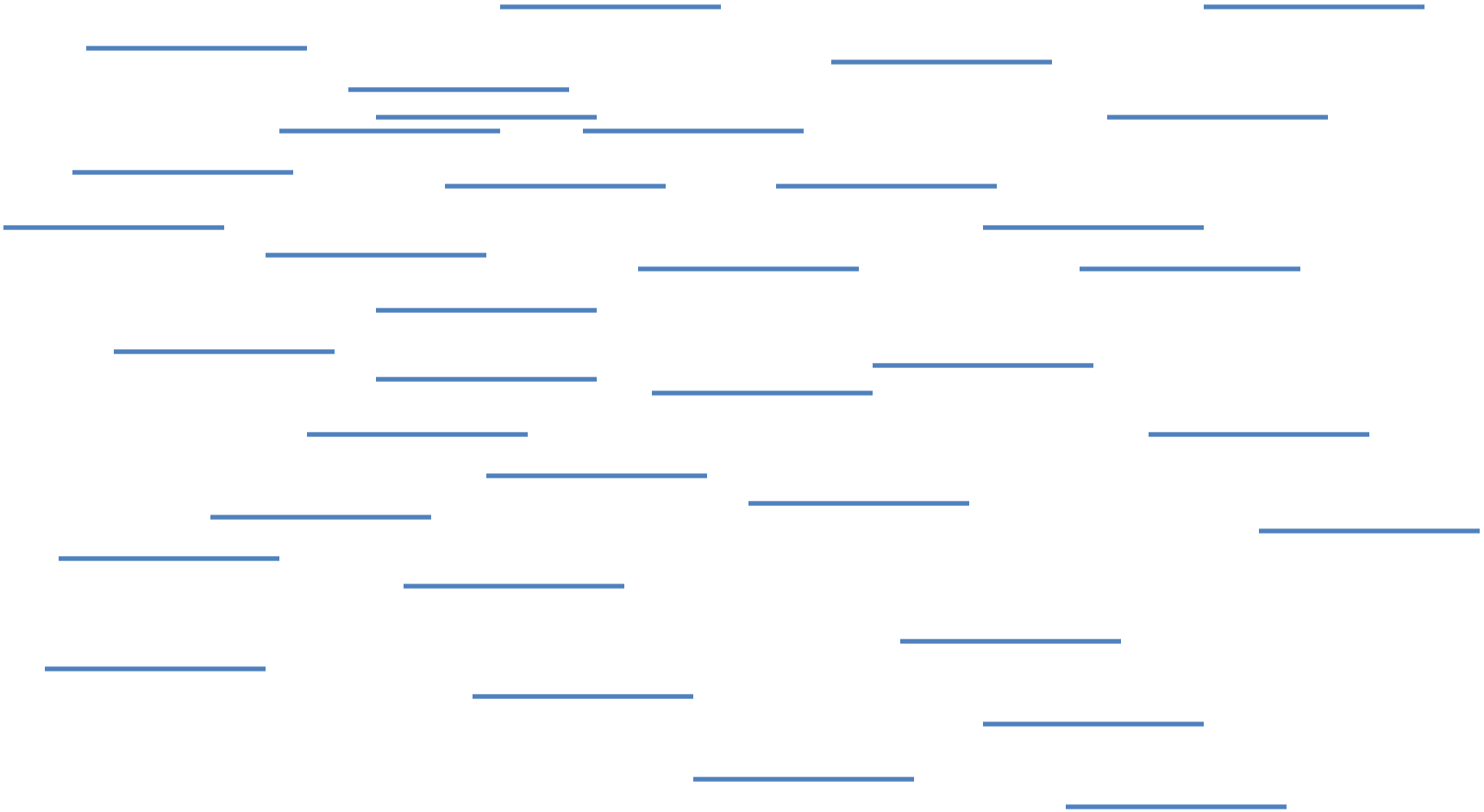
Next Generation Sequencing – The Illumina Platform



Per Med. 2011 May 1;8(3):331-345.

Lots of sequenced reads…

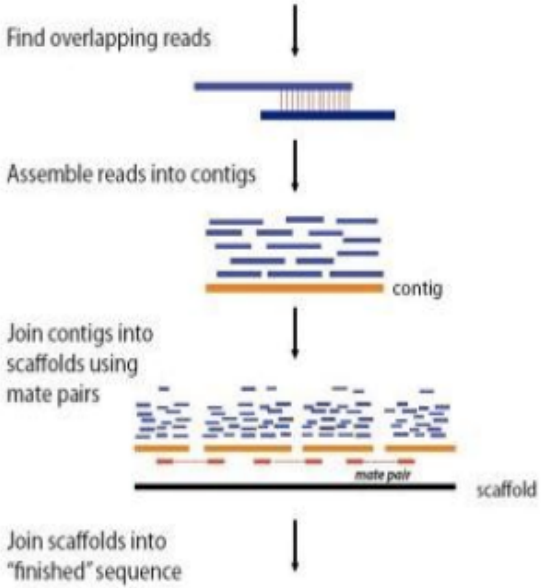
Illumina NovaSeq 6000: up to 20 billion reads, 3.000Gb data, less than 2 days



What do we do with them ?

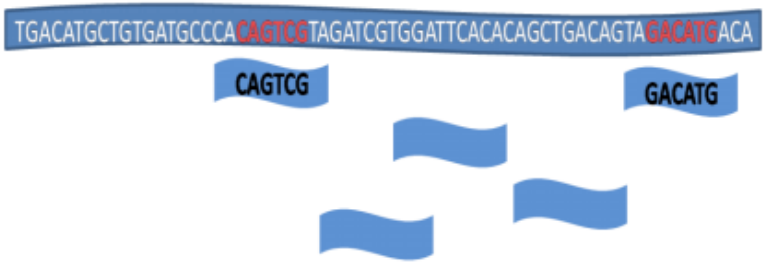
Reads

denovo Assembly



Reference based analysis

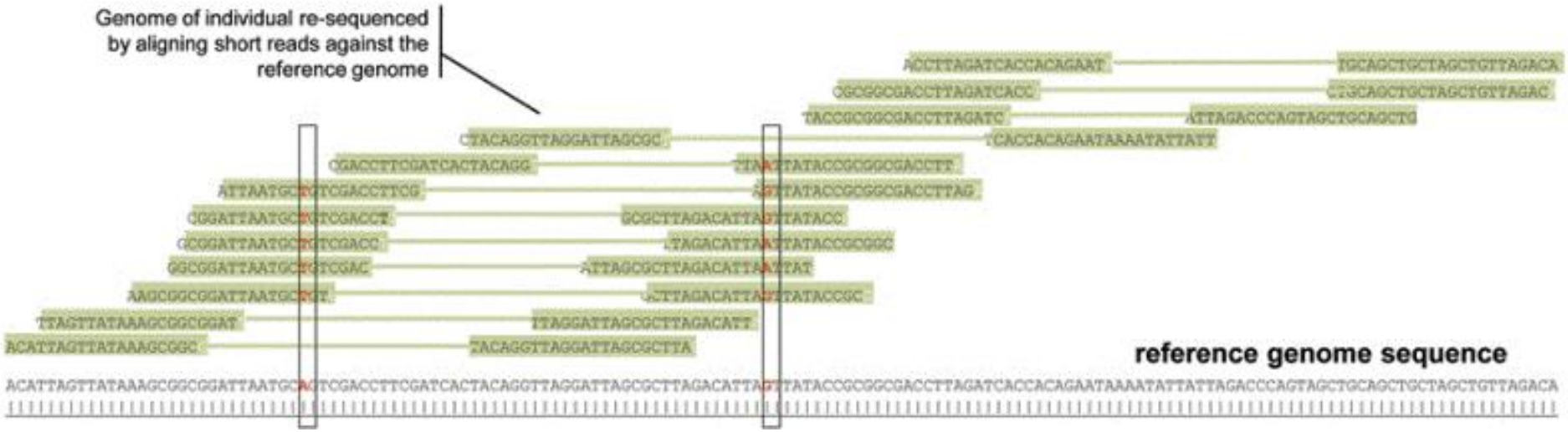
Analysis based on a known reference genome



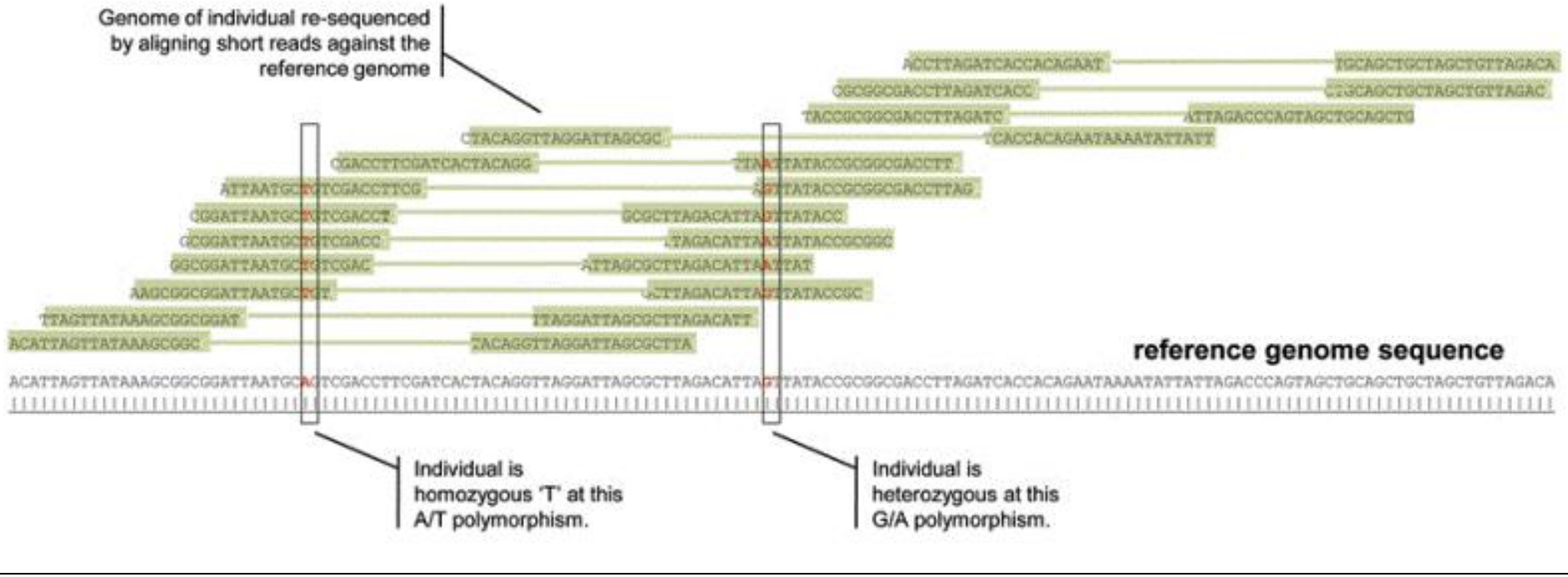
Sequencing of a new organism
Meta-Genomics
Reconstructing cancer genomes

Organism specific experiments
Annotating functional elements

Reads Alignment

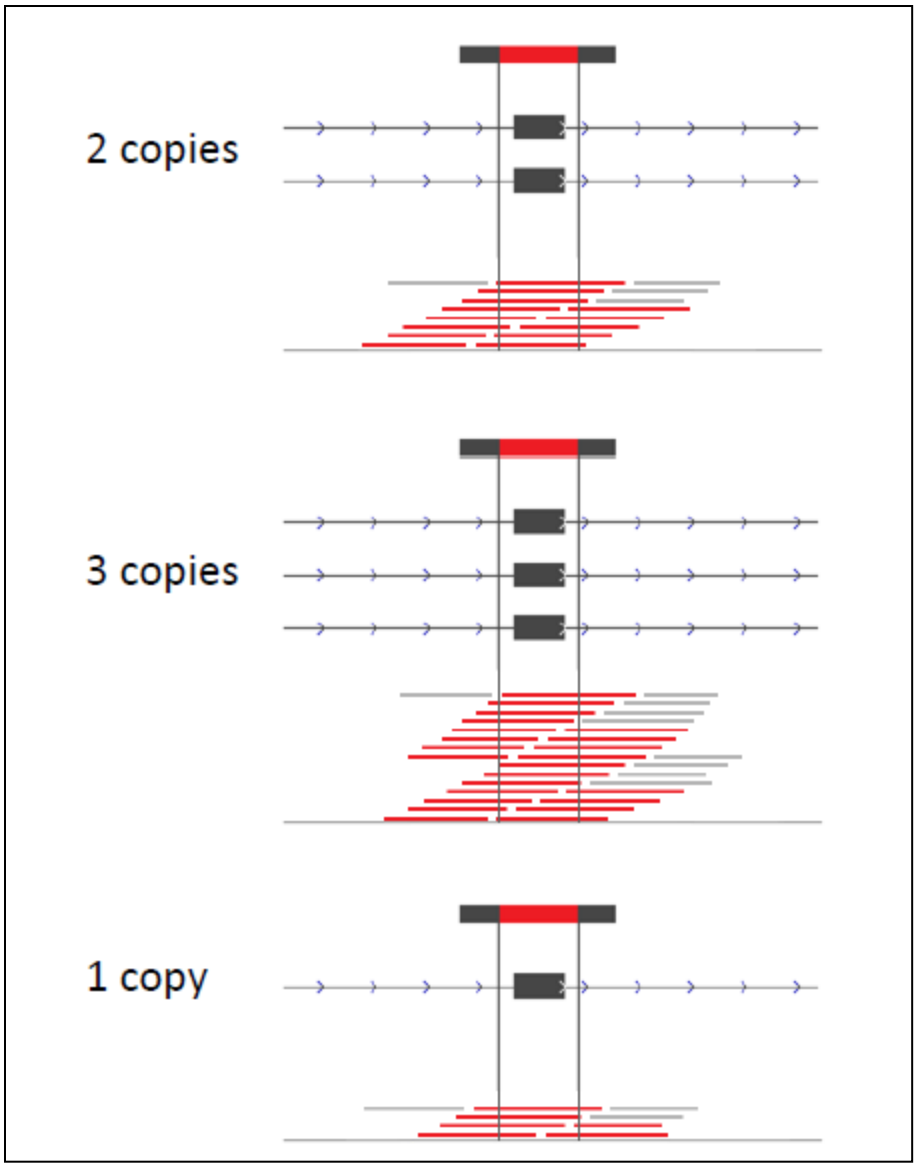


Genome Sequencing: Variant Calling



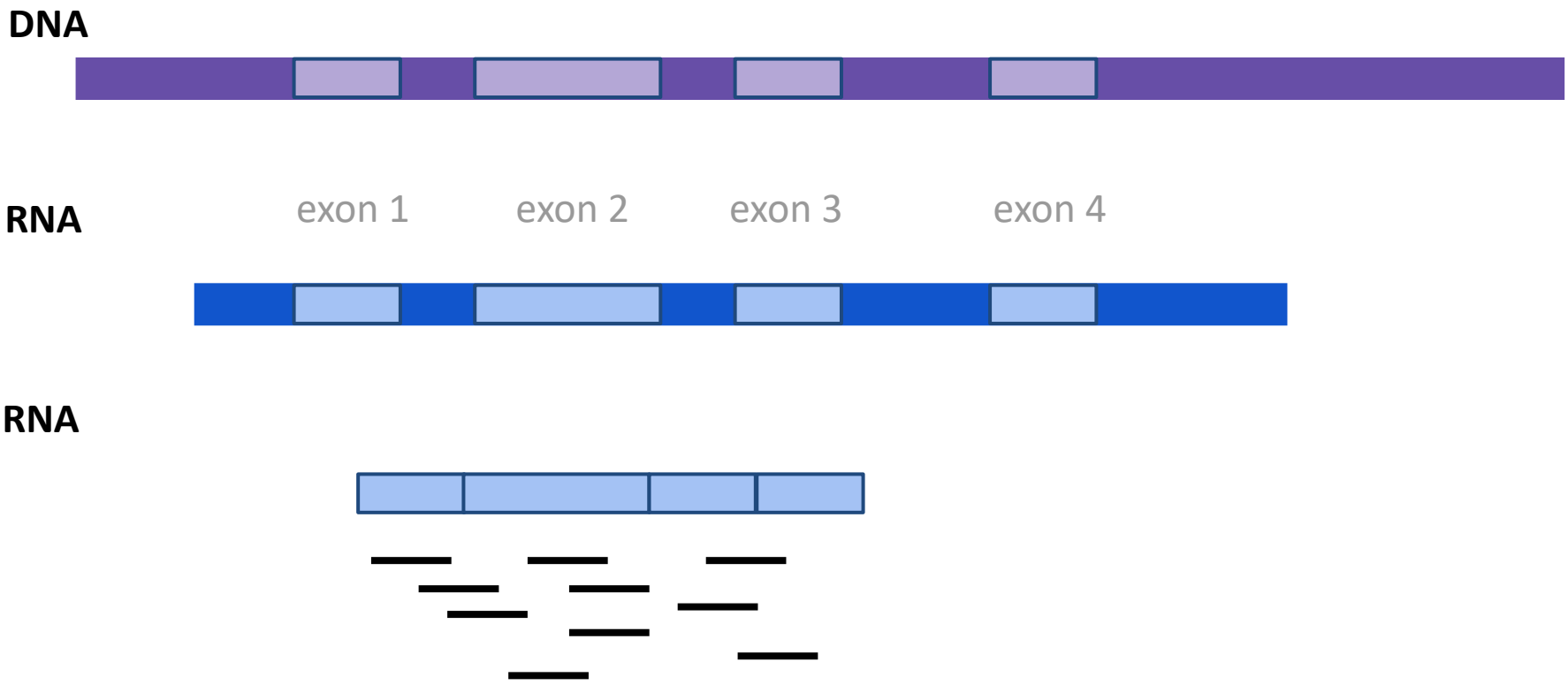
Differences between aligned reads and reference genome can be identified ➔ variants

Genome Sequencing: Copy Numbers



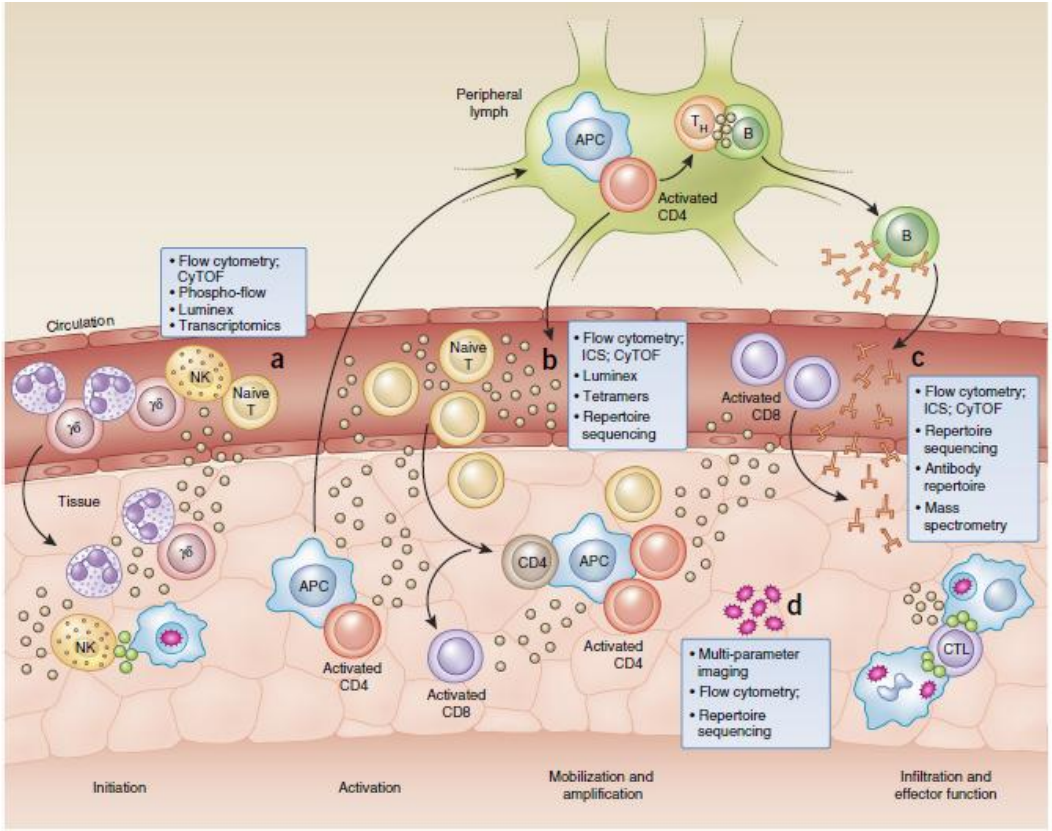
- The number of aligned reads on a given regions is proportional to the number of starting DNA copies of that region
- This information can be used to infer DNA copy number variations (CNVs)

RNA Sequencing

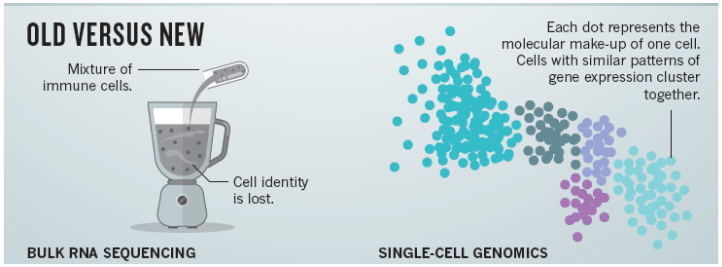


- The number of aligned reads on a given RNA is proportional to the number of its starting molecules
- This information can be used to infer RNA abundances

Immune Systems Heterogeneity



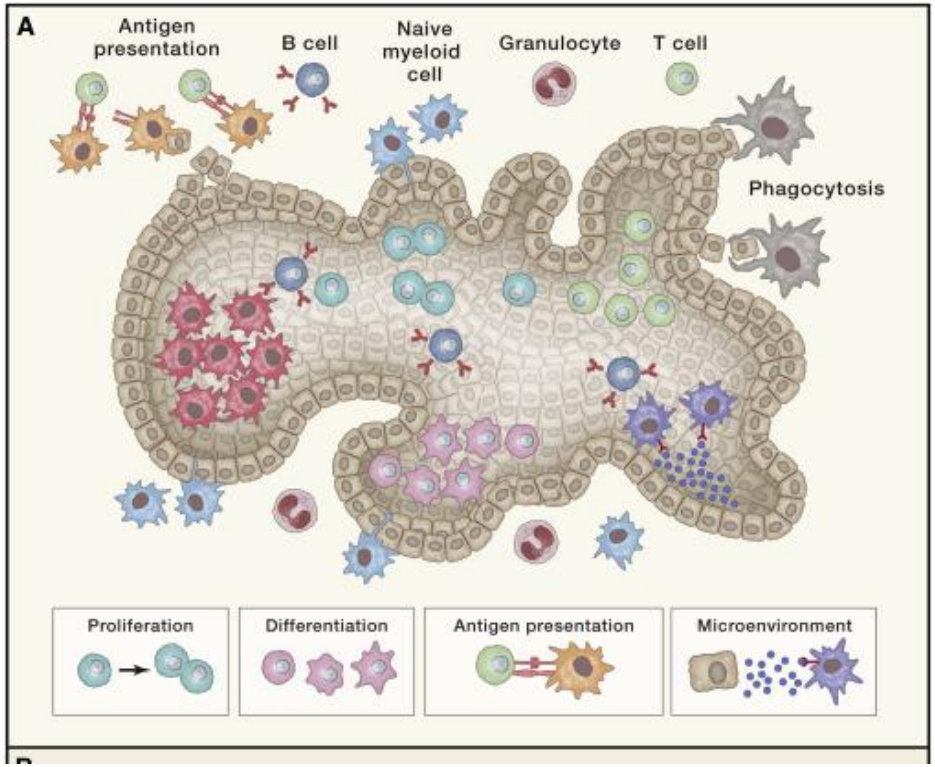
Davis et al. - Nature Immunology **18**, 725–732 (2017)

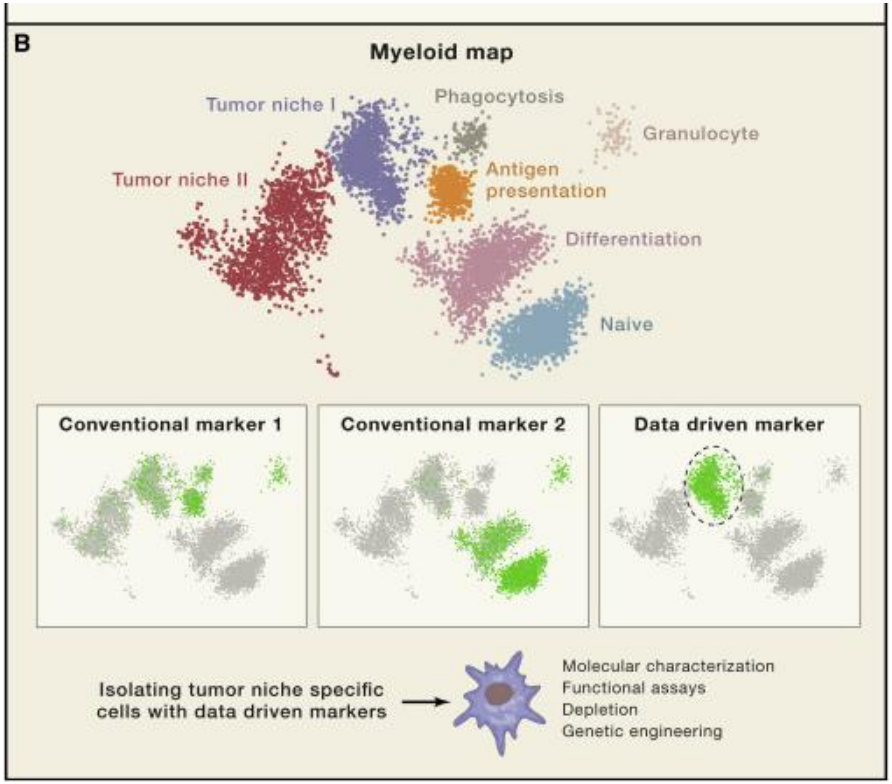


6 JULY 2017 | VOL 547 | NATURE | 27

- Average signals are measured and fails to capture sample heterogeneity
- Individual cell properties and interplay between different cell subtypes can be captured

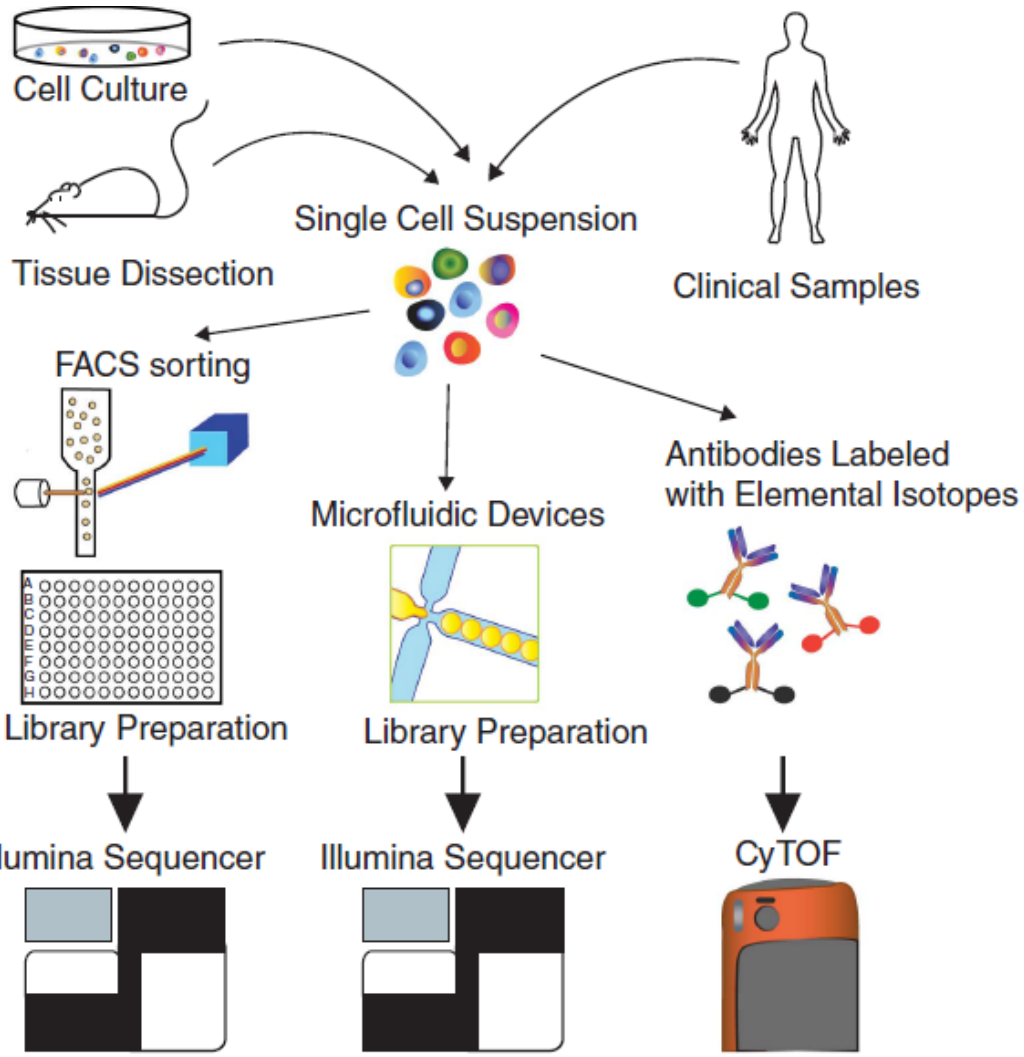
Dissecting tissue heterogeneity at single-cell level





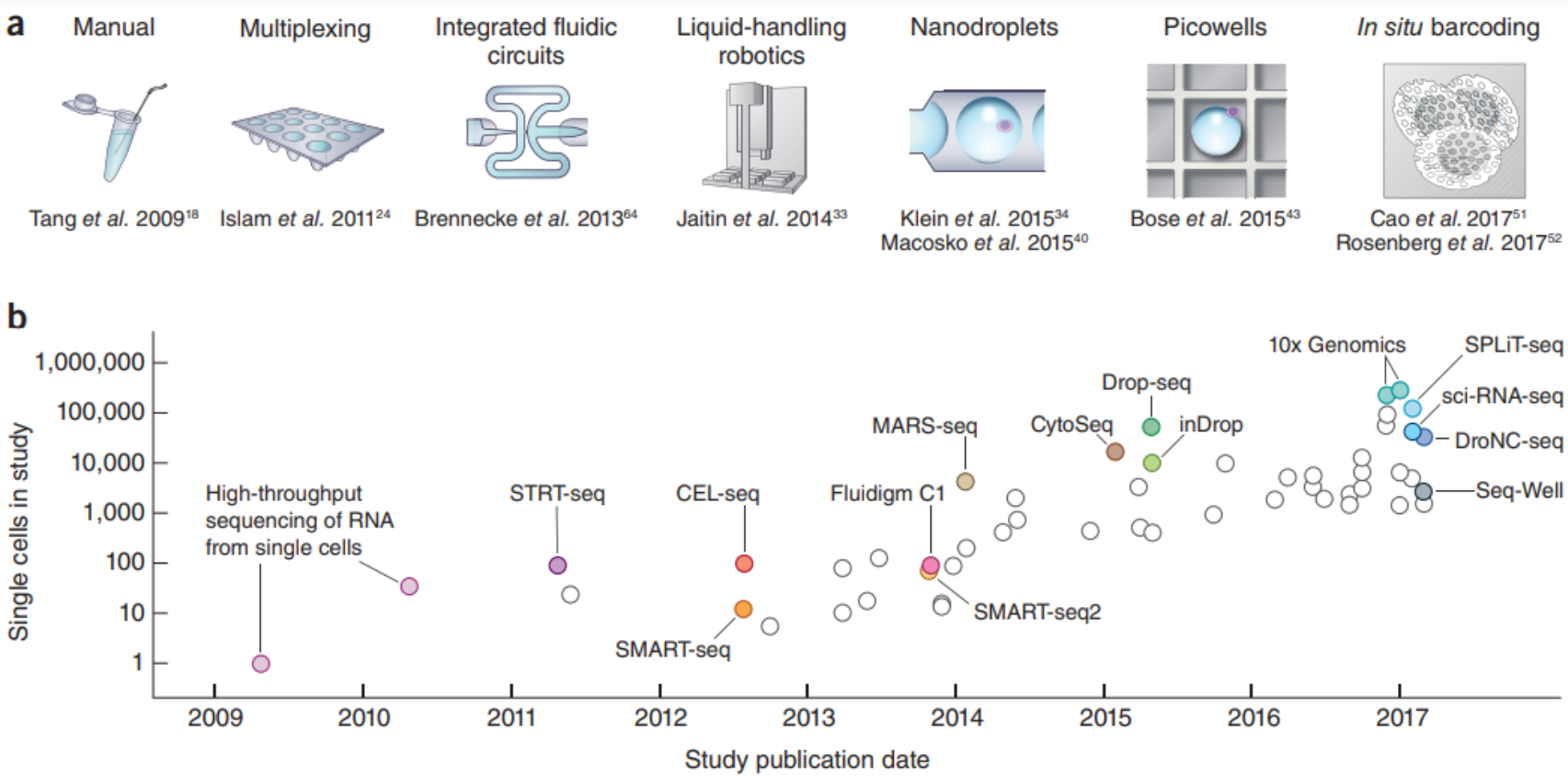
Giladi A, Amit I. Cell. 2018 Jan

Single-Cell Analysis



Imm. and Cell Biol. (2016) 94, 225–229;

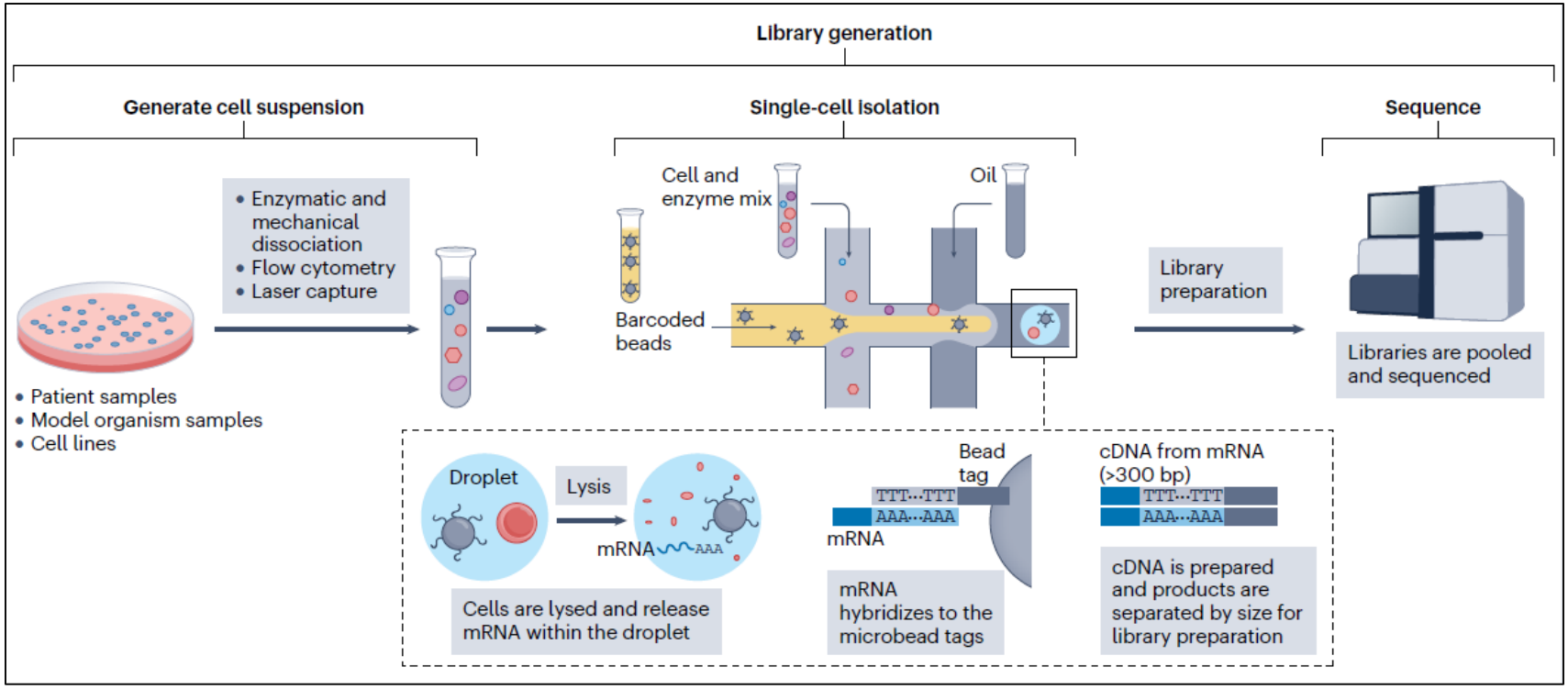
Development of Single-Cell Technologies



Svensson V, Vento-Tormo R, Teichmann SA.Nat Protoc. 2018

Basic workflow in 10X Chromium

Individual cells must be trapped within a space that is not continuous with spaces containing any other individual cells.

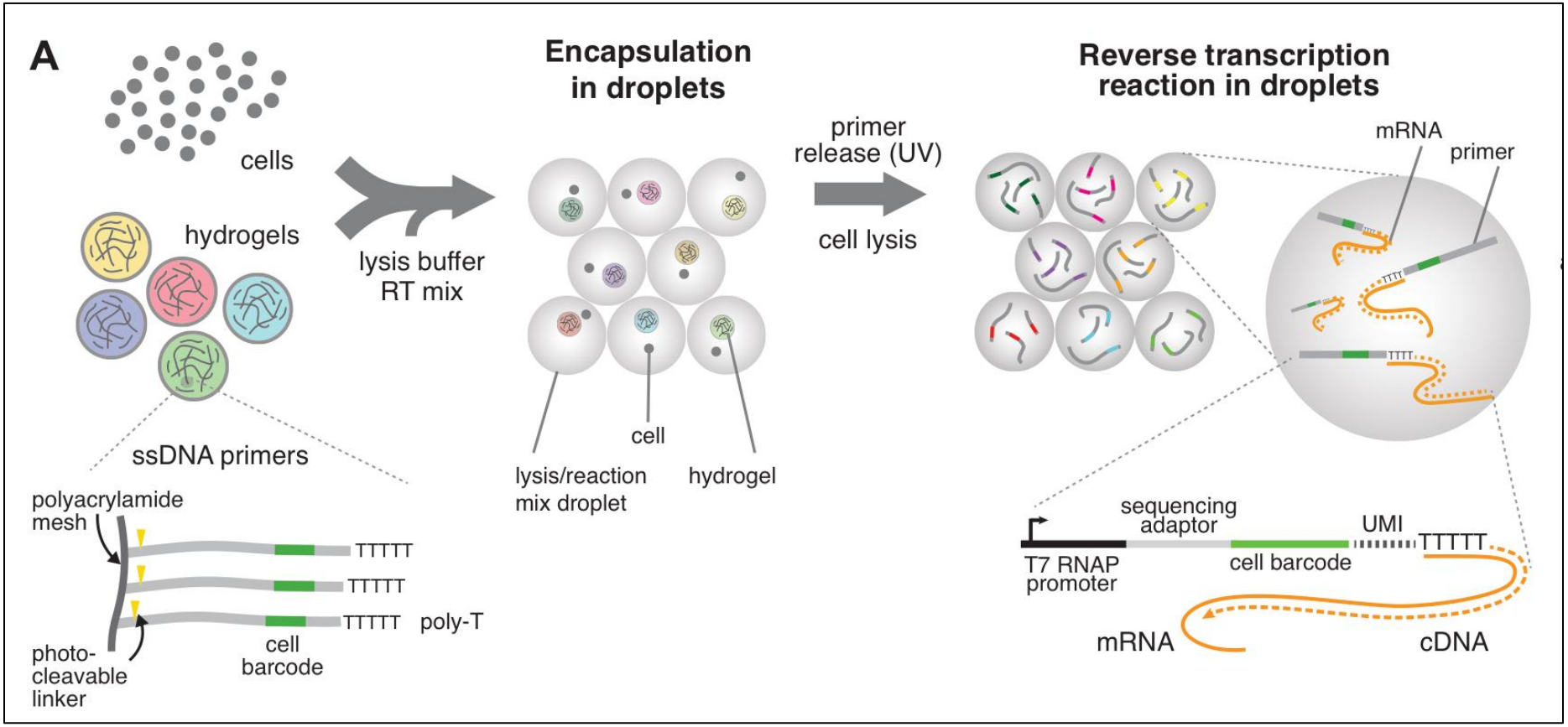


Nature Reviews Drug Discovery volume 22, pages496–520 (2023)

- **10X**: combines an aqueous flow of cells, barcoded primers carried in beads, lysis buffer and reverse transcription enzymes with oil to create microdroplet reaction chambers.
- **Plate-based** technologies perform this step in microwells
- Automated **microfluidic devices** use other forms of microchamber.

