

# OMICS : a new era in medicine

After the development of the human (and mice) genome project in the 90's:

Medicine has become increasingly data-rich, and many of the emerging fields producing such large-scale data designated by adding the suffix '-omics' onto previously used terms.

The importance for the life-science community of such large-scale approaches is reflected in

- number of articles,
- Citations
- companies offering to perform OMICs services
  - using robotic approaches

# Omics fields

Genomics : Study of the genomes of organisms.

Epigenomics : Study of the packaging of the genome (accessibility)

Transcriptomics : of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA, produced in one or a population of cells.

Proteomics : study of the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or system.

Lipidomics : study of the entire complement of cellular lipids,

Glycomics study of the glycome (sugars and carbohydrates)

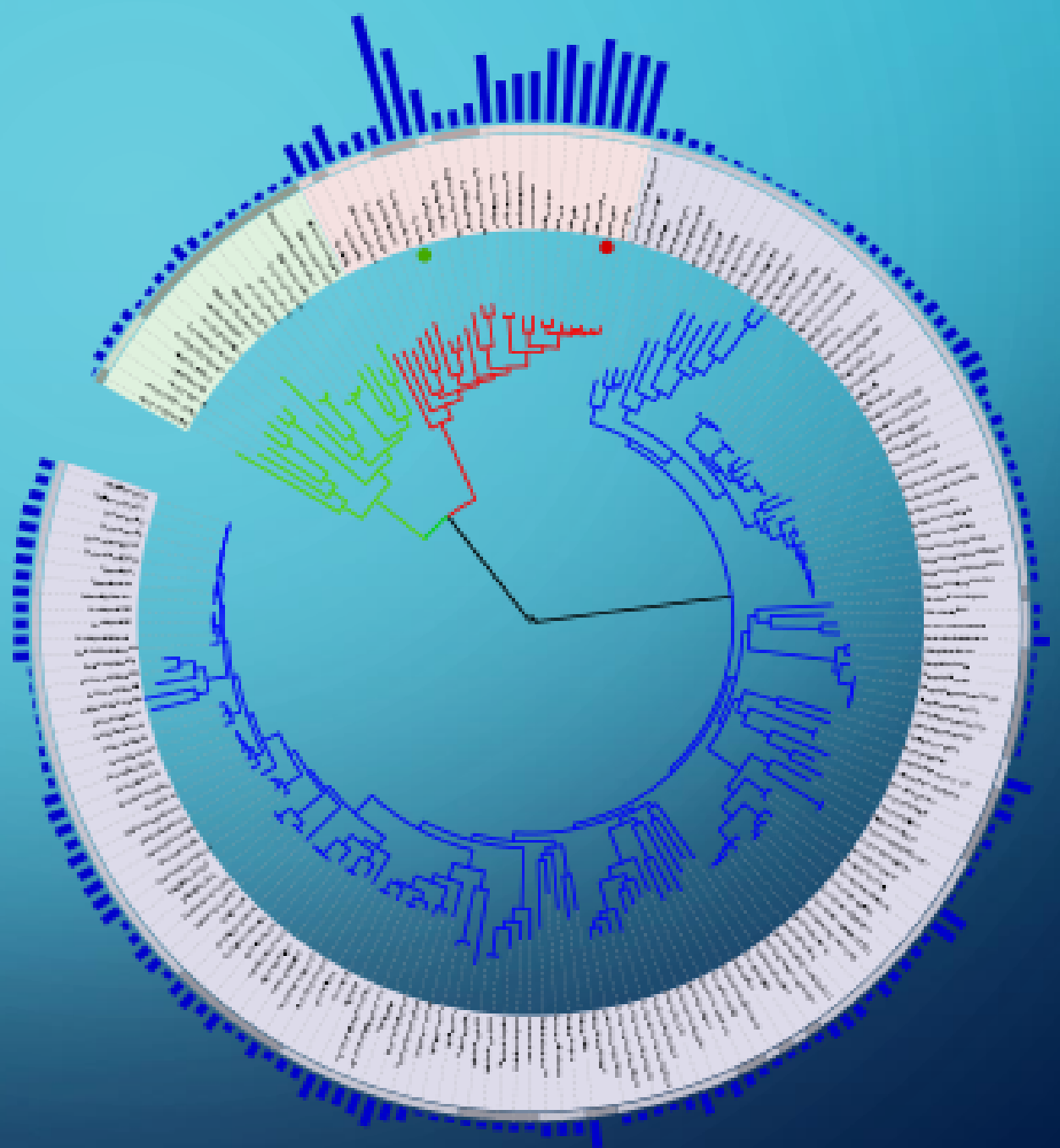
Metabolism : study of chemical processes involving metabolites, quantitative measurement of the dynamic multiparametric metabolic response of living systems

Microbiomics: the study of the communities of microorganisms that live in the digestive tract.

Pharmacogenomics investigates the effect of variations in the human genome on drug's metabolism

# Size of genomes :

Eucaryotes (bacteria, virus)  
Procaryotes (plant, animals)  
Archeo-bacteria



# GENOMICS



In human :

Genome is  $3400 \times 10^6$  bp long.

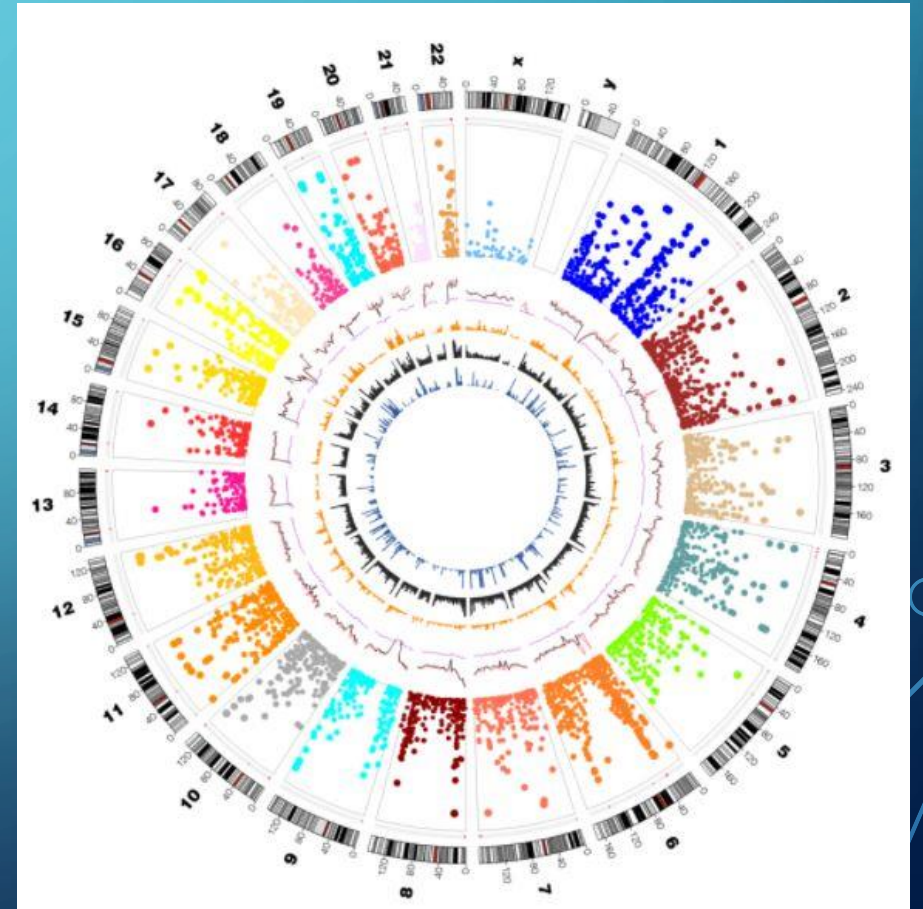
Codes about  $\sim 25\,000$  genes (2%)

23 pairs of chromosomes

Polymorphism are variations in the genome code that, over evolution, allowed the diversification of species

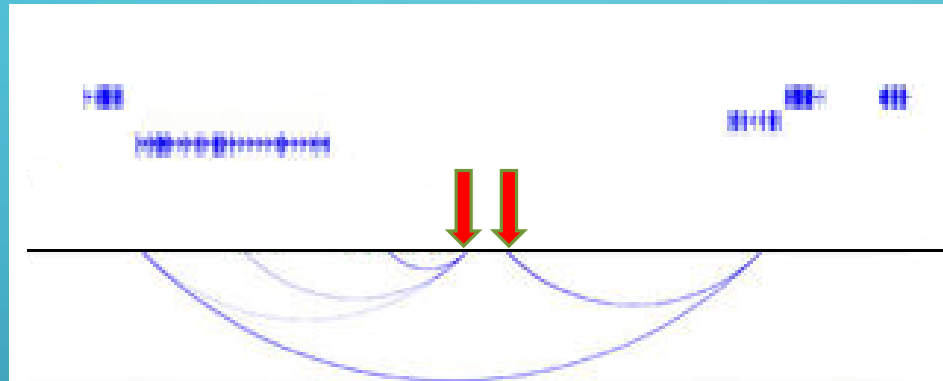
Over  $2.5 \times 10^9$  small changes of a single nucleotide polymorphisms (SNPs) are spread over the 23 pairs of chromosomes.

Analysing the frequency of these SNPs allows association of certain variations with diseases





Most SNPs lie nearby genes (not in the coding part)  
Identifying the functionality that may be altered is very difficult.  
DNA folds in many loops  
Some SNPs may work at a distance



**Genes**

**SNP associated with a disease but in between genes ↓**  
**A “loop” analysis can reveal a remote effect of the SNP on a gene far away**

# EPIGENOMICS

Each tissue (lung, brain, liver....) uses :

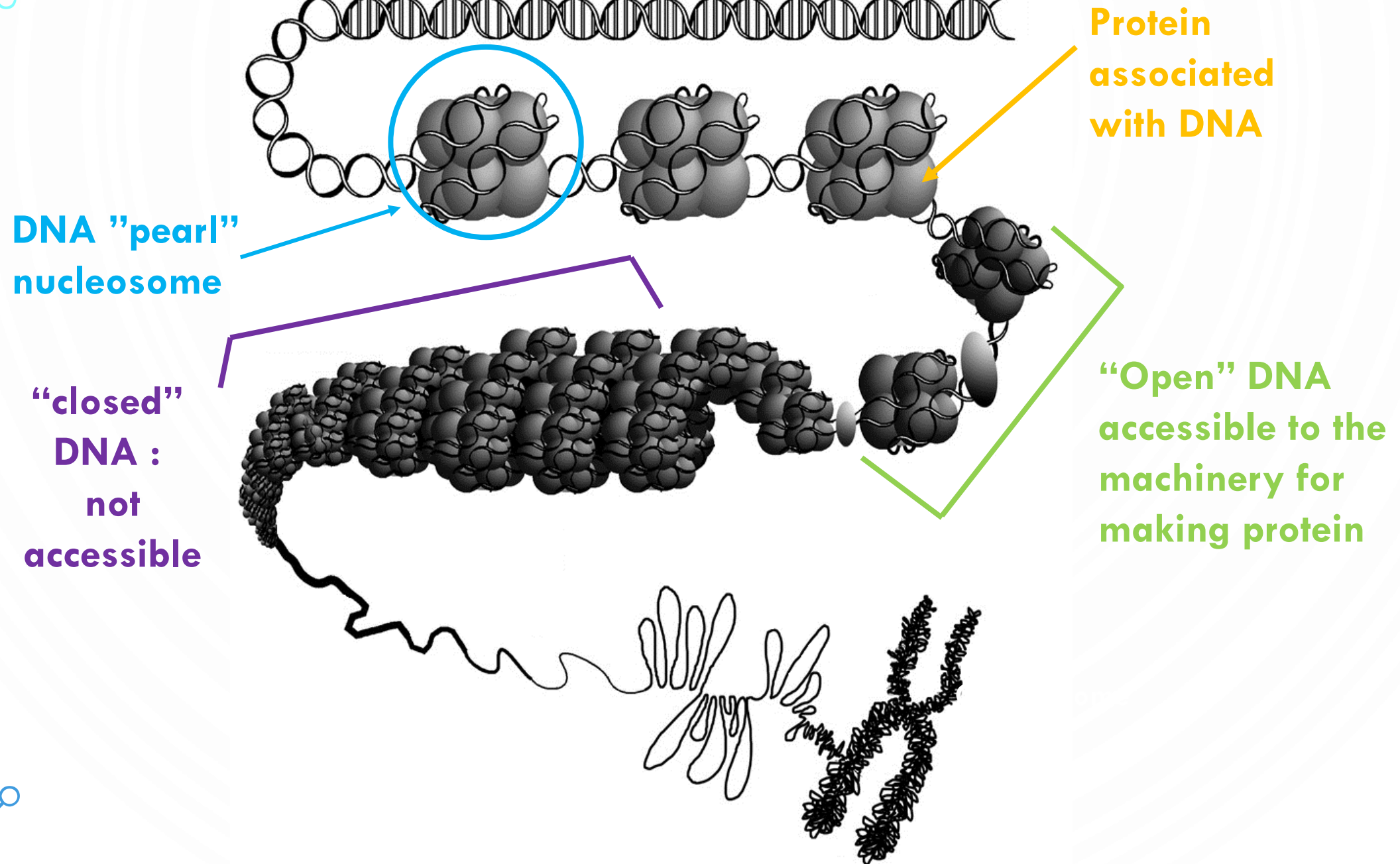
- a small but very highly specific proportion of these genes to fulfil a function
- a common “life support” part

Used part of the genome needs to be accessible

Un-used part need to be “stored” away

- $2.0 \times 10^{13}$  meters
- 70 trips from the earth to the sun and back
- Codes about **~25 000** genes (2%)
- **98%** of our genetic information is not coding for protein.