InDrop: Barcoding process

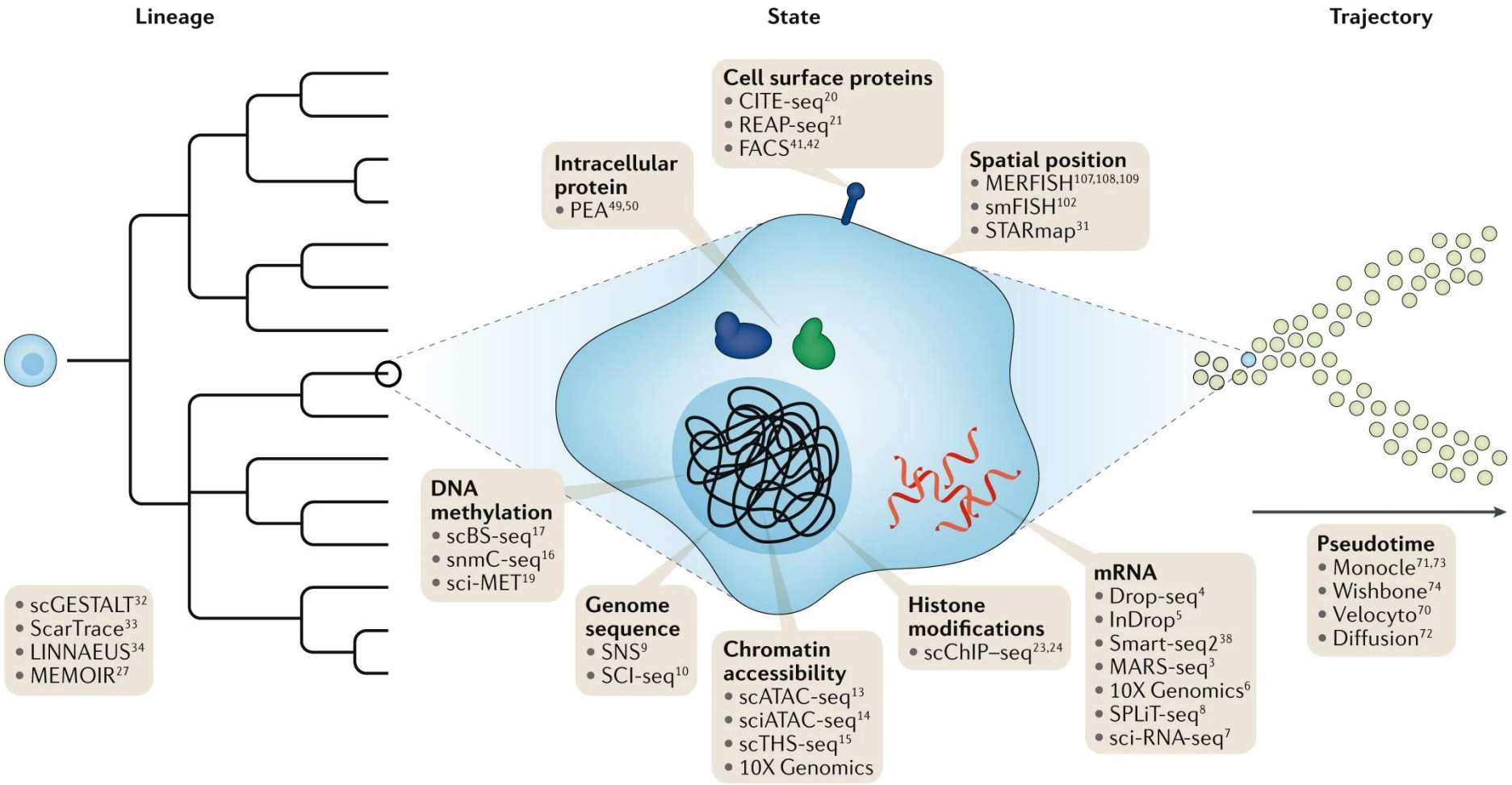
Co-encapsulation of cells, barcodes (delivered by hydrogel bead), and RT–lysis reagents into microfluidic droplets.



Zilionis et al., Nature Protocols 2017

Single-cell multi-omics

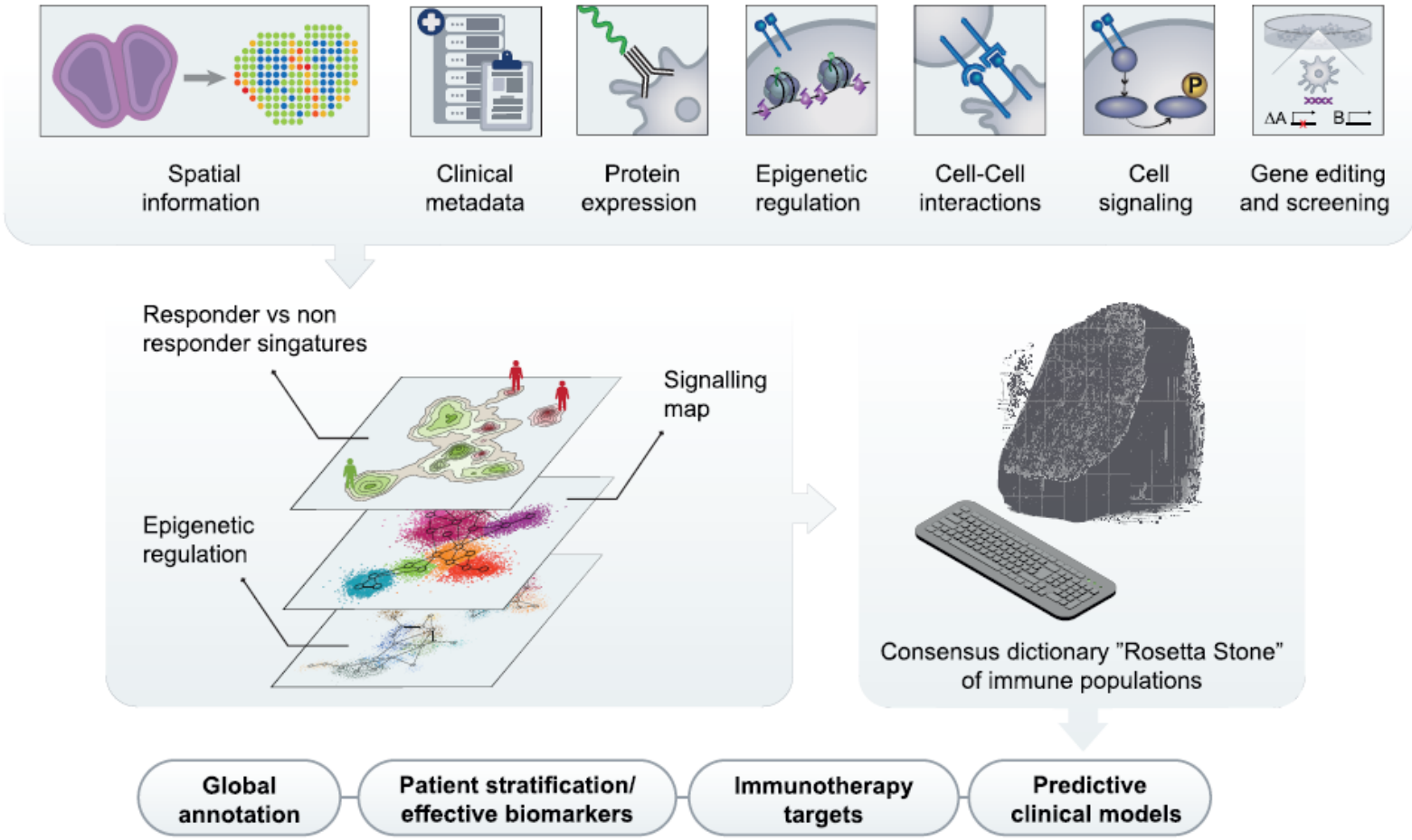
Multimodal and integrative methods for single- cell analyses



Stuart T, Satija R.Nat Rev Genet. 2019

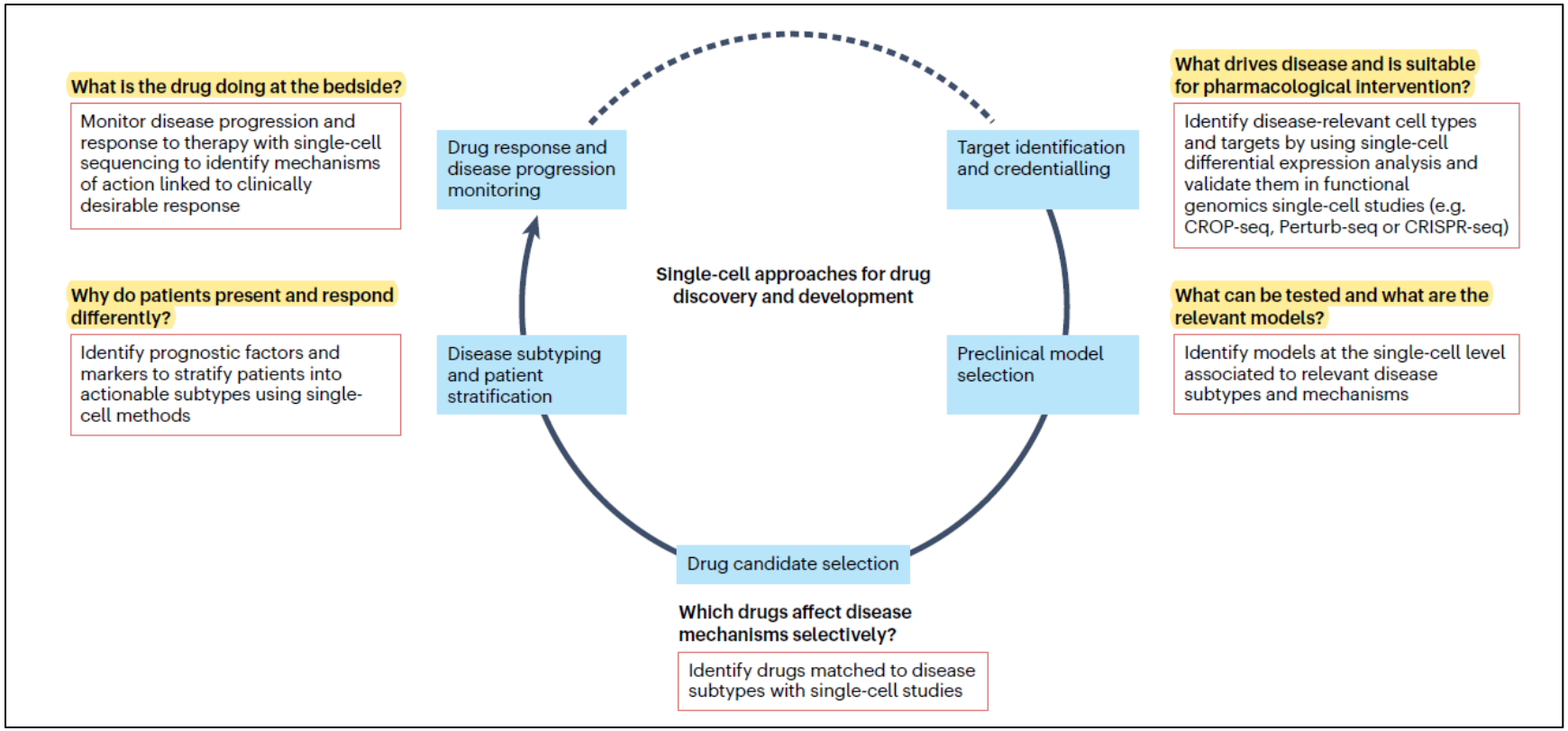
Systems Biology Through Single-Cell Multi-omics

"Rosetta Stone": **deciphering the different immune populations and states across datasets**



Ginhoux F, Yalin A, Dutertre CA, Amit I. Immunity. 2022

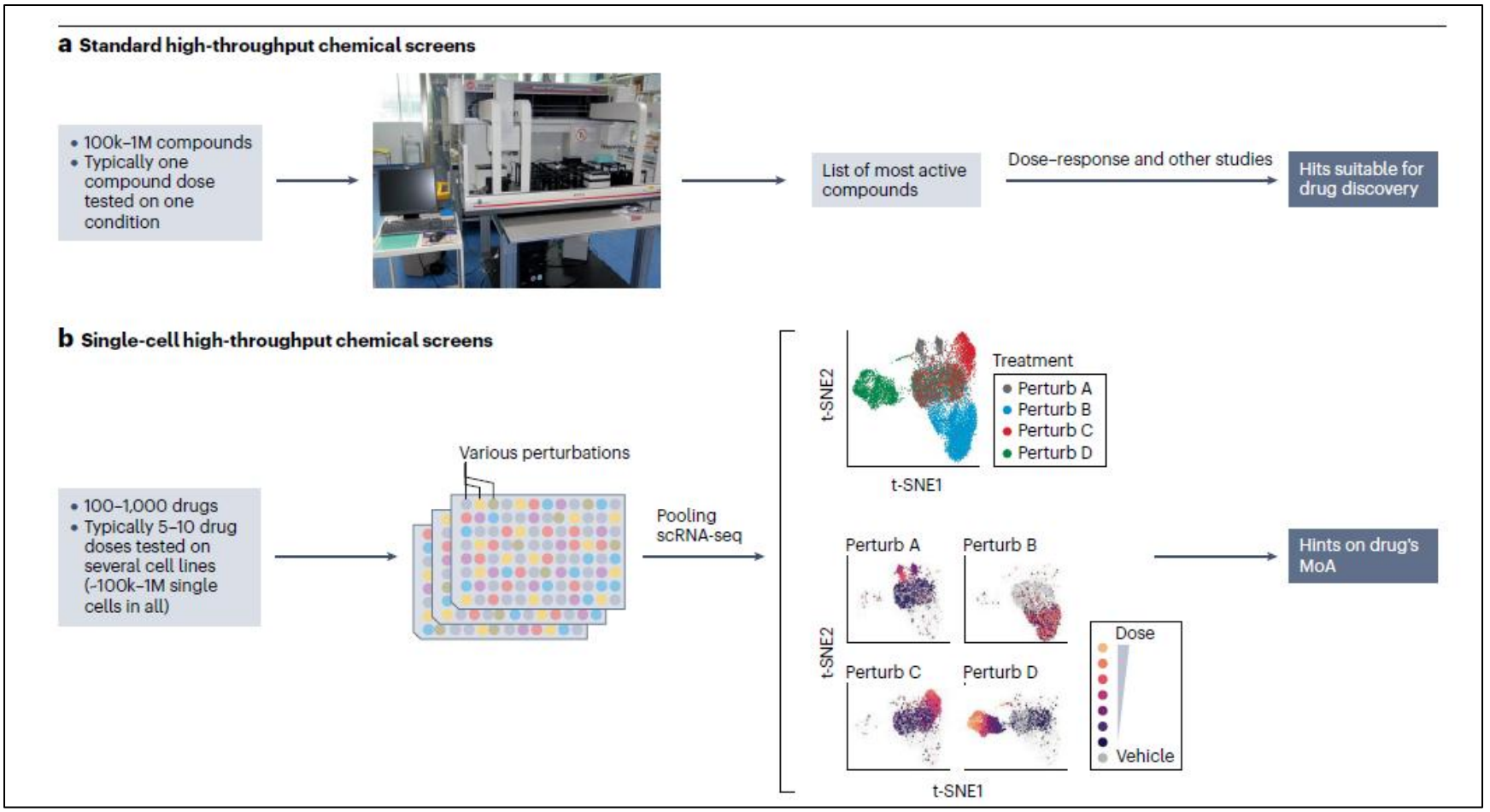
Applications To Drug Discovery



Nature Reviews Drug Discovery volume 22, pages496–520 (2023)

Applications To Drug Discovery

Compound screening



Nature Reviews Drug Discovery volume 22, pages496–520 (2023)

Single-Cell for Biomarker Discovery and Patient Stratification

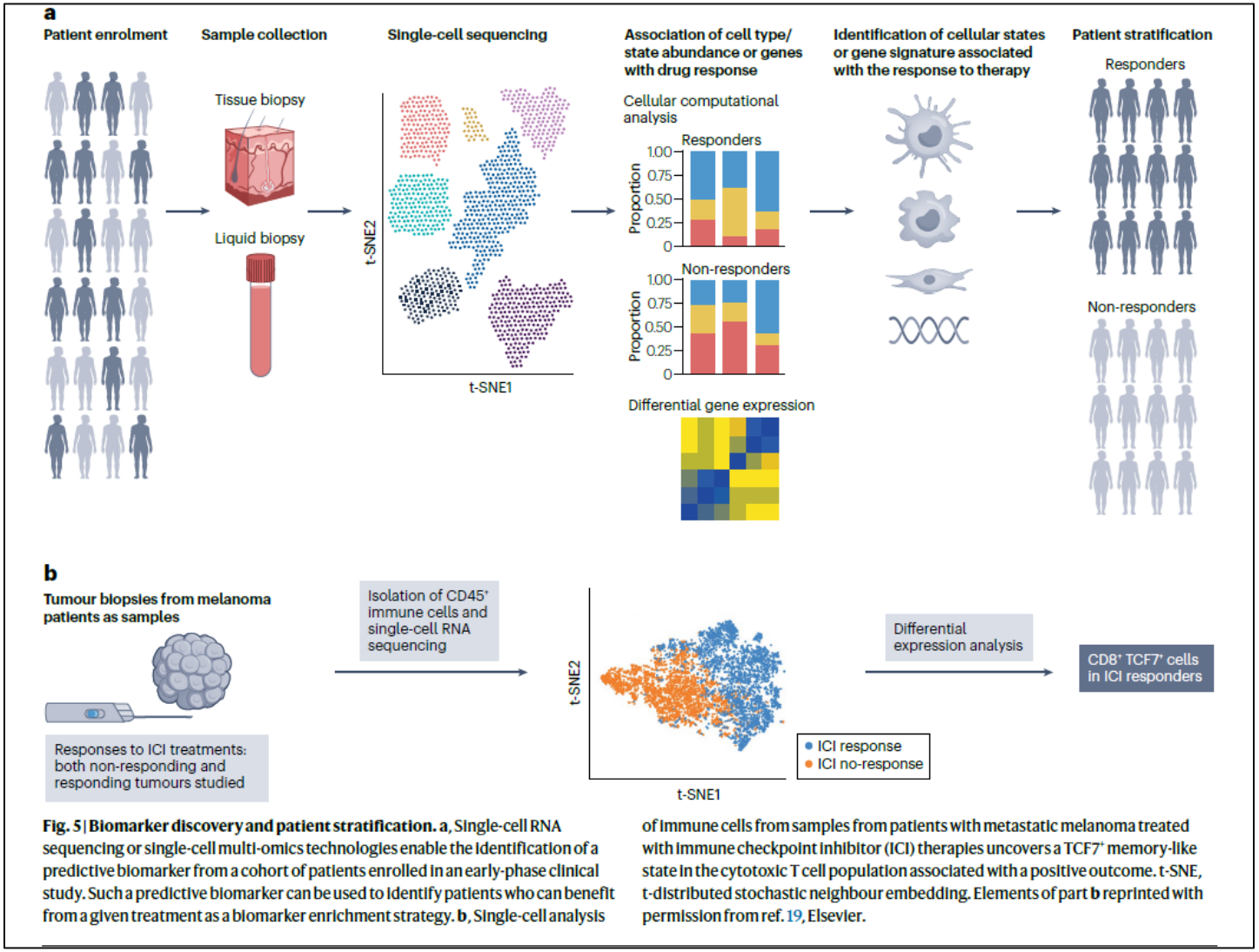
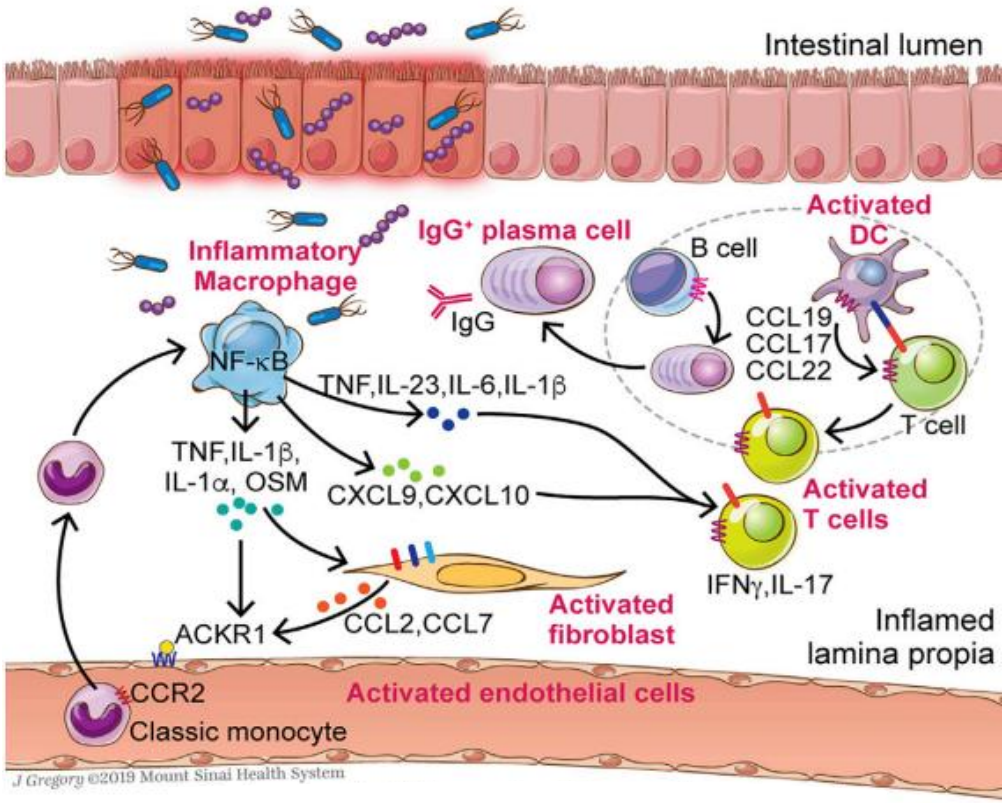


Fig. 5 | Biomarker discovery and patient stratification. a, Single-cell RNA sequencing or single-cell multi-omics technologies enable the identification of a predictive biomarker from a cohort of patients enrolled in an early-phase clinical study. Such a predictive biomarker can be used to identify patients who can benefit from a given treatment as a biomarker enrichment strategy. **b**, Single-cell analysis of immune cells from samples from patients with metastatic melanoma treated with immune checkpoint inhibitor (ICI) therapies uncovers a TCF7$^+$ memory-like state in the cytotoxic T cell population associated with a positive outcome. t-SNE, t-distributed stochastic neighbour embedding. Elements of part **b** reprinted with permission from ref. 19, Elsevier.

Nature Reviews Drug Discovery volume 22, pages496–520 (2023)

Single-Cell-Based Biomarker of response in anti-TNF therapy (CD)

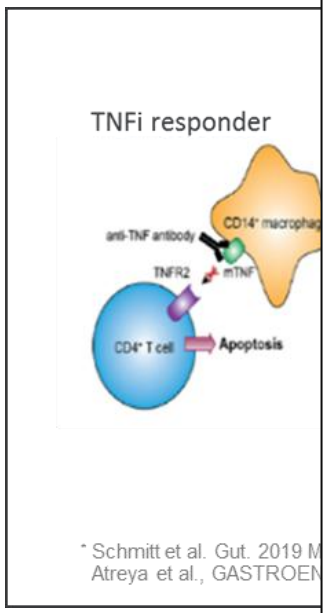
GIMATS^{high} Module in CD inflamed ileums associates with resistance to anti-TNF responder CD patients



- GIMATS=IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and stromal cells
- Existence of two qualitatively distinct subsets of disease, with distinct responses to anti-TNF therapy.

Martin et al., 2019, Cell 178, 1493–1508

Understanding dual targeting: TNF/IL23 in Chron's Disease



* Schmitt et al. Gut. 2019 M
Atreya et al., GASTROEN

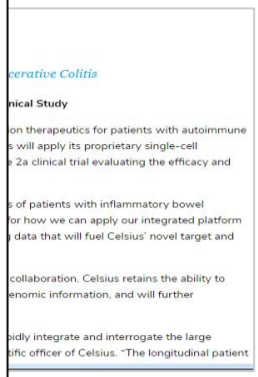
Results of Novel Clinical Study Show Adults with Moderately to Severely Active Ulcerative Colitis Achieved Higher Rates of Clinical Response, Clinical Remission, and Endoscopic Improvement at 12 Weeks with Guselkumab and Golimumab Combination Therapy Versus Either Monotherapy Alone

The VEGA Phase 2a proof-of-concept study shows 83.1 percent of patients who received combination therapy achieved the primary endpoint of clinical response and 36.6 percent of patients achieved clinical remission at week 12

The VEGA study represents a first-of-its-kind biologic combination assessment of an interleukin (IL)-23p19 subunit antagonist with a tumor necrosis factor-alpha (TNFα) antagonist in ulcerative colitis

Janssen
PHARMACEUTICAL COMPANIES OF
Johnson & Johnson

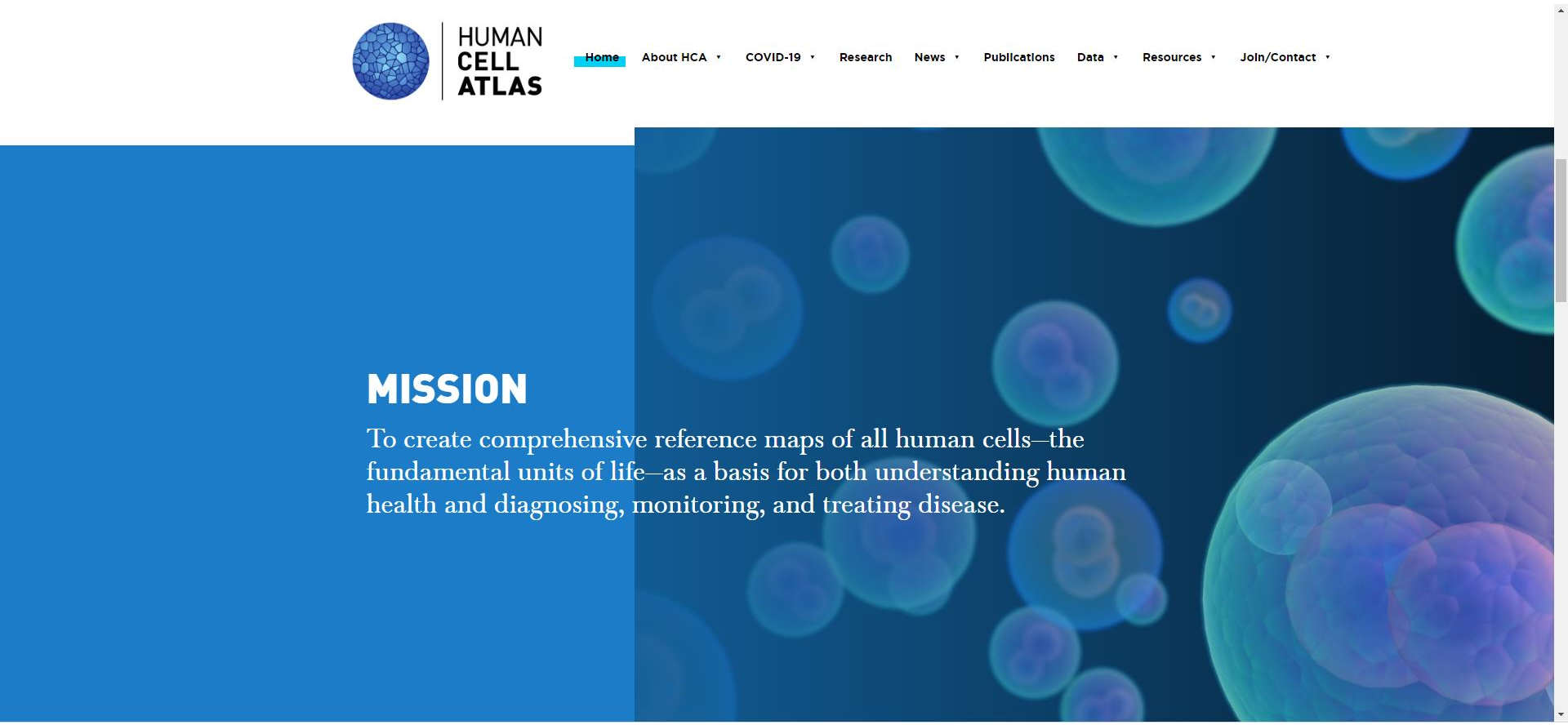
f single-cell
in Phase 2a
apy


cerative Colitis
nical Study
on therapeutics for patients with autoimmune
s will apply its proprietary single-cell
2a clinical trial evaluating the efficacy and
s of patients with inflammatory bowel
or how we can apply our integrated platform
data that will fuel Celsius' novel target and
collaboration. Celsius retains the ability to
enomic information, and will further
idly integrate and interrogate the large
ific officer of Celsius. "The longitudinal patient

Working Hypo
non-
responders

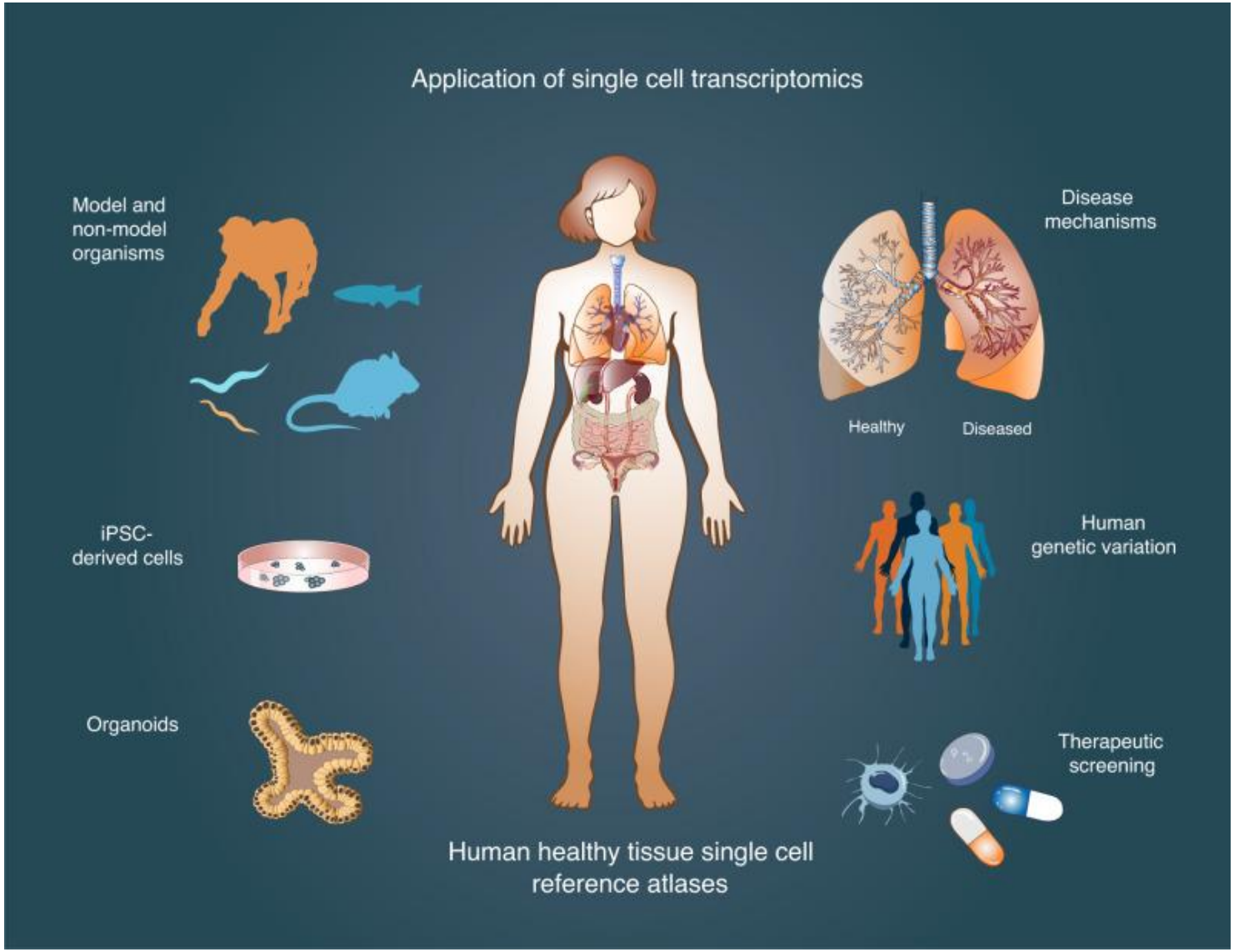
Feagan BG, Sands BE, Sandborn WJ et al. VEGA Study Group. Gusel kumab plus golimumab combination therapy versus guselkumab or golimumab monotherapy in patients with ulcerative colitis (VEGA): a randomised, double-blind, controlled, phase 2, proof-of-concept trial. *Lancet Gastroenterol Hepatol.* 2023;8(4):307–320

Human Cell Atlas



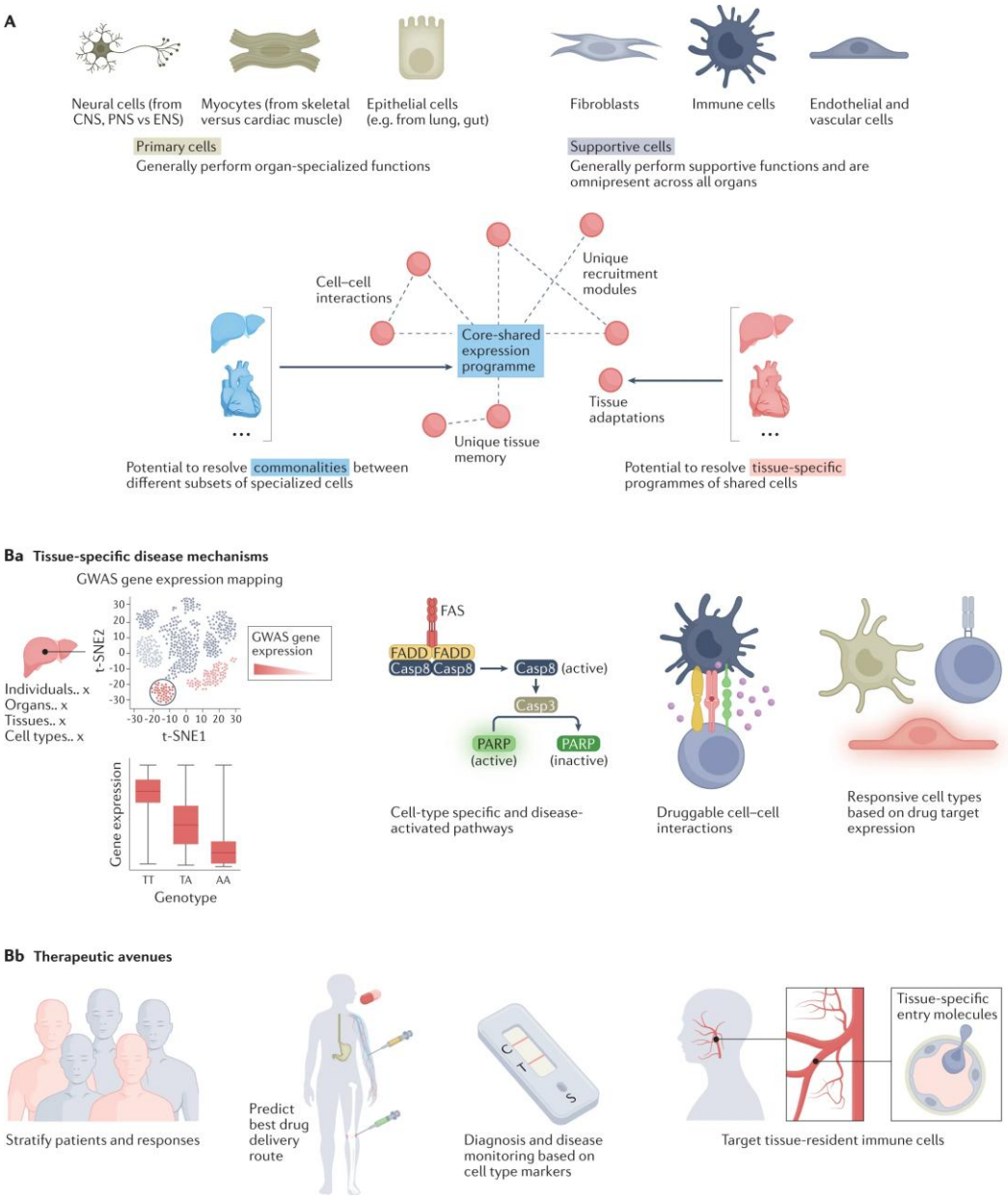
https://www.humancellatlas.org/

Single-Cell Atlas



Aldridge S, Teichmann SA.Nat Commun. 2020 Aug 27;

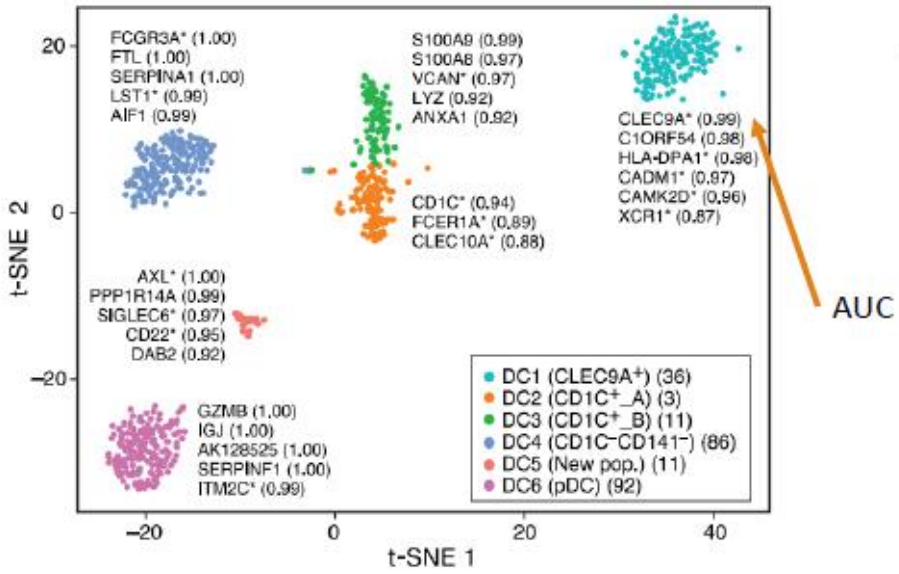
Single-Cell Cross-Tissue Comparisons



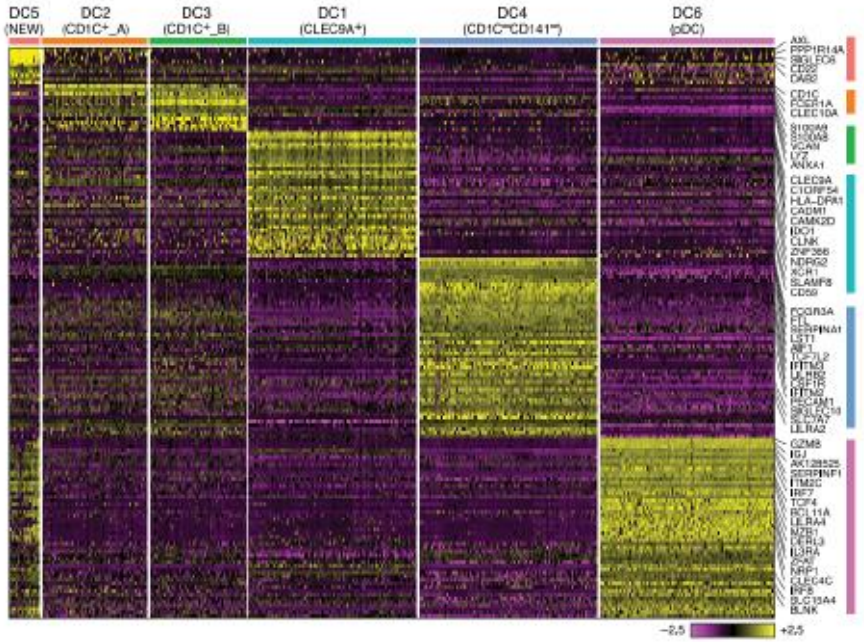
Elmentaite R, Domínguez Conde C, Yang L, Teichmann SA.Nat Rev Genet. 2022 Feb 25.

Immune System Heterogeneity: Reclassification of DCs and monocytes by scRNA-Seq

Enrich and sort cells of interest → Generate counts and QC the data → Identify clusters of cells that are similar to each other → Find discriminative markers per cluster → Isolate cells using newly identified markers → Validate the identify of the cells using scRNA-seq → Confirm the identity of these cells in independent individuals → Perform functional analysis of selected types



AUC

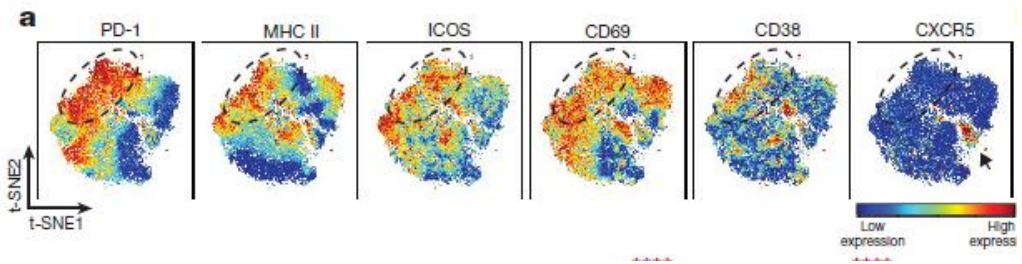


2.400 HLA-DR^{+}lineage^{-} cells

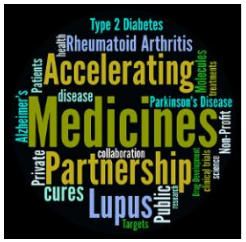
Villani et al. Science 21 Apr 2017

Novel pathological cell types: expanded peripheral Th subsets in RA

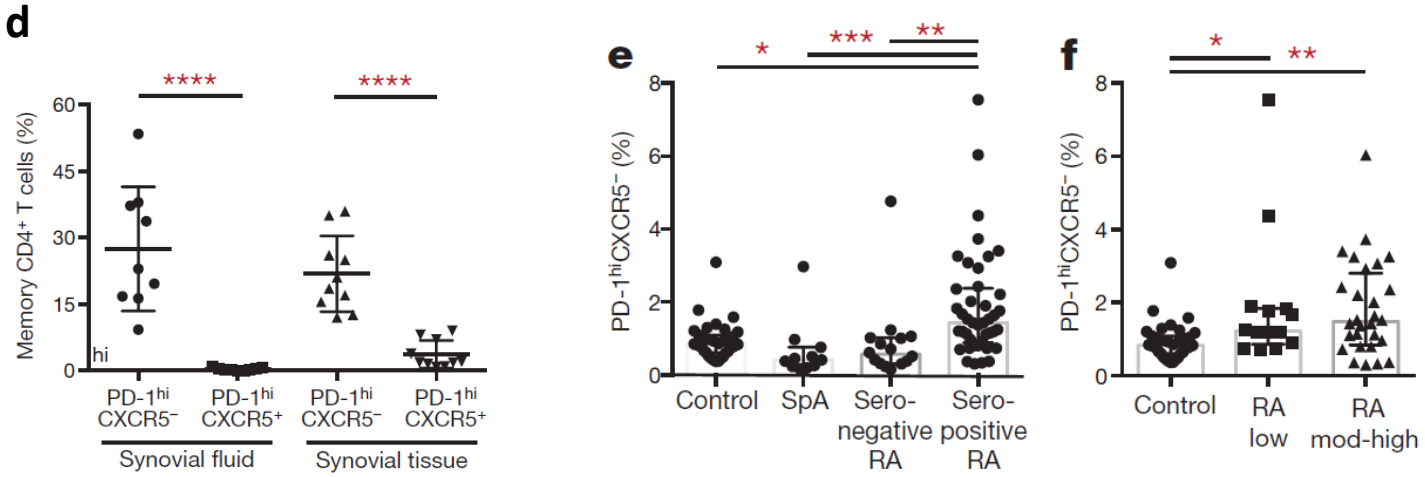
PD-1^{hi} CXCR5^{-} CD4^{+}



- **Pathologically expanded peripheral T helper cell subset drives B cells in RA**
- **Expression of chemokine receptors that direct migration to inflamed sites, such as CCR2, CX3CR1, and CCR5**
- **Synovial PD-1hi CXCR5−CD4+ T cells express factors associated with B-cell help (ICOS, IL21, MAF).**

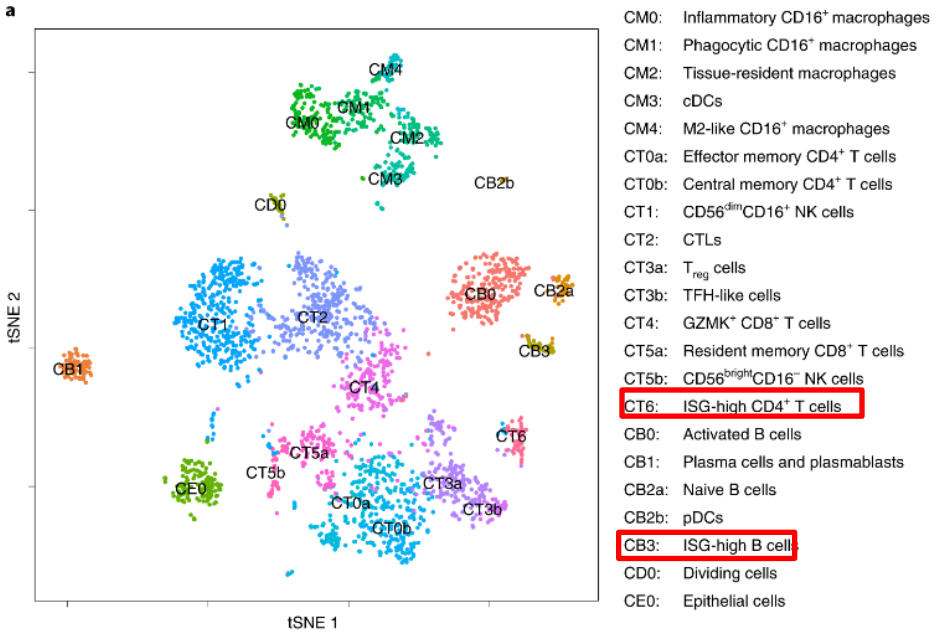


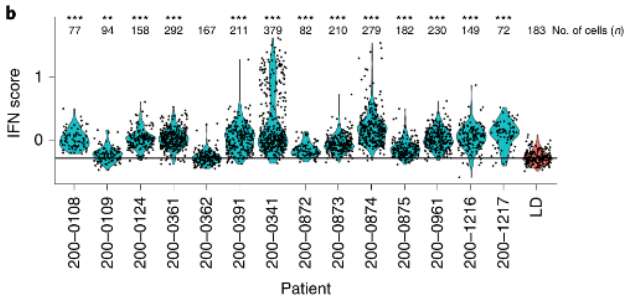
synovial *blood*

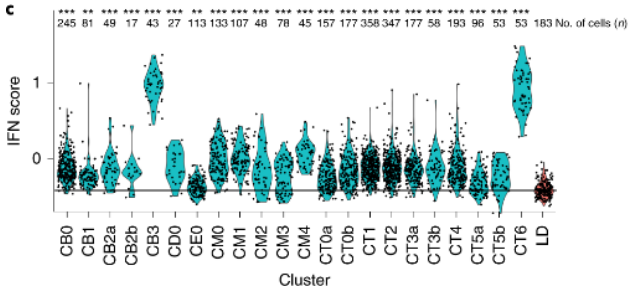


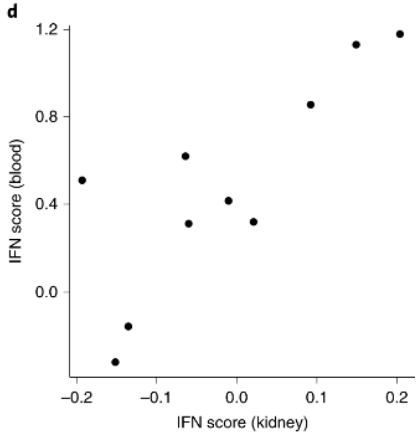
Rao et al. - Nature. 2017 Feb 1;542(7639)

Immune cell landscape in kidneys of patients with lupus nephritis









- First single-cell dissection of LN
- 24 patients, 10 ctrls
- 21 subsets of leukocytes, including clusters of myeloid cells, T-cells, NK, B-cells
- CXCR4, CX3CR1 broadly expressed
- Use of urine liquid biopsies and kidney samples

Arazi et al. Nat Imm 2019

DATA AGGREGATION AND VISUALIZATION

Data Dimensionality

- Gene/protein expression
- Methylation
- Epigenetics markers
- Genetics SNPs
- Metabolomics
- …..

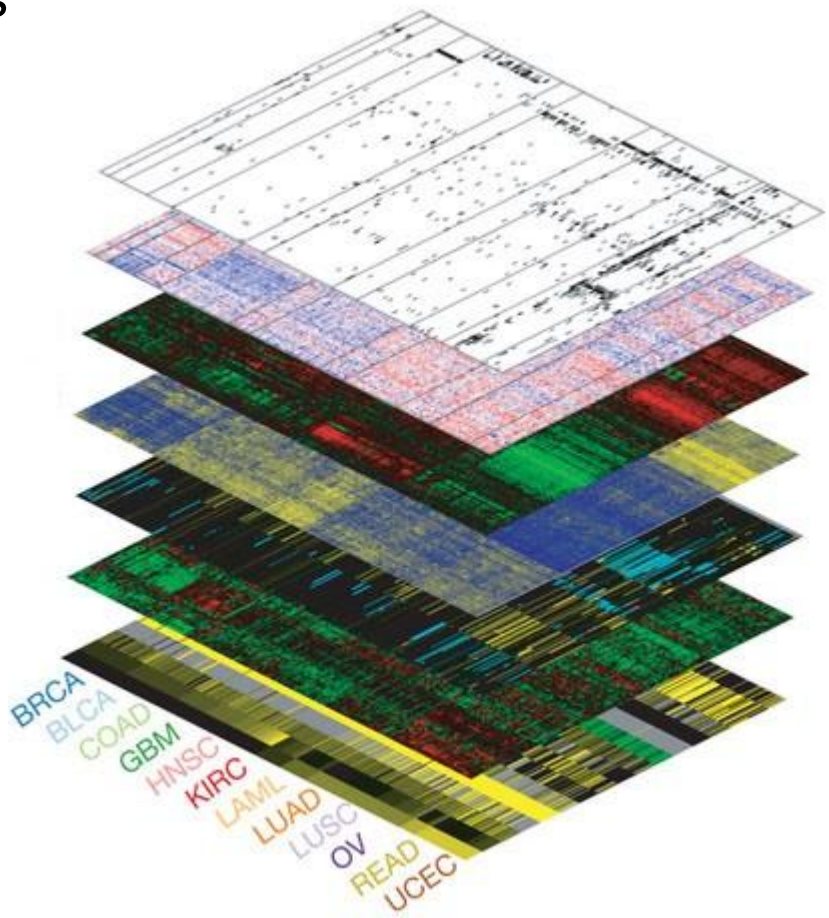
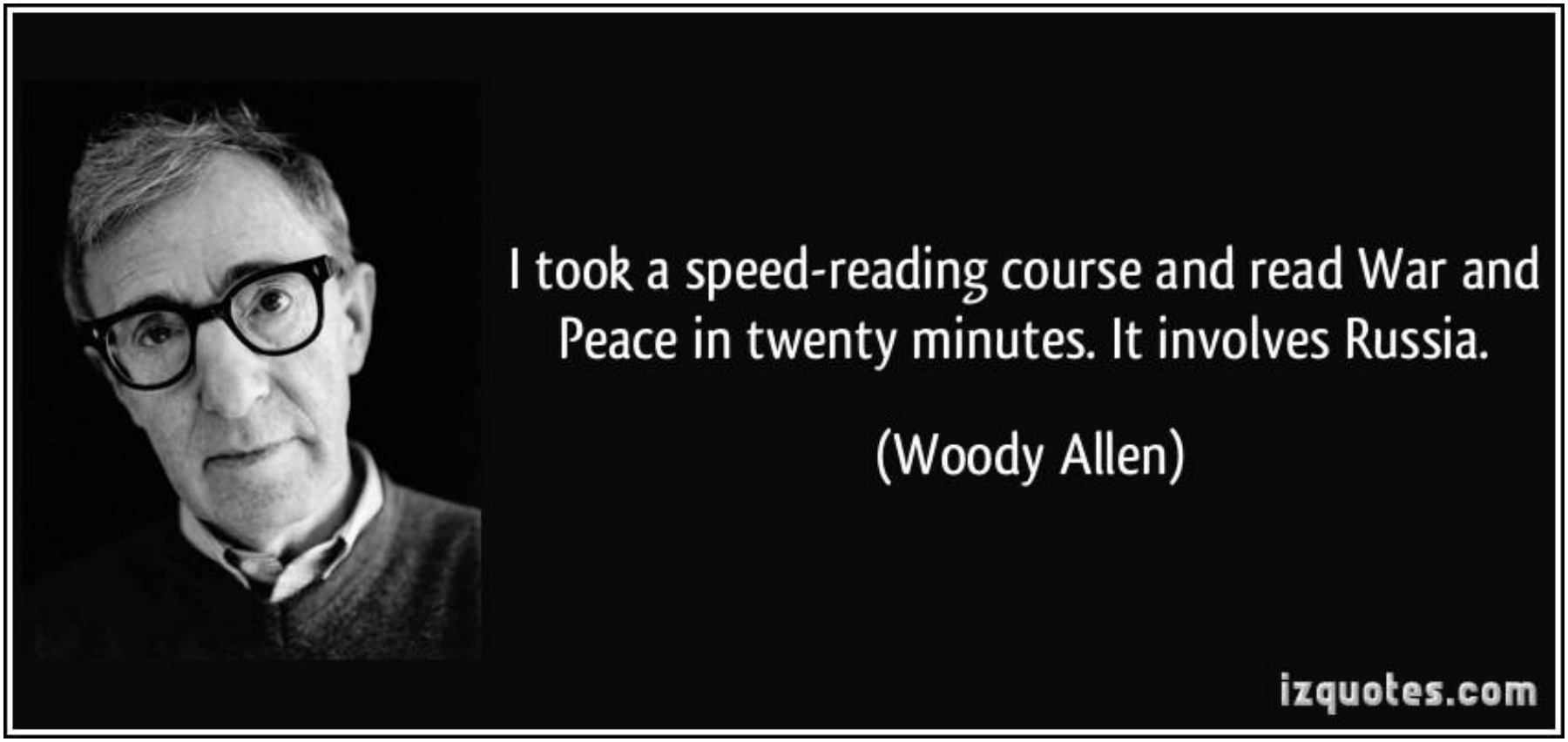


image credit: Nature http://www.nature.com/ng/journal/v45/n10/fig_tab/ng.2764_F1.html

Data aggregation is any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis



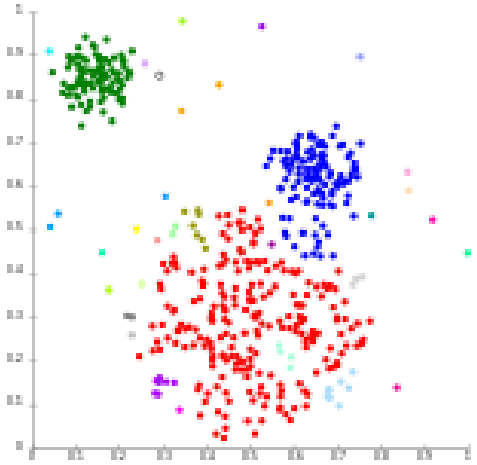
Data Aggregation

- Clustering and Geometrical Representation of Data
- Dimensions Reduction
- Pathways and Gene Sets

Clustering

Finding a partition such that:

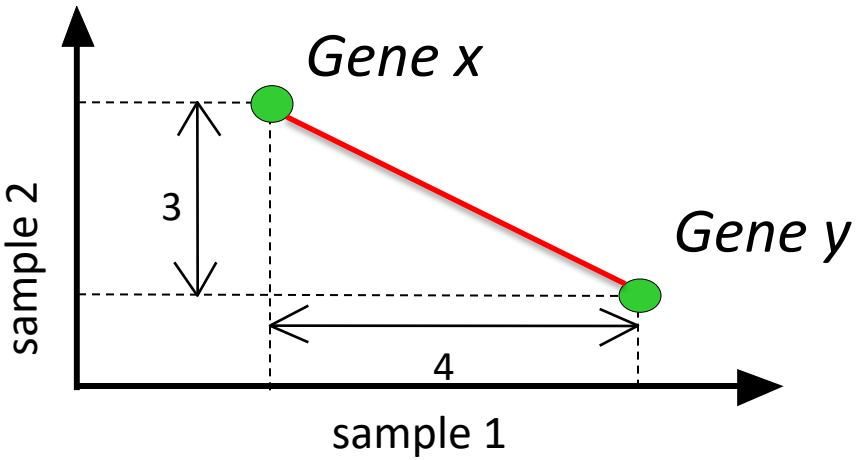
- Distance between objects within the same cluster is **minimised**
- Distance between objects from different clusters is **maximised**



Requires defining a <u>similarity measure</u>

Geometrical Distances as measures of similarity

	Sample 1	Sample 2
Gene X	2	3
Gene Y	5	1



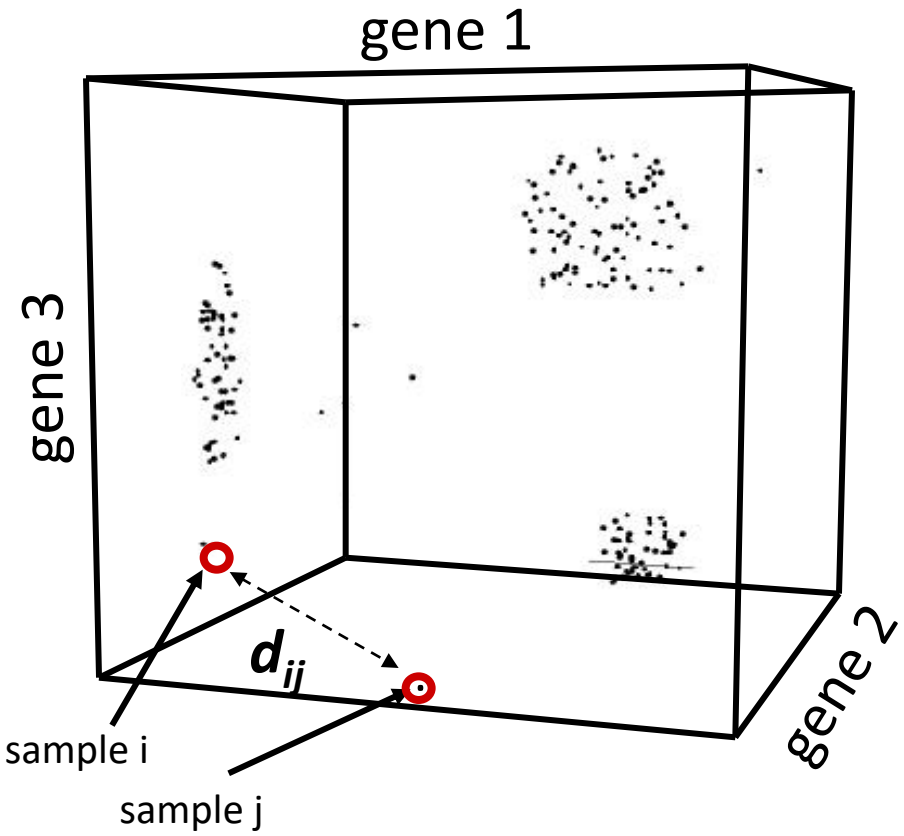
- Euclidean distance : $\sqrt[2]{4^2 + 3^2} = 5$
- Manhattan distance : $4 + 3 = 7$
- "sup" distance : $\max\{4,3\} = 4$

• Similarity among genes/samples is expressed as a mathematical distance

• Genes/samples close in the “expression space” have similar expression profiles

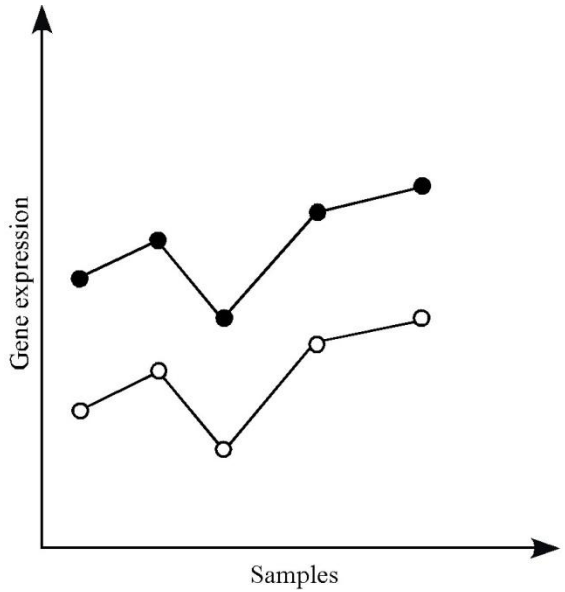
Geometrical Distances as measures of similarity

	Sample i	Sample j
Gene 1	2	3
Gene 2	5	1
Gene 3	7	4

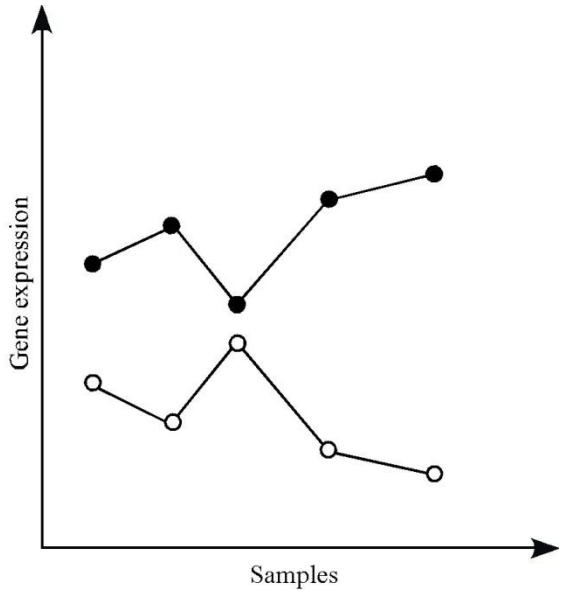


- N genes = N dimensions
- each sample can be represented as a point in the N-dimensional space

Similarity based on correlation



positive correlation



negative correlation

$$\sigma(x, y) = \mathrm{E}\left[(x - \mathrm{E}[x])(y - \mathrm{E}[y])\right],$$

$$\text{- correlation distance} : \frac{\mathrm{cov}(a,b)}{std(a) \cdot std(b)}$$

Unsupervised Methods: Hierarchical Clustering

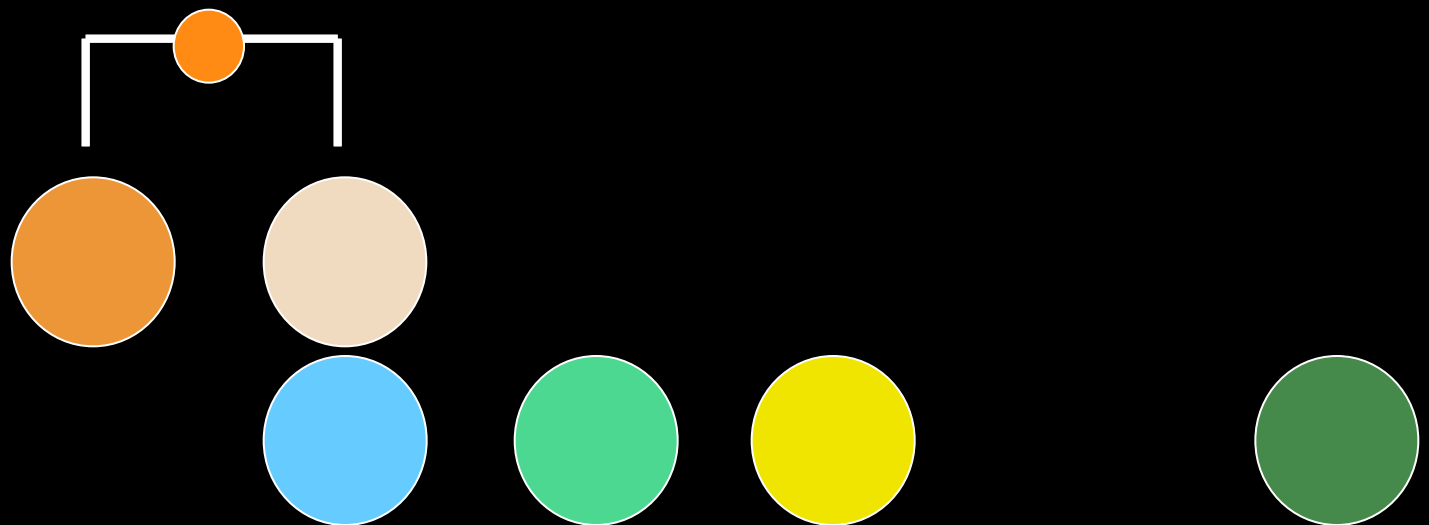
At the beginning, each object (gene) is a cluster. In each of the subsequent steps, the two *closest* clusters are merged into one cluster until there is only one cluster left.

Hierarchical Clustering

Similarity distance: color

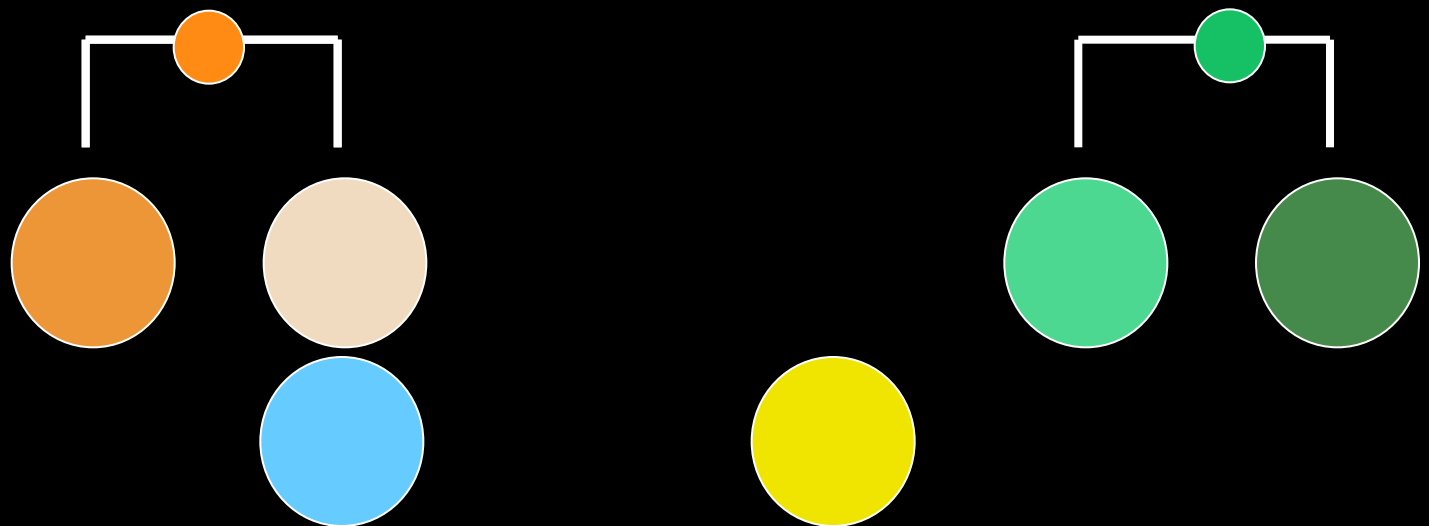
Hierarchical Clustering

Similarity distance: color



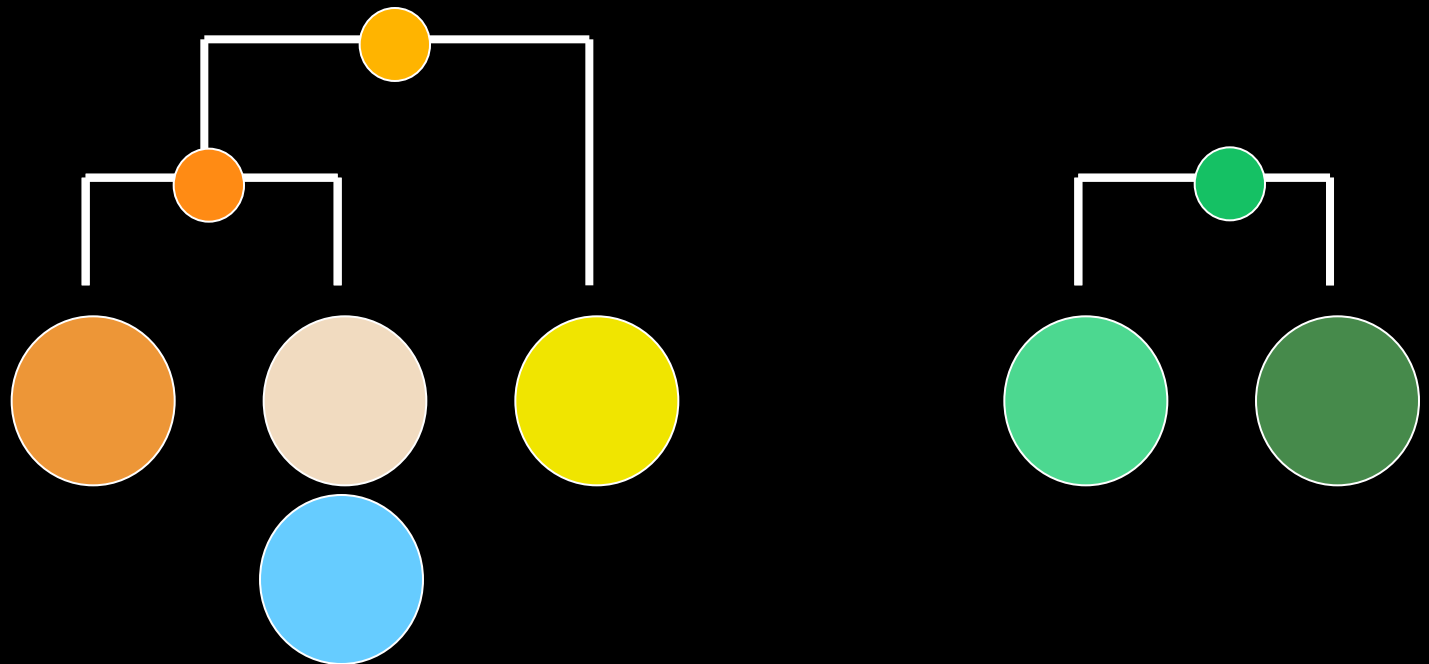
Hierarchical Clustering

Similarity distance: color



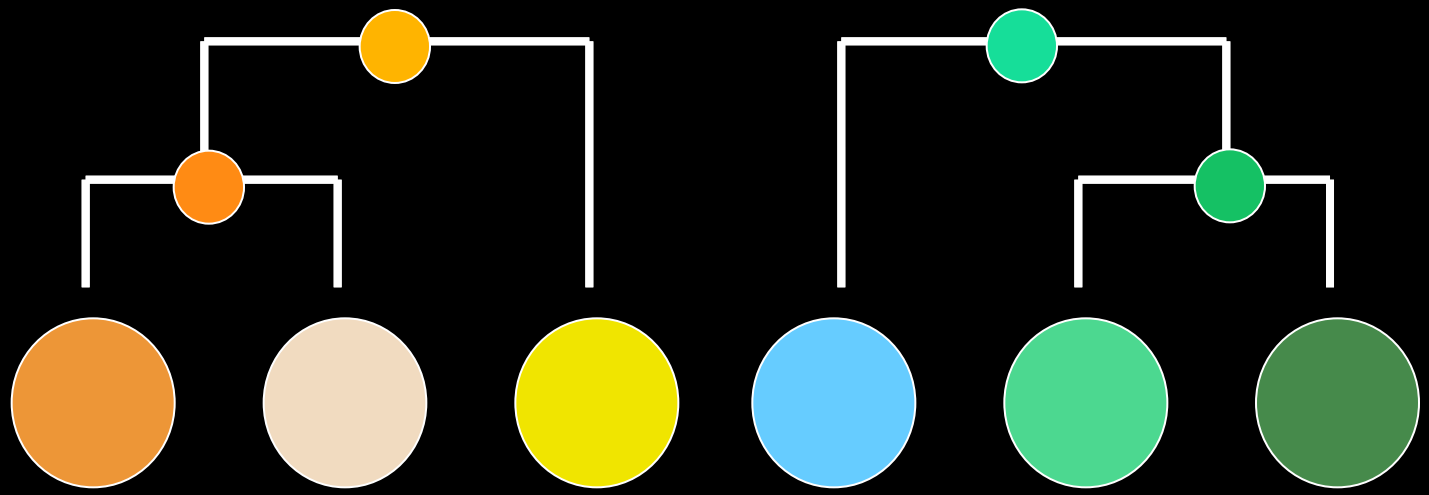
Hierarchical Clustering

Similarity distance: color



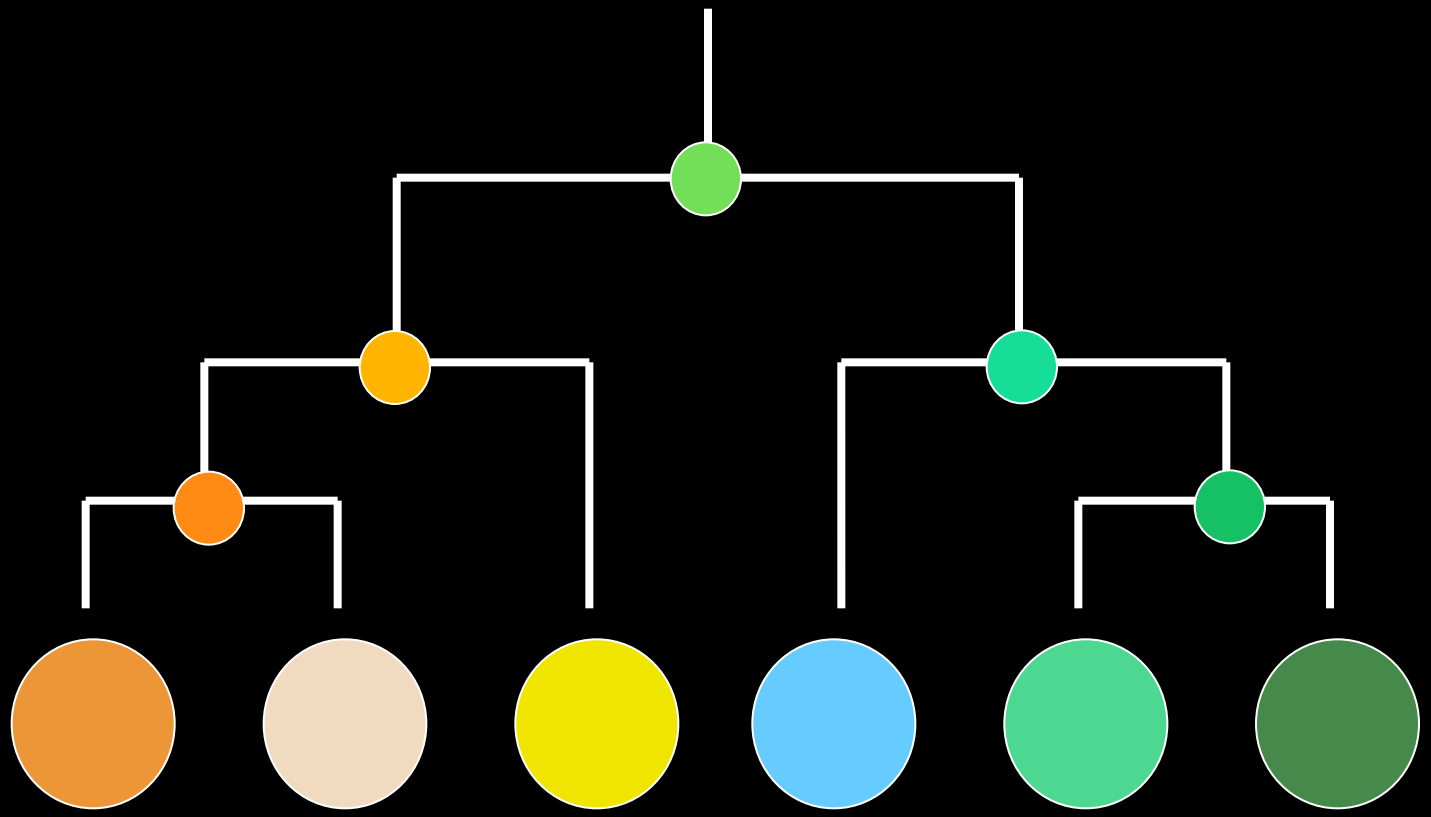
Hierarchical Clustering

Similarity distance: color

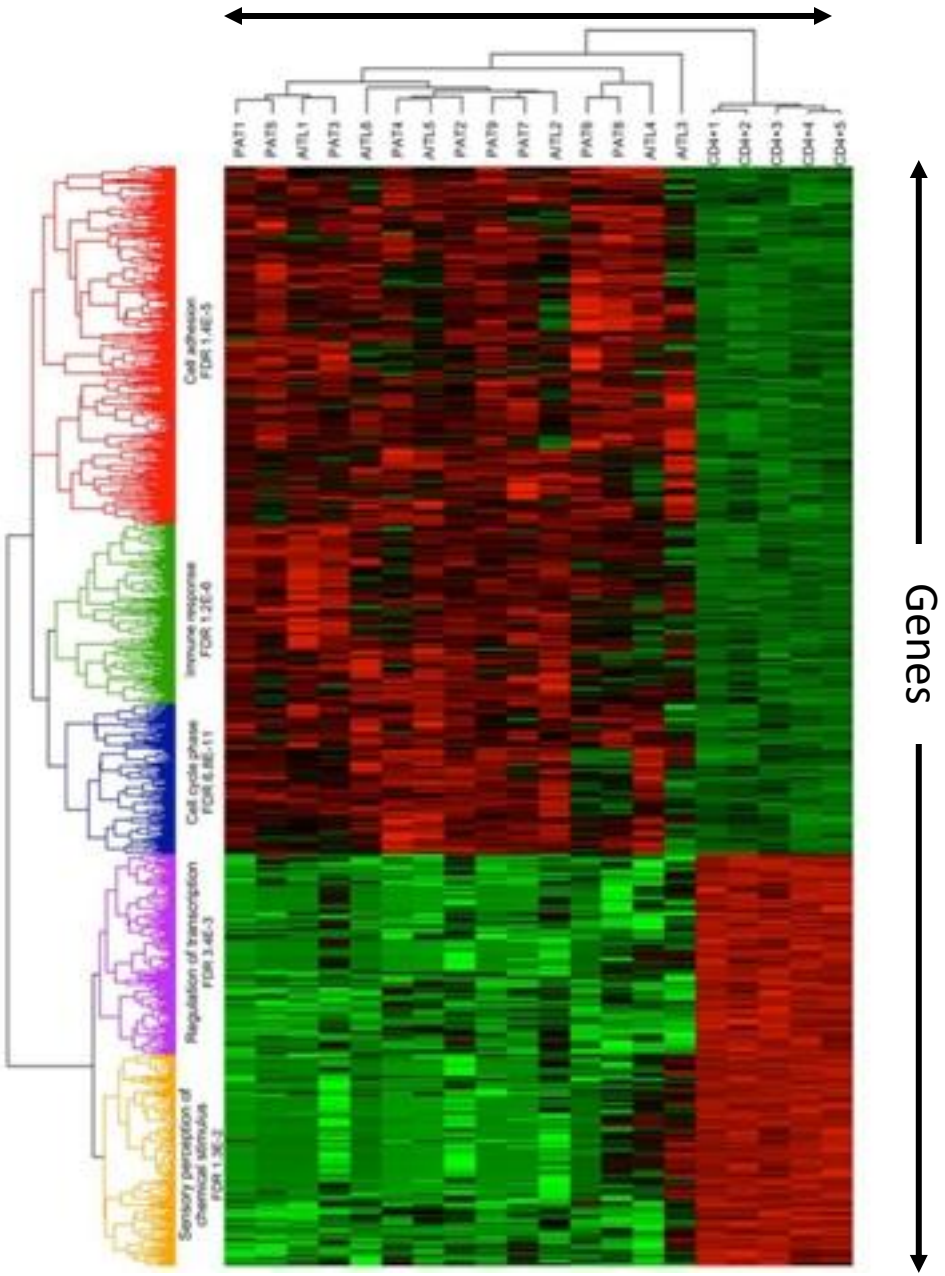


Hierarchical Clustering

Similarity distance: color



Samples



Genes

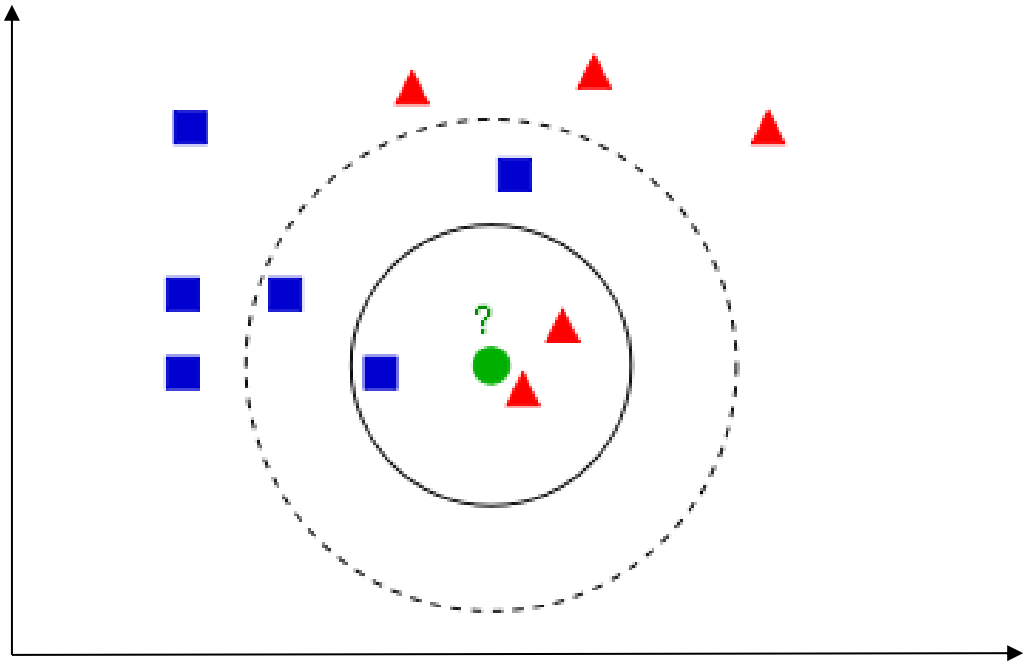
Supervised Clustering (Classification)

- use pre-existing biological information (e.g. tumor type, immune cell type, responders/non-responders etc.)
- Are used to infer which class an unknown sample belongs to
- Machine learning methods: **k-nearest neighbors,** SVM, Random forests, Bayesian networks, Deep Learning

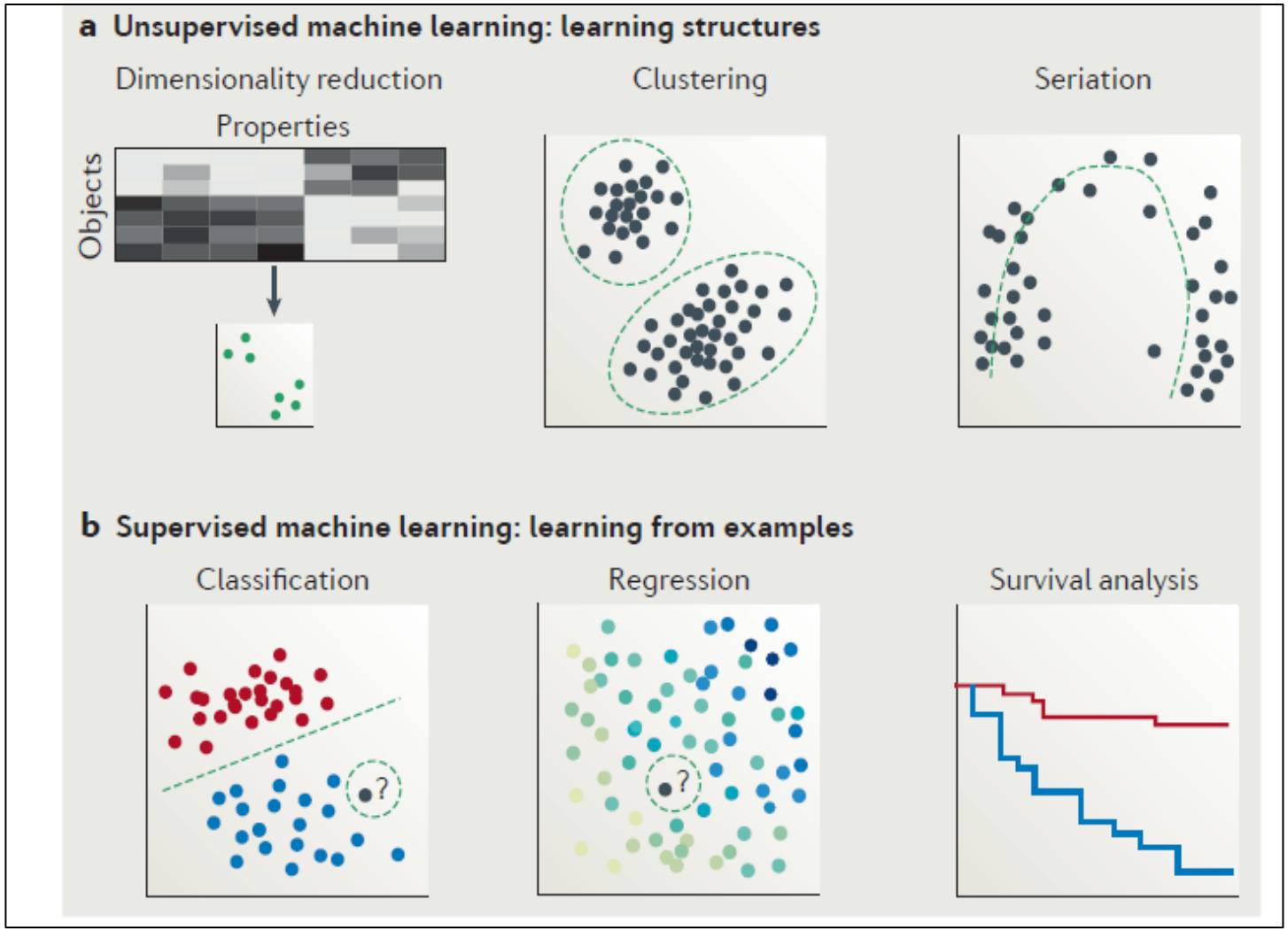
k-nearest neighbors classification (k-NN)

An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small).