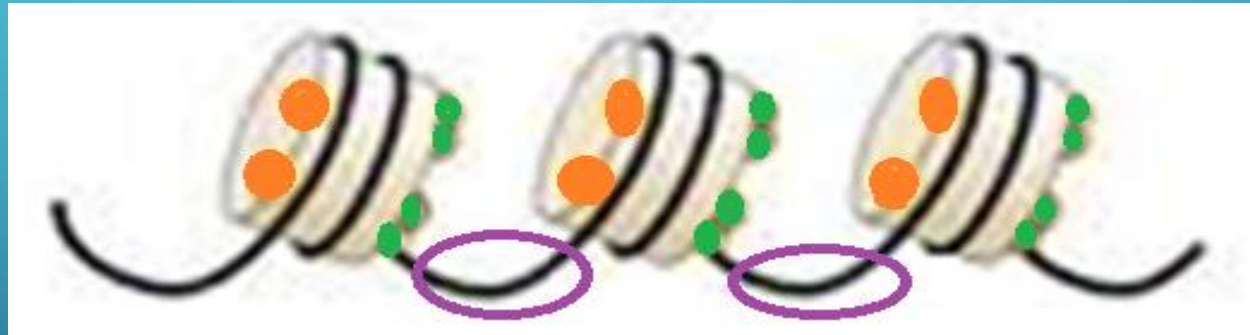
## Double helix DNA





“closed”  
conformation

DNA methylation  
 $-\text{CH}_3$

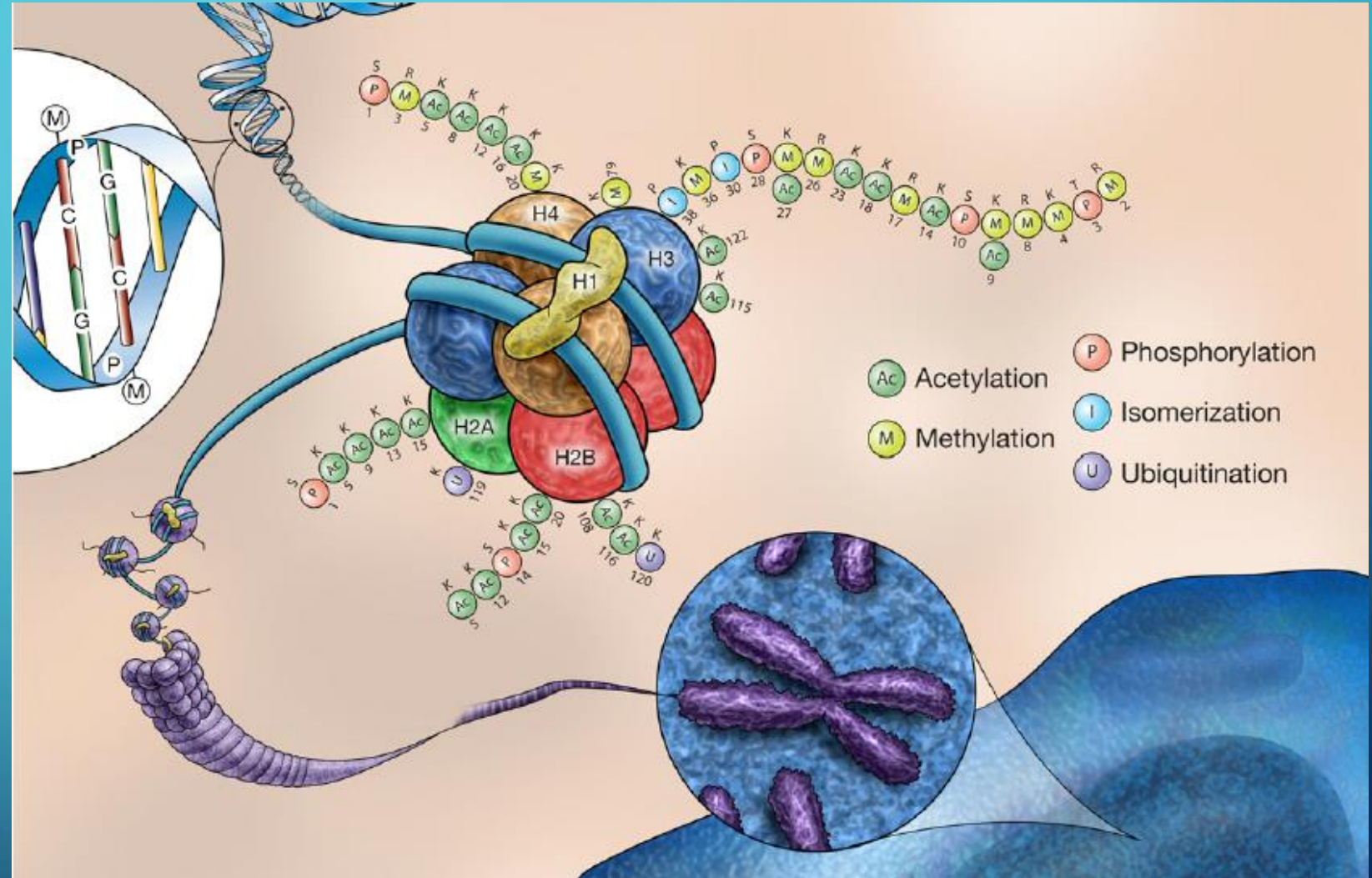


“opened”  
conformation

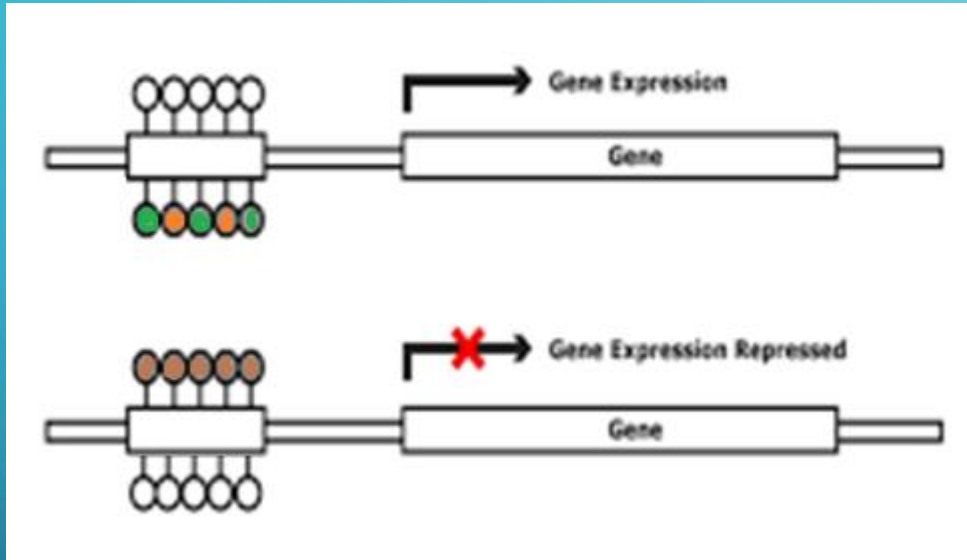
Accessibility



# Histone modifications



No DNA methylation  
+ Histone modification



+ DNA methylation  
No Histone modification

480.000 DNA methylation sites on a chip  
Spread over 25.000 genes regulatory regions  
and inter gene regions

480.000 data point for YES/NO methylation  
status for each cells in the sample

Allow to see if diseases change the accessibility  
of the DNA

Cancer :

Whole region closed : anti-tumour genes

Whole region are open : regions that change  
the function of the tissue

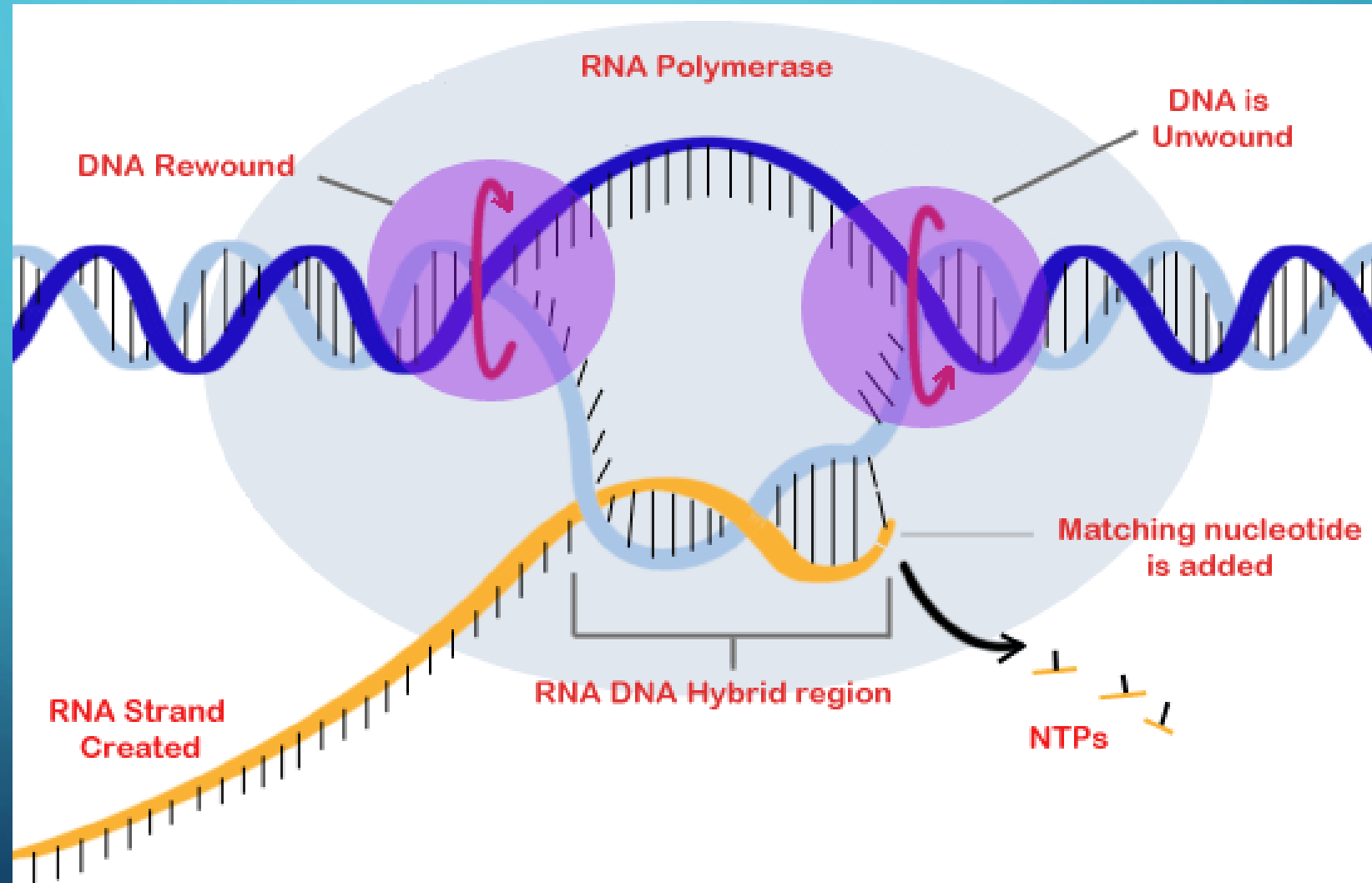
➤ metastasis

25.000 genes , Some active, some silent

Need to transfer “code” from DNA (in the nucleus) to Proteins factories in cells

➤ NEED a MESSENGER : mRNA

## Transcriptomics

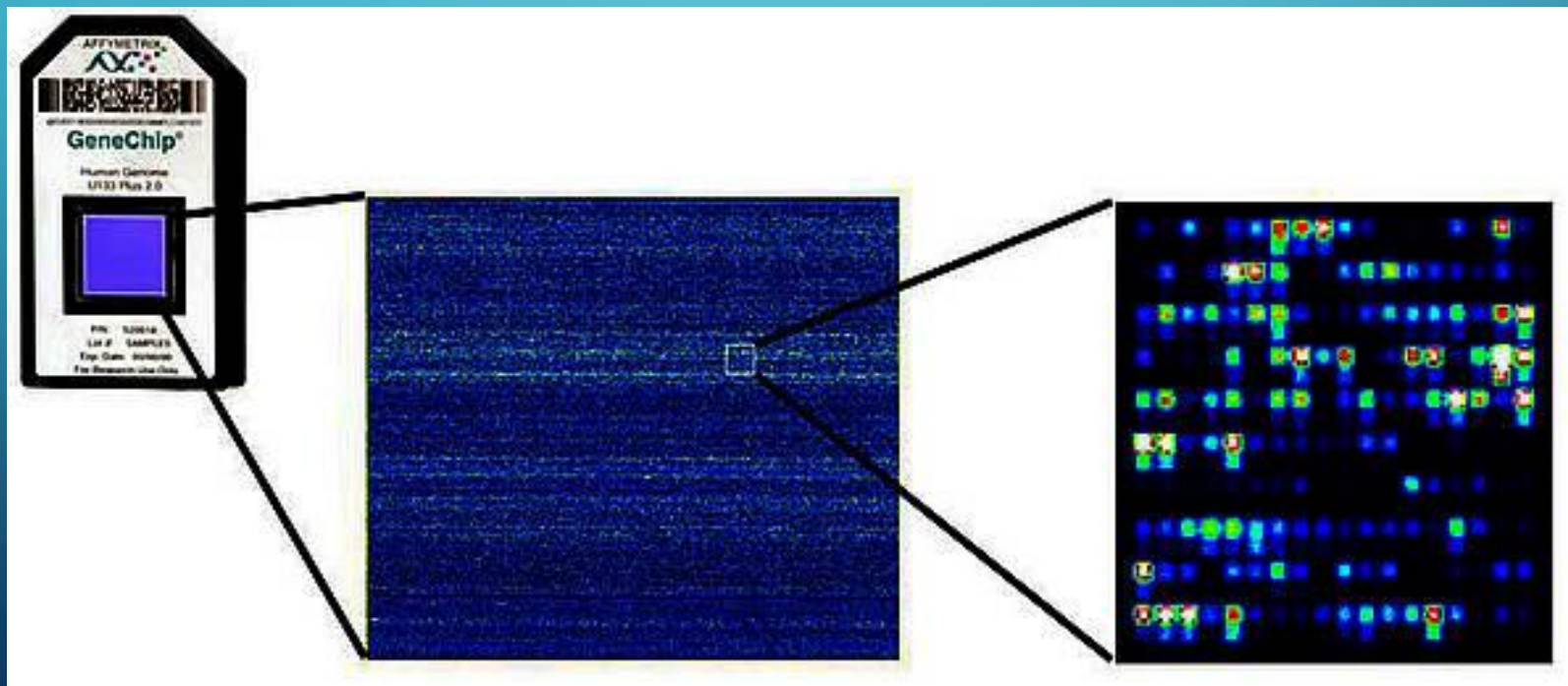


# Transcriptomic : the analysis of 25.000 genes simultaneously in a cells population

8 different region of a mRNA spotted on a chip

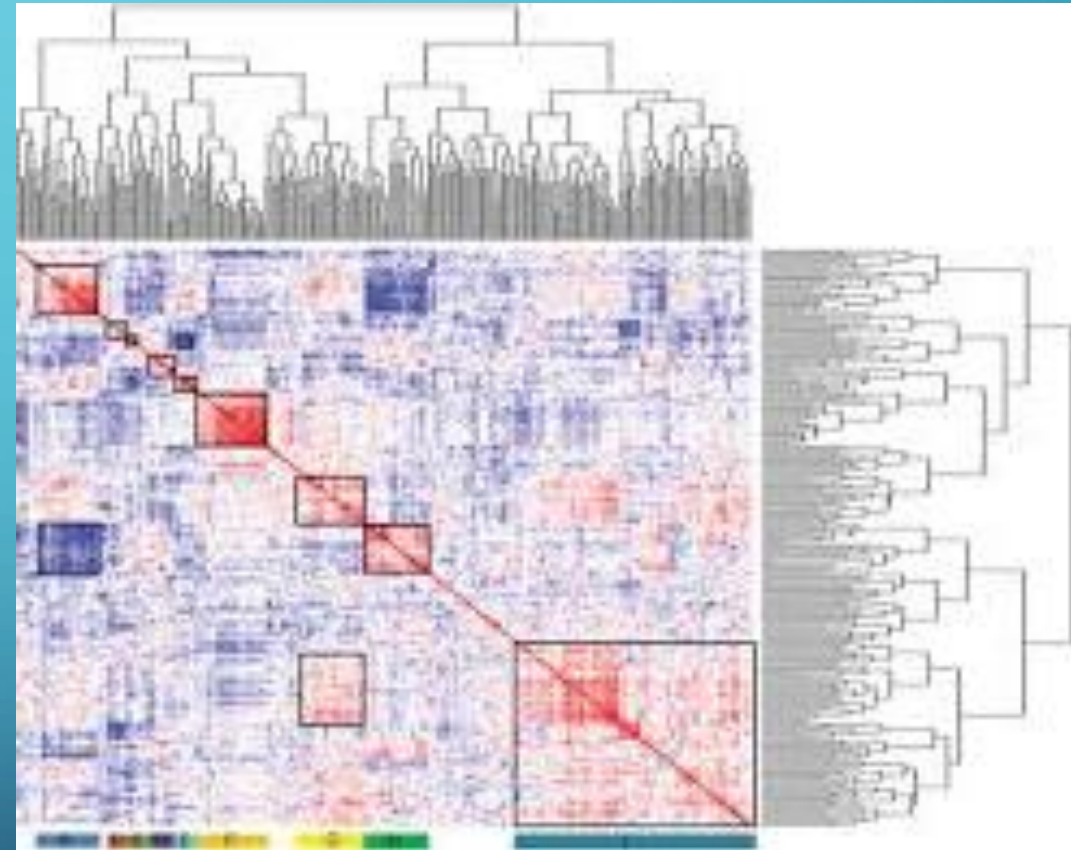
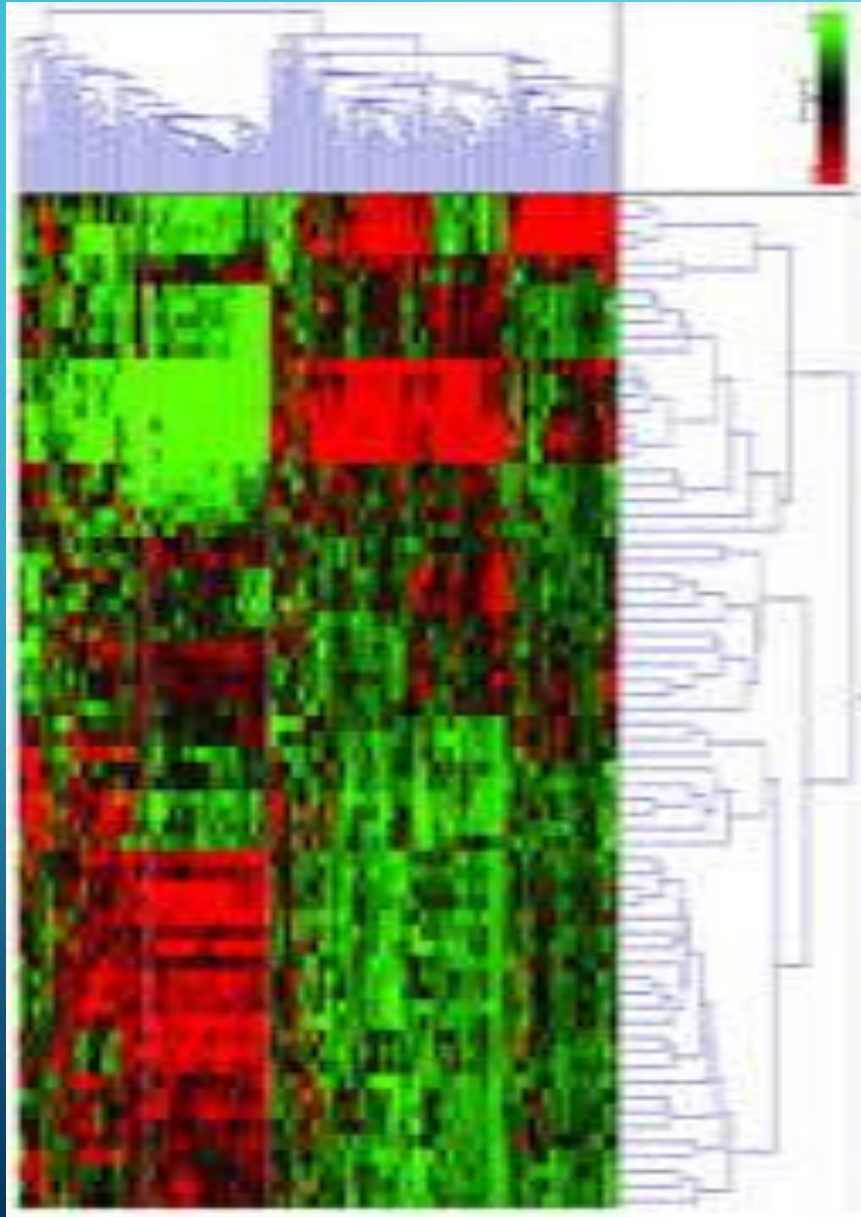
X 25.000 genes

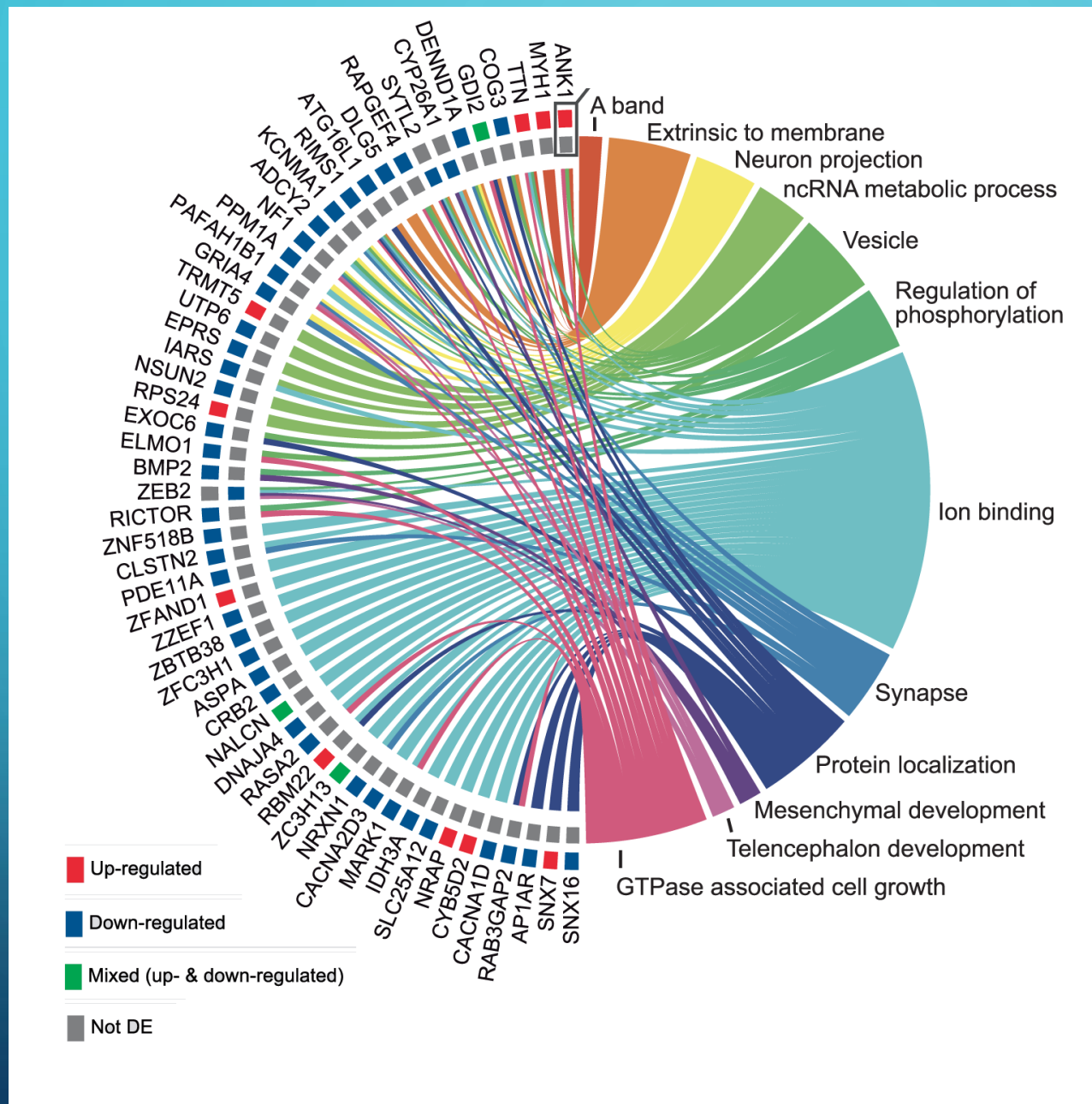
Each giving a fluorescence signal => average 8 signals





Preliminary analytics is similar : PCA, Clustering, heat maps...