If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

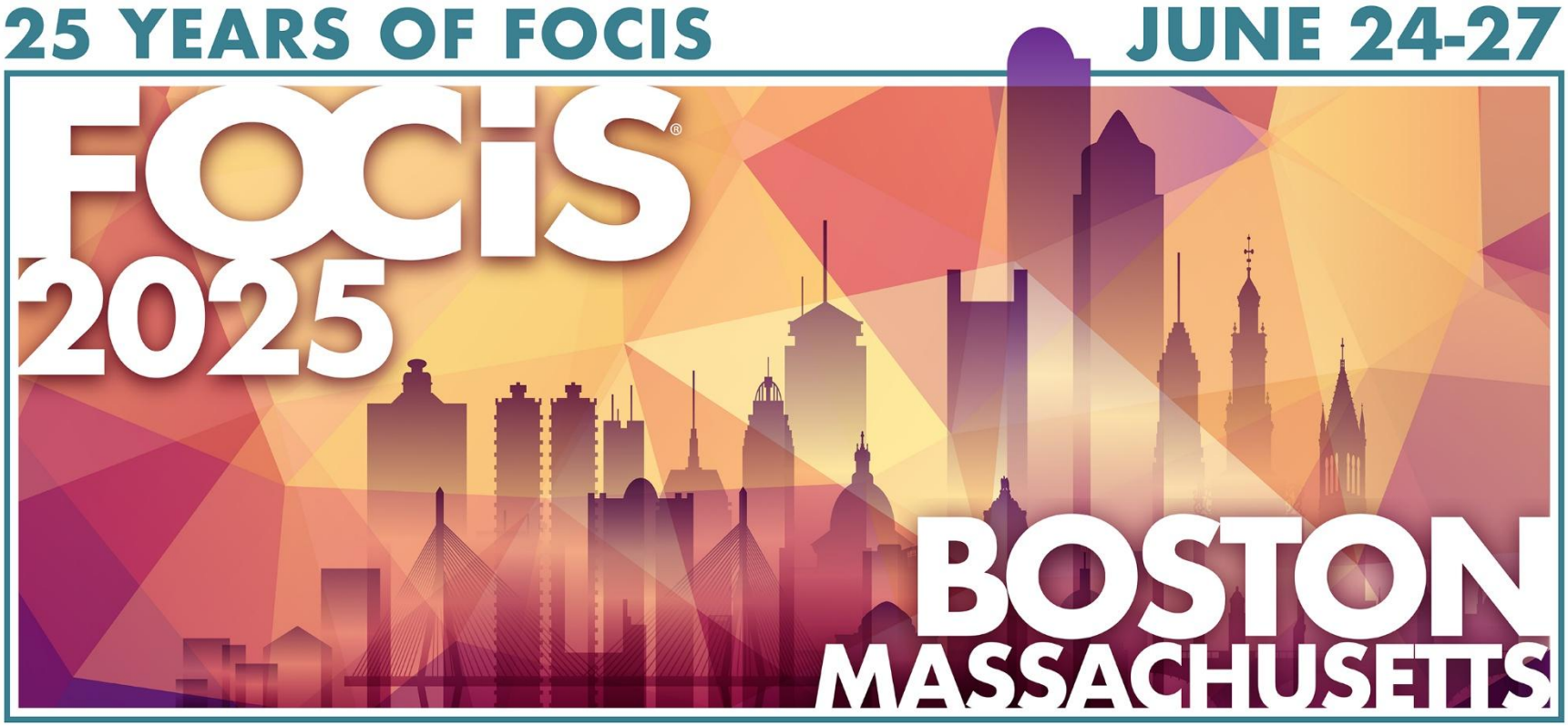


Supervised and Unsupervised Machine Learning



Nature Reviews Immunology 16, 449–462 (2016)

Q & A

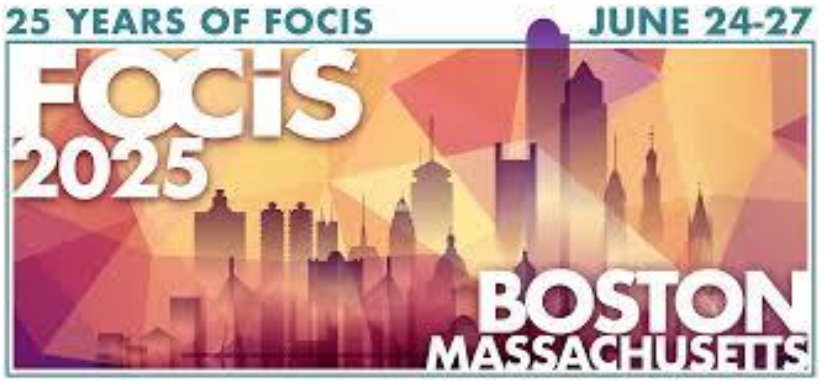

25 YEARS OF FOCIS
JUNE 24-27
FOCIS
2025
BOSTON
MASSACHUSETTS

Deep Dive into Case Studies:

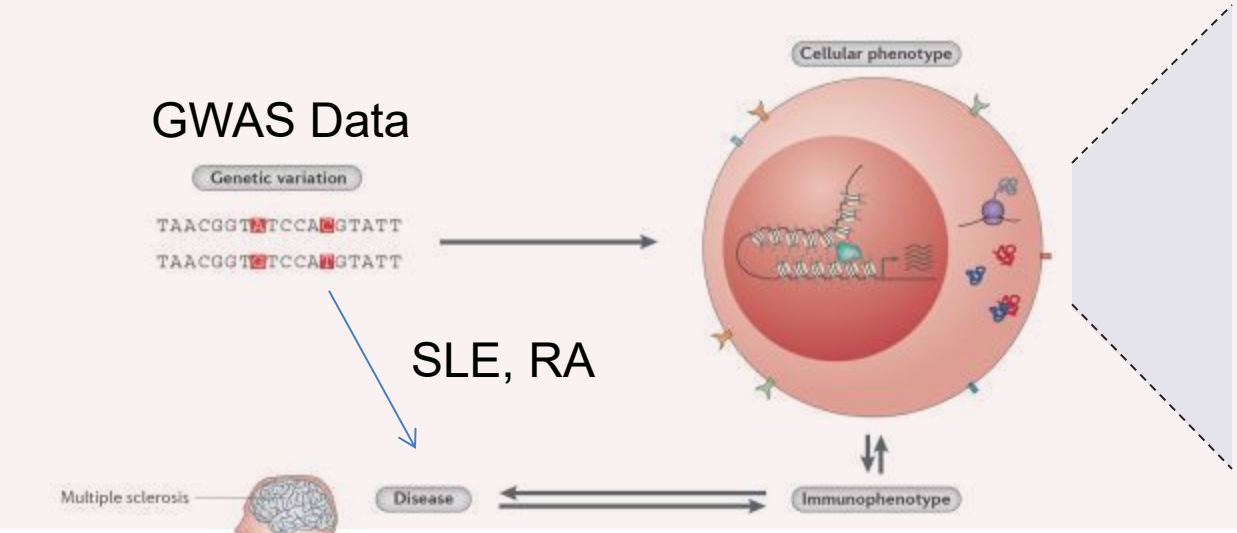
Systems Immunology to Novel Therapeutic Insights

Emanuele de Rinaldis

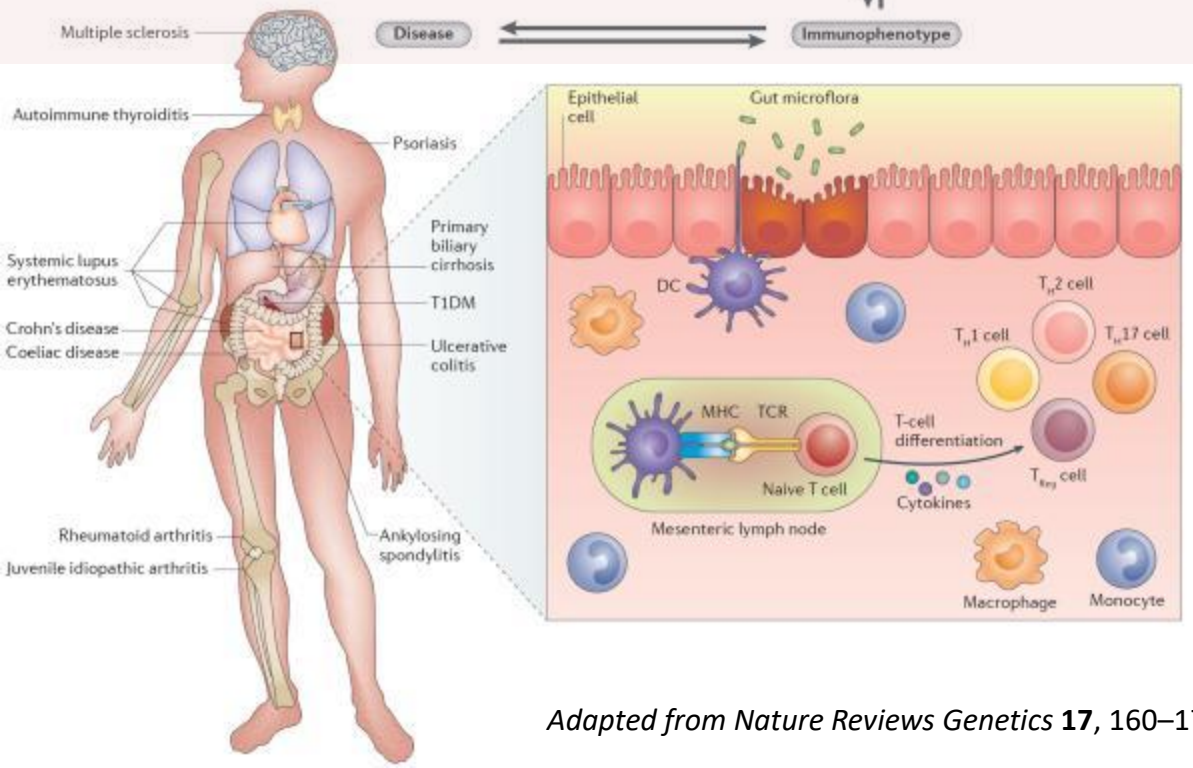
Workshop in Systems Immunology
June 23rd, 2025



Exploring connections between Genetics, Immune Phenotypes and Clinical Phenotypes

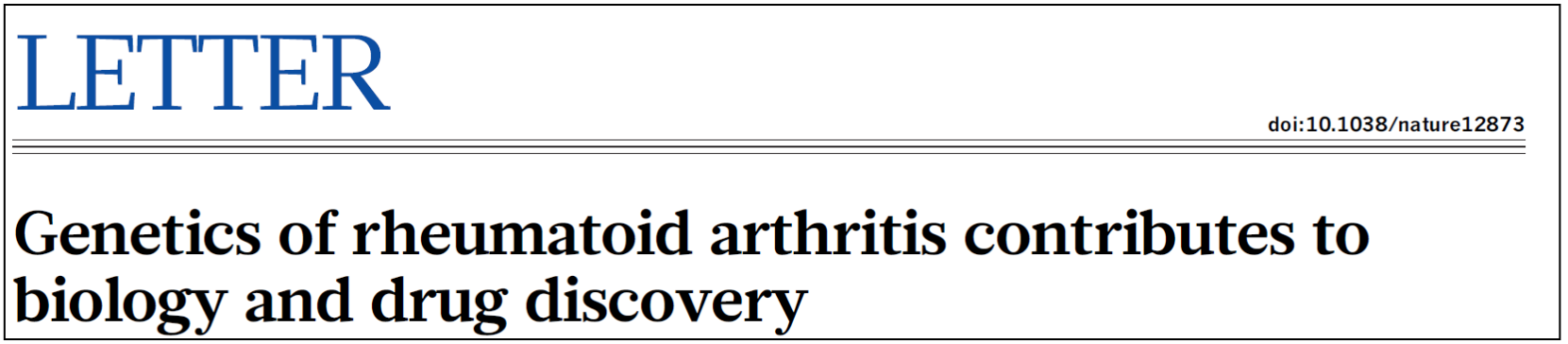


- Gene transcription
- Pathway activation
- Histone modifications
- protein–protein interactions



Adapted from Nature Reviews Genetics **17**, 160–174 (2016)

Leveraging on integration of orthogonal data sets to identify genes of therapeutic interest in RA (R. Plenge's)



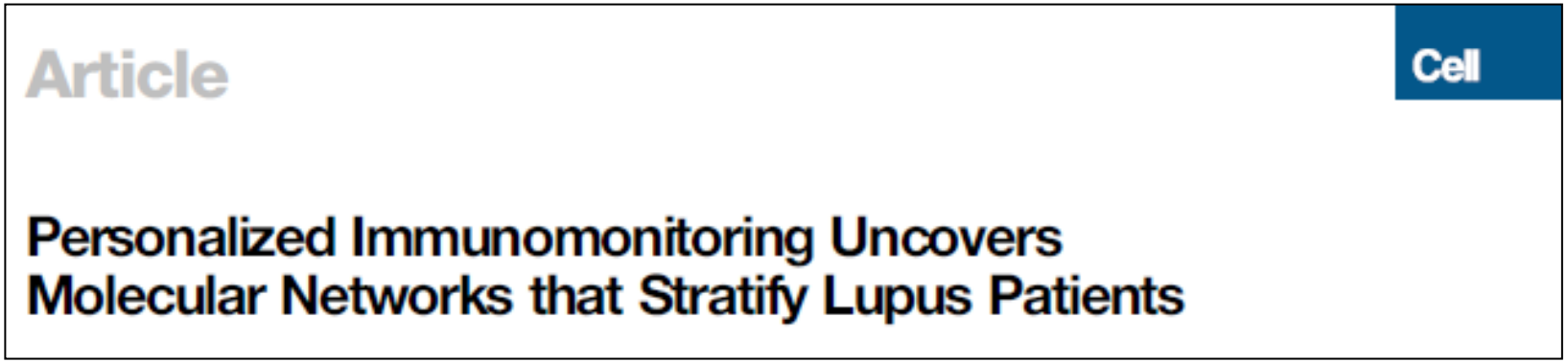
LETTER

doi:10.1038/nature12873

Genetics of rheumatoid arthritis contributes to biology and drug discovery

Nature. 2014 Feb 20;506(7488):376-8

Understanding SLE biology and stratifying patients using blood bulk gene expression data (V. Pascual's)



Article

Cell

Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients

Cell. 2016 Apr 21;165(3):551-6

Genetics and Drug Discovery in RA – Study Workflow

LETTER

doi:10.1038/nature12873

Genetics of rheumatoid arthritis contributes to biology and drug discovery

Identification of SNPs associated to RA

From SNPs to causal genes through gene-mapping

Characterization of results

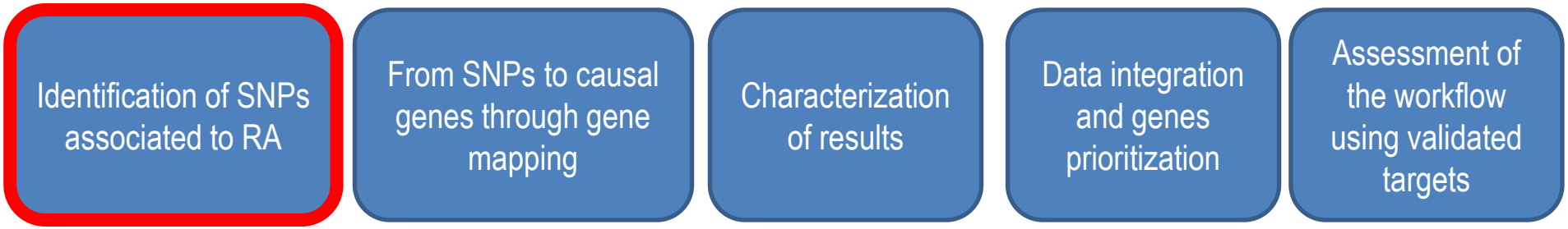
Data integration and genes prioritization

Assessment of the workflow using validated targets

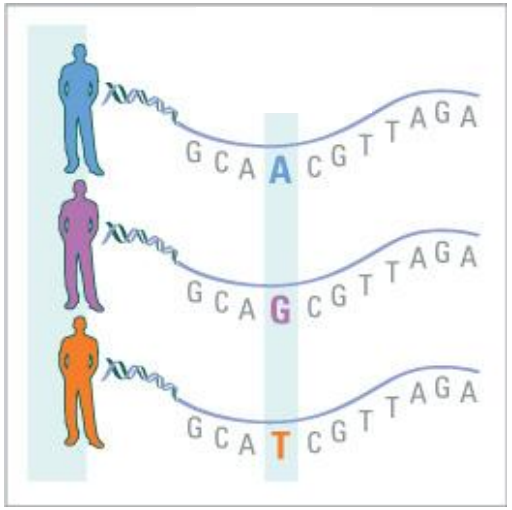
- Novel loci associated to RA
- New hints on disease biology
- Novel candidate targets
- Repositioning of existing drug targets

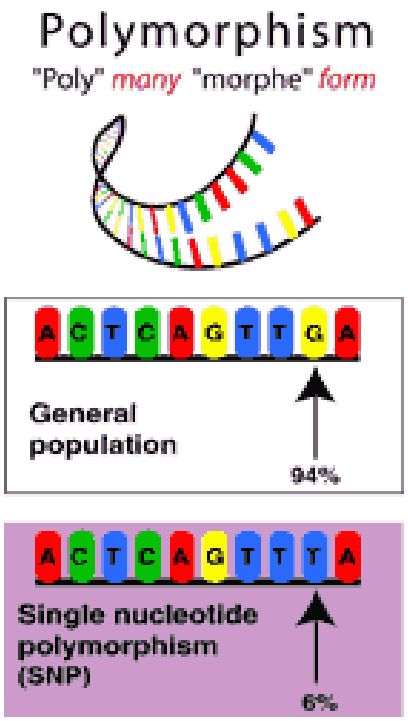
Toolbox

- **Genome Wide Association Studies (GWAS) and Meta-Analysis**
  - Genome variability and SNPs
  - Logistic Regression
  - Linkage Disequilibrium
  - Imputation
  - Manhattan Plots
- **Multiple Testing**
- **Network Analysis**
- **Fine-mapping and data integration**
  - Epigenetics data
  - Transcriptional data ➔ eQTLs
- **Statistical enrichment**

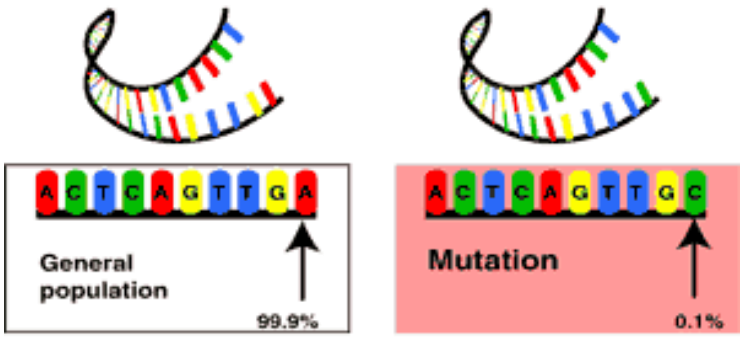

Identification of SNPs associated to RA
From SNPs to causal genes through gene mapping
Characterization of results
Data integration and genes prioritization
Assessment of the workflow using validated targets

Genetic Variability





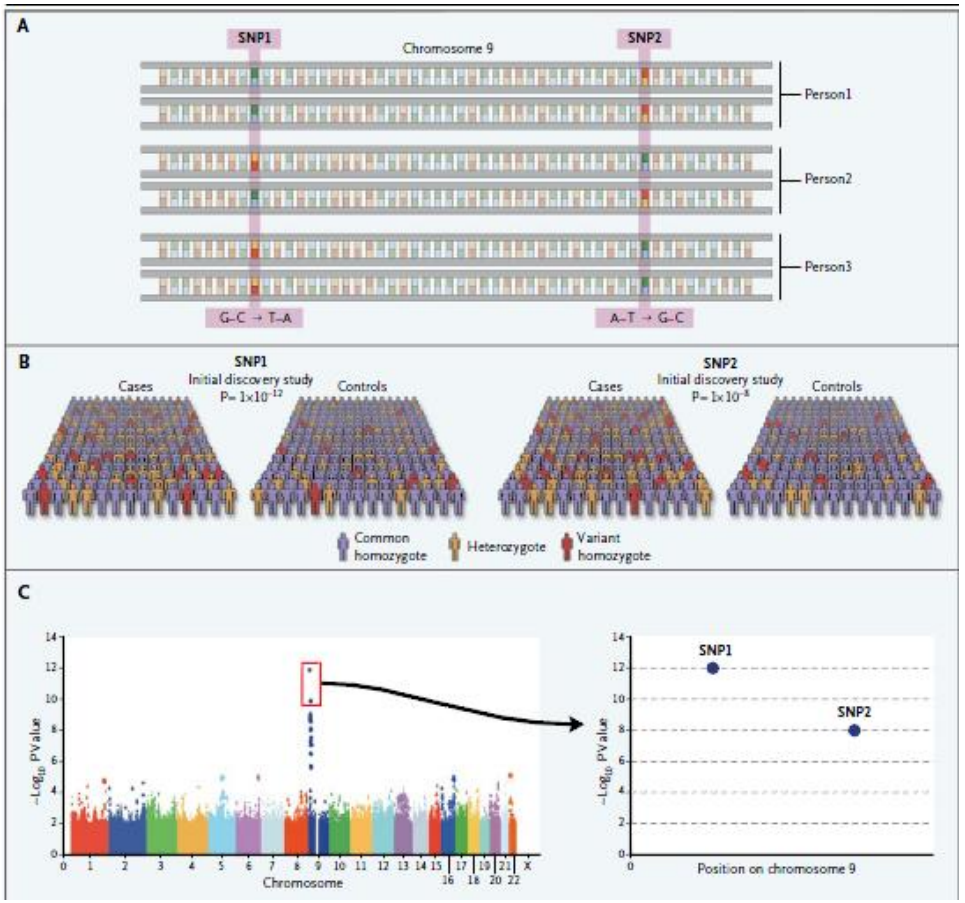
SNP: common (>1%) variant of one **S**ingle **N**ucleotide



Mutation: typically a rare variant, associated with a disease

Genome-Wide Association Studies (GWAS)

Extraction of germline DNA (e.g. blood)



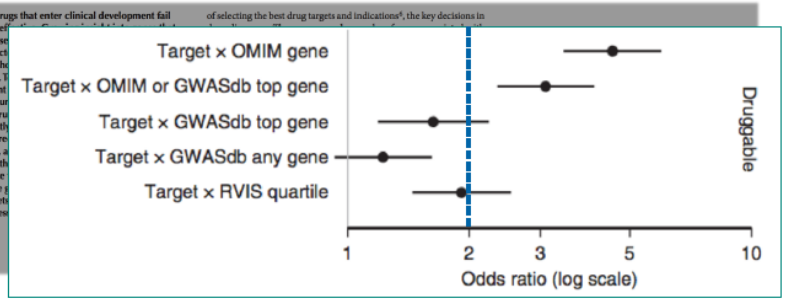
N Engl J Med. 2010 Nov 18;363(21):2076-7.

Genome-wide association study: An approach used in genetics research to look for associations between many - typically hundreds of thousands - specific genetic variations - most commonly single-nucleotide polymorphisms - and particular diseases

Why Pharma Is Investing in Genetics

The support of human genetic evidence for approved drug indications

~2-fold increase in success for genetic targets



RESEARCH ARTICLE

Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval

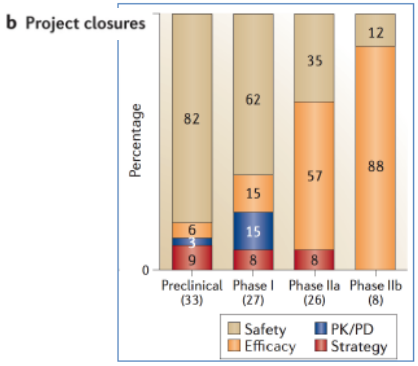
Emily A. King*, J. Wade Davis, Jacob F. Degner

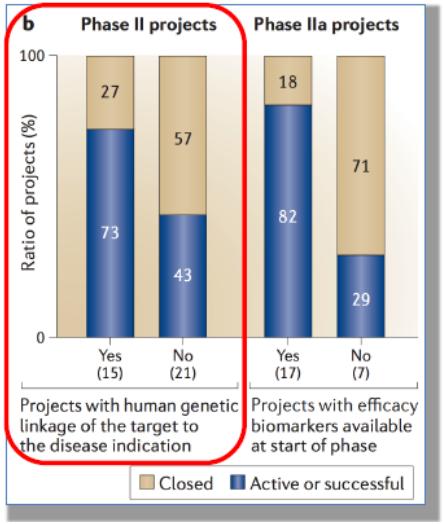
Department of Computational Genomics, AbbVie, North Chicago, Illinois, United States of America

* emily.king@abbvie.com

Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework

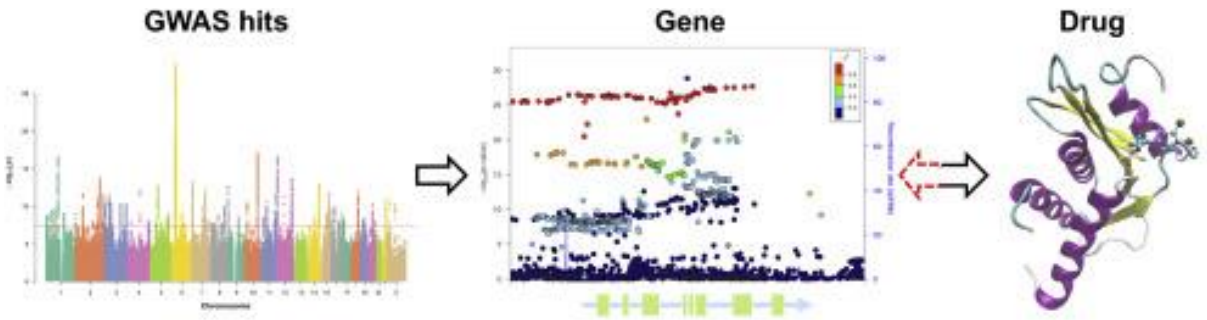
David Cook, Dearg Brown, Robert Alexander, Ruth March, Paul Morgan, Gemma Satterthwaite and Menelas N. Pangalos





- When causality is clear (Mendelian traits or coding variants) fold increase for success greater than 2 folds
- Limited contribution to GWAS genetic evidences not in OMIM ➔ undetermined function

Lessons Learnt from Human Genetics



Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	*SLC30A8/KCNJ11*	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	*PADI4/IL6R*	BB-Cl-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	*TNFR1/PTGER4/TYK2*	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	*IL23A*	Risankizumab
Osteoporosis	*RANKL/ESR1*	Denosumab/Raloxifene and HRT
Schizophrenia	*DRD2*	Anti-psychotics
LDL cholesterol	*HMGCR*	Pravastatin
AS, Ps, Psoriatic Arthritis	*IL12B*	Ustekinumab

Visschner et al. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *AJHG*

sanofi

Genetic Dataset

SNPs

Samples

0	1	…	1	1
1	2	…	2	0
0	1	…	1	2
2	2	…	2	2
0	0	…	0	1
0	1	…	1	2
0	2	…	2	1
0	1	…	1	1
0	2	…	1	1
2	2	…	2	2
0	0	…	0	1
0	1	…	1	2

Number of minor alleles for a given SNP

AA=0
aA or Aa = 1
aa=2

Single test for association

Do people carrying a certain genotype have an increased probability of having the disease?

SNP	S_IBD1	S_IBD2	S_IBD3	S_Cont1	S_Cont2	S_Cont3
rsxxxxx	0	0	1	0	1	0
rsxxxx0	0	0	0	0	0	0
rs...	1	0	0	0	2	1
rs...	0	0	0	1	0	1
rs...	0	0	1	1	1	1
PC1	..					
PC2			..			
PC ..						..

$$Log(\frac{p}{1-p}) = \beta_0 + \beta_1 snp + \beta_2 pc1 + \beta_3 pc2 + \beta_4 pc3 + \ldots$$

The tool used was "plink", What we get back is:
the value of B1, and
a p-value for B1 being different than zero.

11

Other variable of interest like ethnicity, age, sex can be added to the model

Logistic regression

Multiple tests in the same cohort

Test all SNPs for association independently

Correct for multiple testing

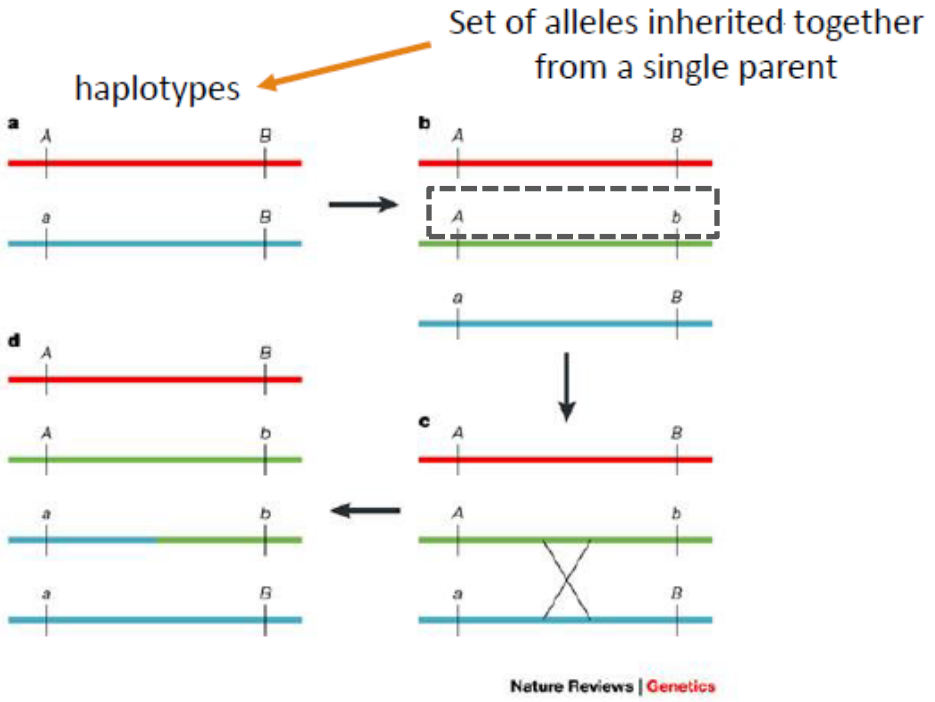
Instead of using 0.05 as a threshold for significance divide it by the total number of independent tests ($0.05/1M=5x10^{-8}$ for genome-wide studies)

Linkage Disequilibrium

- early in the lifetime of the mutation, only three out of the four possible haplotypes will be observed in the population. **The b allele will always be found on a chromosome with the A allele at the adjacent locus**
- The association between alleles at the two loci will gradually be disrupted by recombination between the loci.
- This will result in the creation of the fourth possible haplotype and an eventual decline in LD among the markers in the population as the recombinant chromosome

Non-random association of alleles at different loci (genetic positions)

Genotypes are not independent but correlated



K. G. Ardlie, L. Kruglyak, M. Seielstad, *Nature reviews Genetics* 3, 299-309 (2002)

Multiple tests in the same cohort

Test all SNPs for association independently

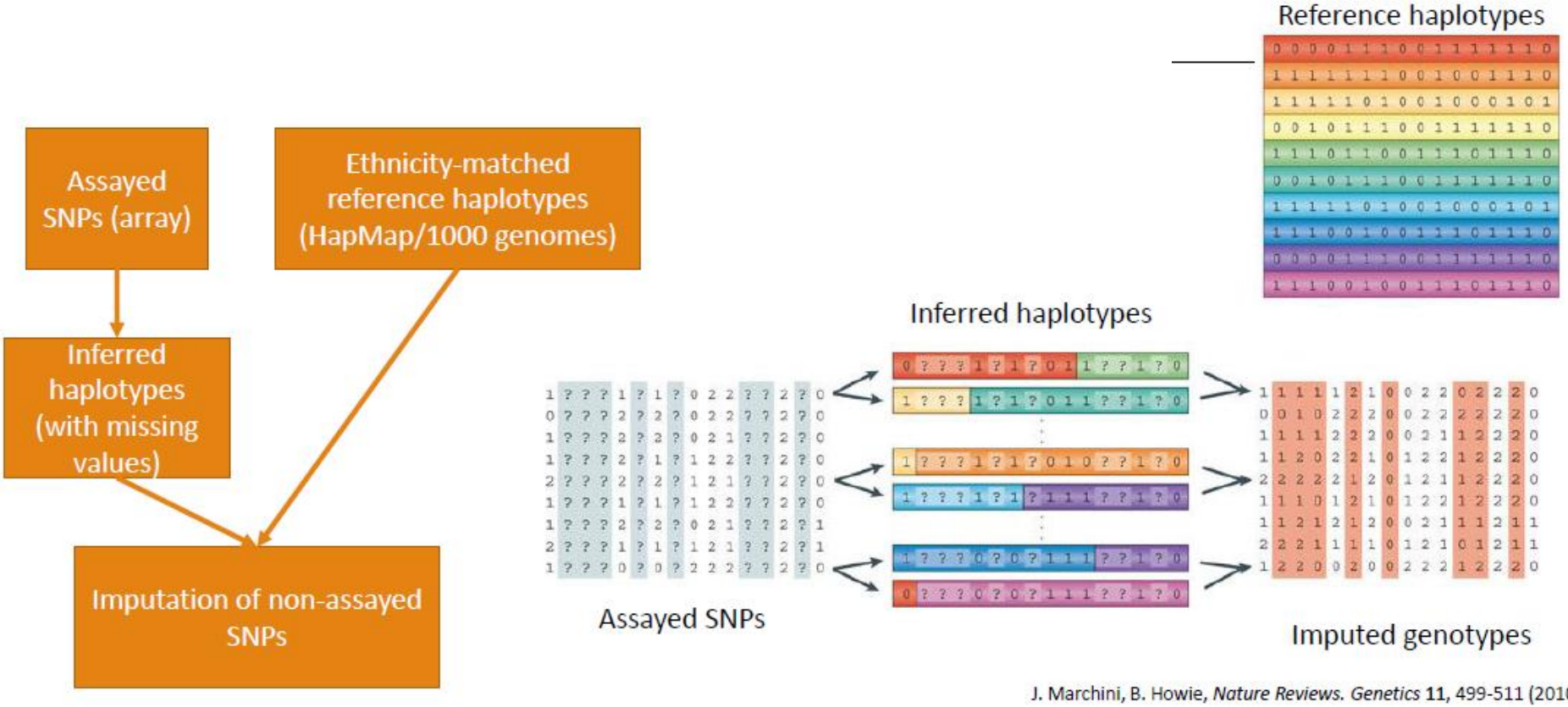
Correct for multiple testing

Divide usual threshold (0.05) by the total number of independent tests

Due to linkage disequilibrium the number of independent SNPs in the genome is estimated at ~1M

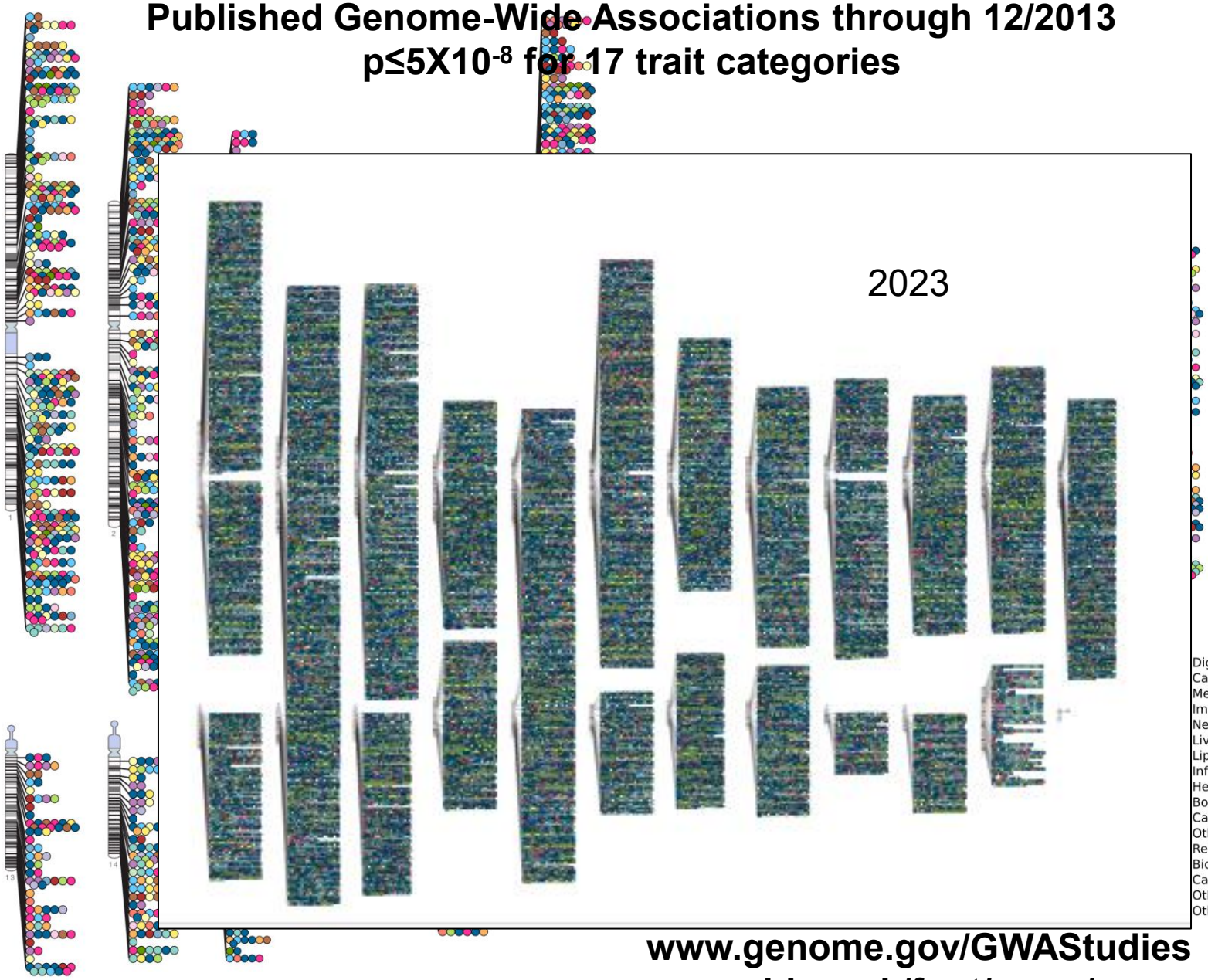
Genome-wide threshold = $5x10^{-8}$

Imputation



J. Marchini, B. Howie, *Nature Reviews. Genetics* **11**, 499-511 (2010)

Published Genome-Wide Associations through 12/2013
$p \leq 5X10^{-8}$ for 17 trait categories