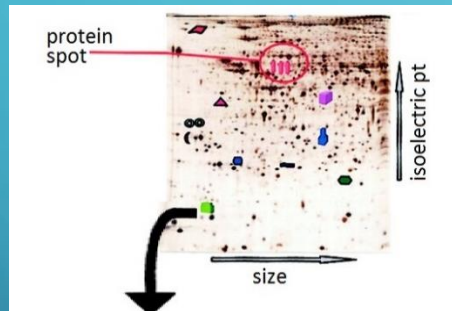


Allocate a function to groups of genes

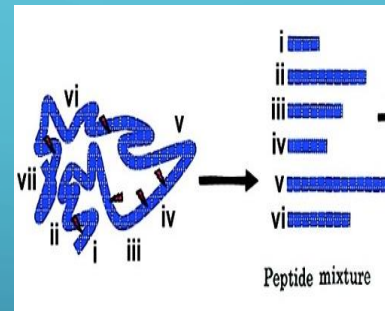
# PROTEOMICS



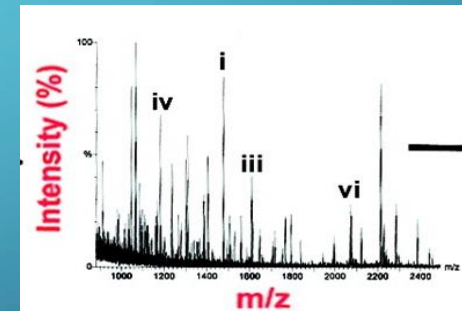
Mixture of different proteins



Spot excision



Digestion into small pieces



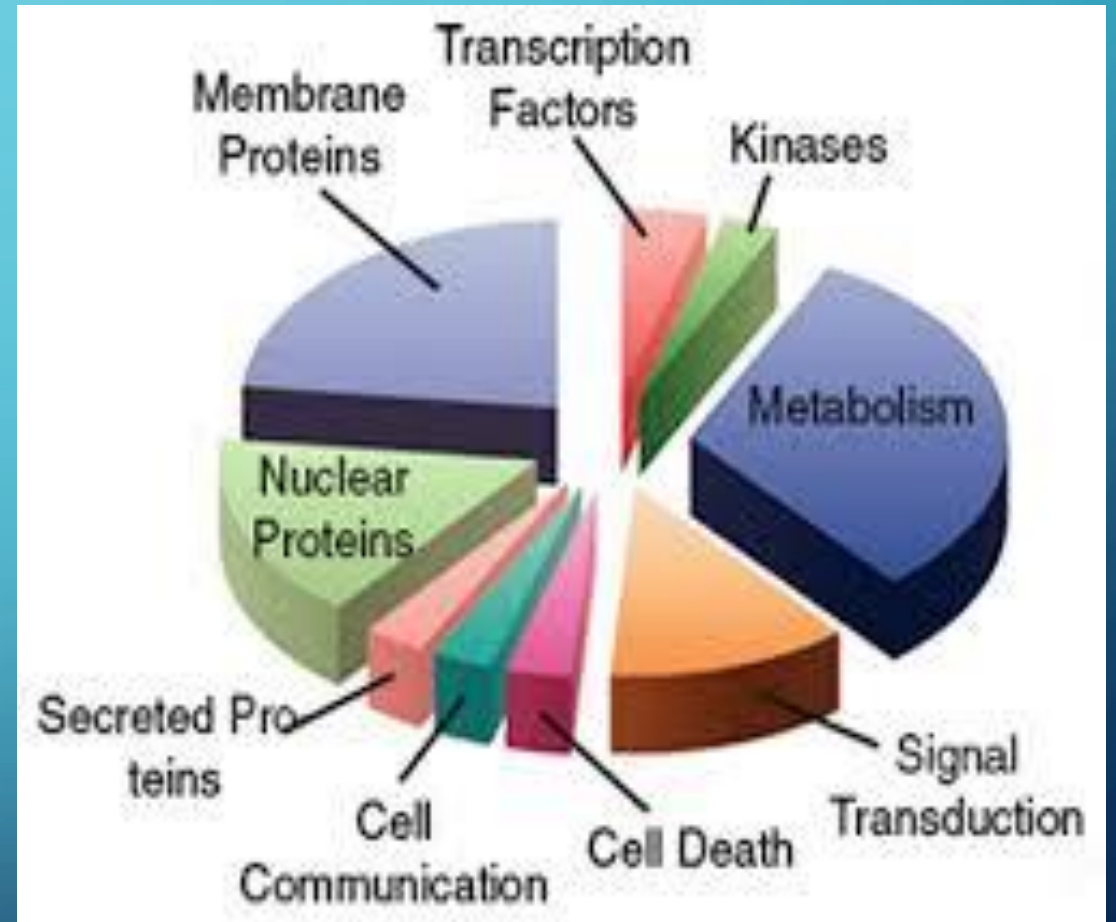
Mass spectrometry



Databases of MS spectres

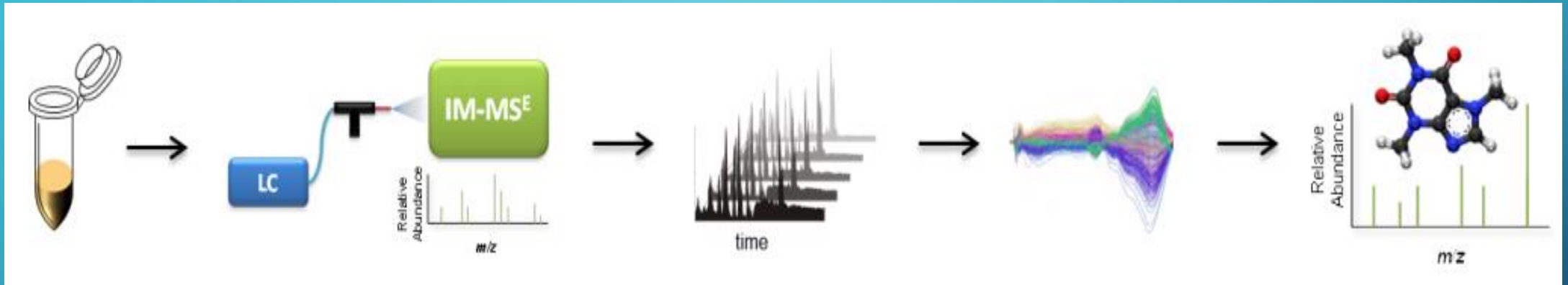


# About 30.000 proteins, with different functions





# Metabolomics : serum or urine



Identification of the individual metabolites

Profiles : qualitative and quantitative data

# Where do that leaves us ?

2.5  $10^9$  SNPs

Millions of potential DNA-loops

>  $10^6$  DNA methylation sites (so far technically measurable)

~ 25.000 pattern of histone modification for each gene regulation region

~ 25.000 levels of expression (further complications) for each gene

Over 30.000 proteins which can then be modified to be activated

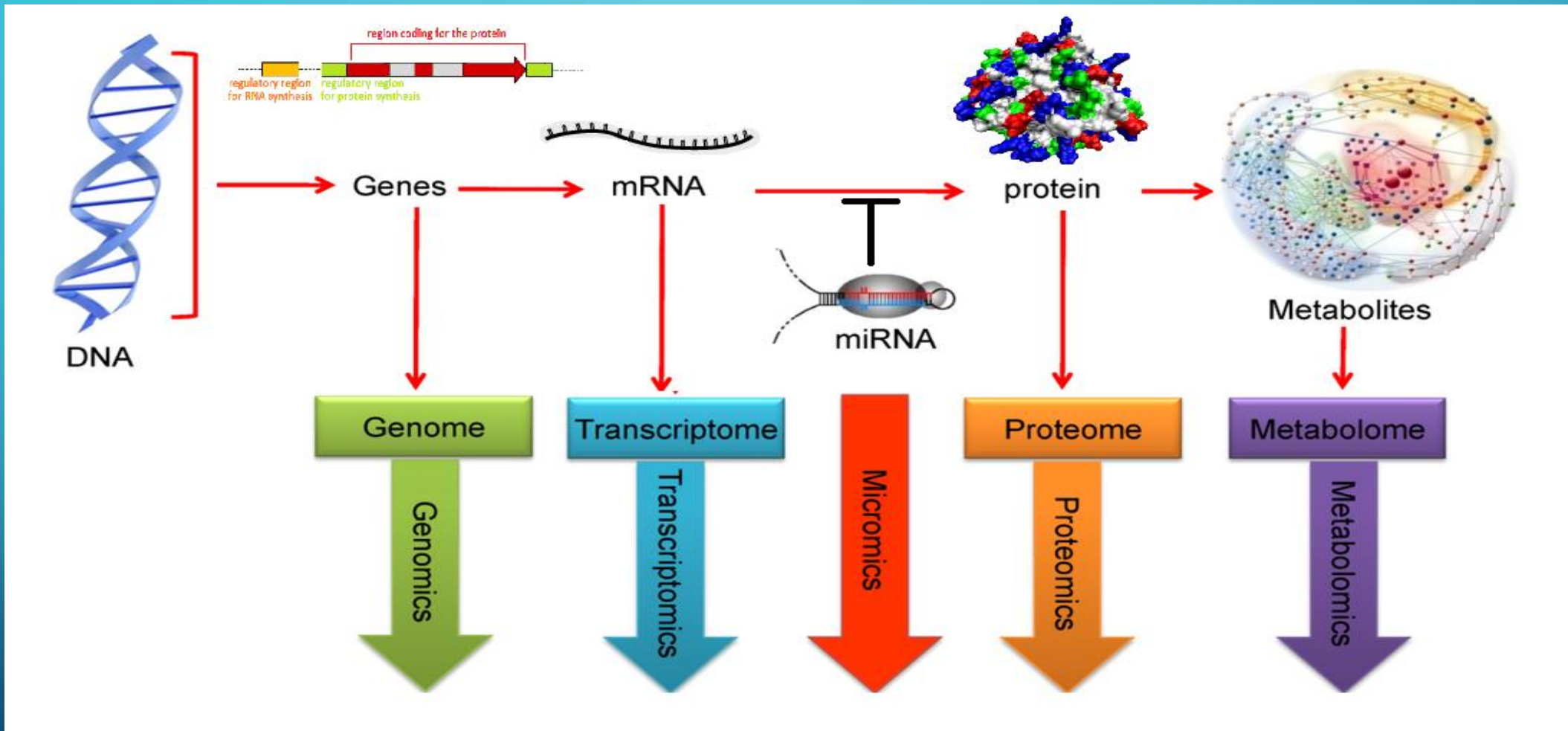
Over 45.000 identified metabolites

**Data integration**

**Data analytics**

**SYSTEM BIOLOGY**

# CHALLENGE TODAY IS



# DATA INTEGRATION

[illegible]