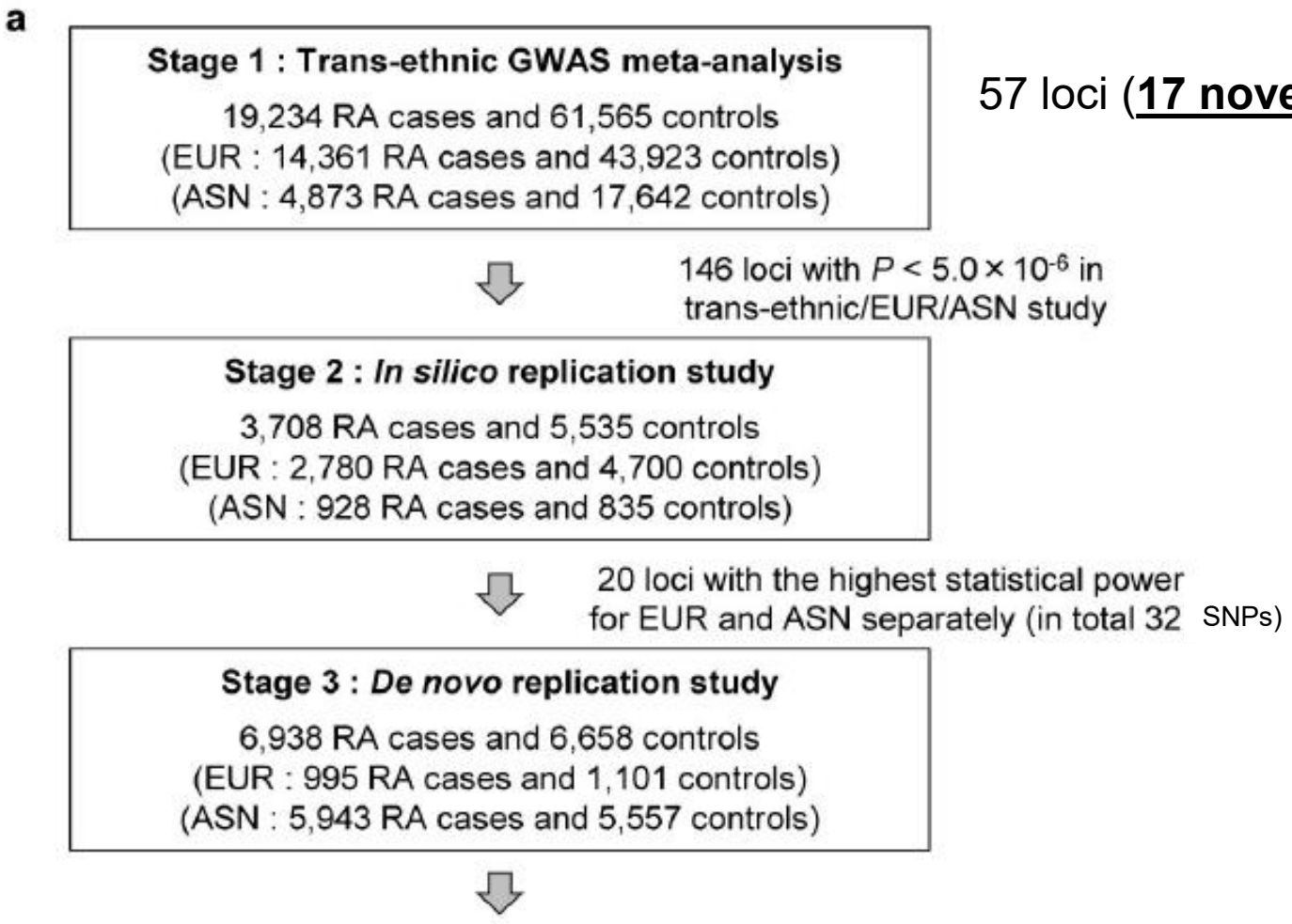


www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

EMBL-EBI

Study Design



57 loci (**17 novel**) pval<10-8

Combining 1-3: **42** novel loci with $P < 5 \times 10^{-8}$

Nature. 2014 Feb 20;506(7488):376-81

100 Total RA risk loci (58 known + 42 novel), including 377 genes

Identification of SNPs associated to RA

From SNPs to causal genes through fine-mapping

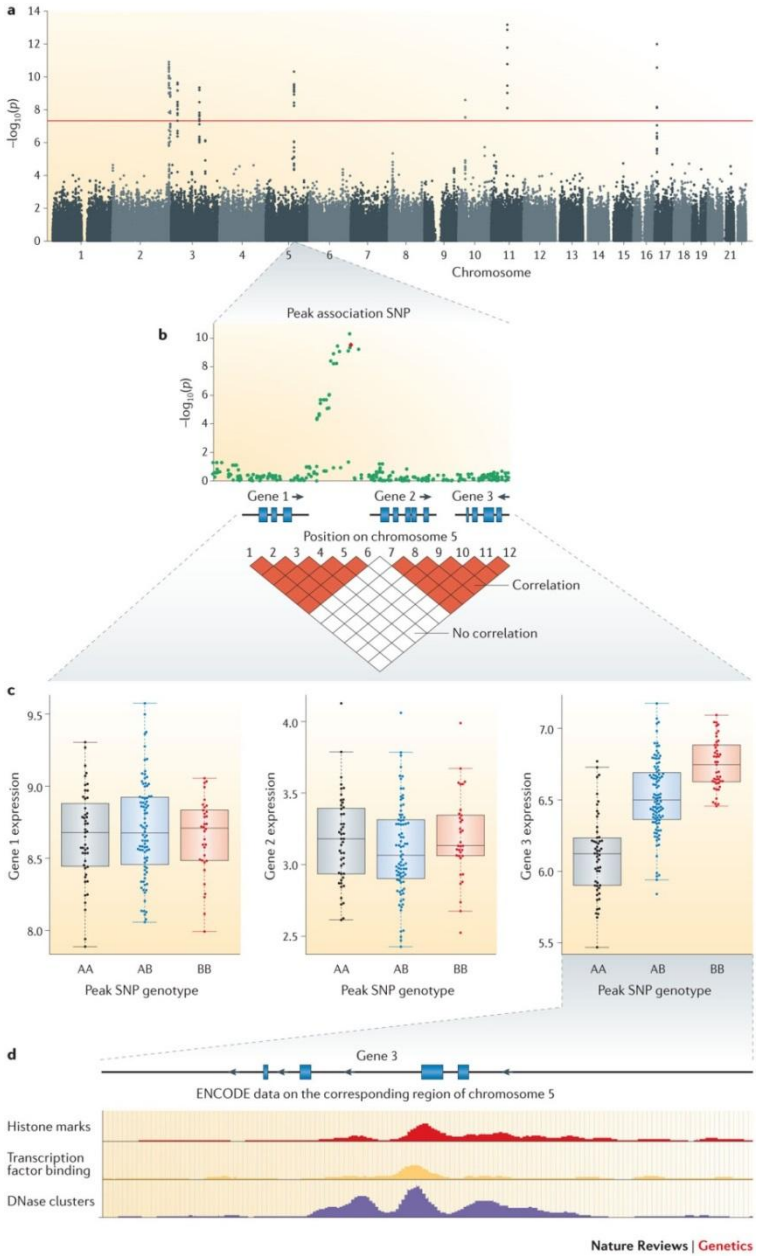
Characterization of results

Data integration and in-silico genes prioritization

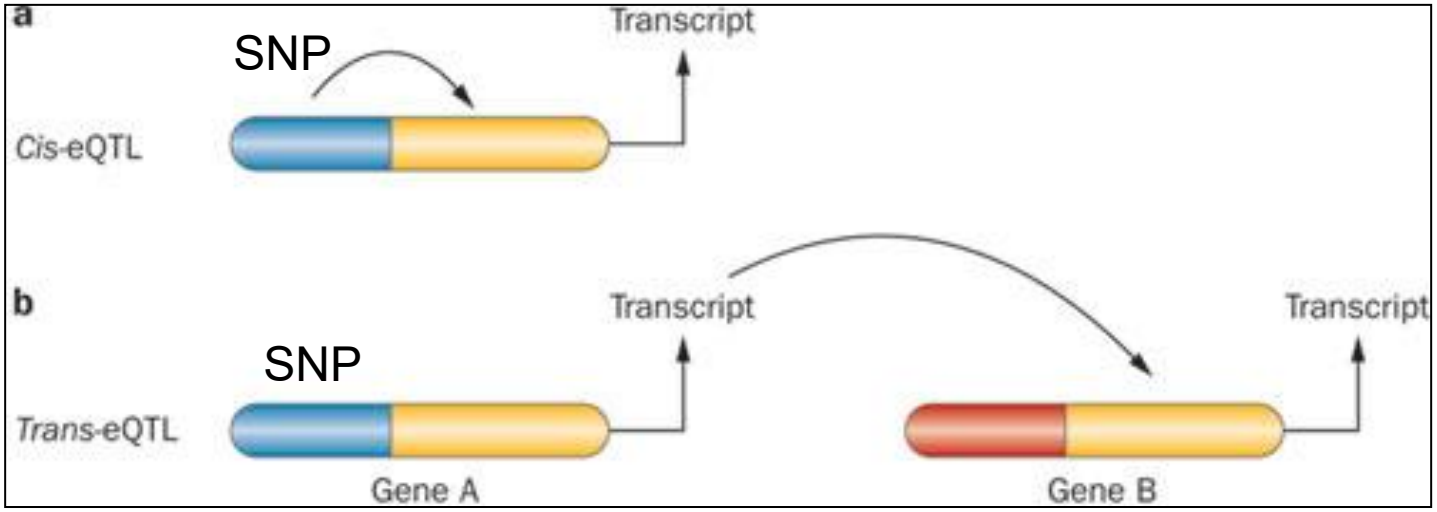
Assessment of the workflow using validated targets

Fine-Mapping: overlaying different information to identify causal genes

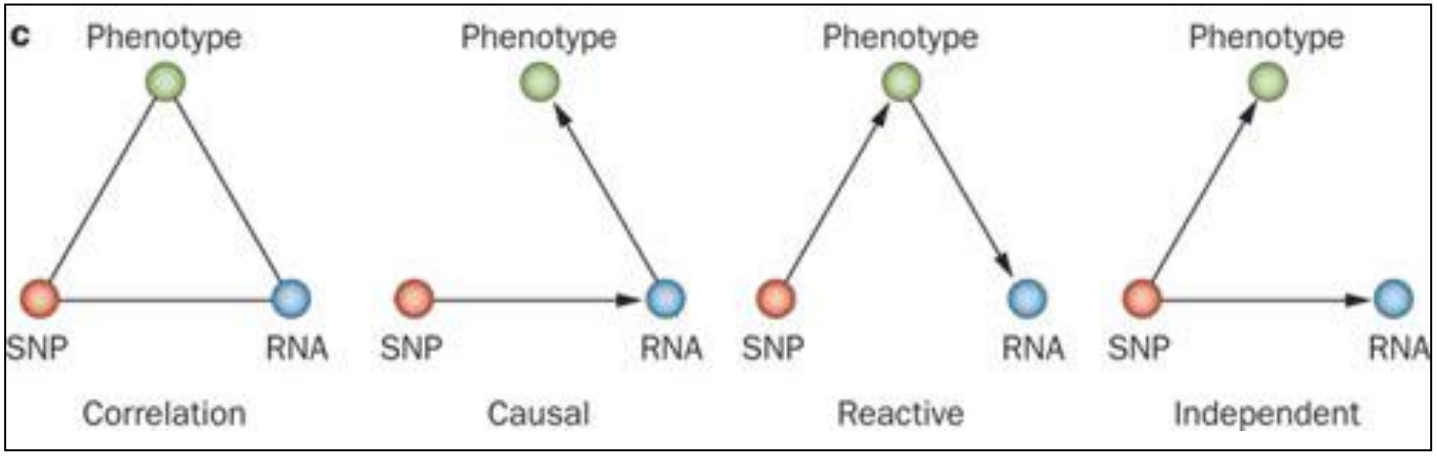
M. Civelek & A. J. Lusis – Nat. Rev. Genetics (2013)



Combining DNA and RNA information – eQTLs and Causal Networks



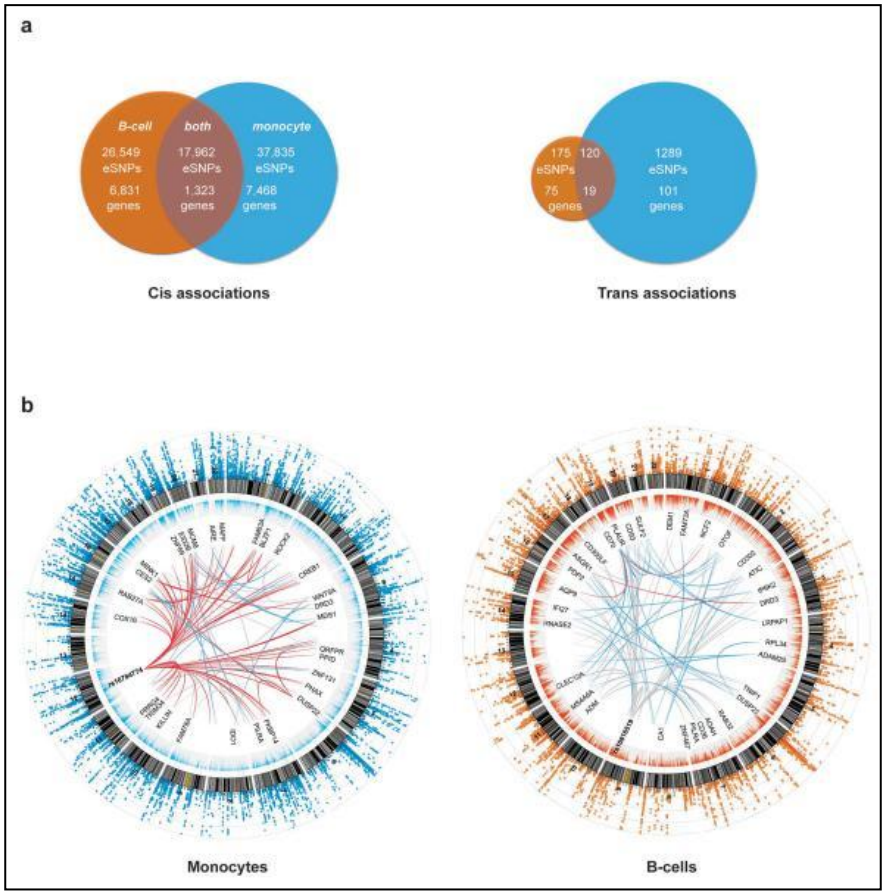
- DNA/RNA combined analysis can highlight eSNPs
- eSNPs exert an effect on genes' transcription

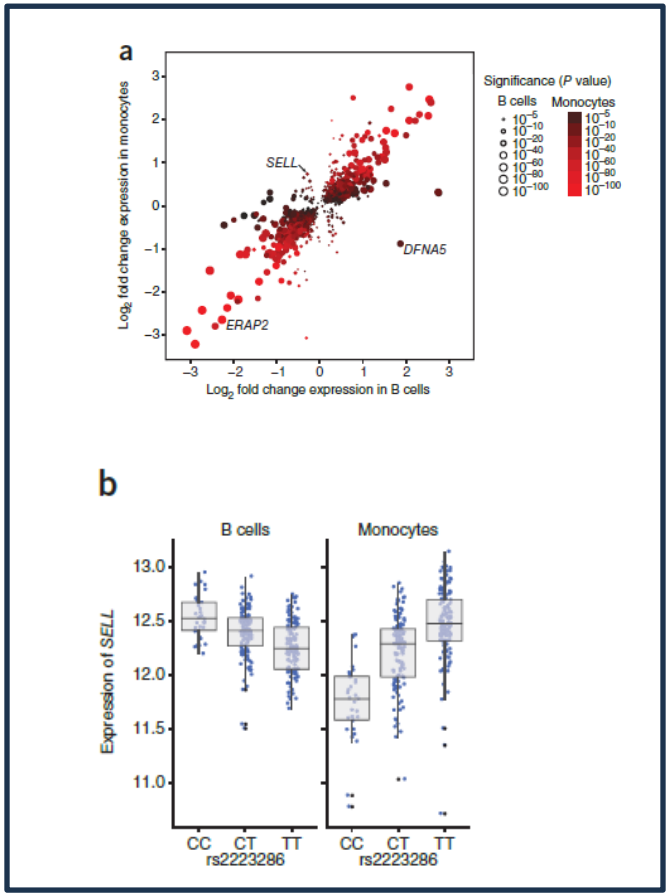


Using the SNP as a 'causal anchor', causal relationships between the three can be modelled: Causal Networks

MacLellan WR, et al. Nat Rev Cardiol. 2012

Cell type-specific eQTLs in B-cell and monocytes



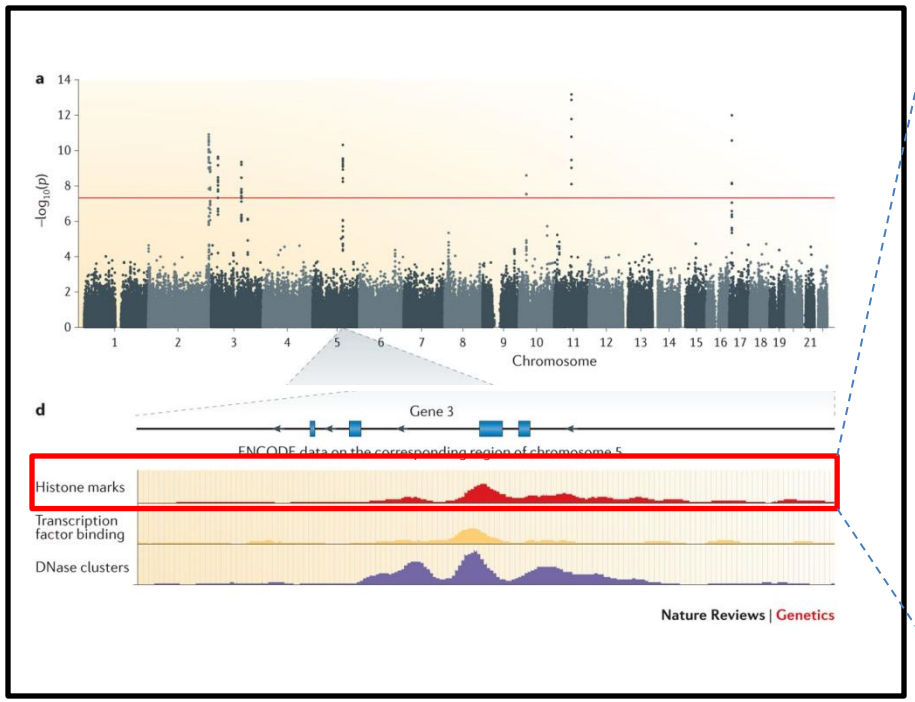


Circos Plots

[J. Knight et al. - Nat Genet.](#) *2012 Mar 25;44(5):502-10.*

Assessment of enrichment of 100 non-MHC RNA risk loci in epigenetic chromatin marks

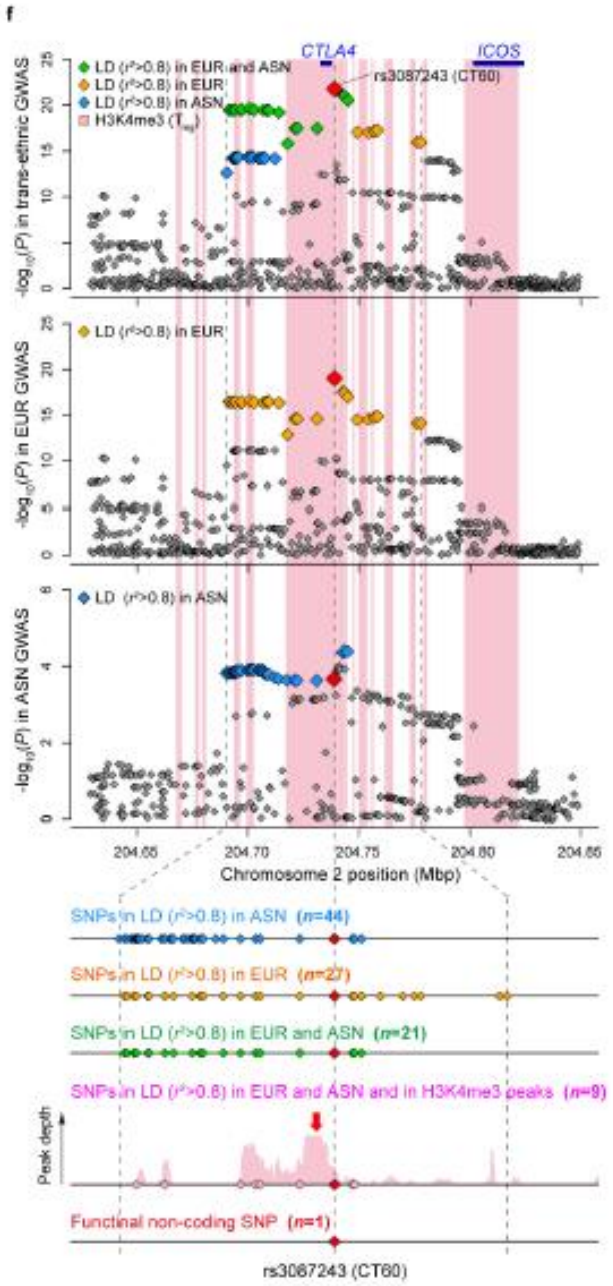
100 RA risk loci



d

Cell types	*P* for H3K4me3 enrichment
T_{reg} primary cells	$\leq 1.0\times10^{-5}$
$CD4^+$ memory primary cells	3.0×10^{-5}
$CD4^+$ naive primary cells	0.0041
$CD8^+$ memory primary cells	0.0065
Smooth muscle, rectal	0.034
Mucosa, colon	0.038
$CD8^+$ naive primary cells	0.12
Mucosa, stomach	0.13
$CD34^+$ primary cells	0.18
$CD34^+$ cultured cells	0.19
Mobilized $CD34^+$ primary cells	0.19
$CD19^+$ primary cells	0.24
$CD3^+$ primary cells	0.30
Mucosa, duodenum	0.40
Muscle satellite cultured cells	0.46
Cingulate gyrus (brain)	0.53
Skeletal muscle	0.77
Mucosa, rectal	0.77
Smooth muscle, colon	0.79
Mesenchymal stem cells (adipose)	0.81
Adipose nuclei	0.84
Smooth muscle, duodenum	0.85
Mid frontal lobe (brain)	0.86
Hippocampus middle (brain)	0.91
Mesenchymal stem cells (bone marrow)	0.91
Pancreatic islets	0.93
Inferior temporal lobe (brain)	0.93
Substantia nigra (brain)	0.93
Adult kidney	0.94
Adult liver	0.95
Mesenchymal stem cells (adipocyte)	0.98
Mesenchymal stem cells (chondrocytes)	0.99
Anterior caudate (brain)	0.99
Smooth muscle, stomach	0.99

Example: Fine Mapping of CTLA4



Regional (trans-ethnic, European, Asian) SNP associations of the CTLA4 locus in stage 1 GWAS meta-analysis

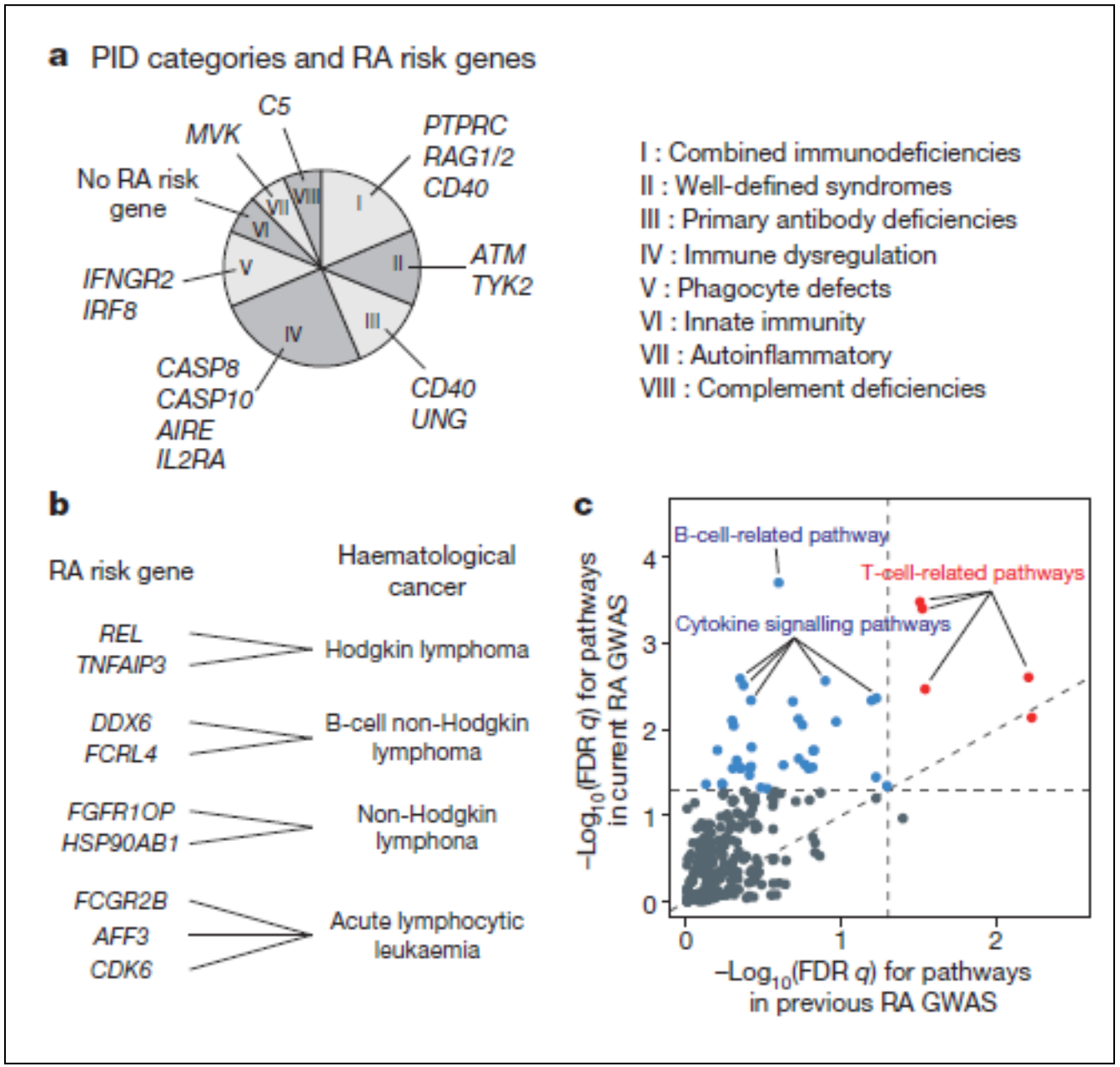
1. functional non-coding variant of CT60 (rs3087243) showed the most significant association with RA.

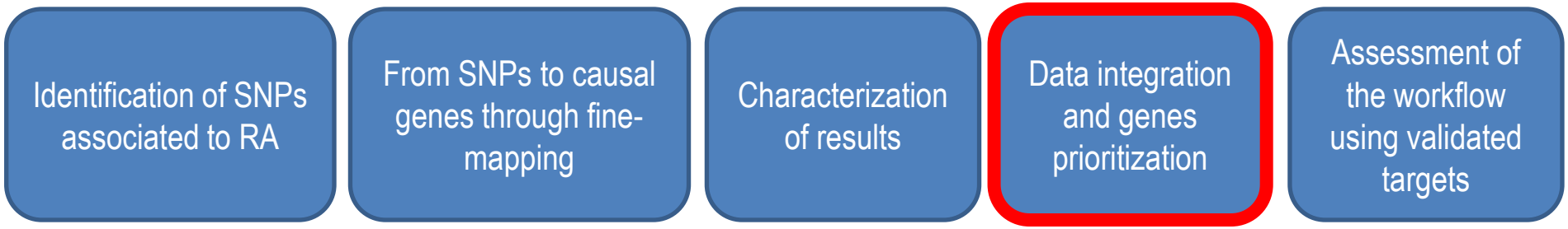
2. Trans-ethnic fine mapping of candidate causal variants decreased the number of candidate variants from 44 (LD in Asians) and 27 (LD in Europeans) to 21 (LD in both populations).

3. **Selected the 9 candidate variants included in Treg H3K4me3 peaks, including CT60 (close to H3K4me3 summit)**

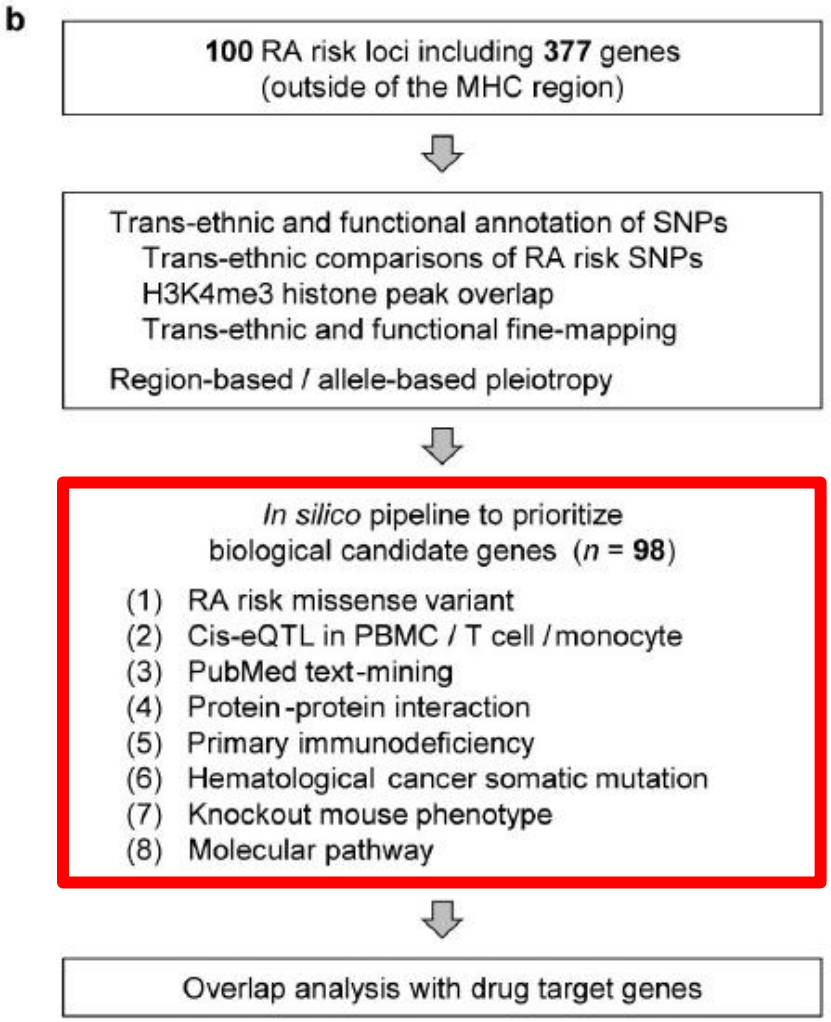

Identification of SNPs associated to RA
From SNPs to causal genes through fine-mapping
Characterization of results
Data integration and genes prioritization
Assessment of the workflow using validated targets

Characterization of Results (100 loci, 377 genes)




Identification of SNPs associated to RA
From SNPs to causal genes through fine-mapping
Characterization of results
Data integration and genes prioritization
Assessment of the workflow using validated targets

In-silico genes prioritization



Nature. 2014 Feb 20;506(7488):376-81

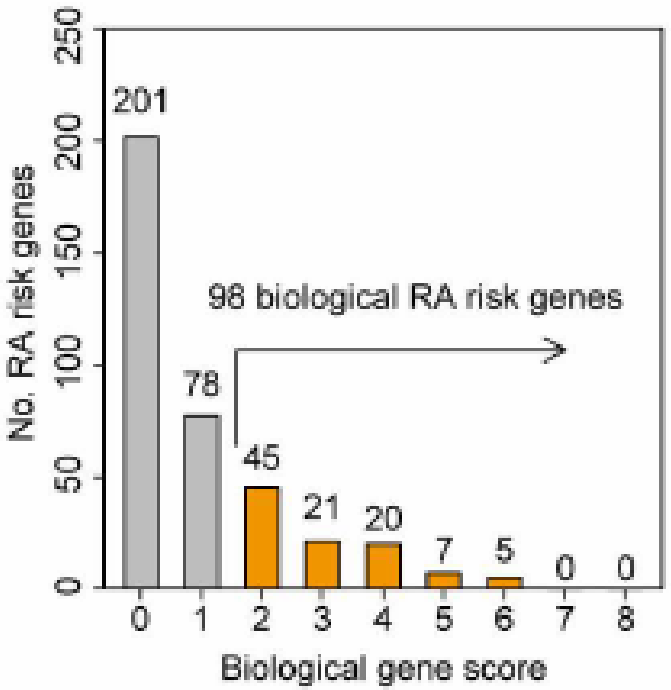
Prioritization of biological candidate genes from RA risk loci

a

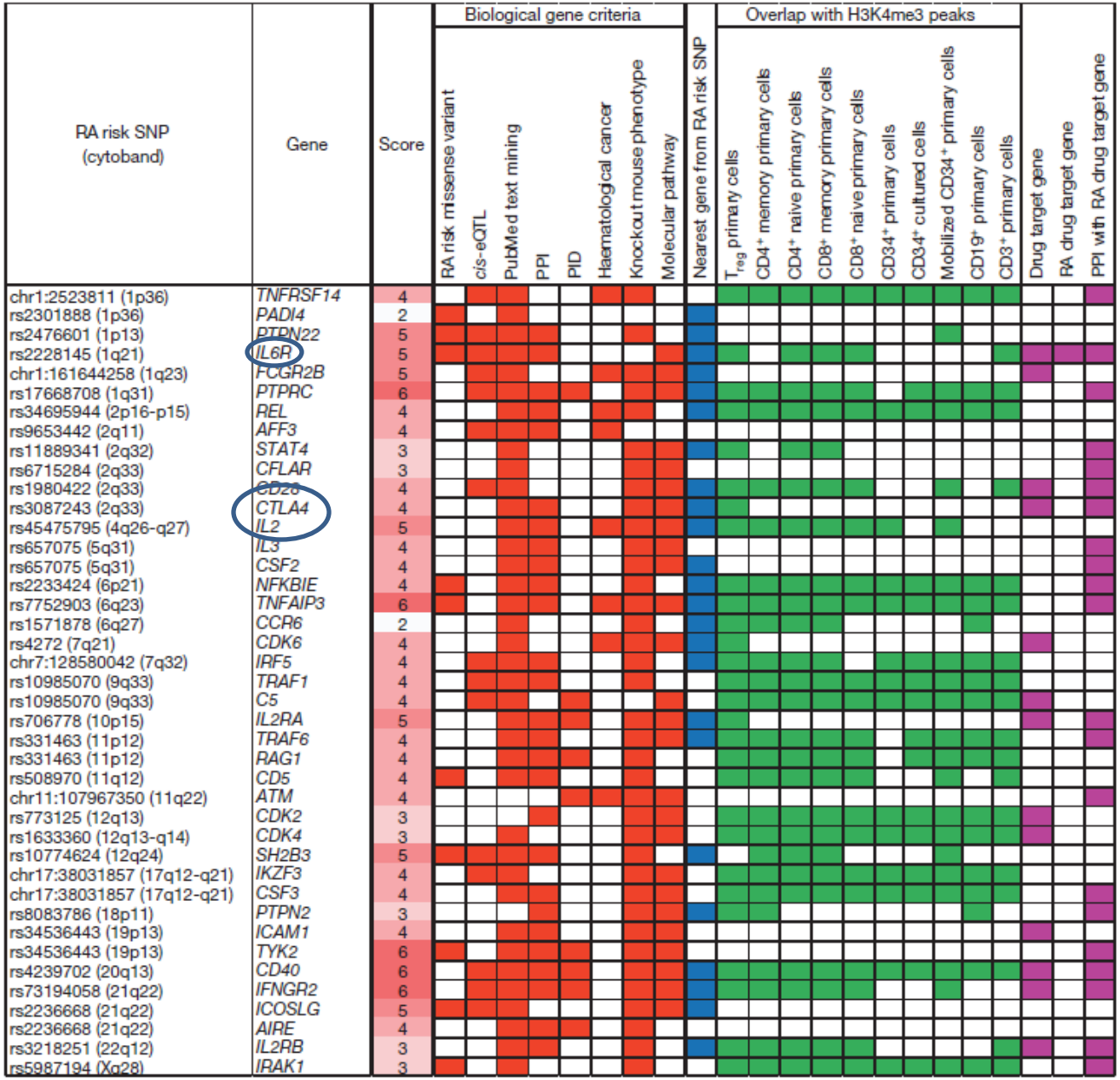
Biological RA risk gene prioritization criteria

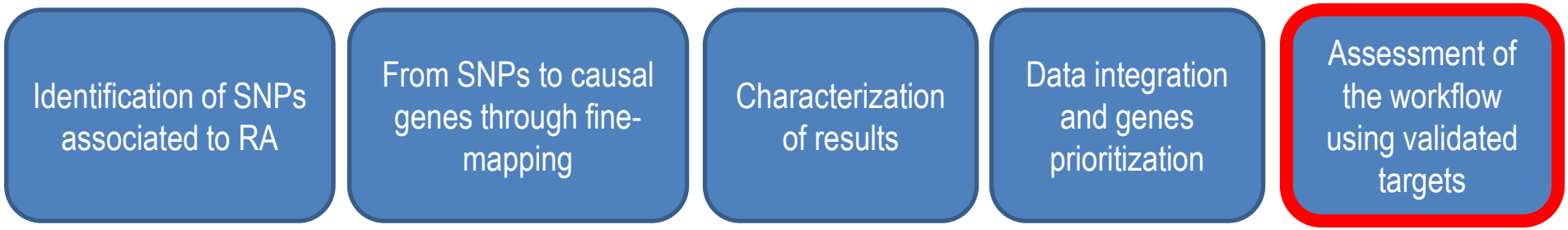
(1) RA risk missense variant (n = 19)
(2) Cis-eQTL (n = 51)
(3) PubMed text-mining (n = 90)
(4) Protein-protein interaction (n = 63)
(5) Primary immunodeficiency (n = 15)
(6) Hematological cancer (n = 17)
(7) Knockout mouse phenotype (n = 86)
(8) Molecular pathway (n = 35)

b



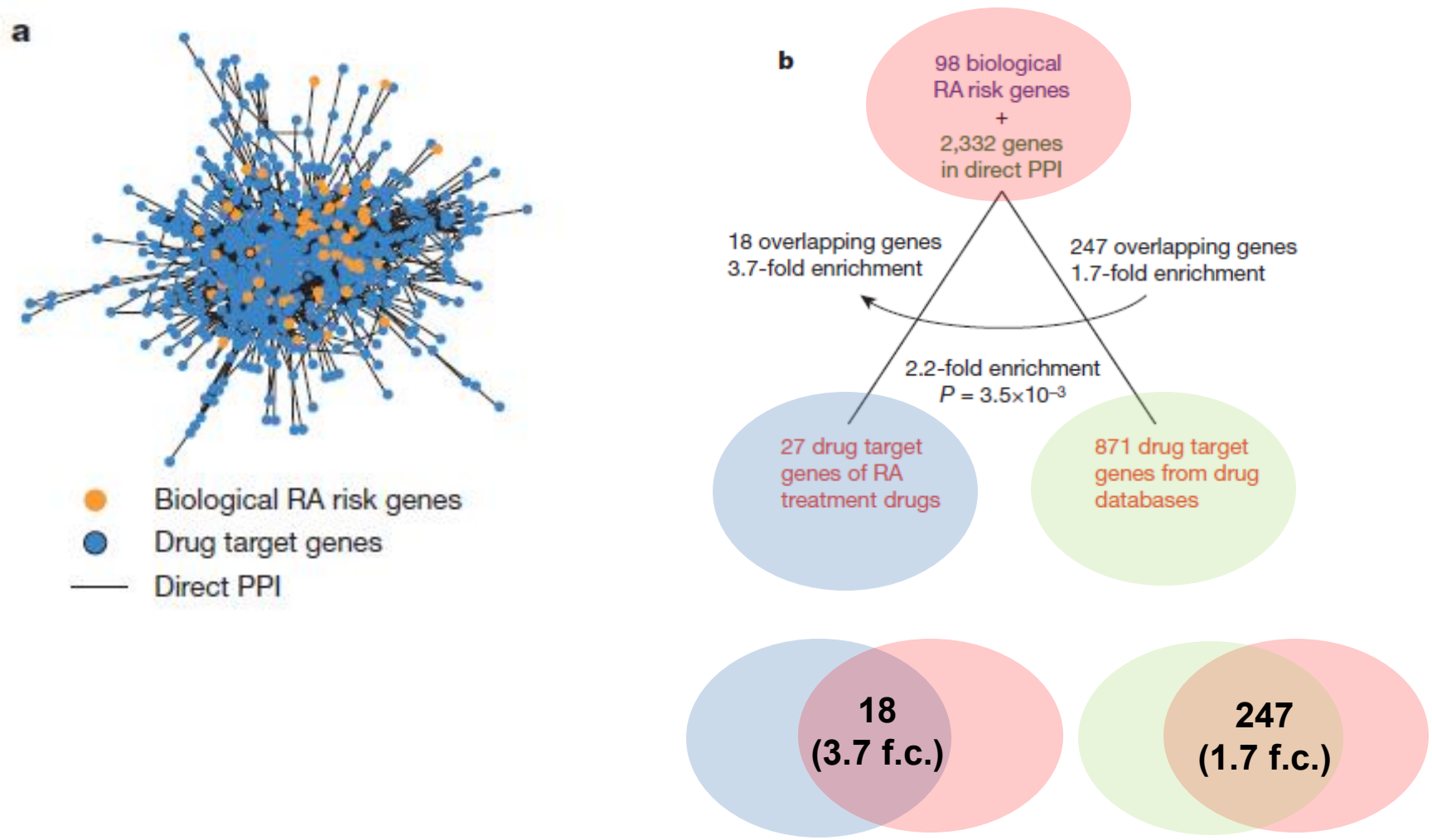
Data Integration and Gene Prioritization