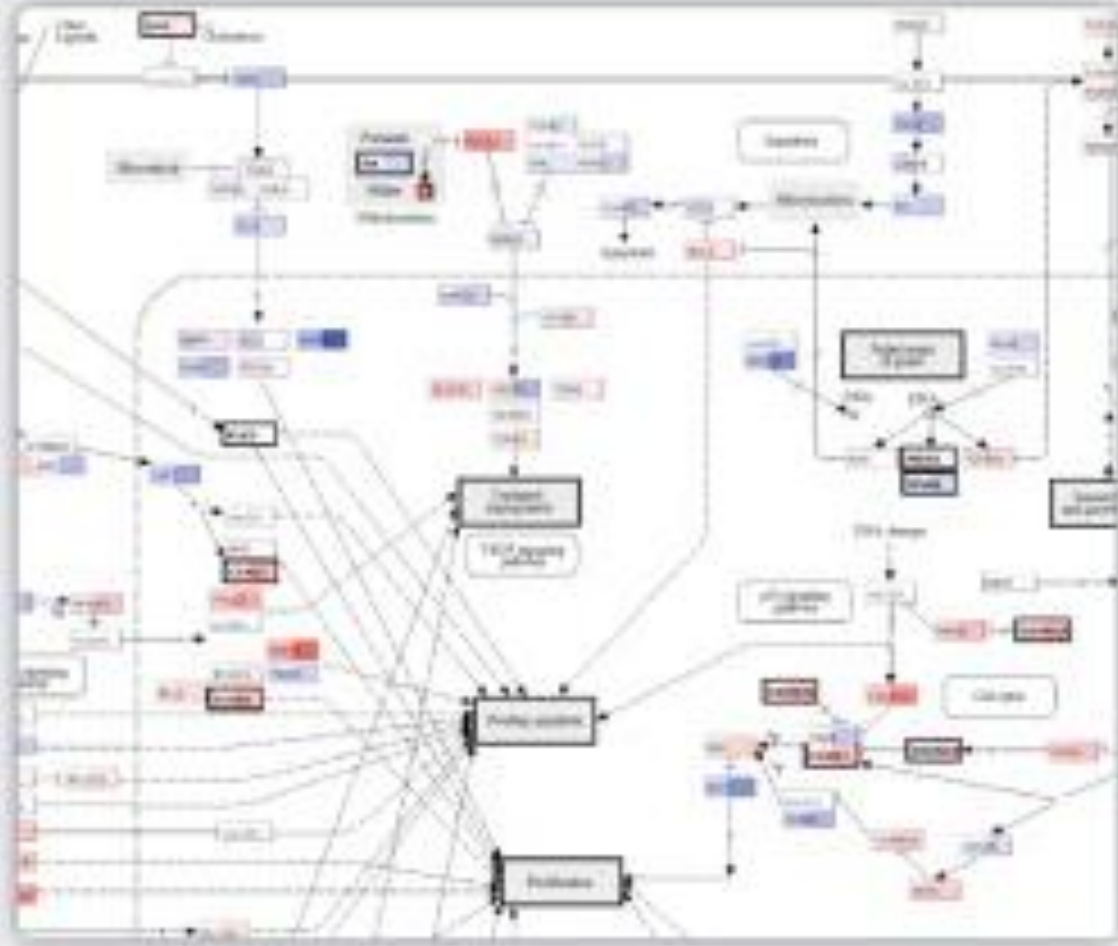
# Creation of resources

## Repository and

## Free accessibility to data

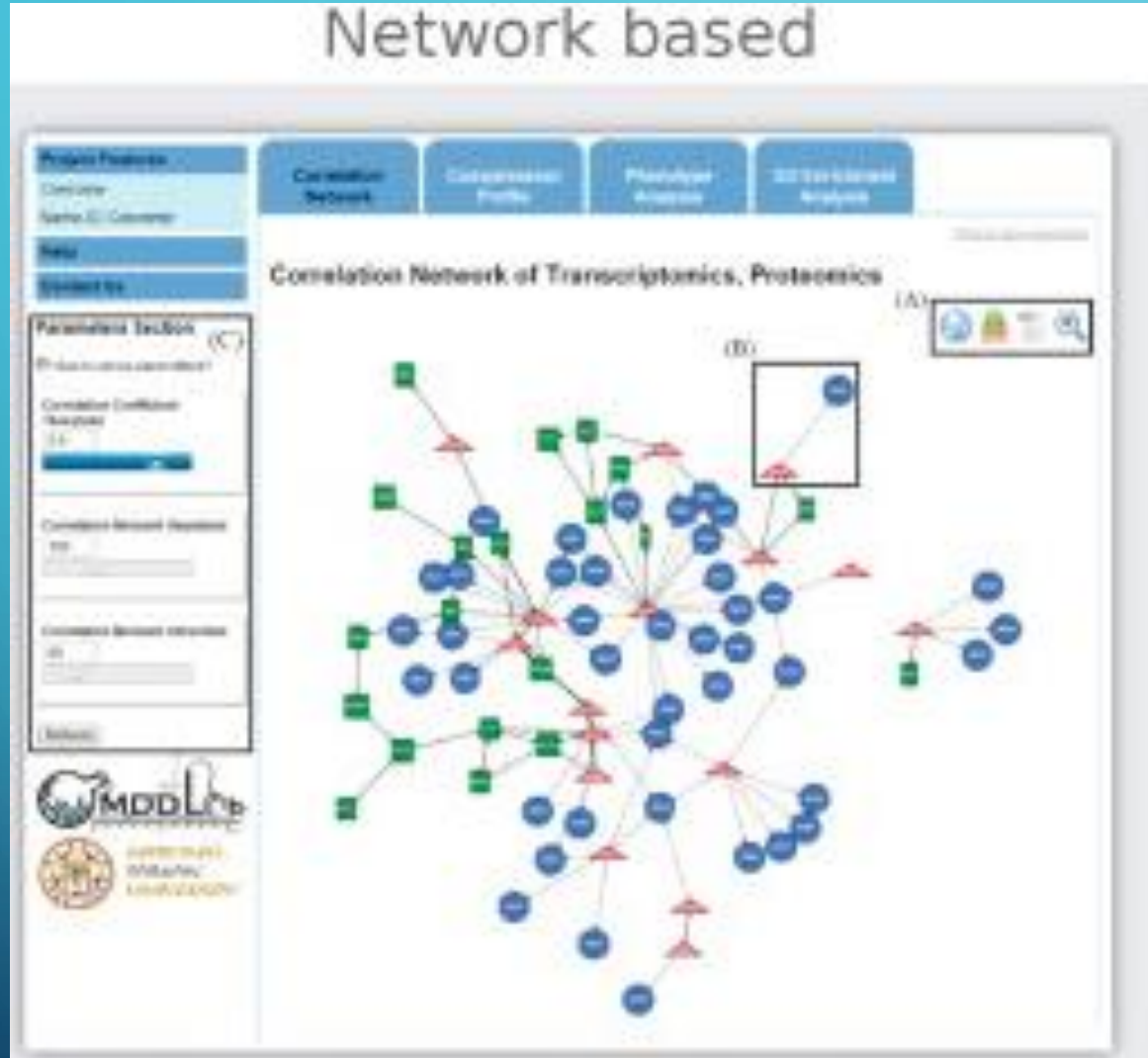


## Pathway based

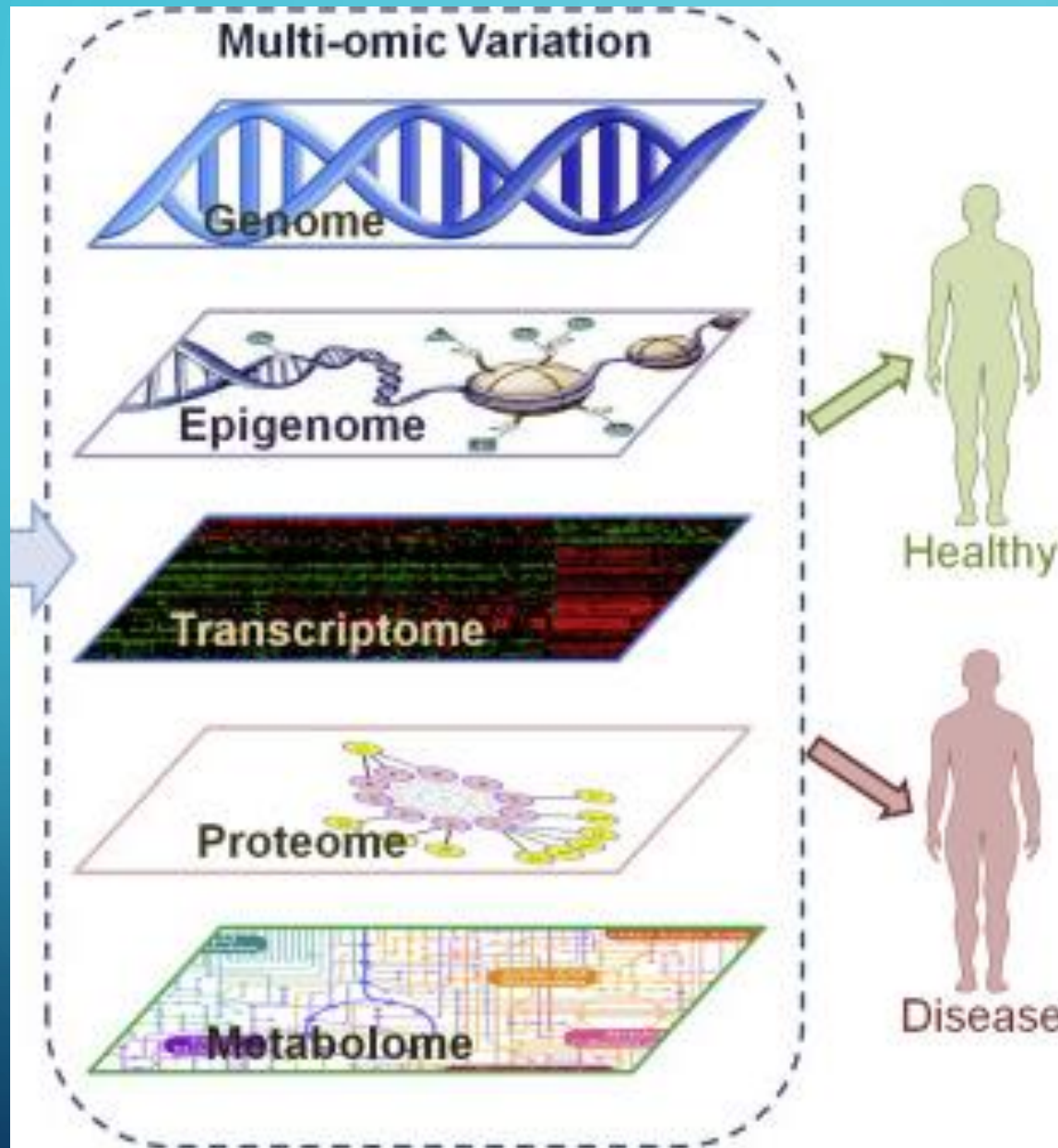


Development of databases  
integrating knowledge's  
On functional pathways

## Network based



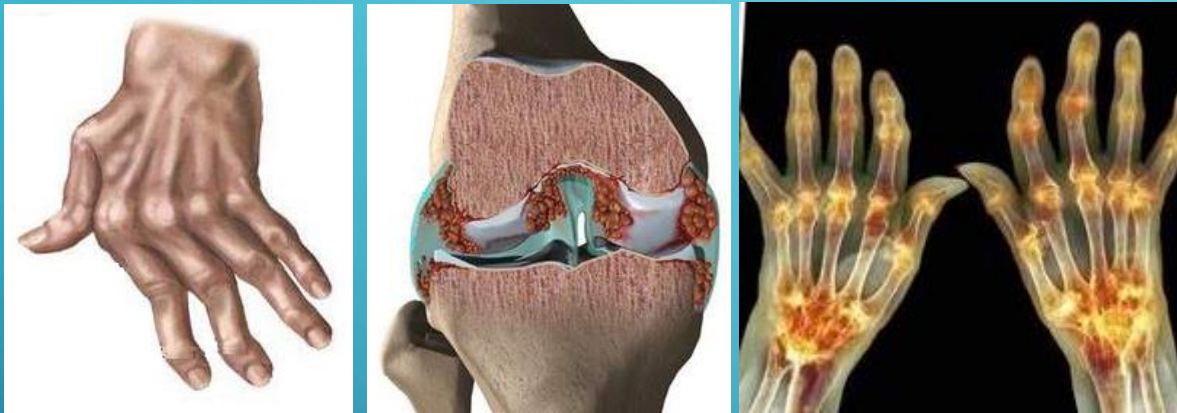
Development of algorithms  
Allowing multi-platform data  
to be analysed and  
searching databases for  
relationships



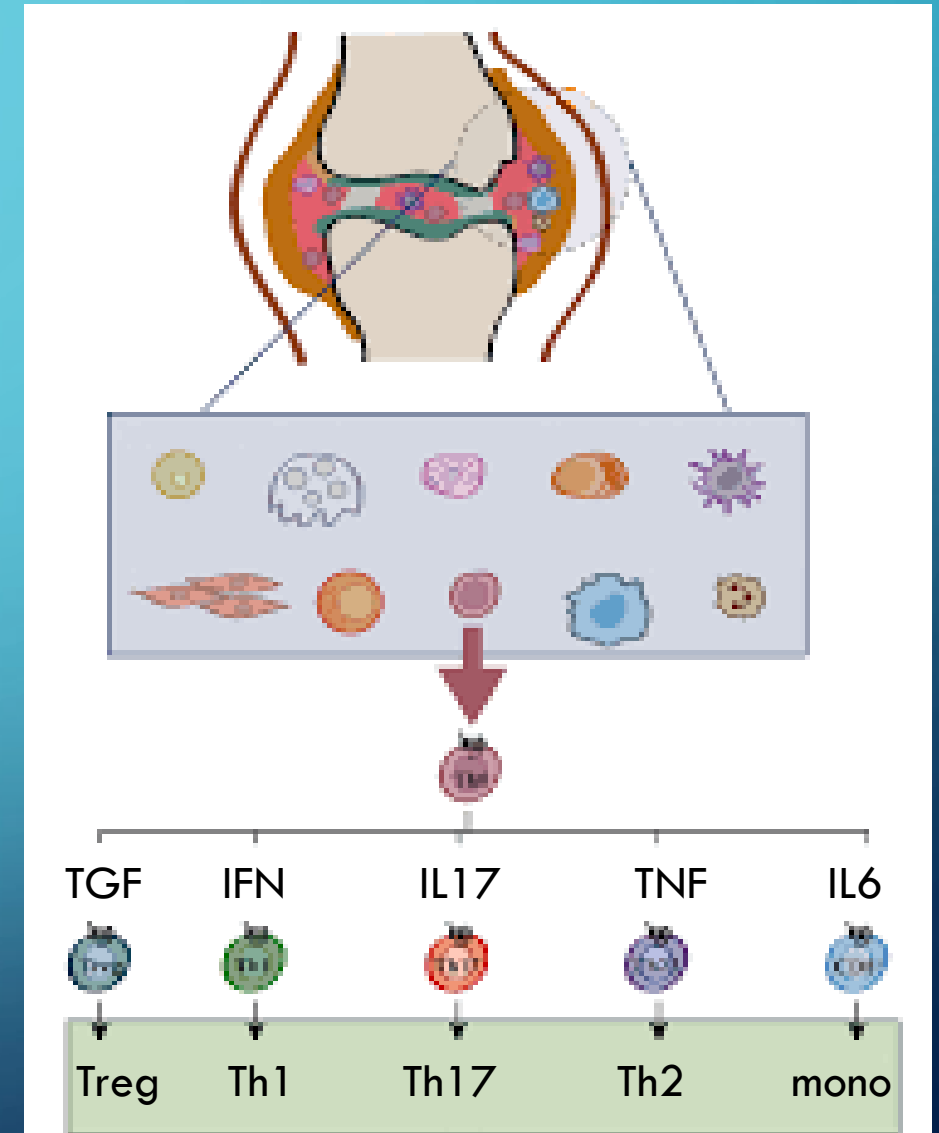
Personalised Medicine  
Drug design  
Prevention

# Rheumatoid arthritis

RA is a chronic autoimmune disorder that primarily affects joints.



1-2% of the population  
Reduces life expectancy  
Major pain and disability burden  
Major co-morbidities (cardiovascular problem)  
Effective biological therapy cost ~£15-25 K/year



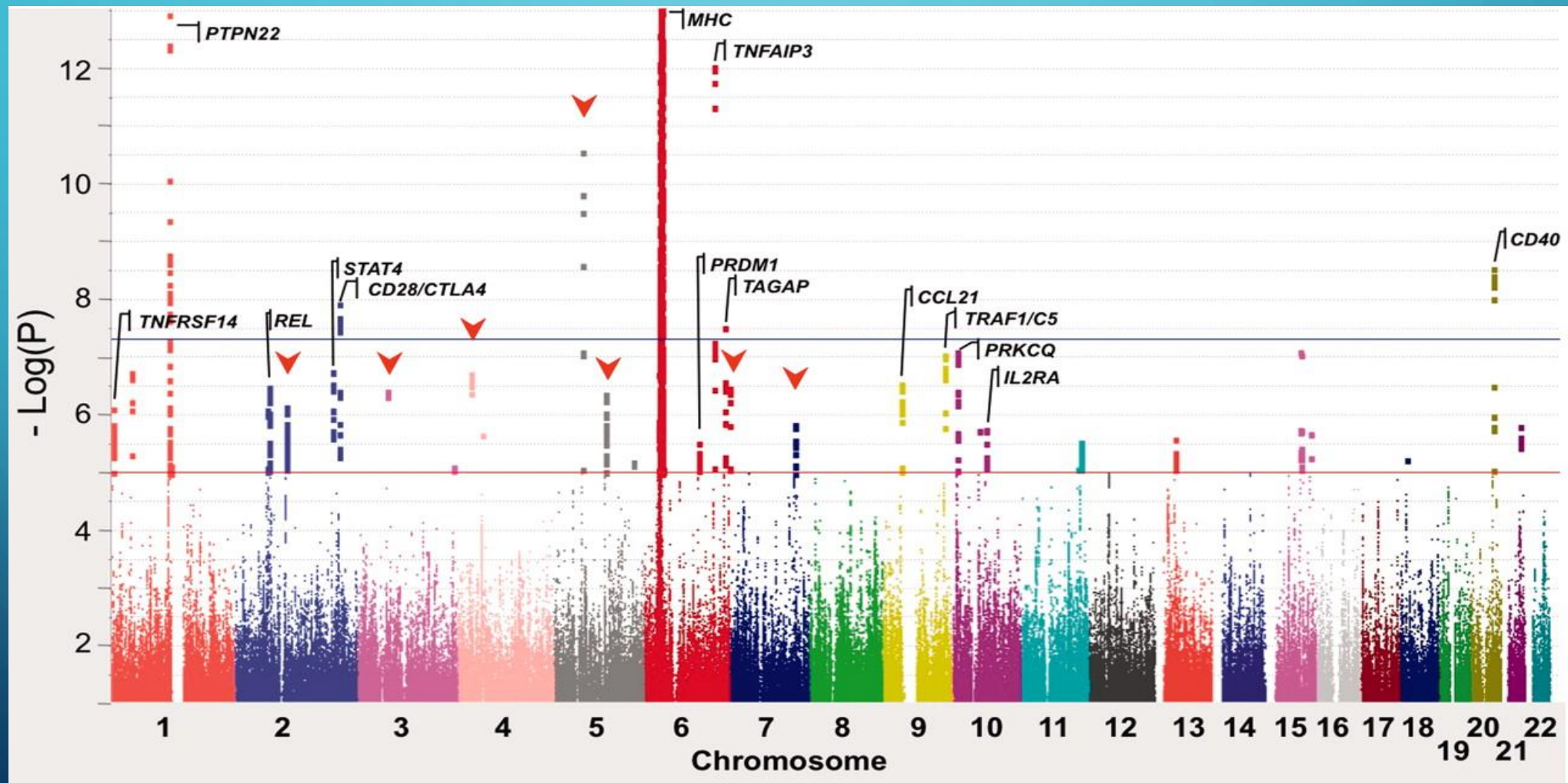


RA has a strong genetic component estimated at 60% of the risk while 40% come from environment and life style.