Identification of SNPs associated to RA
From SNPs to causal genes through fine-mapping
Characterization of results
Data integration and genes prioritization
Assessment of the workflow using validated targets

Assessment of the workflow using validated targets

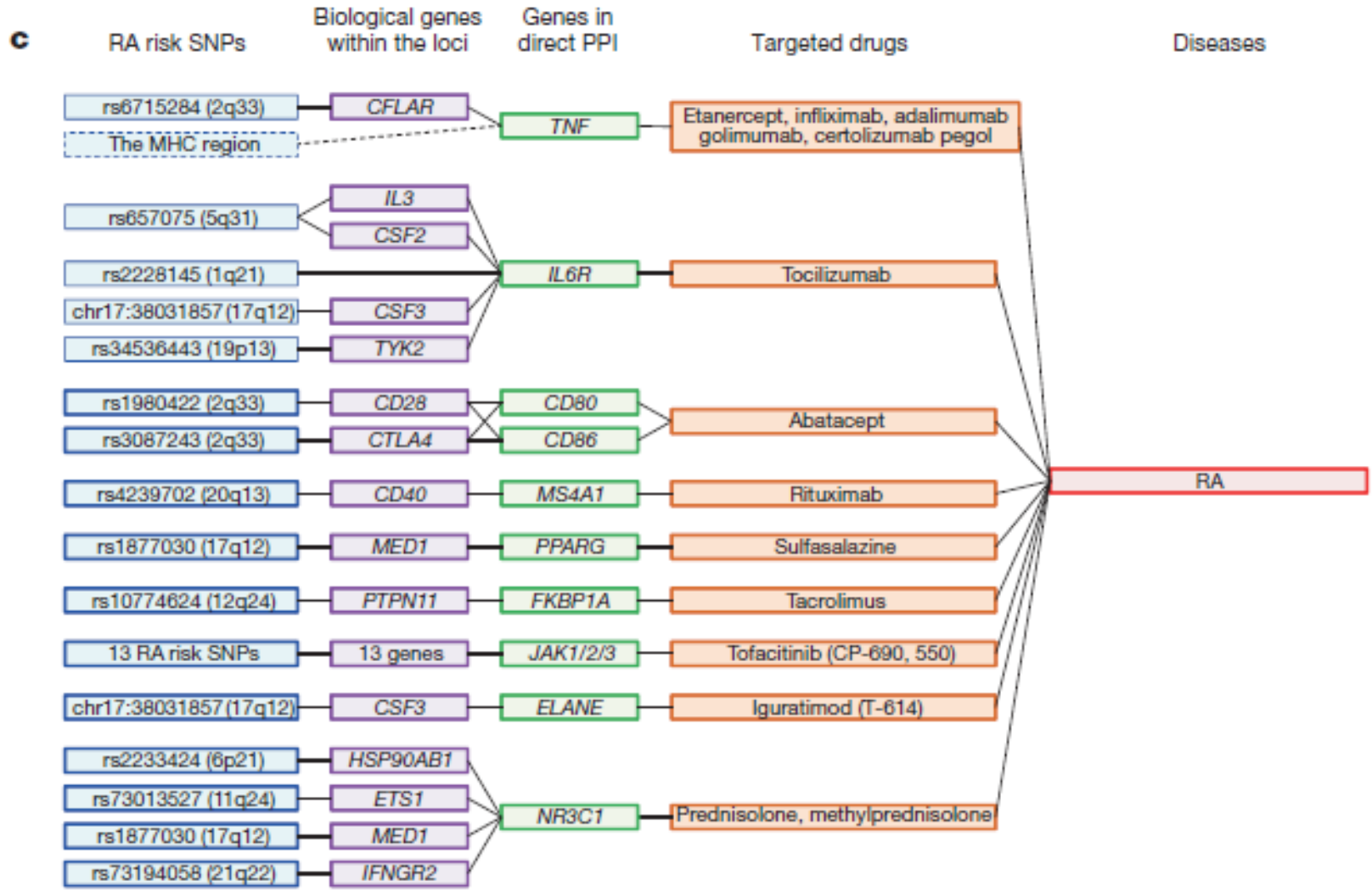
- 98 biological RA risk genes (score >=2)
- +2.322 genes in PPI



➔ Strong enrichment for RA approved drugs (extracted from DrugBank TherapeuticTargets Database - TTD)

Mapping RA risk SNPs to drug targets

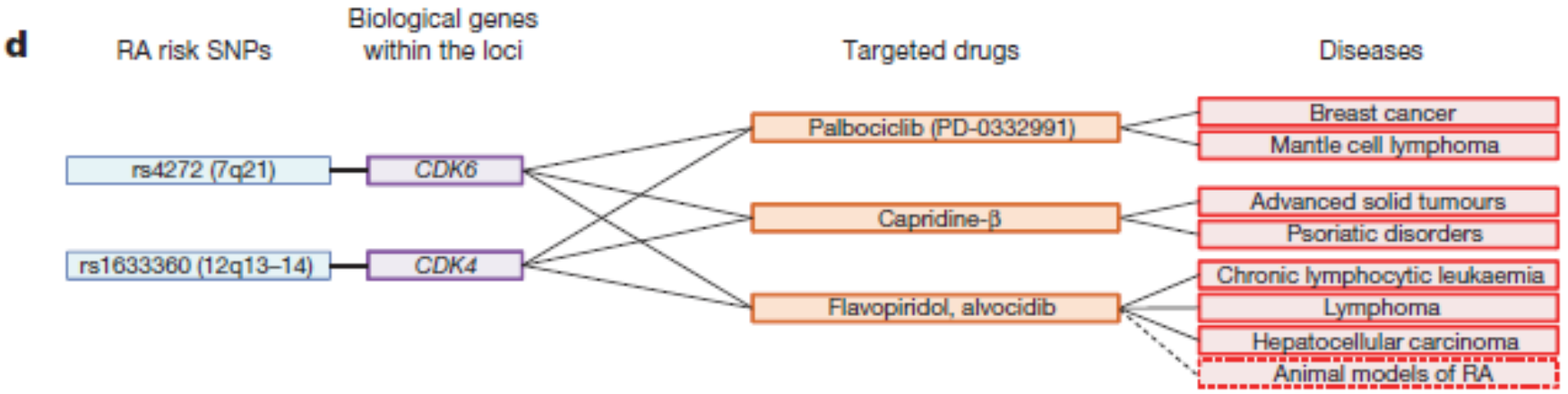
Example: PPI link to known RA drugs



- 98 biological RA risk genes (score >=2)

Drug Repurposing

Connections between RA genes and drugs indicated for other diseases



Study Summary

- Comprehensive genetic study with.100,000 subjects:
  - identified 42 novel RA risk loci
  - provided novel insights into RA pathogenesis.
  - demonstrated role of genetics for drug discovery

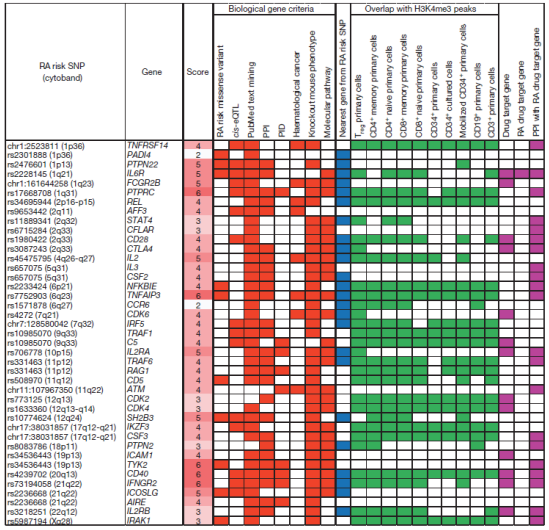
- Systematic approach to derive disease biological insights and novel drug candidates by integrating human genetic data with different layers of orthogonal information

Take Home Messages

- GWAS can be used to identify candidate regions and genes associated with risk of RA
- Testing millions of hypotheses implies hunting for very low p-values
- Using a series of strategies and additional data to understand the functions of associated genes to disease and prioritize them as possible targets
  - Epigenetics
  - PPI Network analysis
  - Link genetics to intermediate phenotypes (e.g. eQTLs)
- Enrichments for existing RA and other drugs supports the pipeline
- If new candidate RA genes can be targeted by existing drugs, drug repositioning opportunities can be evaluated

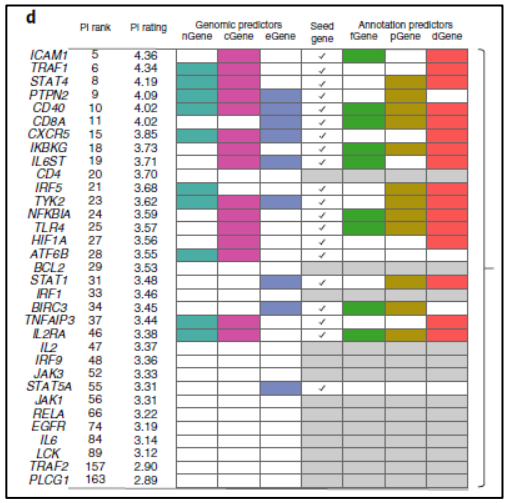
Similar Approaches to Data Integration

Genetics, functional genomics, network connectivity. RA



R. Plenge Nature 2014

Genetics, functional genomics, immune-related annotations, network connectivity. 30 immune traits



J.Knight Nat Genetics 2019

Genetics, animal models, text-mining, druggability, animal models, text mining, pathways. All diseases

Open Targets

https://www.targetvalidation.org/

SLE Molecular Immune Monitoring – Study Workflow

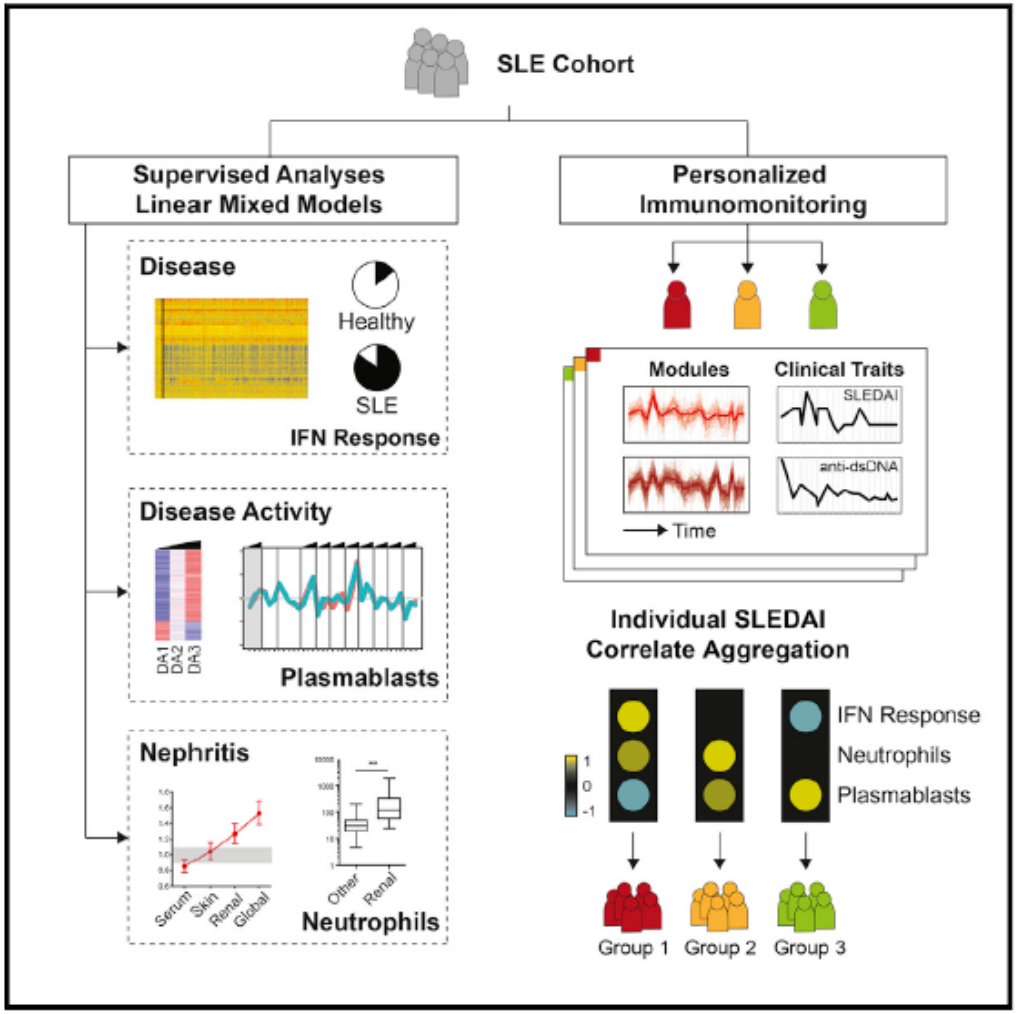
Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients

Romain Banchereau,[1,7] Seunghee Hong,[1,7] Brandi Cantarel,[1] Nicole Baldwin,[1] Jeanine Baisch,[1] Michelle Edens,[1] Alma-Martina Cepika,[1] Peter Acs,[1] Jacob Turner,[1] Esperanza Anguiano,[1] Parvathi Vinod,[1] Shaheen Khan,[2] Gerlinde Obermoser,[1] Derek Blankenship,[1] Edward Wakeland,[2] Lorien Nassi,[2,3] Alisa Gotte,[2,3,4] Marilynn Punaro,[2,3] Yong-Jun Liu,[1,5] Jacques Banchereau,[6] Jose Rossello-Urgell,[1] Tracey Wright,[2,3] and Virginia Pascual[1,3,*]

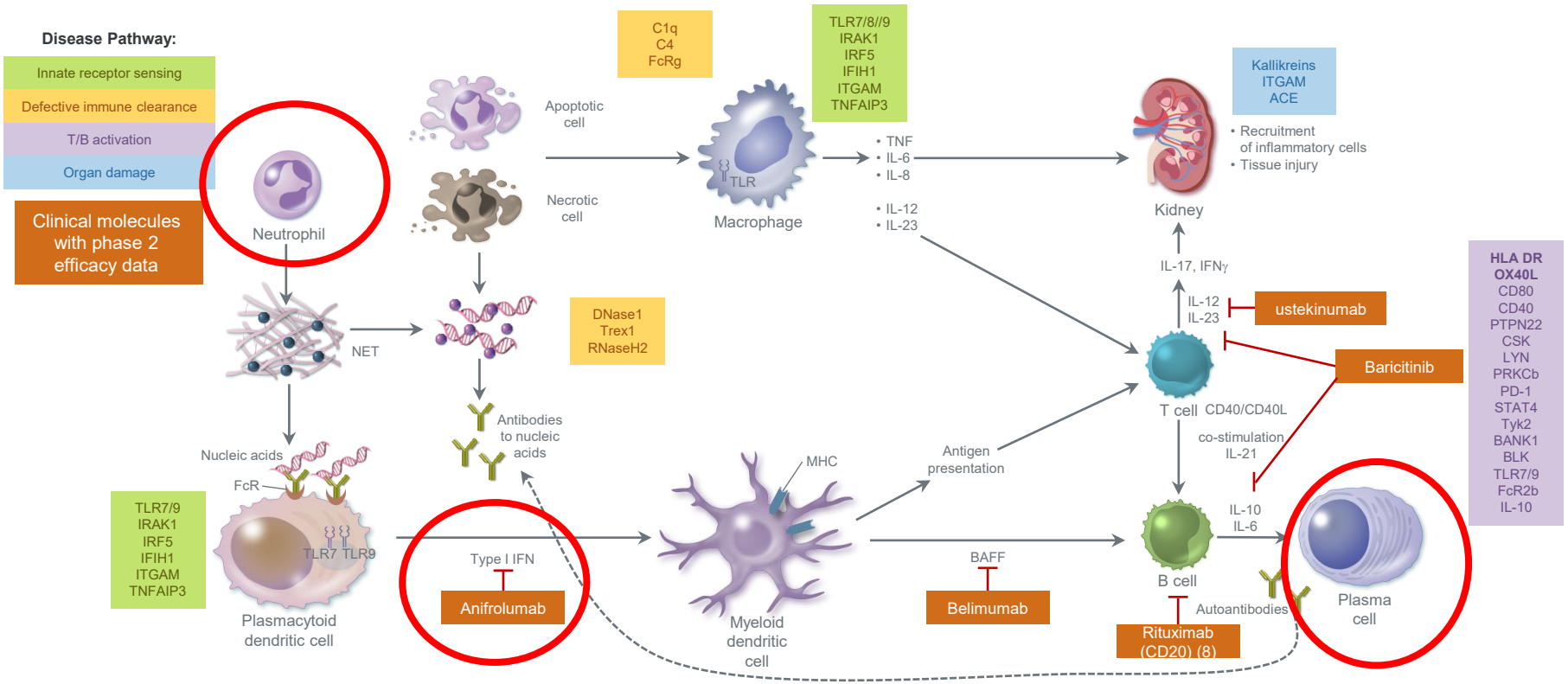
[1]Baylor Institute for Immunology Research, Dallas, TX 75204, USA
[2]UT Southwestern Medical Center, Dallas, TX 75235, USA
[3]Texas Scottish Rite Hospital for Children, Dallas, TX 75219, USA
[4]Vanderbilt University School of Medicine, Nashville, TN 37232, USA
[5]MedImmune, Gaithersburg, MD 20878, USA
[6]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA
[7]Co-first author
*Correspondence: virginia.pascual@bswhealth.org
http://dx.doi.org/10.1016/j.cell.2016.03.008

Cell. 2016 Apr 21;165(3):551-65

Clinical and transcriptional profiling of 158 lupus pediatric patients, up to a period of 4 years



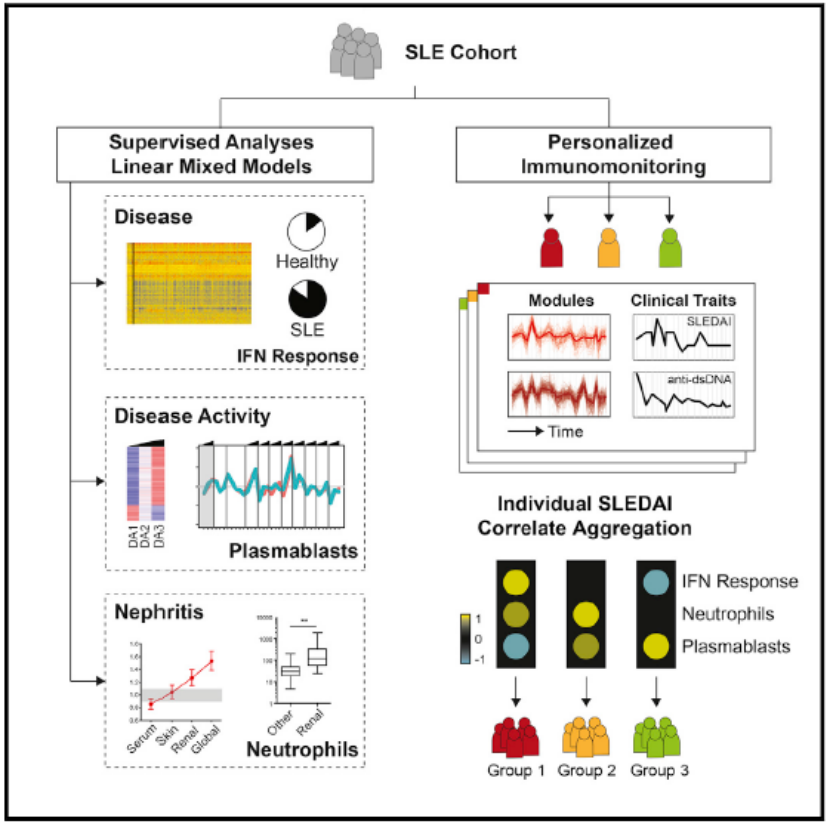
Cells and Pathways Driving SLE



Lih-Ling Lin - Adapted from Tsokos et al., *Nat Rev Rheumatol*. 2016

SANOFI

Analytical Study Workflow



Clinical and transcriptional profiling of 158 lupus pediatric patients, up to a period of 4 years

Genes associated to SLE, DA, Race, Treatment, disease subtypes

From genes to blood modules

WGCNA modules for each patient and correlation with SLEDAI

Linking WGCNA to blood modules

Stratification of Patients into Groups

Toolbox

- **Gene Expression Analysis**
  - Multivariate linear regression modelling
  - Heat maps and Hierarchical Clustering
  - Aggregating gene expression through modules
    - WGCNA modules
    - Blood modules
  - Gene Set Enrichment Analysis

SLE Blood Transcriptional Fingerprint

Genes associated to SLE, DA, Race, Treatment, disease subtypes

From genes to ”blood modules”

WGCNA modules for each patient and correlation with SLEDAI

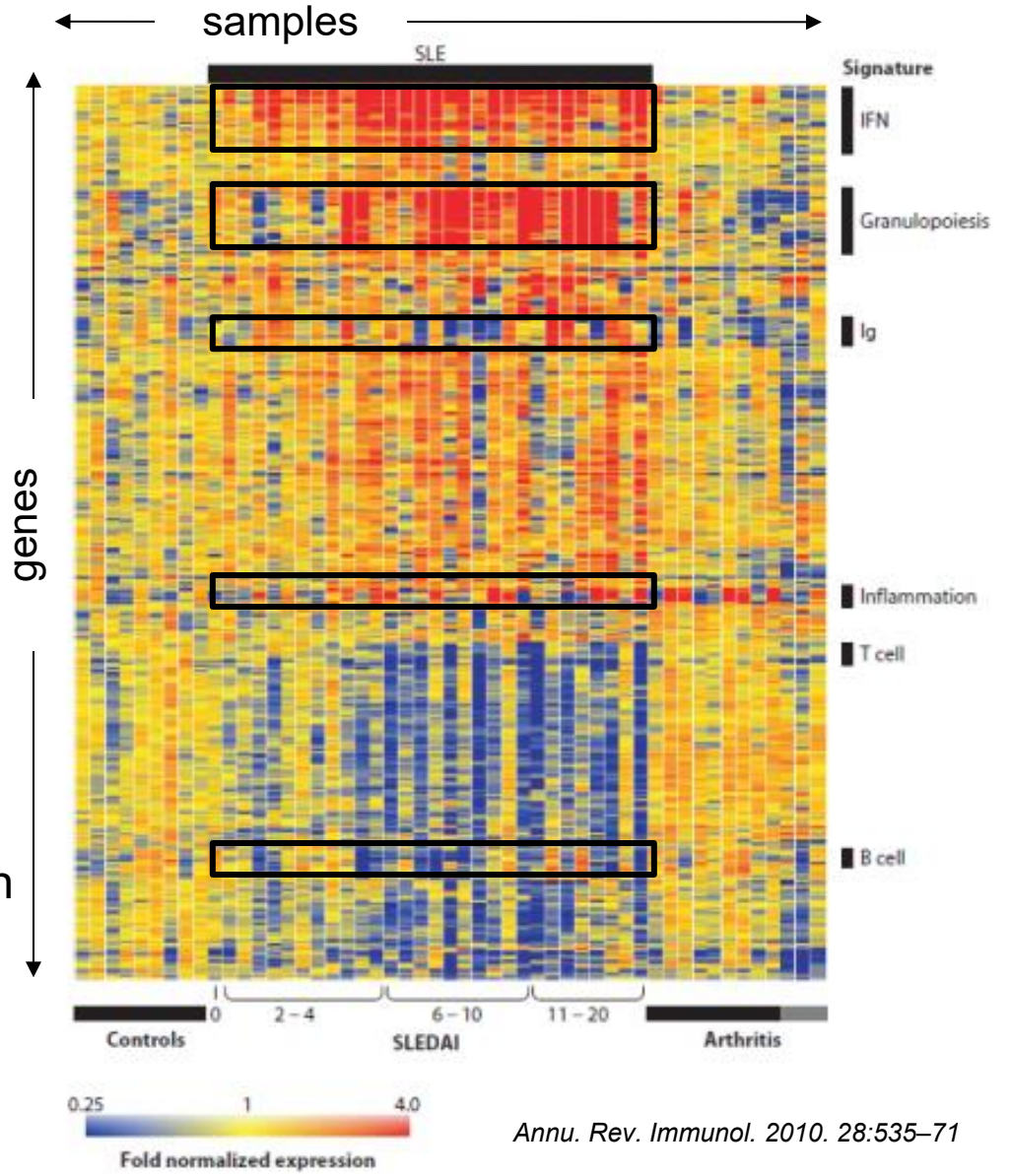
Linking WGCNA to blood modules for inference of biological function

Stratification of patients into groups

Gene Signatures

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Gene 1	1235	546	943	263	136	314
Gene 2	1266	32	556	435	687	2718
Gene 3	947	2829	389	3820	2039	1414
Gene 4	392	2398	84	829	4392	512
...	...	...	...	...	...	...

- Group of genes whose expression values, altogether, are associated with a given feature.
- Genes in signatures often show coordinated expression levels although this is not a requirement.



Annu. Rev. Immunol. 2010. 28:535–71

Differential Expressed Genes: Multivariate Linear Model

Example: 2 groups – SLE and Healthy Controls

$$y_{gi} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_g$$

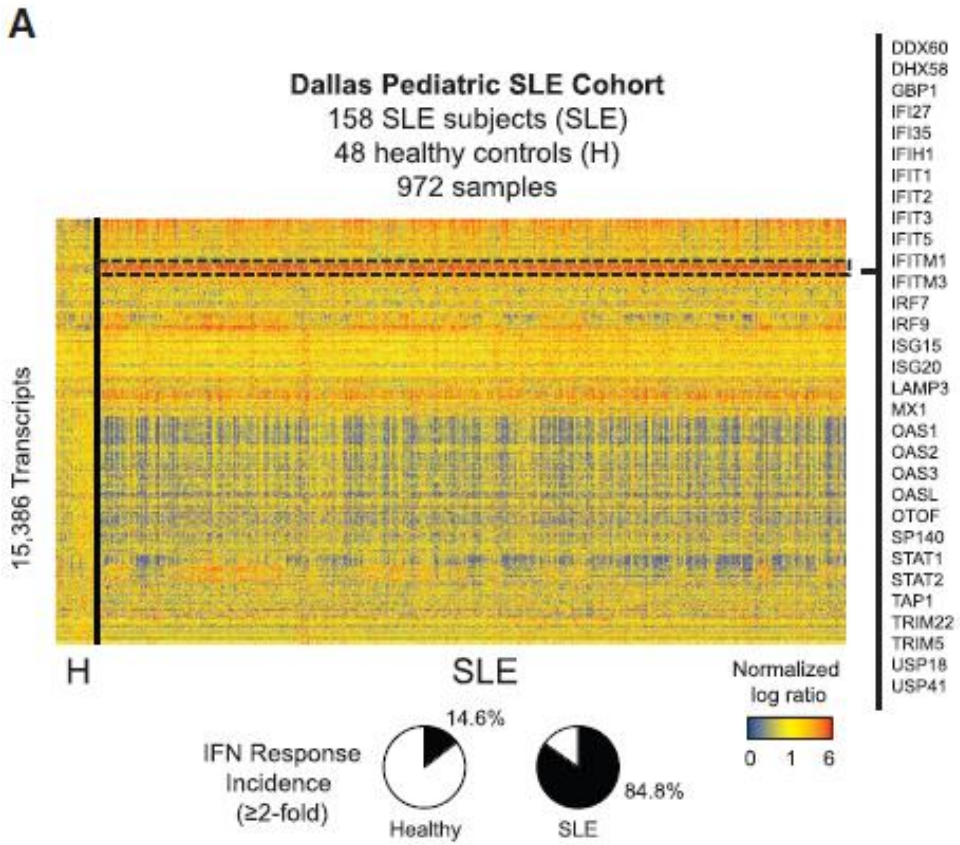
Explanatory Variables (e.g. SLE/Healty, Disease Activity, Treatment etc.)

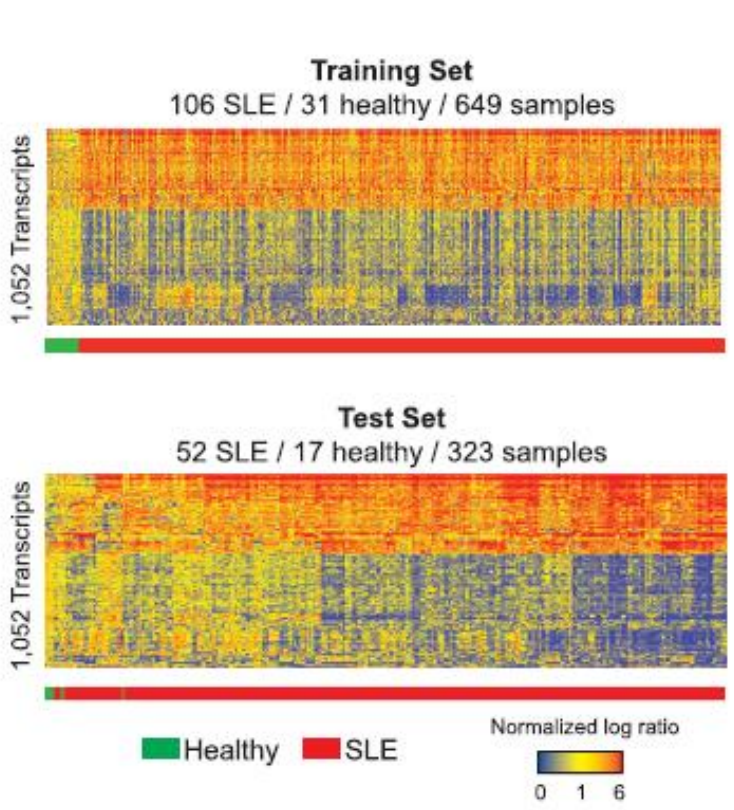
Expression of gene A

- Values of B coefficients
- P-value of B coefficients being different than 0

Genes Associated with SLE

Interferon response

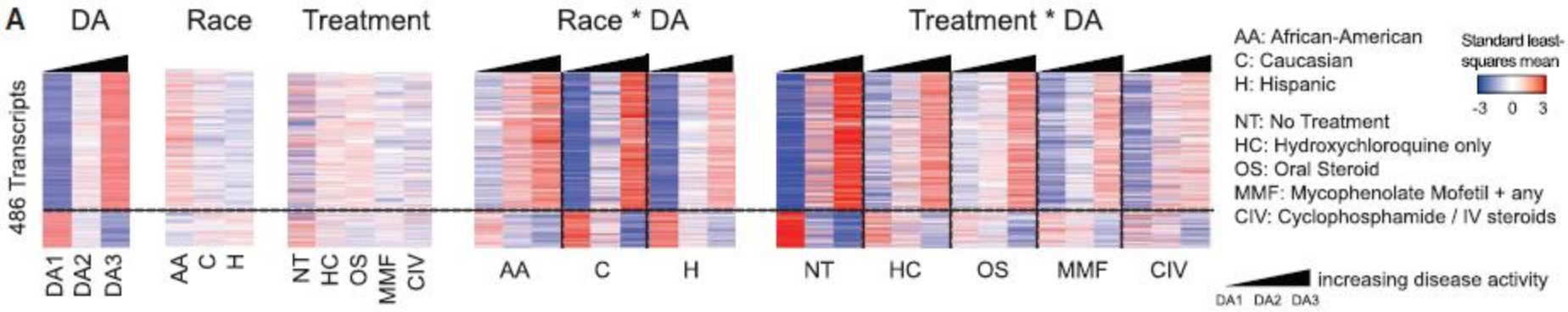




Cell. 2016 Apr 21;165(3):551-65

➔ 1.052 genes differentially expressed in SLE vs Healthy

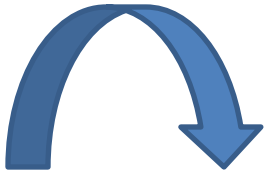
Genes Associated With SLE Disease Activity



- 486 Transcripts Differentially Expressed between DA1 (SLEDAI: 0-2) and DA3 (SLEDAI >7)
- Results stratified by Race, Treatment

SLE Blood Transcriptional Fingerprint

- % of genes up/down
- QuSage fold-change



Genes associated to SLE, DA, Race, Treatment, disease subtypes

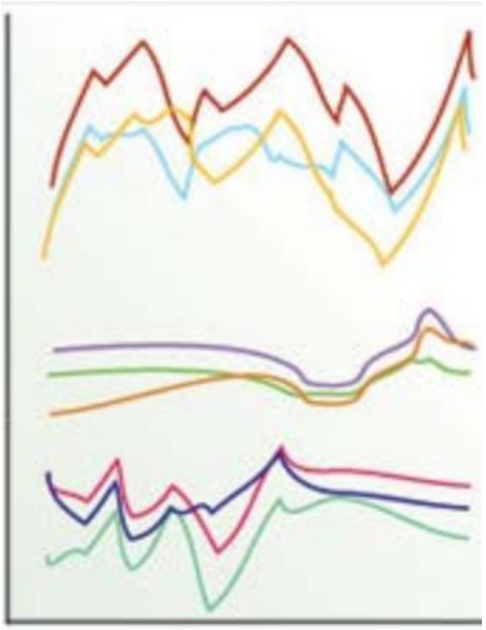
From genes to "Blood Modules"

WGCNA modules for each patient and correlation with SLEDAI

Linking WGCNA to blood modules for inference of biological function

Stratification of patients into groups

Building Transcriptional Blood Modules



Blood Samples (239)

Blood Module 1

Blood Module 2

Blood Module 3

239 Blood Samples:
- systemic juvenile idiopathic arthritis (n = 47)
- systemic lupus erythematosus (n = 40)
- type I diabetes (n = 20)
- metastatic melanoma (n = 39)
- acute infections (Escherichiacoli [n = 22]
- Staphylococcus aureus [n = 18], Influenza A [n = 16])
- liver-transplant recipients undergoing immunosuppressive therapy (n = 37).

260 blood modules identified

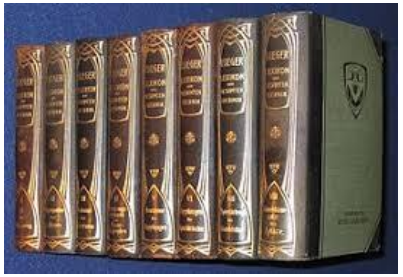


Table 1. Functional Interpretation of Transcriptional Modules

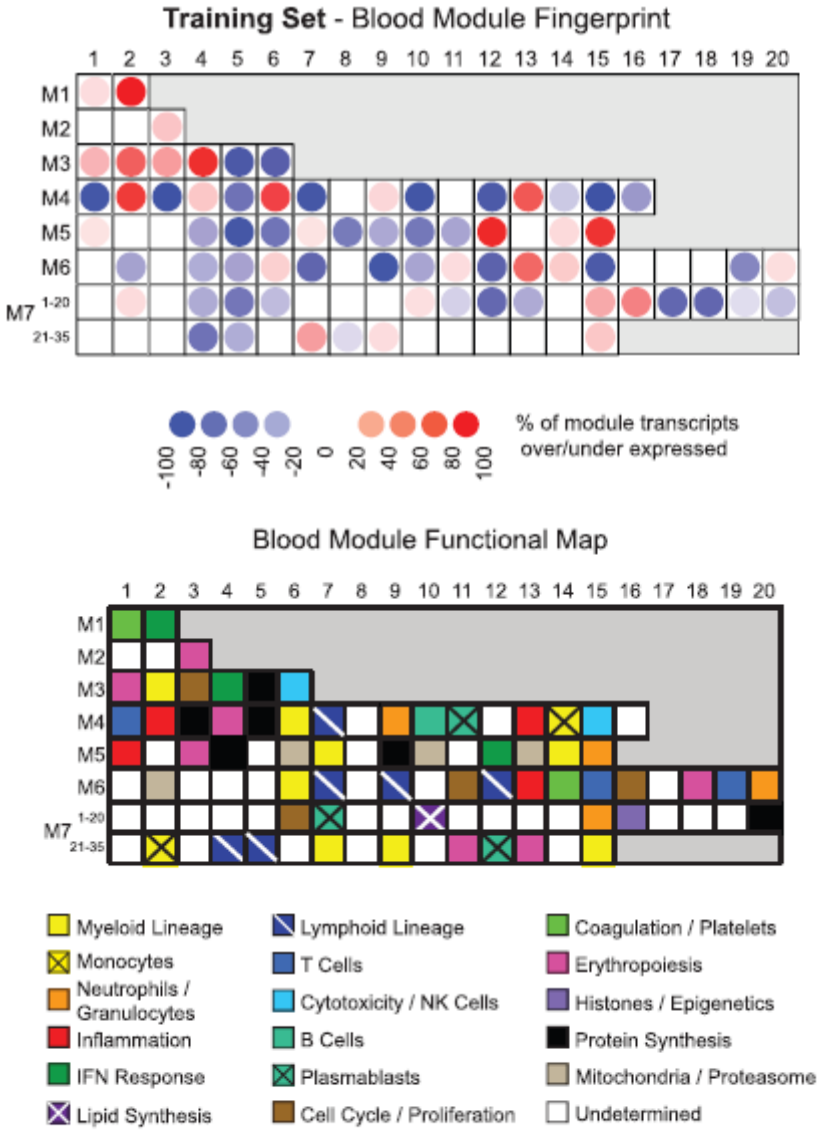
Module I.D.	Number of Probe Sets	Keyword Selection	Interpretation
M 1.1	76	Ig, Immunoglobulin, Bone, Marrow, PreB, IgM,Mu.	Plasma cells. Includes genes coding for Immunoglobulin chains (e.g., IGHM, IGJ, IGLL1, IGKC, IGHD) and the plasma cell marker CD38.
M 1.2	130	Platelet, Adhesion, Aggregation, Endothelial, Vascular	Platelets. Includes genes coding for platelet glycoproteins (ITGA2B, ITGB3, GP6, GP1A/B) and platelet-derived immune mediators such as PPPB (pro-platelet basic protein) and PF4 (platelet factor 4).

Immunity. 2008 Jul 18;29(1):150-64

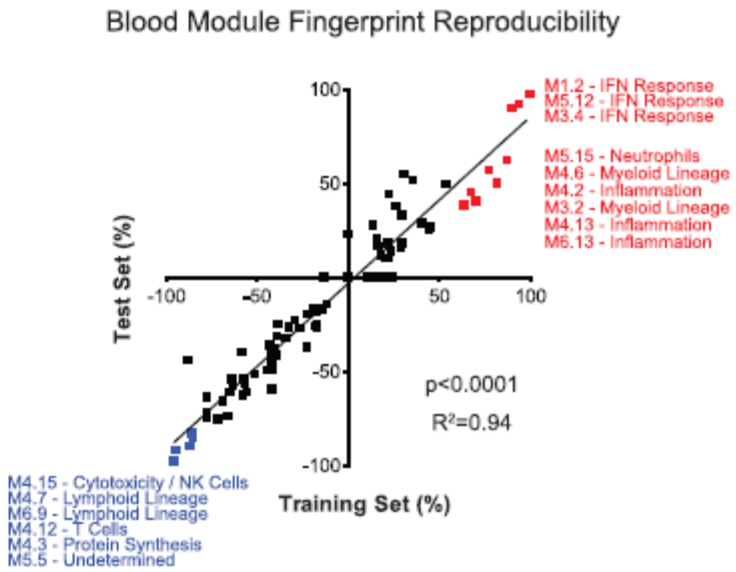
SLE vs Healthy: From Genes to Modules

C

Modules perturbed in SLE vs Healthy



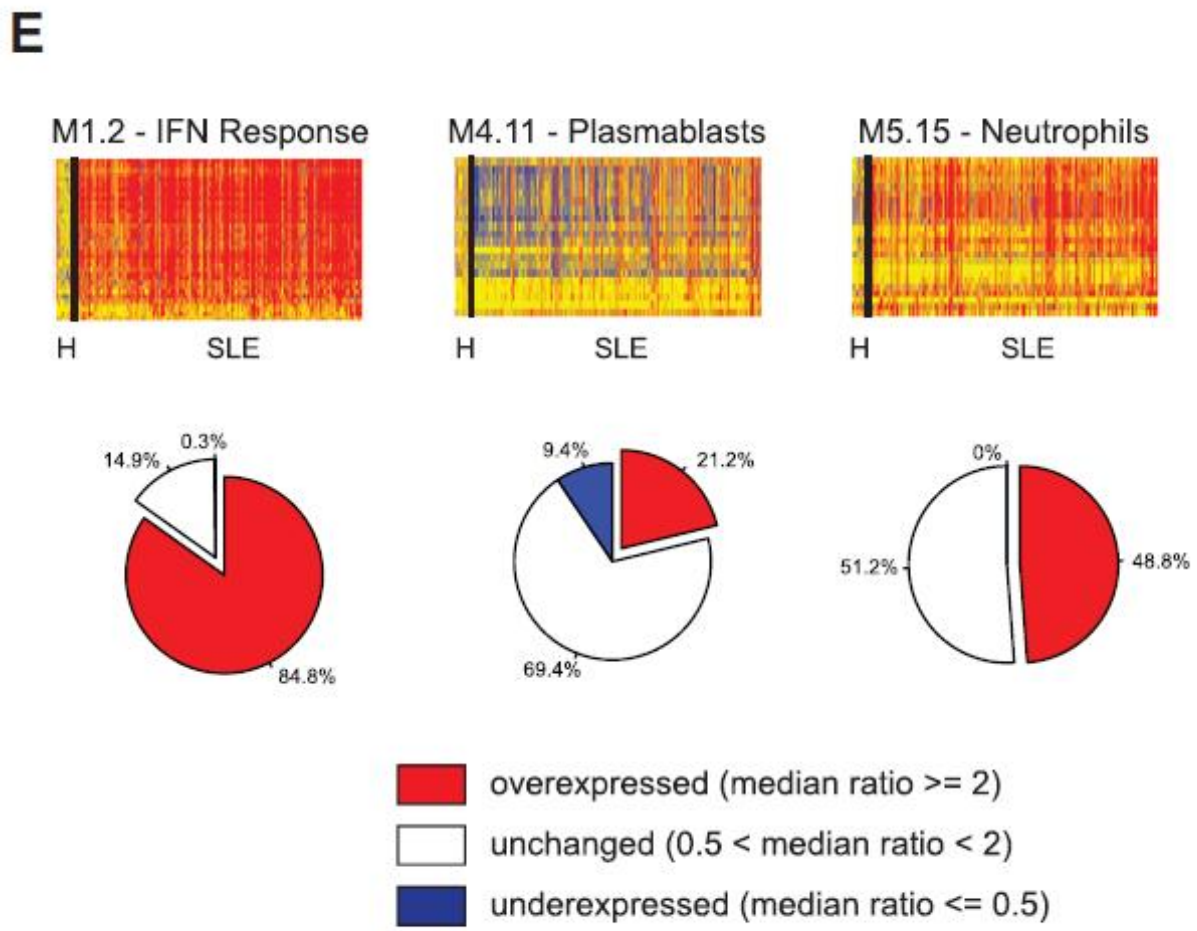
D



- **Over-expression of IFN response, neutrophil, inflammation, cell cycle, erythropoiesis, and histone modules.**
- **Down-regulation of NK cell/cytotoxicity, lymphoid lineage, B cells, T cells, and protein synthesis**

Cell. 2016 Apr 21;165(3):551-65

Over-expression of IFN, plasmablast and neutrophil module genes



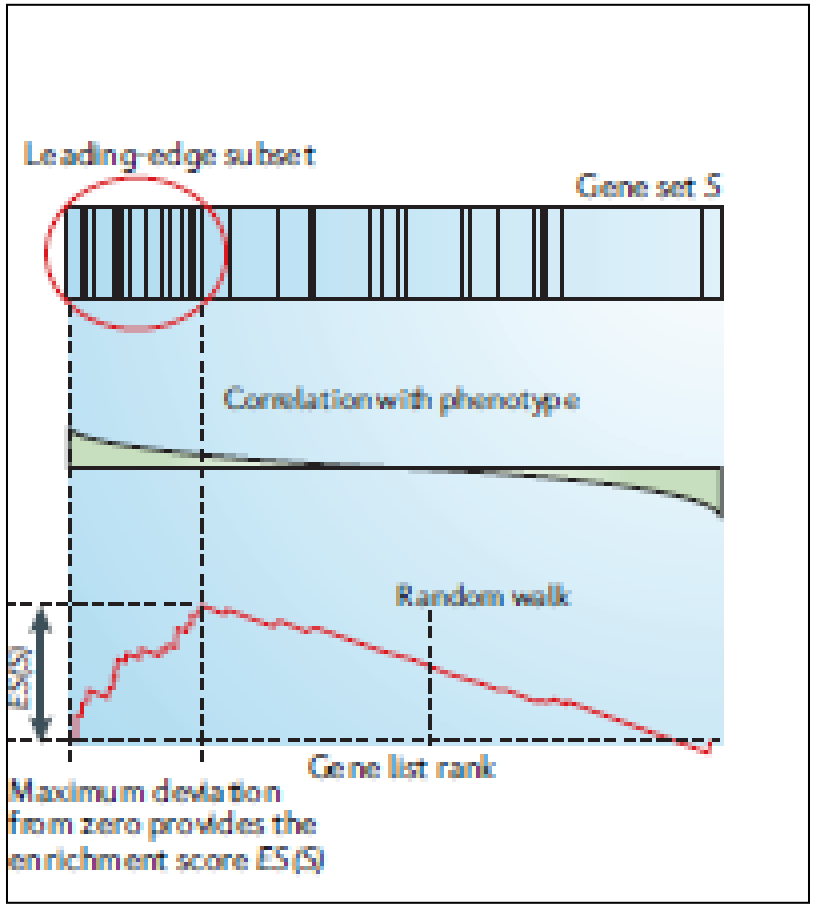
Cell. 2016 Apr 21;165(3):551-65

Gene Set Enrichment Analysis (GSEA)

Gene1
Gene2
Gene3
Gene4
Gene5
Gene6
Gene7
Gene8
Gene9
Gene10
Gene11
Gene12
Gene13
Gene14
Gene15
Gene16
Gene17
Gene18
Gene19
Gene20

Genes are ranked according to their fold-change

Gene set databases
KEGG
GenMAPP
BioCarta
Expression signatures
Cytogenetic loci
Amplifications
Etc.



An enrichment score is calculated for each pathway, taking into account the directionality of the input list

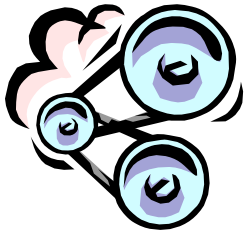
Aravind Subramanian et al, PNAS 2005

From Gene Expression To Gene Set Scores

Gene1
Gene2
Gene3
Gene4
Gene5
Gene6
Gene7
Gene8
Gene9
Gene10
Gene11
Gene12
Gene13
Gene14
Gene15
Gene16
Gene17
Gene18
Gene19
Gene20

List of perturbed genes (differentially expressed between class A and B)

Gene set databases
KEGG
GenMAPP
BioCarta
Expression signatures
Cytogenetic loci
Amplifications
Etc..

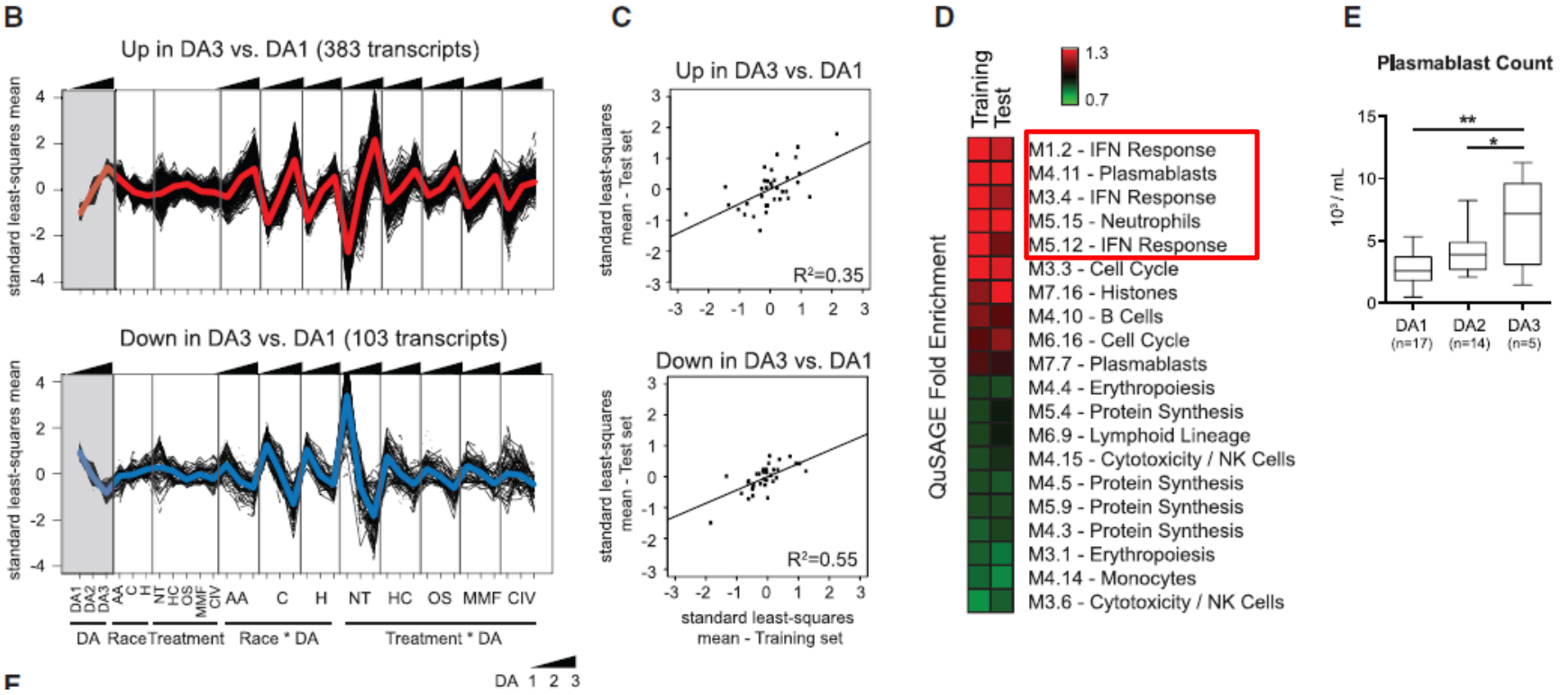


Bioinformatics analysis using pathway DBs

- **Apoptosis** (p-val=0.001)
- **Cell Cycle**(p-val=0.00004)

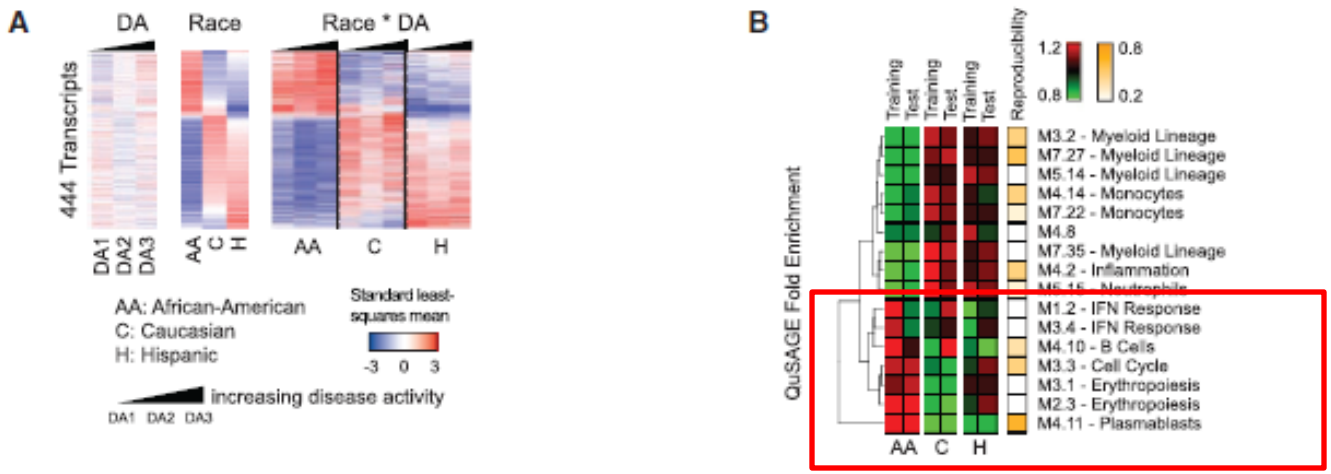
List of perturbed pathways

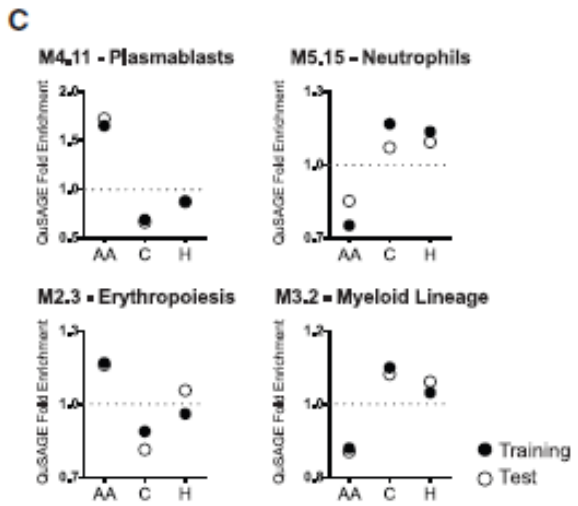
From Genes to Modules Associated with Disease Activity



➔ Genes associated with DA are enriched for **IFN and Plasmablast modules**

Genes and Modules Associated With Race

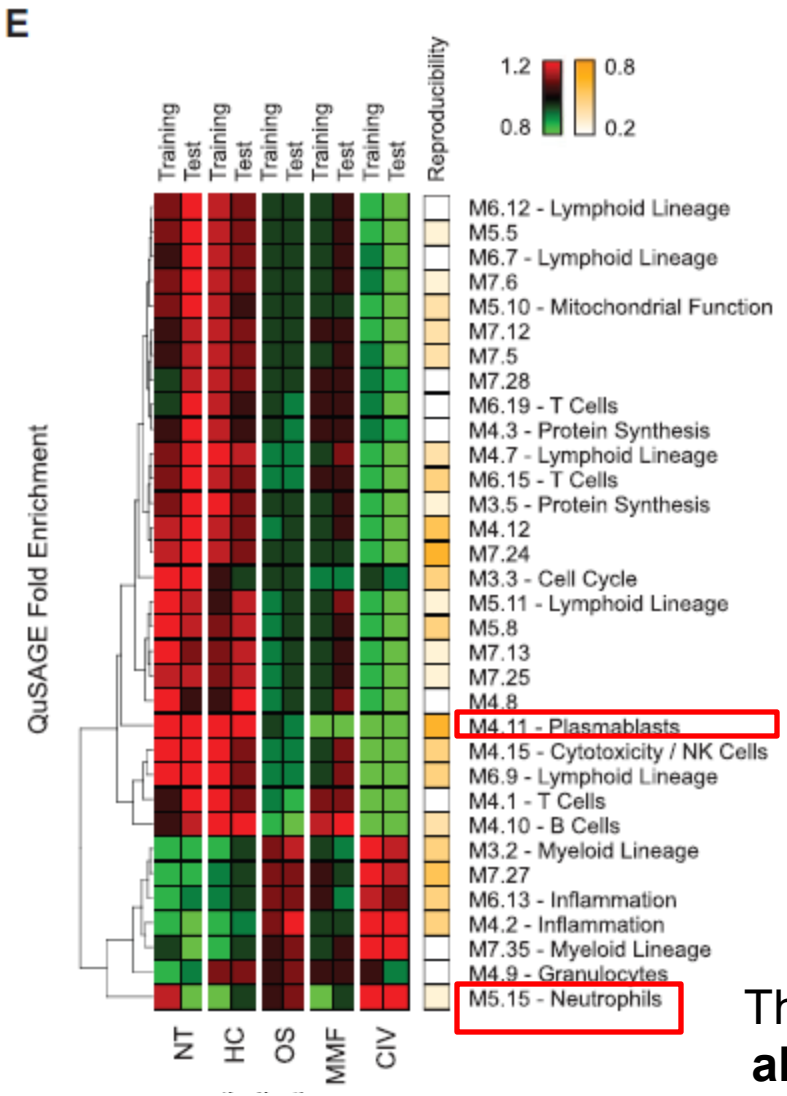




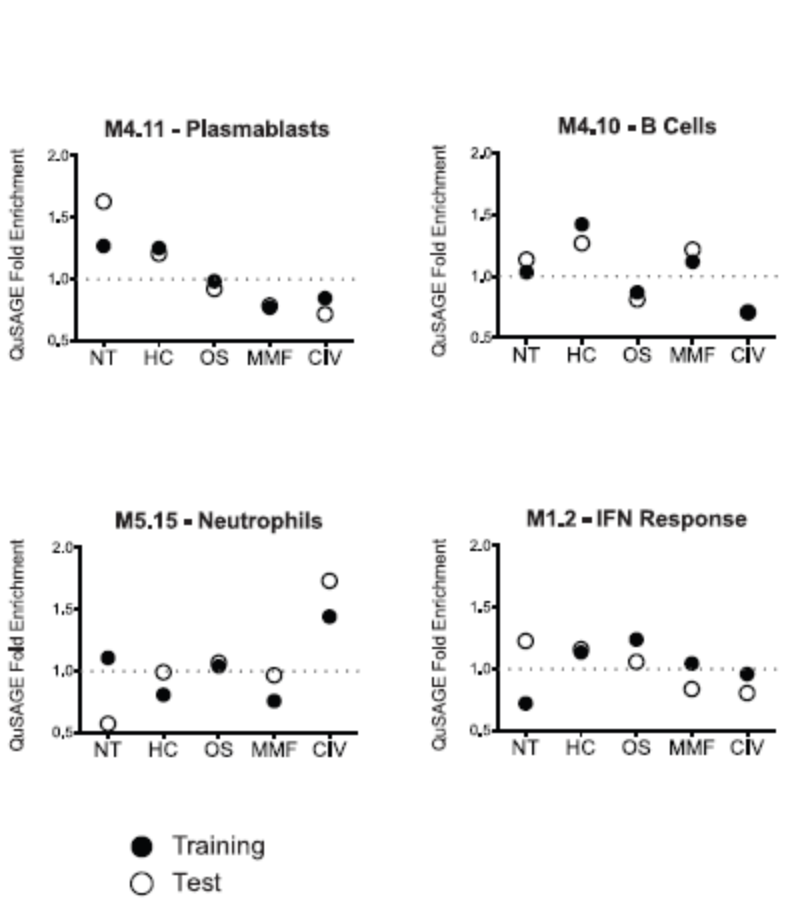
Cell. 2016 Apr 21;165(3):551-65

Increased Plasmablast Responses in African-American Patients

Modules Associated With Treatment



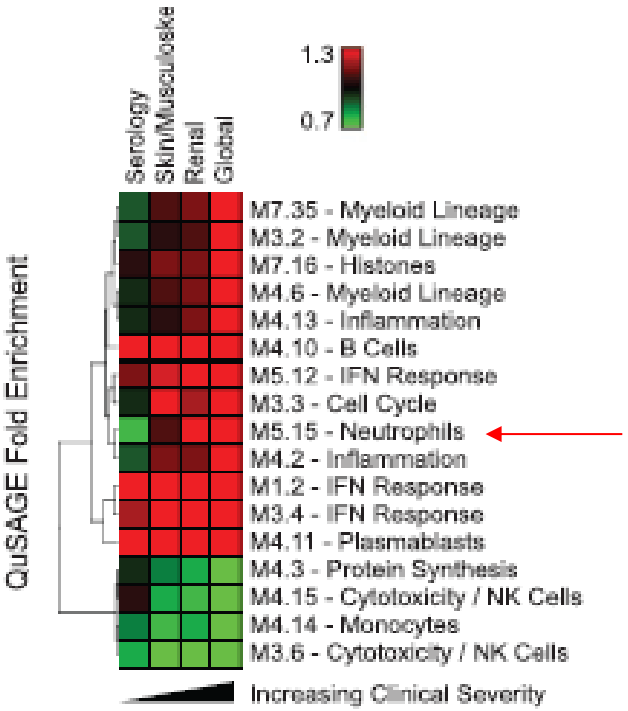
NT: No Treatment
HC: Hydroxychloroquine only
OS: Oral Steroid
MMF: Mycophenolate Mofetil + any
CIV: Cyclophosphamide / IV steroids

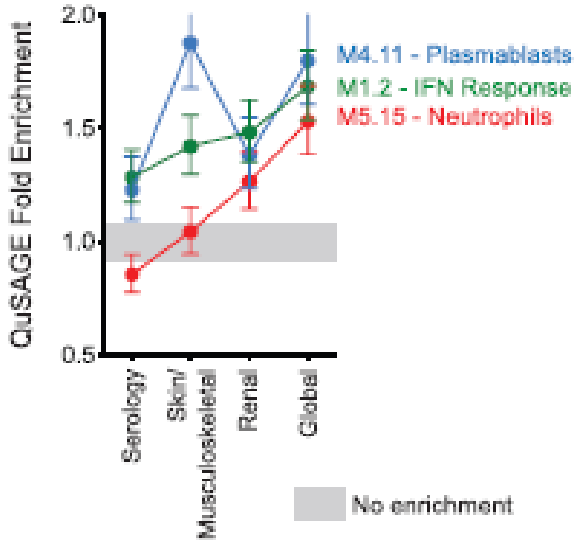


The **plasmablast signature was decreased by all treatments compared to no treatment**, but most strongly by (MMF) and CIV, two cytostatic drugs that suppress activated lymphocytes

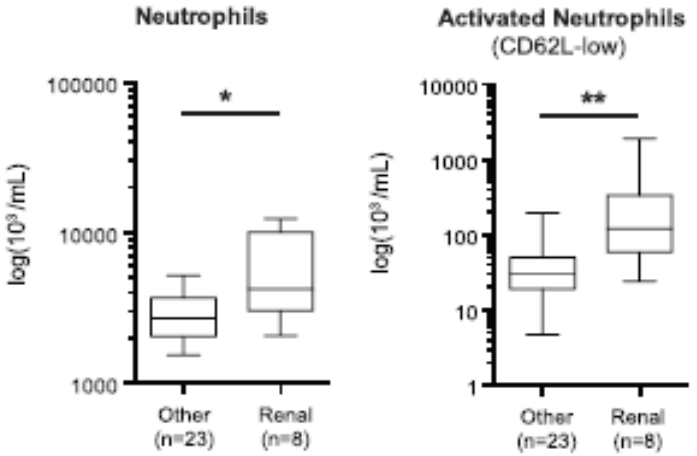
Cell. 2016 Apr 21;165(3):551-65

Modules associated with Disease Types



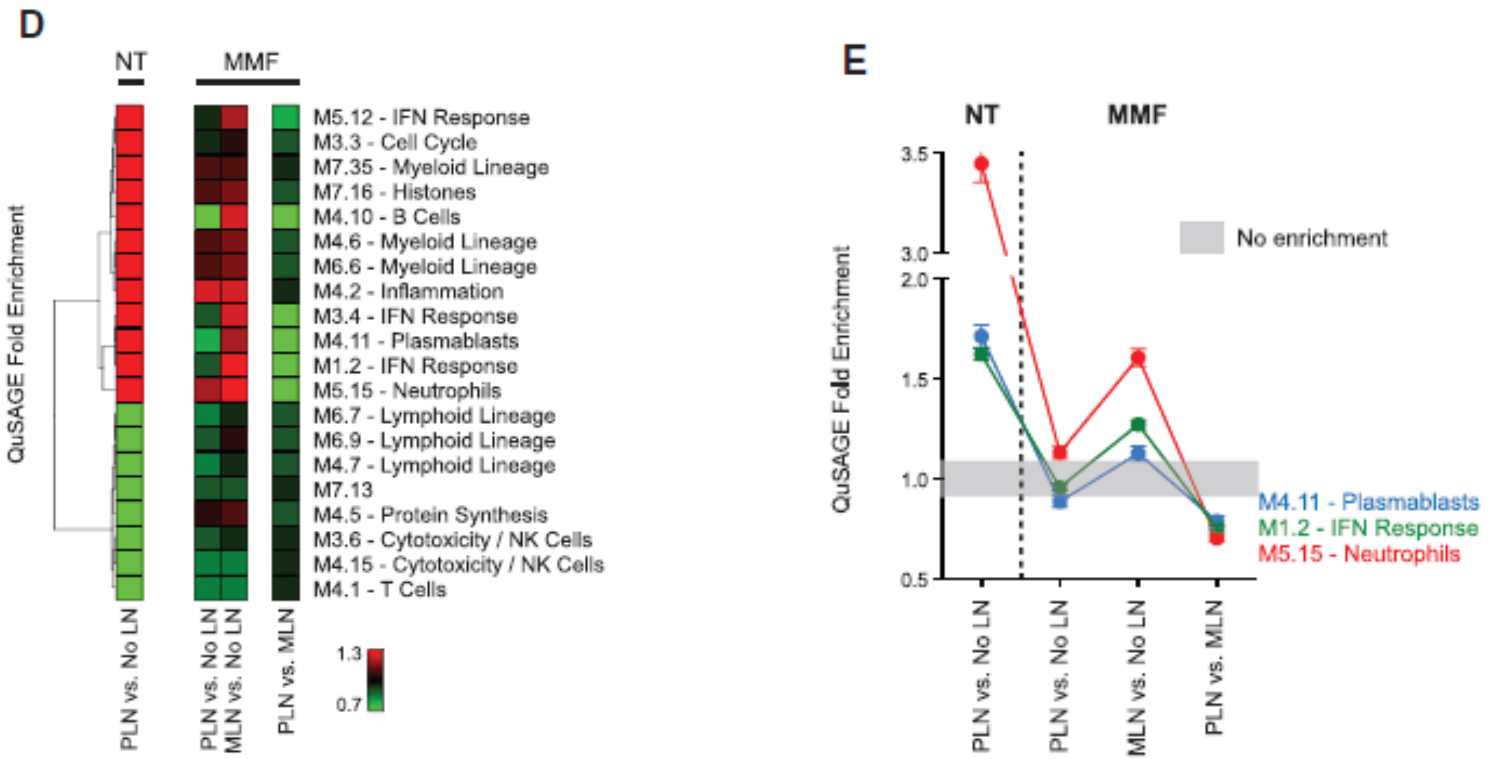


Neutrophil module is mostly associated with Lupus Nephritis



Cell. 2016 Apr 21;165(3):551-65

Modules associated with treatment in different nephritis sub-classes



Distinct treatment response signatures across different nephritis subclasses: PLN (proliferative nephritis) vs MLN (membranous nephritis) , treated with MMF (mycophenolate mofetil)

Cell. 2016 Apr 21;165(3):551-65

SLE Blood Transcriptional Fingerprint

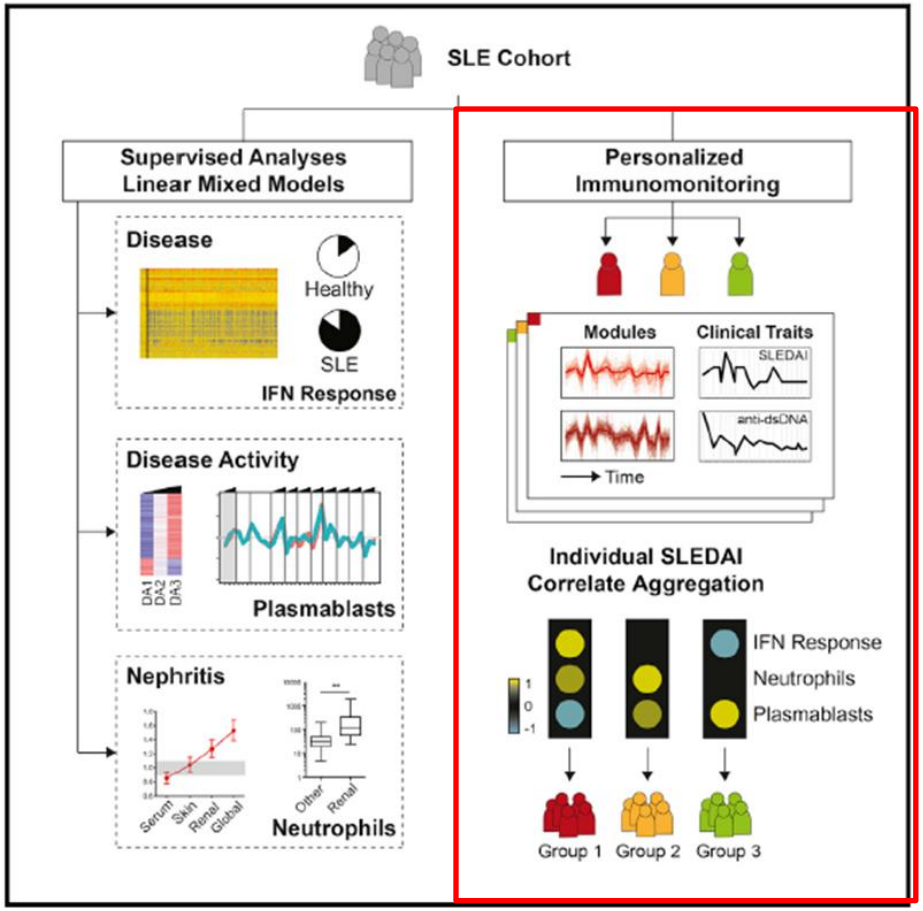
Genes associated to SLE, DA, Race, Treatment, disease subtypes

From genes to "blood" modules

WGCNA modules for each patient and correlation with SLEDAI

Linking WGCNA to blood modules for inference of biological function

Stratification of patients into groups

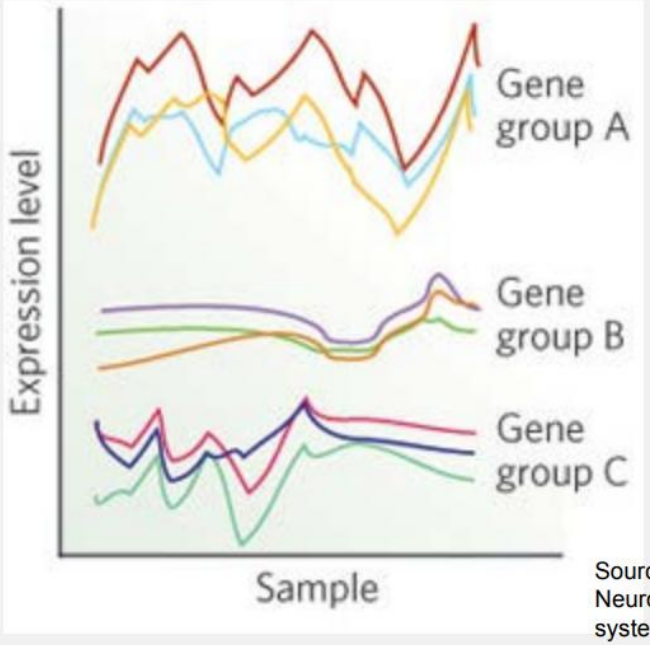


WGCNA – Weighted Gene Correlation Network

Aim of WGCNA: summarizing individual genes into modules, based on correlation

Modules found in WGCNA:

Groups of co-expressed genes (with similar expression profiles over a large group of individuals)



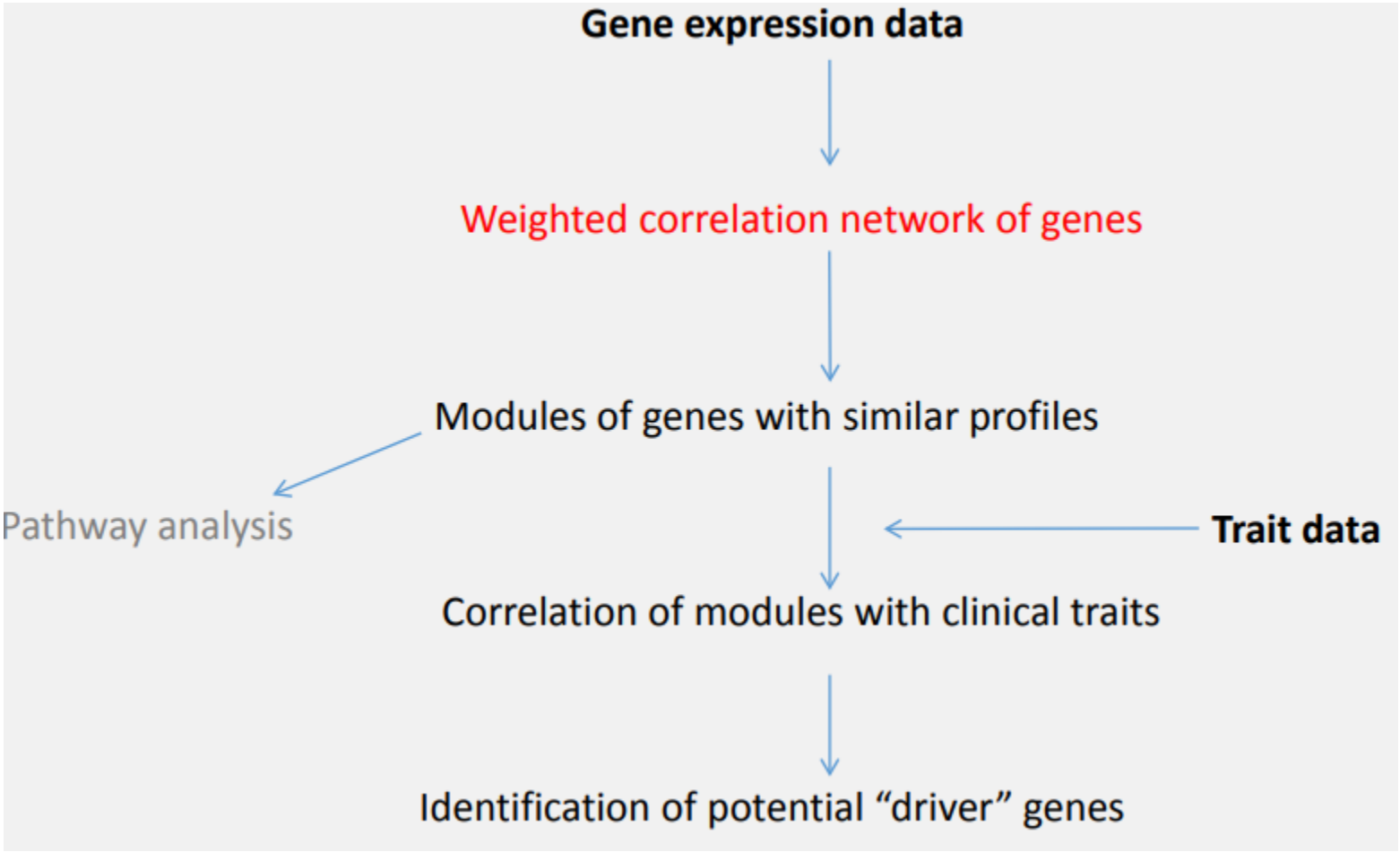
Source: Daniel H. Geschwind & Genevieve Konopka. Neuroscience in the era of functional genomics and systems biology, Nature 461, 908-915

Central Hypothesis:

Genes with **similar expression patterns** are of interest because they may be

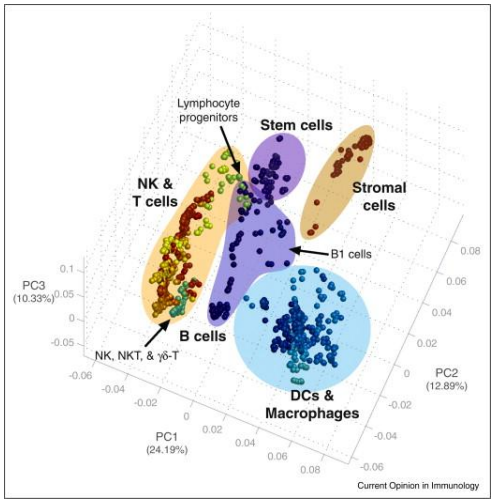
- tightly co-regulated
- functionally related
- members of the same pathway

WGCNA – Workflow



Measuring Modules with One Metrics: Eigengenes

	WGCNA MODULE X							
	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Gene 6	Gene 7	Eigengene Value
sample 1	17	55	80	41	3	70	70	A
sample 2	43	100	56	91	72	22	2	B

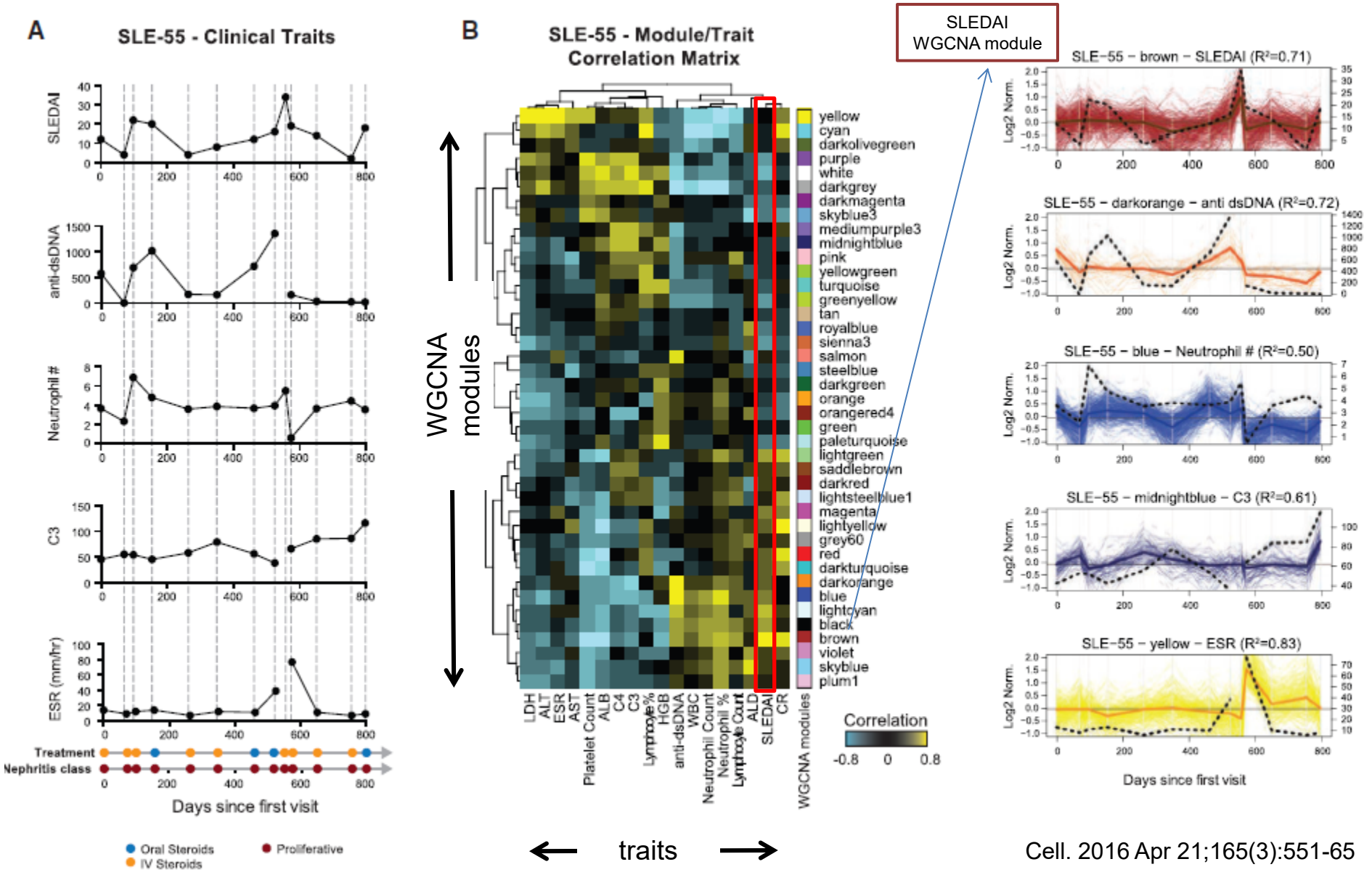


Kim and Lanier – Curr Opin Immun 2013

- Module's "eigengene" is the first principal component of the expression matrix of the corresponding module
- It can be used as summary score for a vector of genes in a sample

Patient 55:Linking WGCNA modules to clinical traits by correlation

Patient- specific modules of co-expressed transcripts over time are identified by WGCNA.
Analysis run on WGCNA modules (eigengenes) correlated with continuous clinical traits



Cell. 2016 Apr 21;165(3):551-65

Patient 55: Linking WGCNA Modules to Blood Modules

Genes associated to SLE, DA, Race, Treatment, disease subtypes

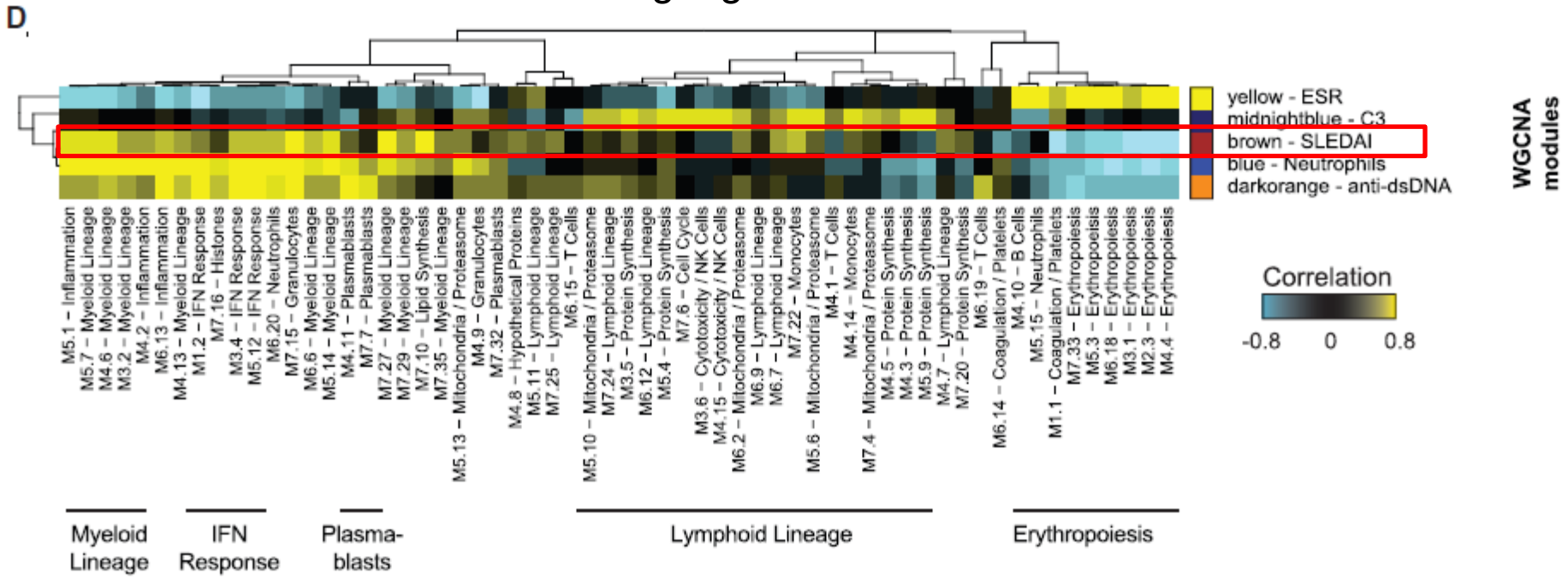
From genes to "blood modules"