>100 genes were associated with RA

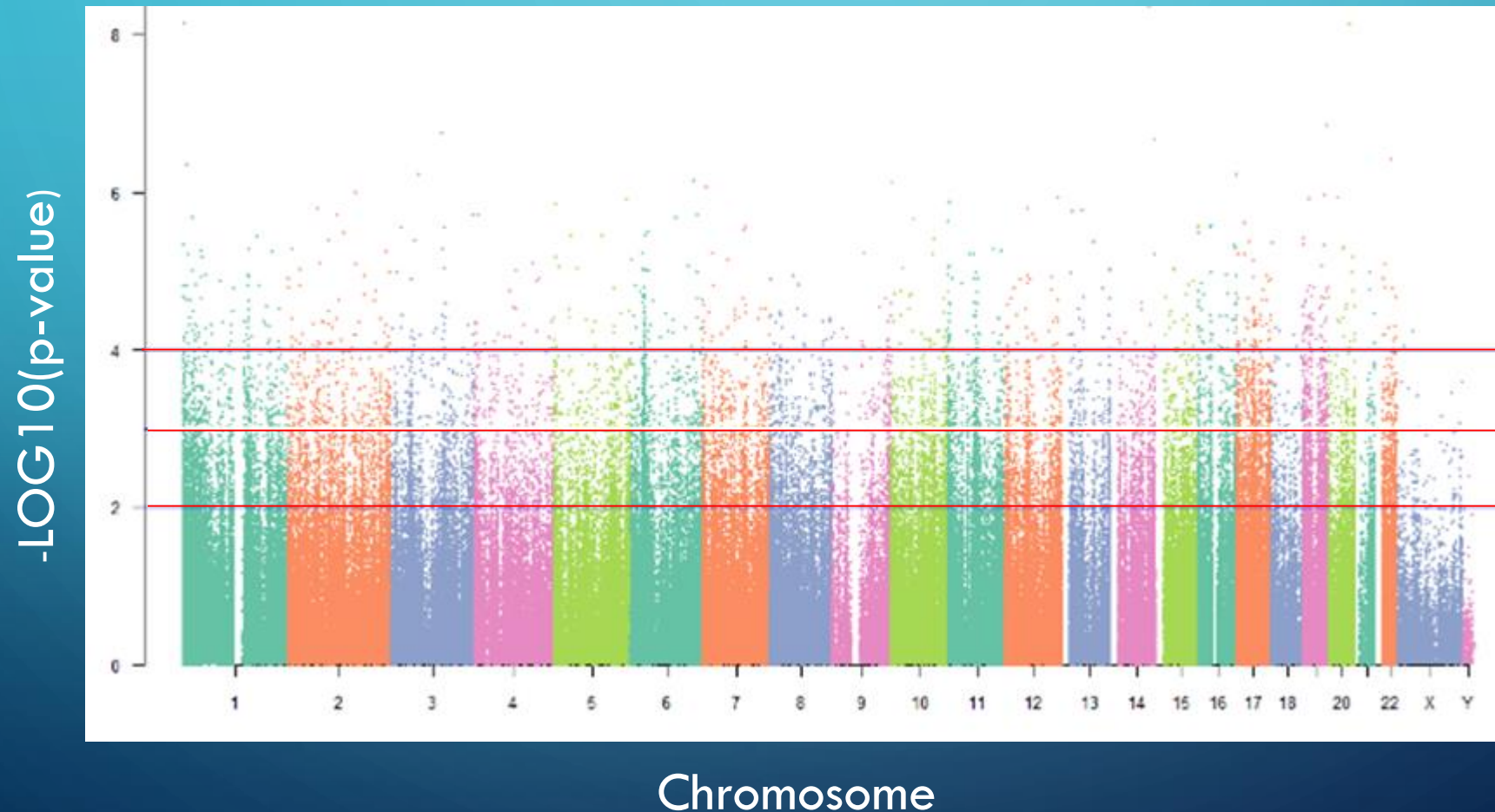
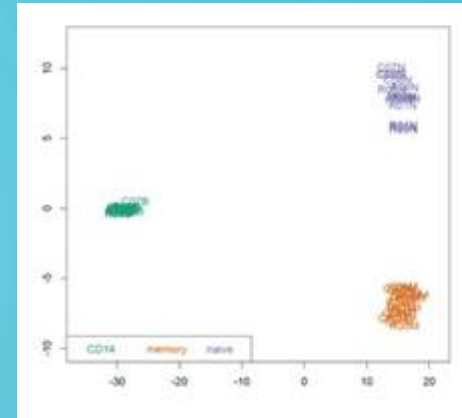
Chromosome	SNP	Gene	Ethnicity	Chromosome	SNP	Gene	Ethnicity
1	rs2228145	IL6R	Caucasians	8	rs4840565	BLK	Caucasians
1	rs2240336	PADI4	Caucasians	8	rs678347	GRHL2	Caucasians
1	rs883220	POU3F1	Caucasians	8	rs1516971	PVT1	Caucasians
1	rs798000	CD2	Caucasians	8	rs998731	TPD52	Caucasians
1	rs2843401	MMEL1	Caucasians	9	rs10739580	TRAF1	Caucasians
1	rs2014863	PTPRC	Caucasians	9	rs2812378	CCL21	Caucasians
1	rs2476601	PTPN22	Caucasians	10	rs12764378	ARID5B	Caucasians
1	rs10494360	FCGR2A	Caucasians	10	rs2275806	GATA3	Caucasians
1	rs2105325	Loc100506023	Caucasians	10	rs947474	PRKCQ	Caucasians
1	rs28411352	MTF1-INPP5B	Caucasians	10	rs10795791	IL2RA	Caucasians
1	rs227163	TNFRSF9	Asian	10	rs12413578	10p14	Caucasians
1	rs10430455	FCRL3	Asians	10	rs726288	SFTPD	Asian
1	rs6682654	CD244	Asians	10	rs2671692	WDFY4	Caucasians
2	rs10209110	AFF3	Caucasians	10	rs793108	ZNF438	Caucasians
2	rs34695944	REL	Caucasians	11	rs595158	CD5	Caucasians
2	rs6546146	SPRED2	Caucasians	11	rs4938573	DDX6	Caucasians
2	rs13426947	STAT4	Caucasians	11	rs570676	TRAF6	Caucasians
2	rs1980422	CD28	Caucasians	11	rs3781913	PDE2A-ARAP1	Asians
2	rs11571302	CTLA4	Caucasians	11	chr11:107967350	ATM	Caucasians
2	rs11900673	B3GNT2	Asians	11	rs4409785	CEP57	Caucasians
2	rs6732565	ACOXL	Caucasians	11	rs73013527	ETS1	Caucasians
2	rs6715284	CFLAR-CASP8	Caucasians	11	rs968567	FADS1-FADS2-FADS3	Caucasians
2	rs10175798	LBH	Caucasians	12	rs10683701	KIF5A	Caucasians
3	rs35677470	DNASE1L3	Caucasians	12	rs773125	CDK2	Caucasians
3	rs3806624	EOMES	Caucasians	12	rs10774624	SH2B3-PTPN11	Caucasians
3	rs9826828	IL20RB	Caucasians	13	rs9603616	COG6	Caucasians
3	rs4452313	PLCL2	Caucasians	14	rs911263, rs1950897	RAD51L1/RAD51B	Caucasians
4	rs78560100	IL2, IL21	Caucasians	14	rs2841277	PLD4	Asians
4	rs932036	RBPJ	Caucasians	14	rs3783782	PRKCH	Asian
4	rs2867461	ANXA3	Asians	15	rs8043085	RASGRP1	Caucasians
4	rs13142500	CLNK	Caucasians	15	rs8026898	TLE3	Caucasians
4	rs2664035	TEC	Caucasians	16	rs13330176	IRF8	Caucasians
5	rs71624119	ANKRD55	Caucasians	16	rs4780401	TXNDC1	Caucasians
5	rs39984	GIN1	Caucasians	17	rs12936409	IKZF3	Caucasians
5	rs657075	CSF2	Asians	17	rs72634030	C1QB	Caucasians
6	rs59466457	CCR6	Caucasians	17	rs1877030	MED1	Caucasians
6	rs6911690	PRDM1	Caucasians	18	rs7234029	PTPN2	Caucasians
6	rs629326	TAGAP	Caucasians	18	rs2469434	CD226	Asian
6	rs6920220	TNFAIP3	Caucasians	19	rs34536443	TYK2	Caucasians
6	rs72928038	BACH2	Caucasians	19	chr19:10771941	ILF3	Caucasians
6	rs12529514	CD83	Asians	20	rs6032662	CD40	Caucasians
6	rs2233434	NFKBIE	Asians	21	rs2834512	RCAN1	Caucasians
6	rs2234067	ETV7	Caucasians	21	rs9979383	RUNX1	Caucasians
6	rs9378815	IRF4	Caucasians	21	rs73194058	IFNGR2	Caucasians
6	rs9373594	PPIL4	Asian	21	rs1893592	UBASH3A	Caucasians
6	amino acid position 9	HLA-DPB1	Caucasians	21	rs2075876	AIRE	Asians
6	amino acid position 9	HLA-B	Caucasians	22	rs3218251	IL2RB	Caucasians
6	amino acid position 74	HLA-DRB1	Caucasians	22	rs909685	SYNGR1	Caucasians
6	amino acid position 71	HLA-DRB1	Caucasians	22	rs1043099	GATSL3	Caucasians
6	amino acid position 11	HLA-DRB1	Caucasians	22	rs5754217	UBE2L3	Caucasians
6	amino acid position 77	HLA-A	Caucasians	X	rs13397	IRAK1	Caucasians
7	rs3807306	IRF5	Caucasians	X	chrX:78464616	P2RY10	Asian
7	rs4272	CDK6	Caucasians				
7	rs67250450	JAZF1	Caucasians				