Extract WGCNA modules for each patient and correlation with SLEDAI

Linking WGCNA to blood modules for inference of biological function

Stratification of patients into groups

Correlation between patient SLE-55 WGCNA and blood modules using eigengenes



Cell. 2016 Apr 21;165(3):551-65

SLE Blood Transcriptional Fingerprint

Genes associated to SLE, DA, Race, Treatment, disease subtypes

From genes to “blood” modules

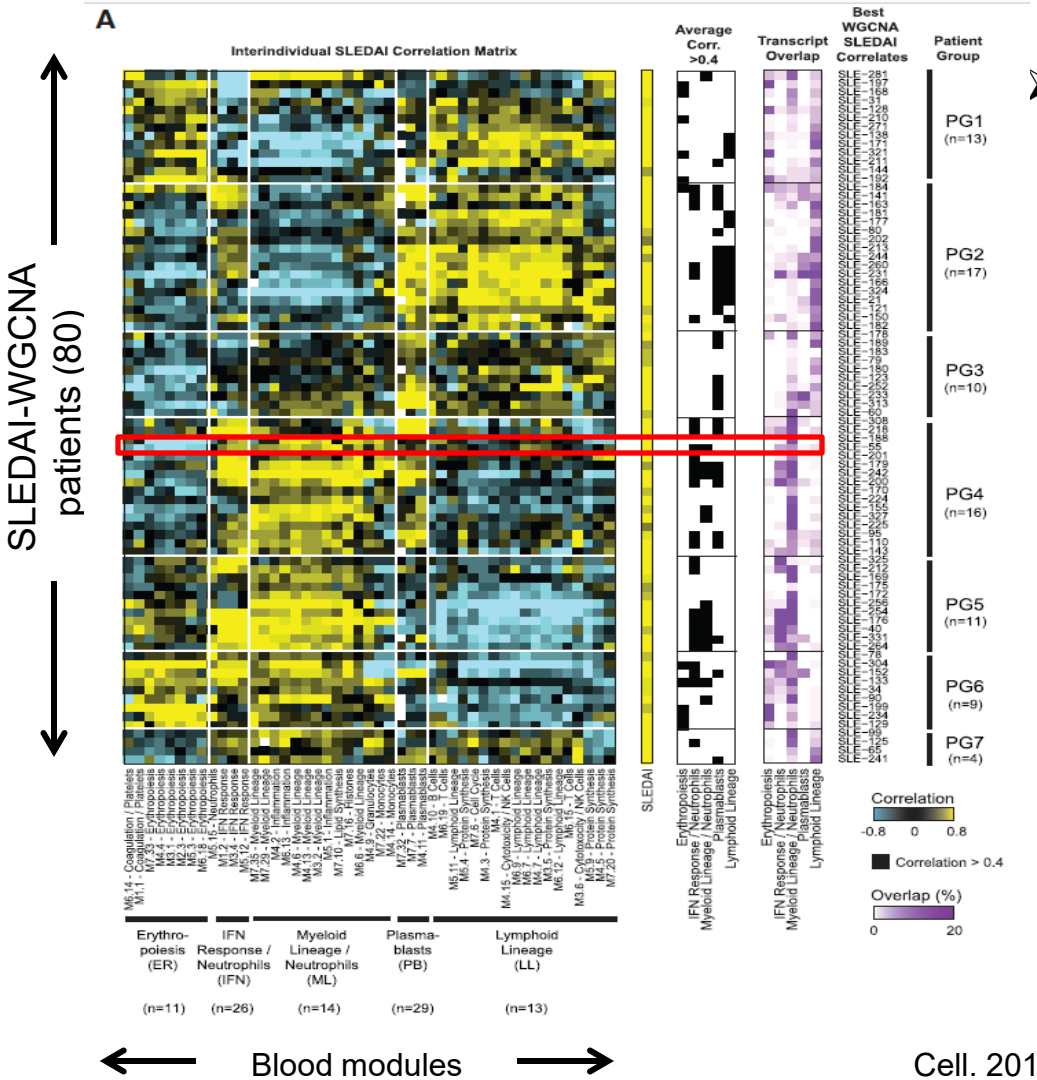
WGCNA modules for each patient and correlation with SLEDAI

Linking WGCNA to blood modules for inference of biological function

Stratification of patients into groups

Stratification of SLE Patients Based on Transcriptional Correlates of SLEDAI

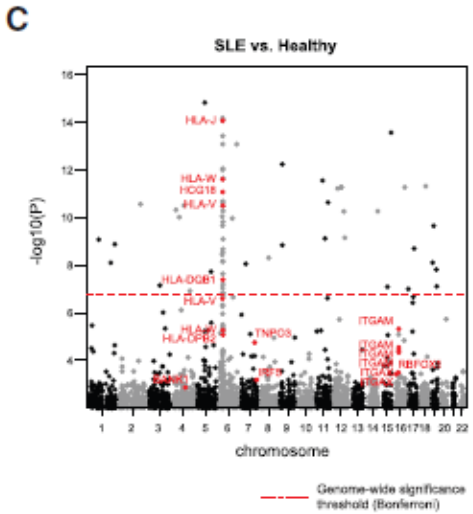
matrix of correlation between the SLEDAI WGCNA and blood modules, for all patients



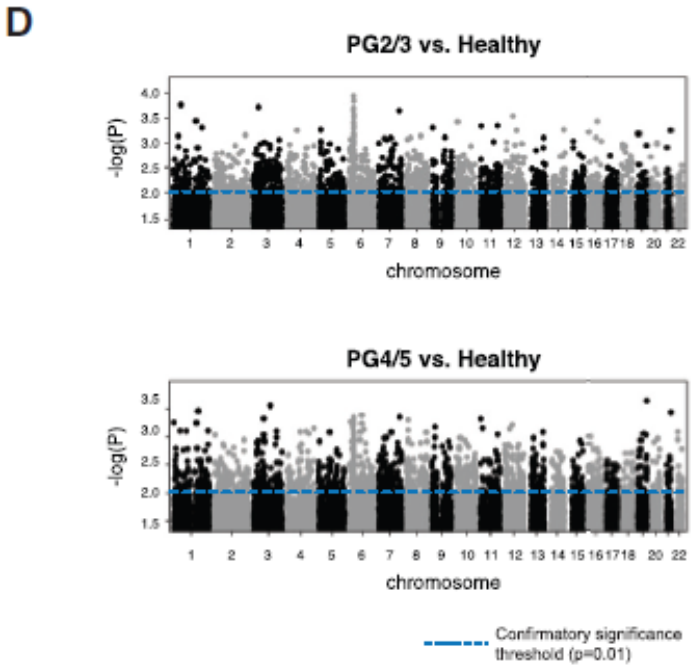
- 7 patient groups identified based on the combination of blood immune signatures that best correlate with the SLEDAI-WGCNA.

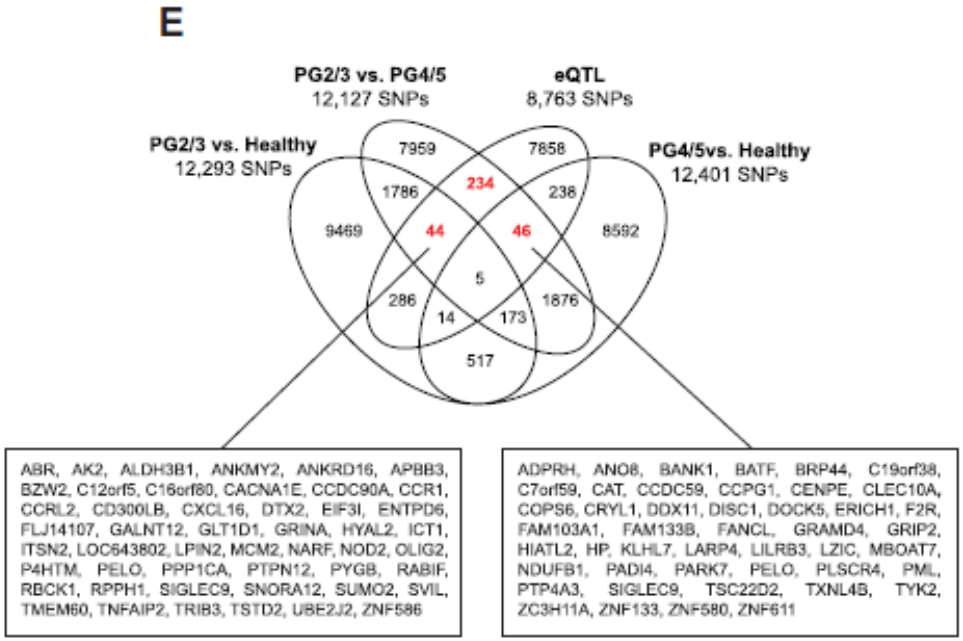
Cell. 2016 Apr 21;165(3):551-65

Genetic Analysis of Patients Groups



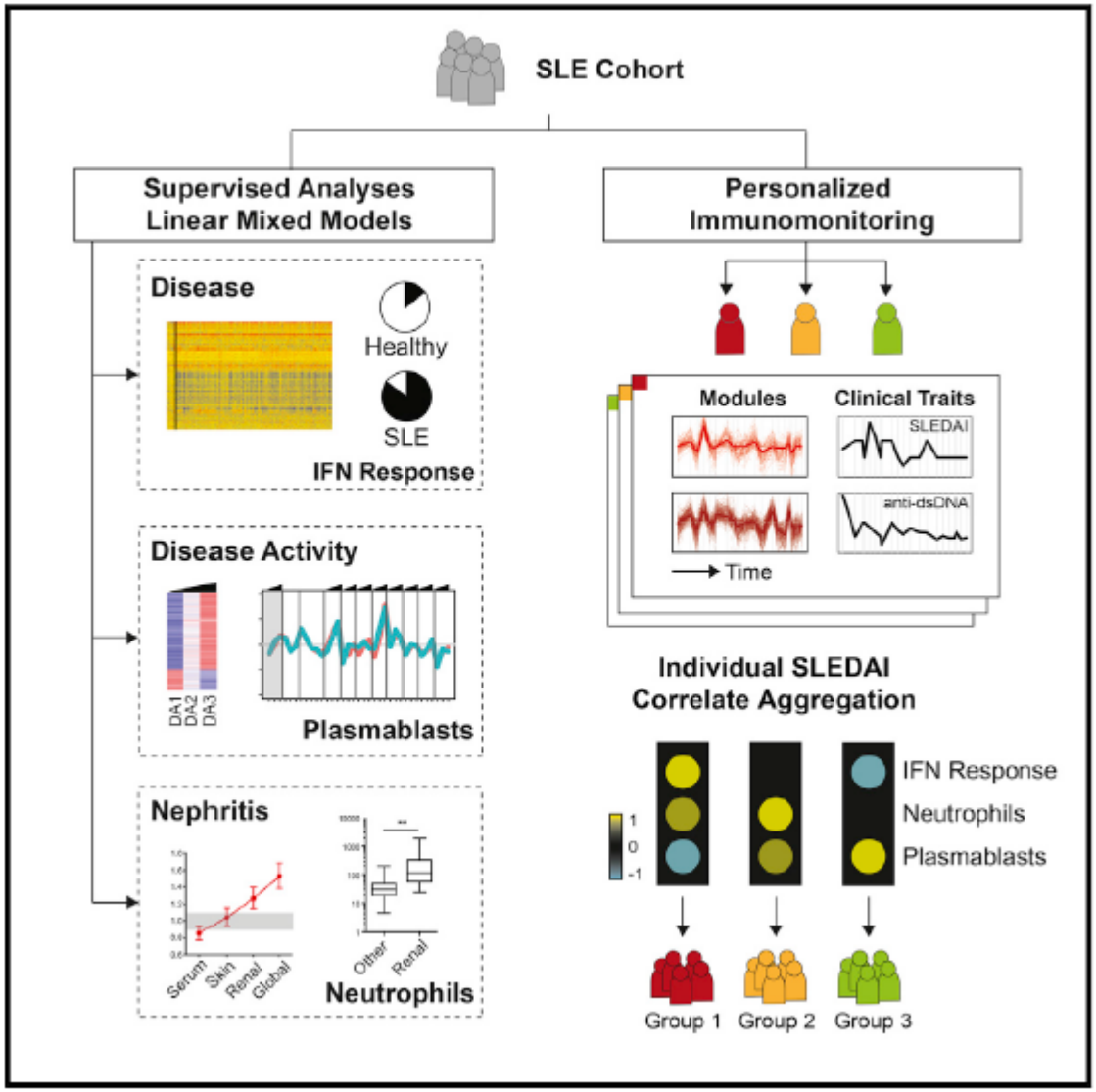
- To find a genetic basis for these clusters, SNP analysis was conducted (135 patients + HCs)
- SNPs differentiating between PG2/3 and PG/4/5 were found
- Intersection with eQTL SNPs (SNPs associated with DA genes expression) points to IFN-inducible genes





Cell. 2016 Apr 21;165(3):551-65

Let's Recap

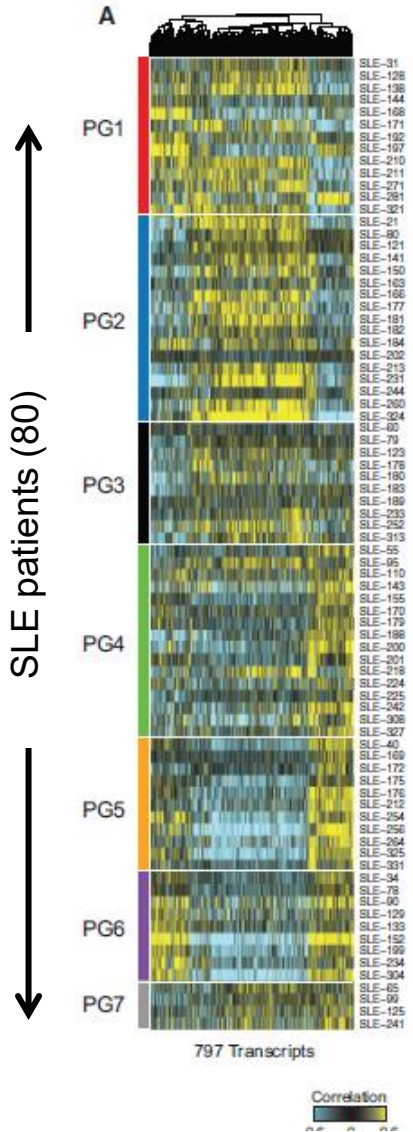


- Supervised analysis, linear models and blood module analyses led to identification of: i) IFN response module in SLE ii) Plasmablasts associated to DA iii) neutrophils associated to nephritis
- Different immune signatures correlating with SLEDAI (7 groups)
- Groups supported by different genotypes
- Supports the development of customized treatment strategies.

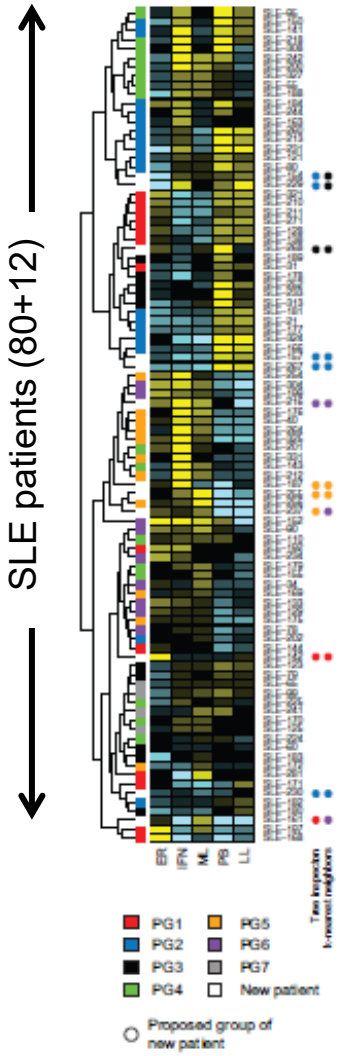
A Targeted Panel for SLE Patient Stratification

Can we find a gene panel which allows direct patients stratification ?

- Hierarchical clustering of the 797 transcripts differentially correlating with SLEDAI between the seven patient groups



- Assignment of novel 12 patients to each of the groups “guilt-by-association”



Cell. 2016 Apr 21;165(3):551-65

Study Summary

- Clinical and transcriptional profiling of 158 lupus patients up to a period of 4 years
- IFN, Plasmablast and Netrophil signatures driving SLE
- Neutrophil-related signatures associate with progression to active nephritis
- Molecular correlates of disease activity stratify patients into seven major groups
- Molecular stratification may improve the outcome of clinical trials in SLE

Cell. 2016 Apr 21;165(3):551-65

Take Home Messages

- Interpreting biological meaning of ~25K genes in a heterogeneous sample such as blood – starting from bulk data – is extremely complex!

- One way to address this is to aggregate this information into functional building blocks (i.e. dimensionality reduction):
  - "blood modules" : of general use, obtained from previous analysis of 239 disease blood samples
  - "WGCNA modules": patient specific and obtained from correlation analysis of the SLE study data set
- Blood modules can be (tentatively) assigned a biological meaning looking at their gene content

- Patient-specific WGCNA modules can be associated to clinical phenotypes based on longitudinal correlation

- For each sample, each module can be measured with one number (e.g. % of genes up- or down-regulated in the module, or eigenvalue). This approach has the statistical advantage of reducing number of tests

- Patients' SLEDAI can be associated with different modules. On this basis patients can be stratified into classes, presumably having different underlying biology and needing different treatments

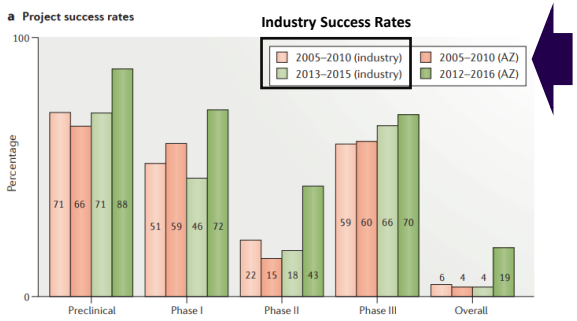
Single-cell RNA sequencing of human tissue supports successful drug targets

Emma Dann[*1,2], Erin Teeple[*3], Rasa Elmentaite[2], Kerstin B Meyer[1], Giorgio Gaglia[3], Frank Nestle[4], Virginia Savova[3], Emanuele de Rinaldis[3+], Sarah A Teichmann[1,2,5]

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
2. Ensocell Therapeutics, BioData Innovation Centre, Wellcome Genome Campus, Hinxton, Cambridge, UK
3. Precision Medicine & Computational Biology, Sanofi Research US, Cambridge, MA, USA
4. Research & Development, Sanofi, Cambridge, MA, USA
5. Theory of Condensed Matter, Cavendish Laboratory/Dept Physics, University of Cambridge, JJ Thomson Ave, Cambridge, UK

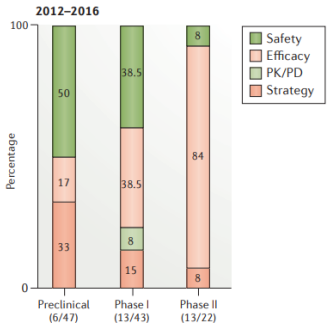
*Authors contributed equally +Co-Corresponding and senior authors

Motivation: how can single cell data support decision making in target discovery?



AstraZeneca (AZ) performance improvement after instituting a new drug development framework focusing on 5 priority areas: target, tissue, safety, patient, potential (5Rs). Note that Phase II success rates remain below half of projects.

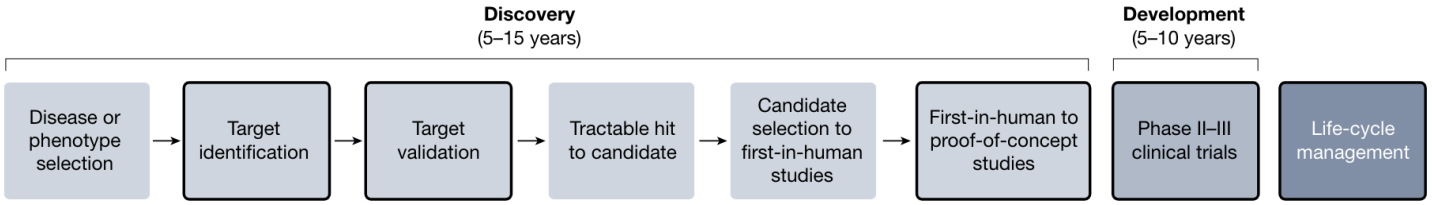
Morgan et al. *Nat Rev Drug Discov* 2018



2012-2016 reasons for failure by project phase. There exists a persistent need for new information to inform drug development work.

Morgan et al. *Nat Rev Drug Discov* 2018

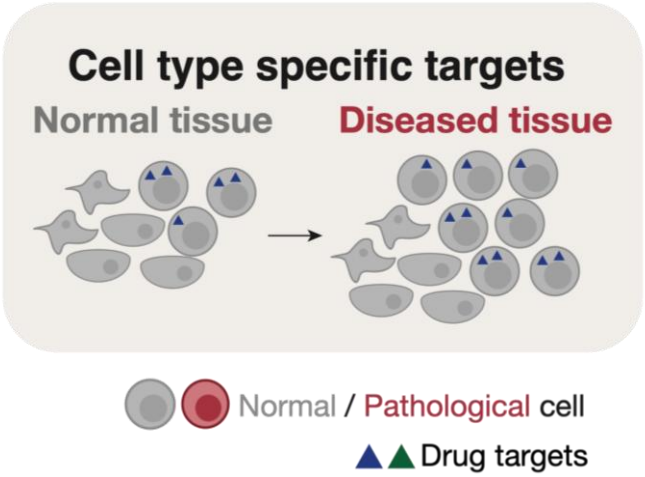
90% of drug development failures stem from inadequate target selection for a disease

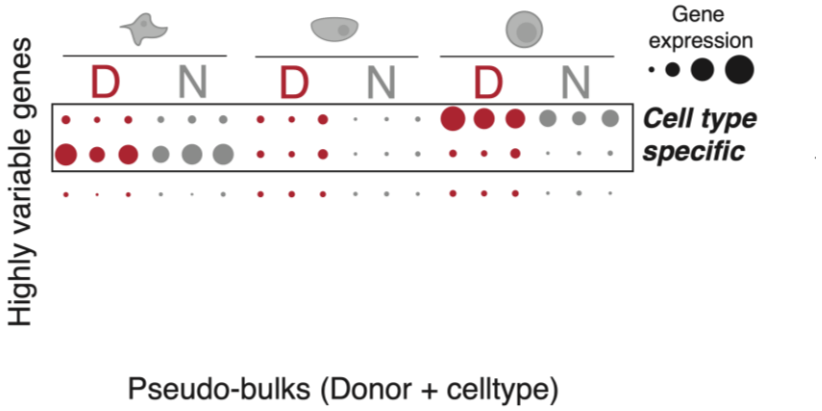


How can single-cell RNA-seq boost the selection of promising drug targets ?

sanofi

scRNAseq support: cell type-specific expression in disease-relevant tissue

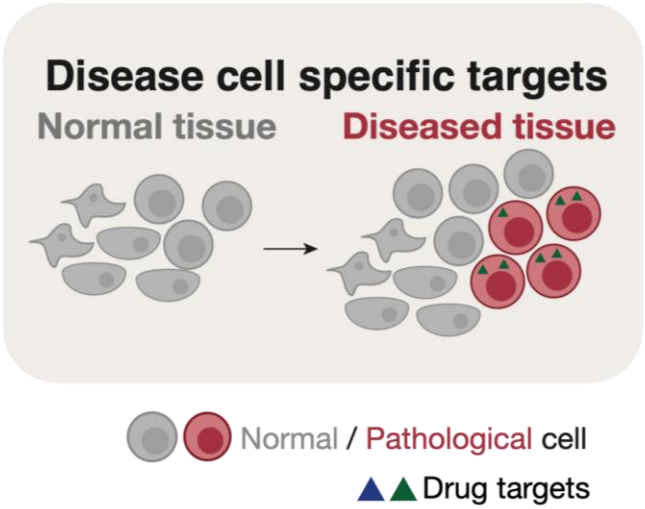


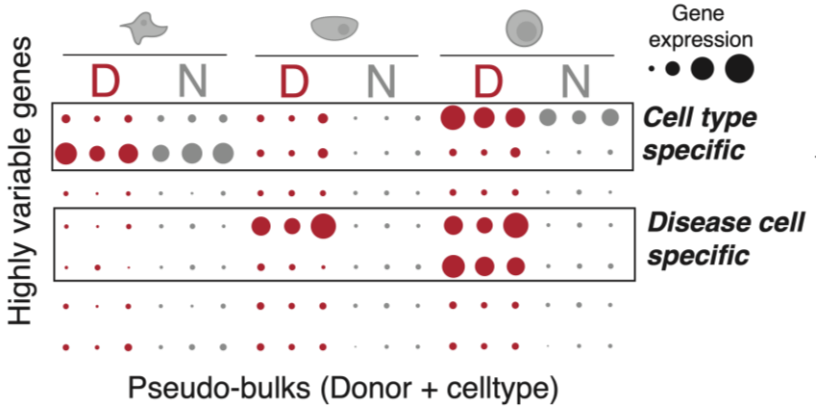


Cell type specific support defined as differential expression across cell types in disease.

sanofi

scRNAseq support: cell type-specific expression in Disease (D) versus Normal (N)

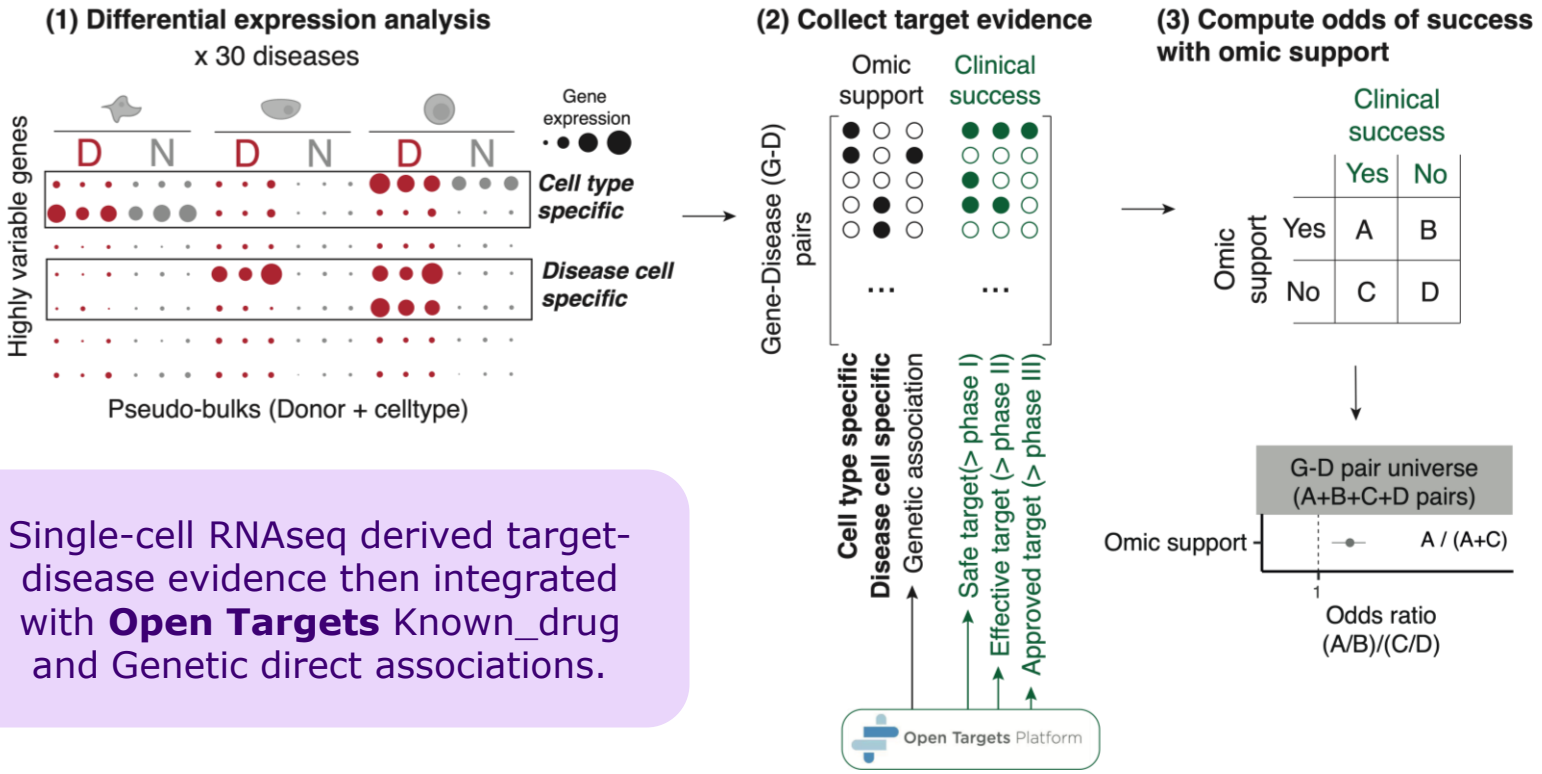




Disease cell specific support defined as differential expression in cell types in disease vs normal.

sanofi

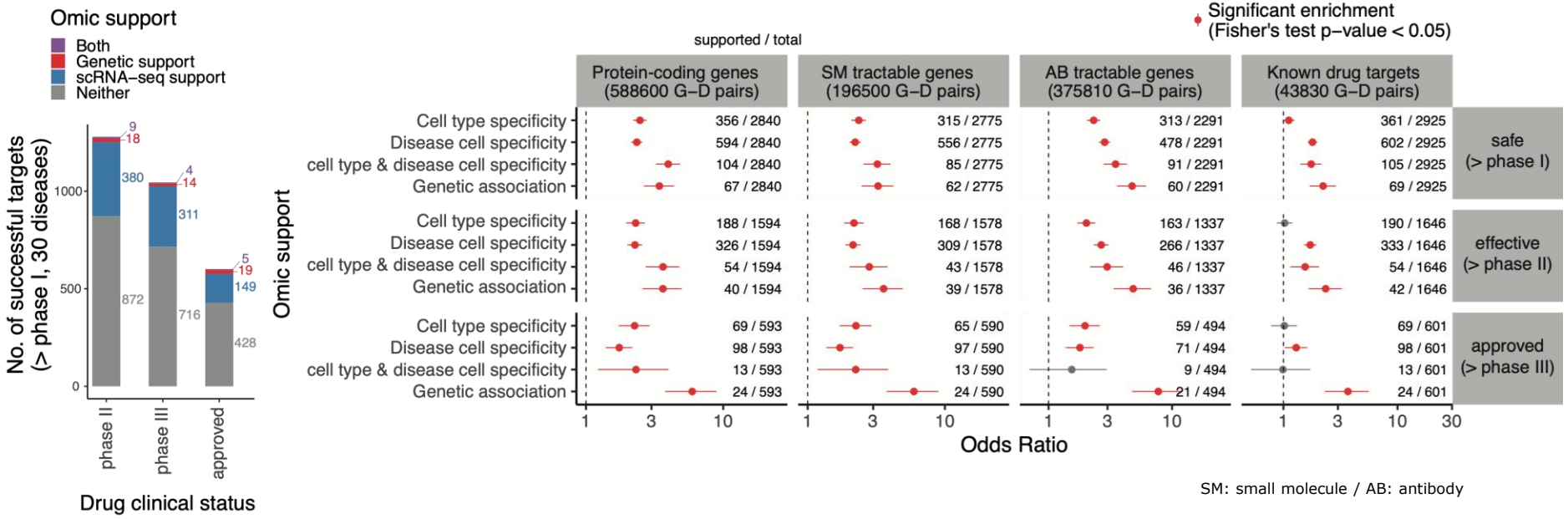
Approach: Evaluating scRNA-seq support for known drug targets



Single-cell RNAseq derived target-disease evidence then integrated with **Open Targets** Known_drug and Genetic direct associations.

sanofi

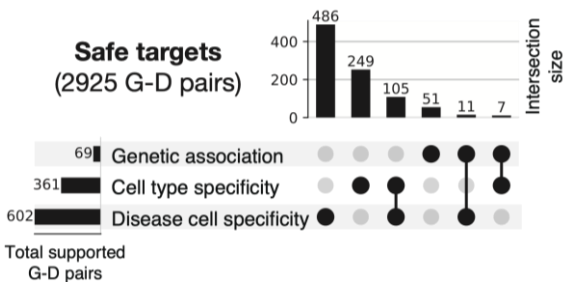
Targets with single-cell support have 3x higher odds of clinical success

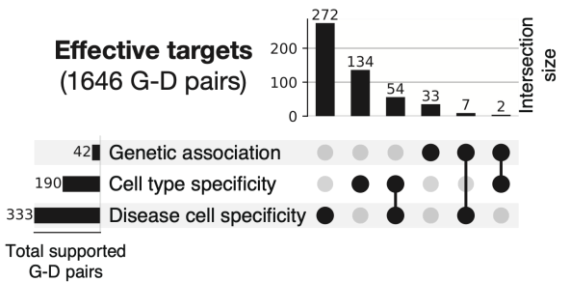


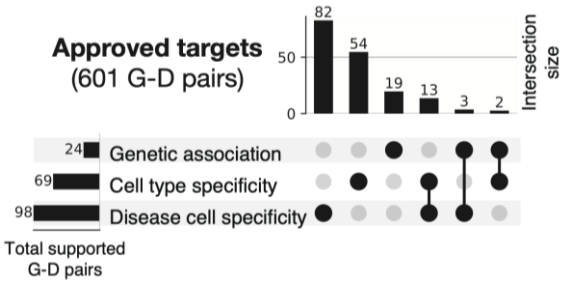
SM: small molecule / AB: antibody

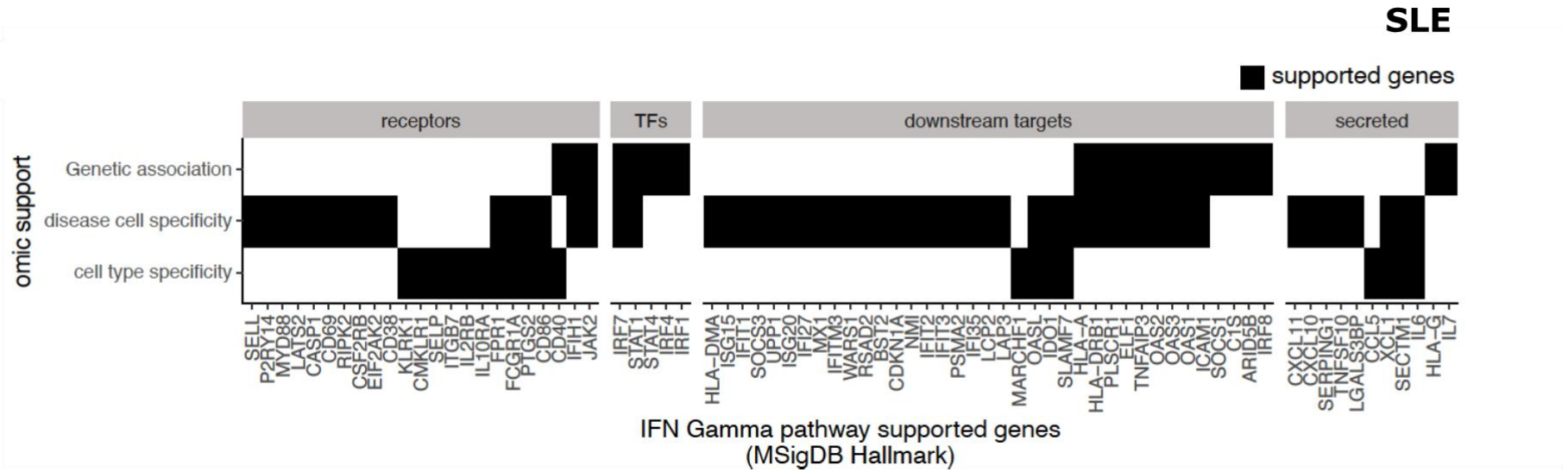
sanofi

Single-cell support identifies a target space complementary to direct genetic evidence









sanofi

Summary

- Retrospective analysis of association between cell type-specific expression and target success.
- scRNA-seq support is found to be associated with greater odds of clinical trial success
- scRNA-seq support identifies a target space complementary to genetics, with distinctive molecular and druggability characteristics.

- **Next steps:**
  - Expand the number of diseases included in the analysis by matching to disease-relevant tissue (not all will have disease versus control evidence analysis)
  - scRNAseq is a rich data source – additional association evidence approaches can be considered

sanofi

Q & A

My Best Reference for Systems Immunology: ImmunoScope

Explore

Settings

ImmunoSCOPE
1,418 Tweets

Follow

ImmunoSCOPE
@Immunoscope

Building an immune compendium #immunology

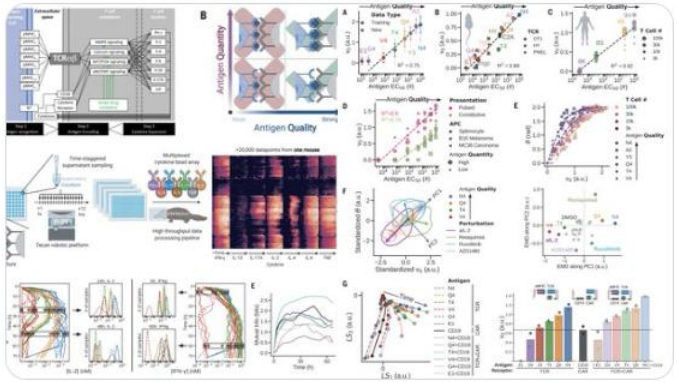
Paris, France Joined May 2018

1,328 Following **7,792** Followers

Tweets Tweets & replies Media Likes

ImmunoSCOPE @Immunoscope · May 19

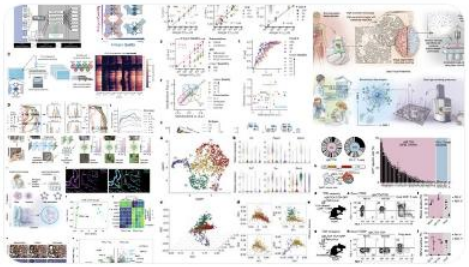
Importance of antigen quality rather than quantity in predicting dynamic T cell responses | #CD8 #TCR #antigen #timecourse | Grégoire Altan-Bonnet @g_altanbonnet @p_r_francois @trademaker17 @McGillUPhysics @NCIResearchCtr | Science @ScienceMagazine bit.ly/3wvUUko



4 36

Search Twitter

Privacy Policy, including Cookie Use.



You might like

Nature Immunology @NatImmunol — Follow

AAI @ImmunologyAAI — Follow

Science Immunology @SciImmunology — Follow

Show more

What's happening

WNBA · LIVE

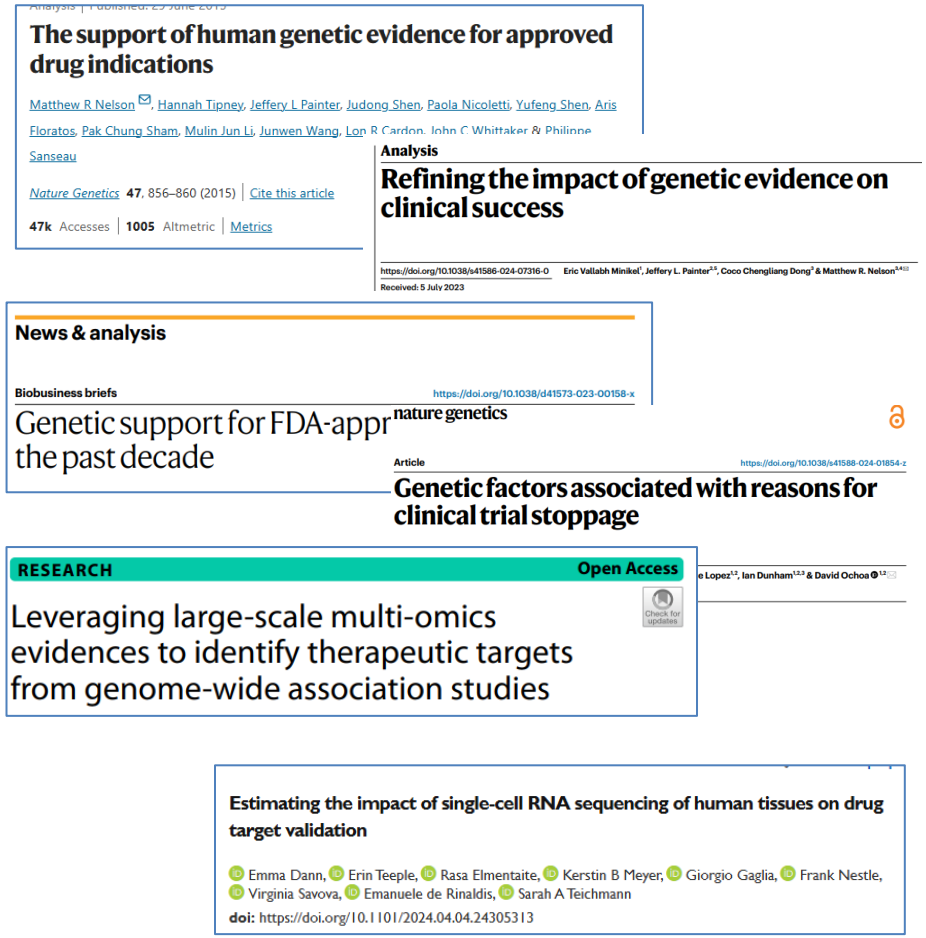
Don't miss what's happening
People on Twitter are the first to know.

Log in Sign up

Why Pharma Is Investing in Genetics

Population Genetics & Genomics

Key Pillars of Target ID Strategy – 2015-2025



- Genetically supported targets have twice the chances of success in clinic
- 2/3 FDA approved drugs are supported by genetics
- Genetics predicts safety of drug targets
- Genetics predicts drug target direction of effect
- Single cell data identifies targets more likely to enter in clinical development