Over 100 genes associated with RA :  
many with T-cell related functions



# DNA methylation patterns in RA: Very early, drug naïve patients 3 particular cell types :

Naïve T-cells Memory T-cells and monocytes



Highly significant differences

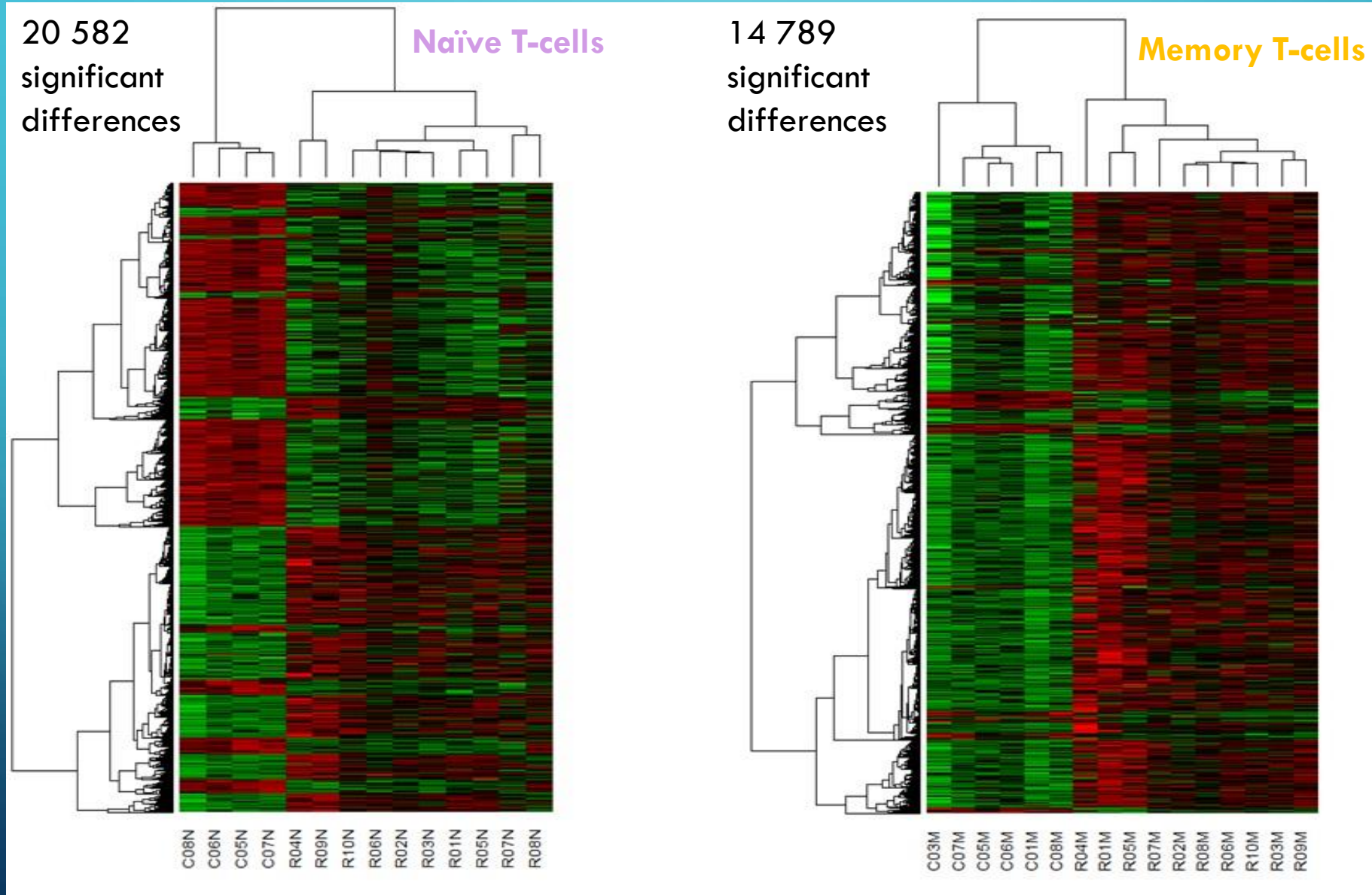
Medium significance

Low significance



# Difference in pattern of methylation between 2 cell types

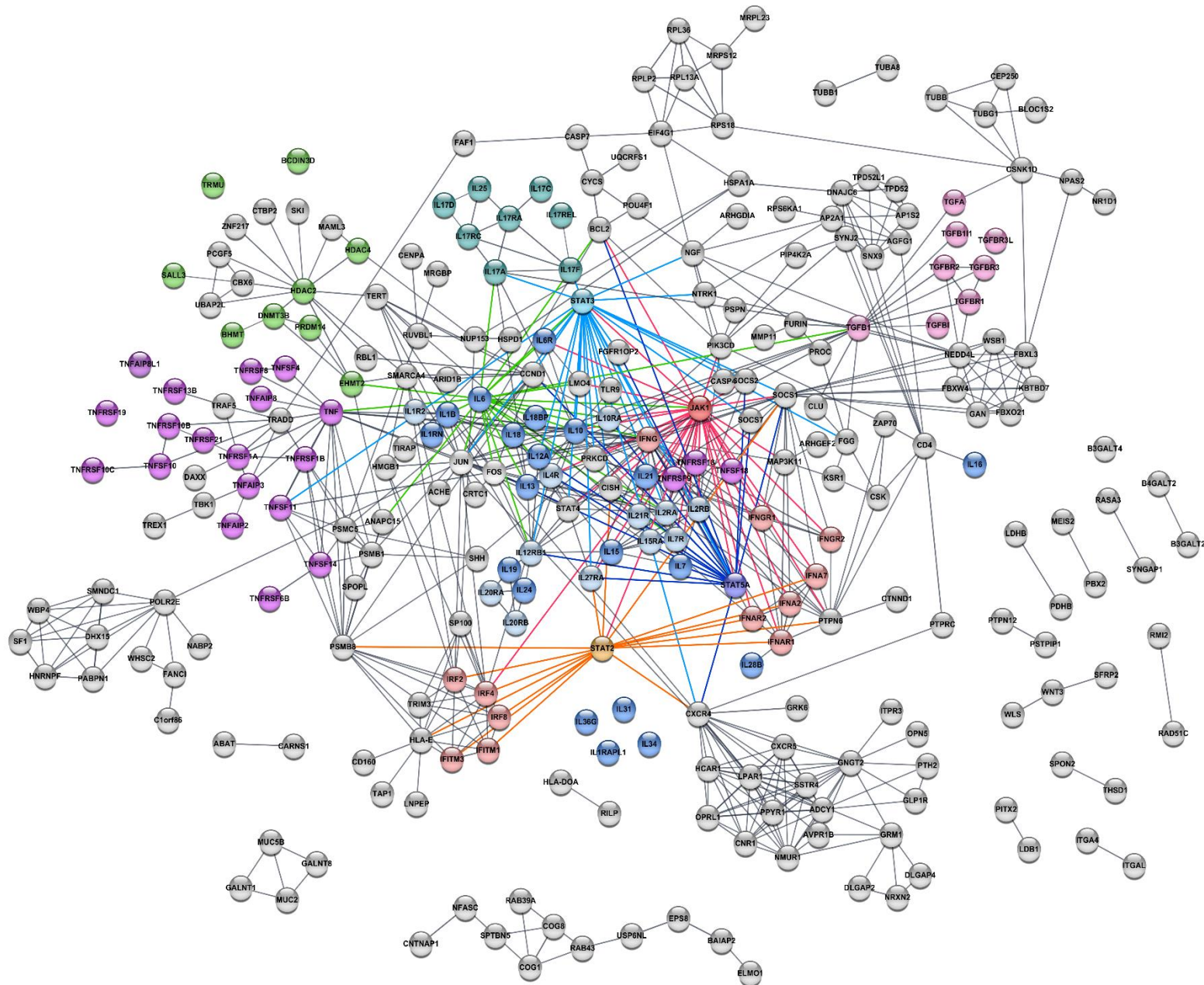
## Hierarchical clustering



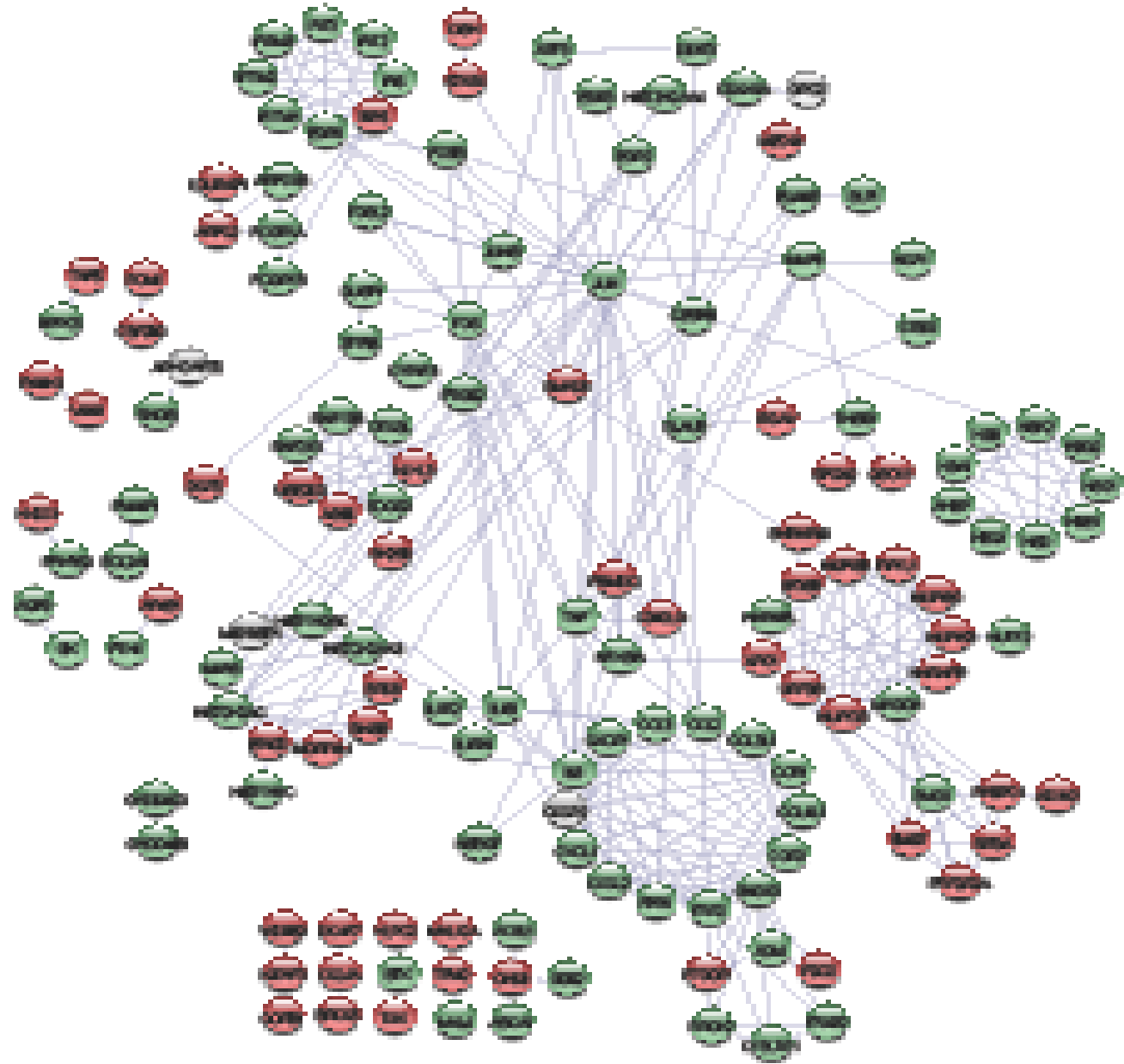


608 genes with differential DNA methylation patterns between health and RA

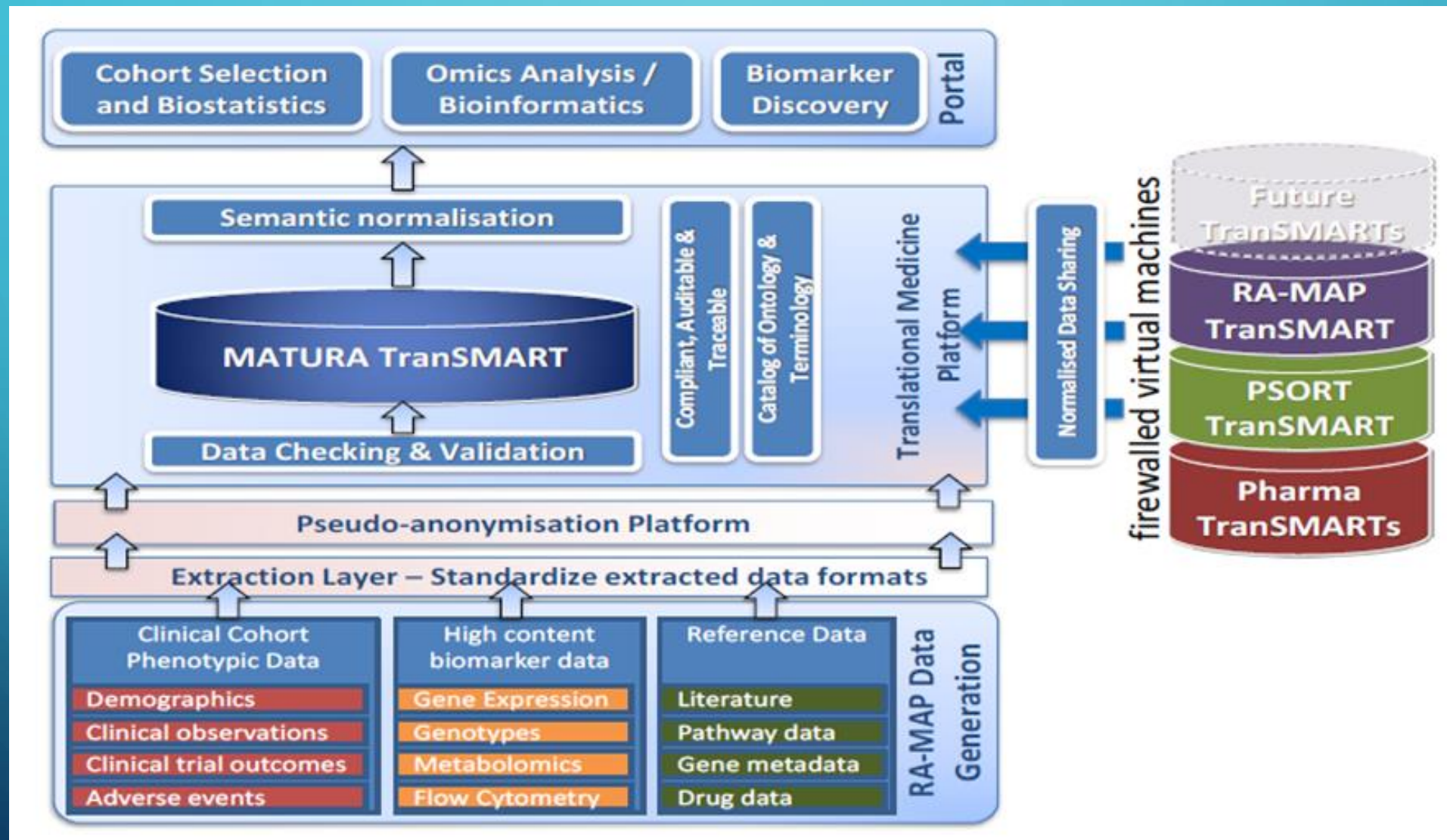
The figure displays a complex network diagram representing interactions between 608 genes. The nodes are colored in various shades (green, blue, red, orange, purple, grey) and connected by lines of different colors (blue, red, orange, green, black). The network is highly interconnected, with many nodes having multiple connections. The layout is circular, with nodes arranged in concentric rings. The central part of the network is the most dense, with many overlapping connections. The outer rings are less dense, with nodes more clearly visible. The overall structure suggests a highly complex and interconnected gene network.



**xxx genes with  
differential gene  
expression  
patterns between  
health and RA**

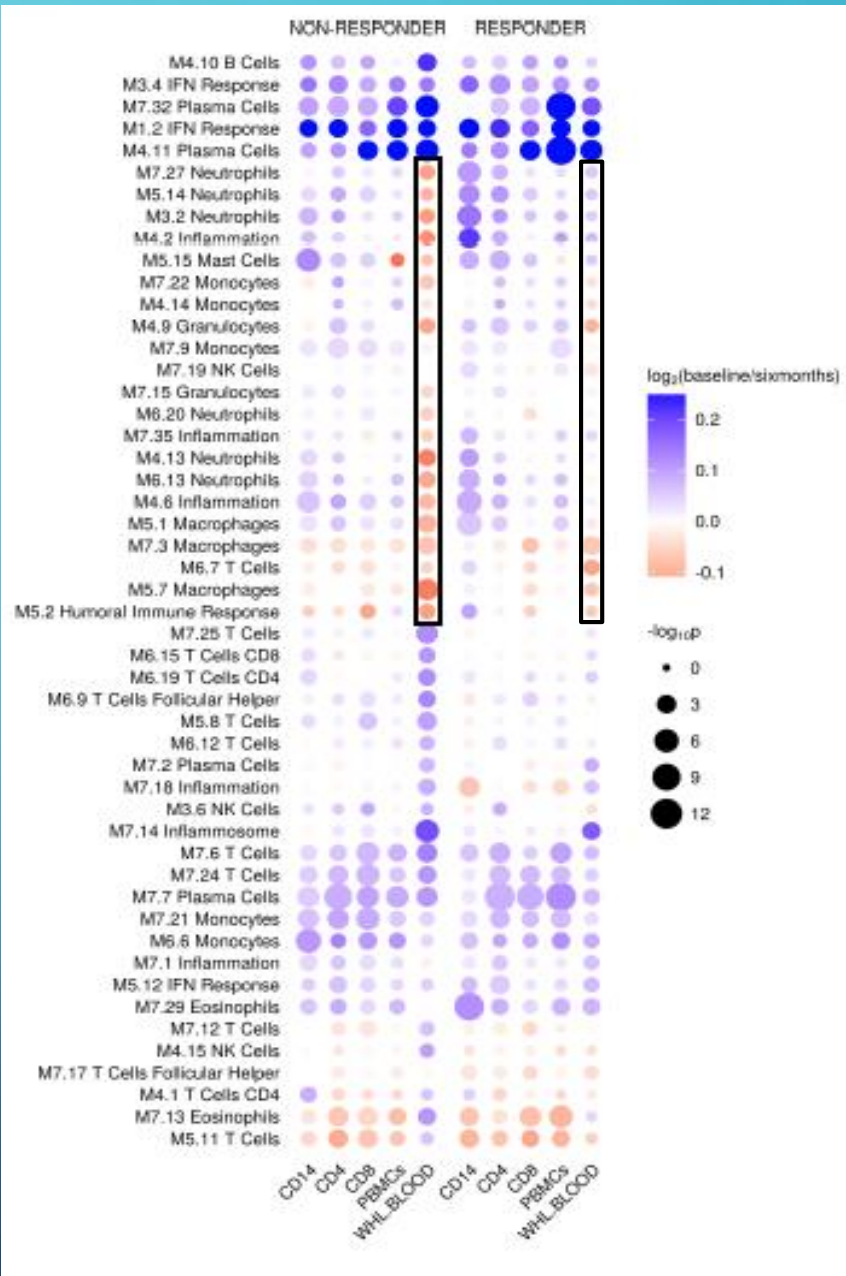
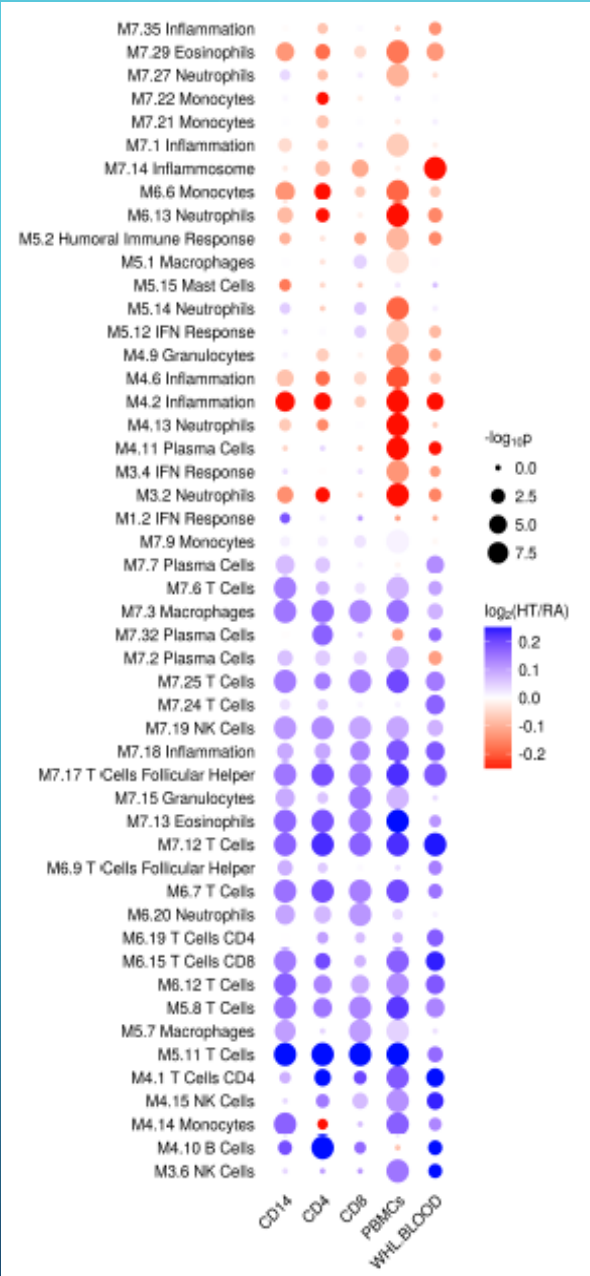


# MRC multiplatform analysis project : RA-MAP consortium



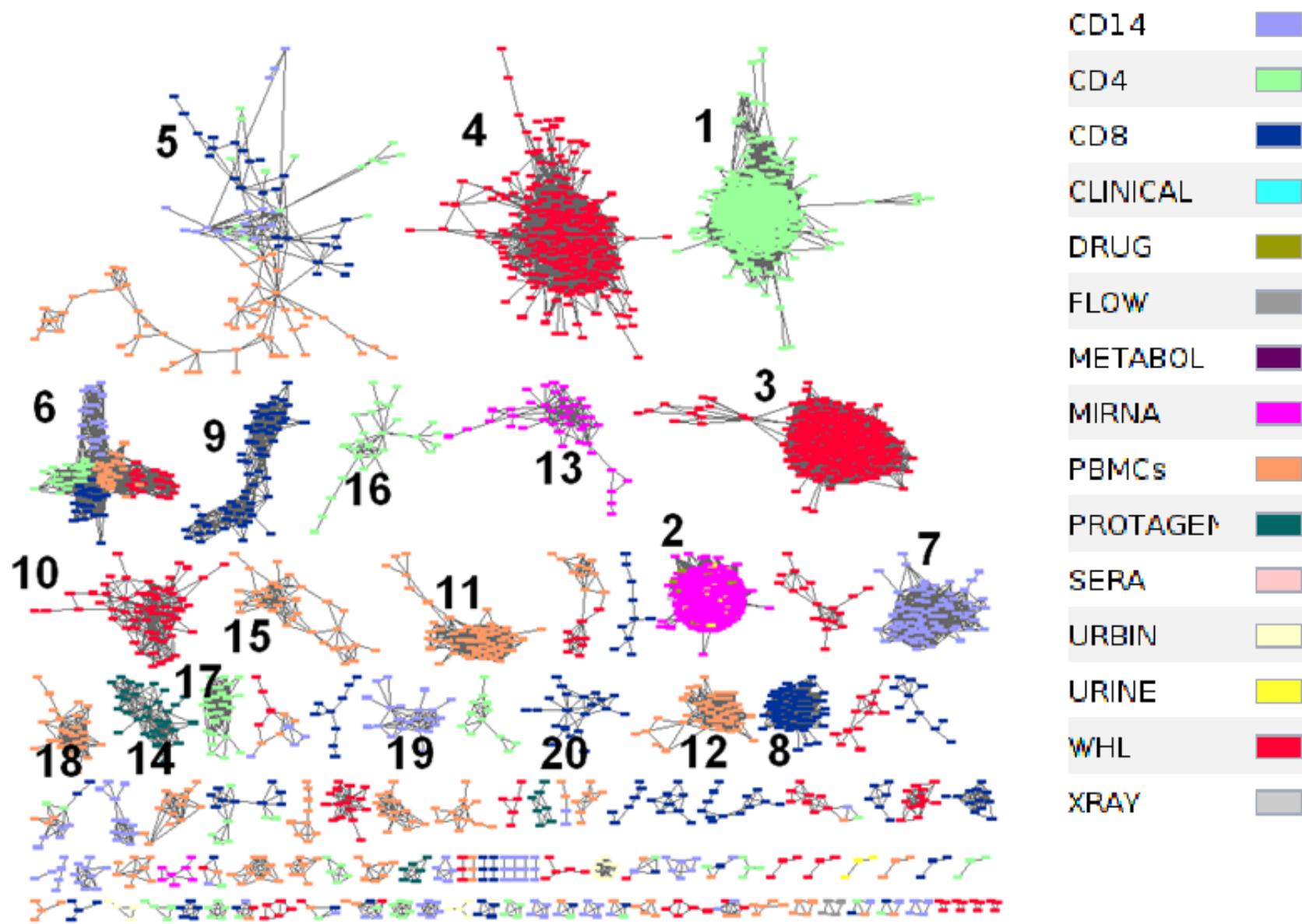


# Gene expression data

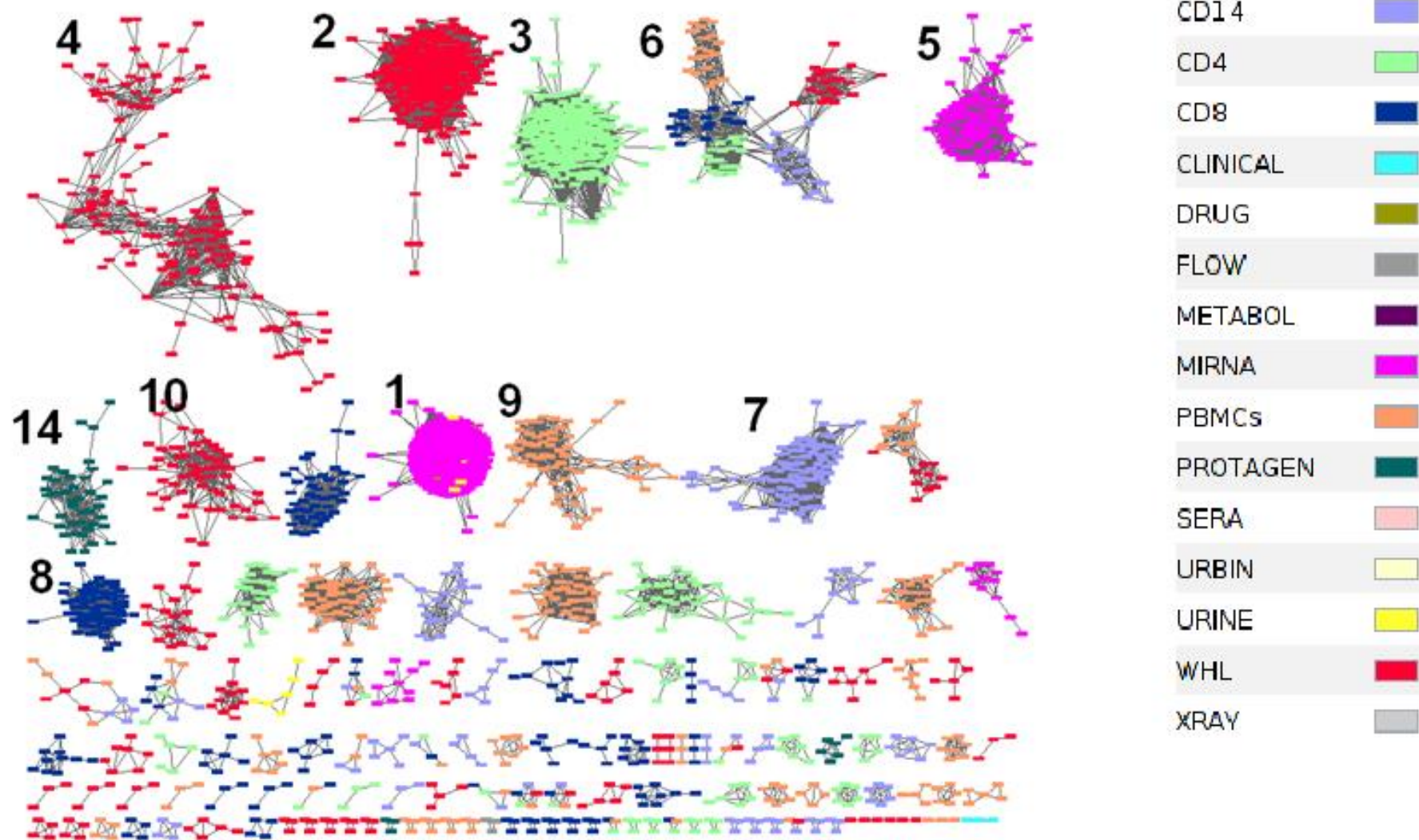




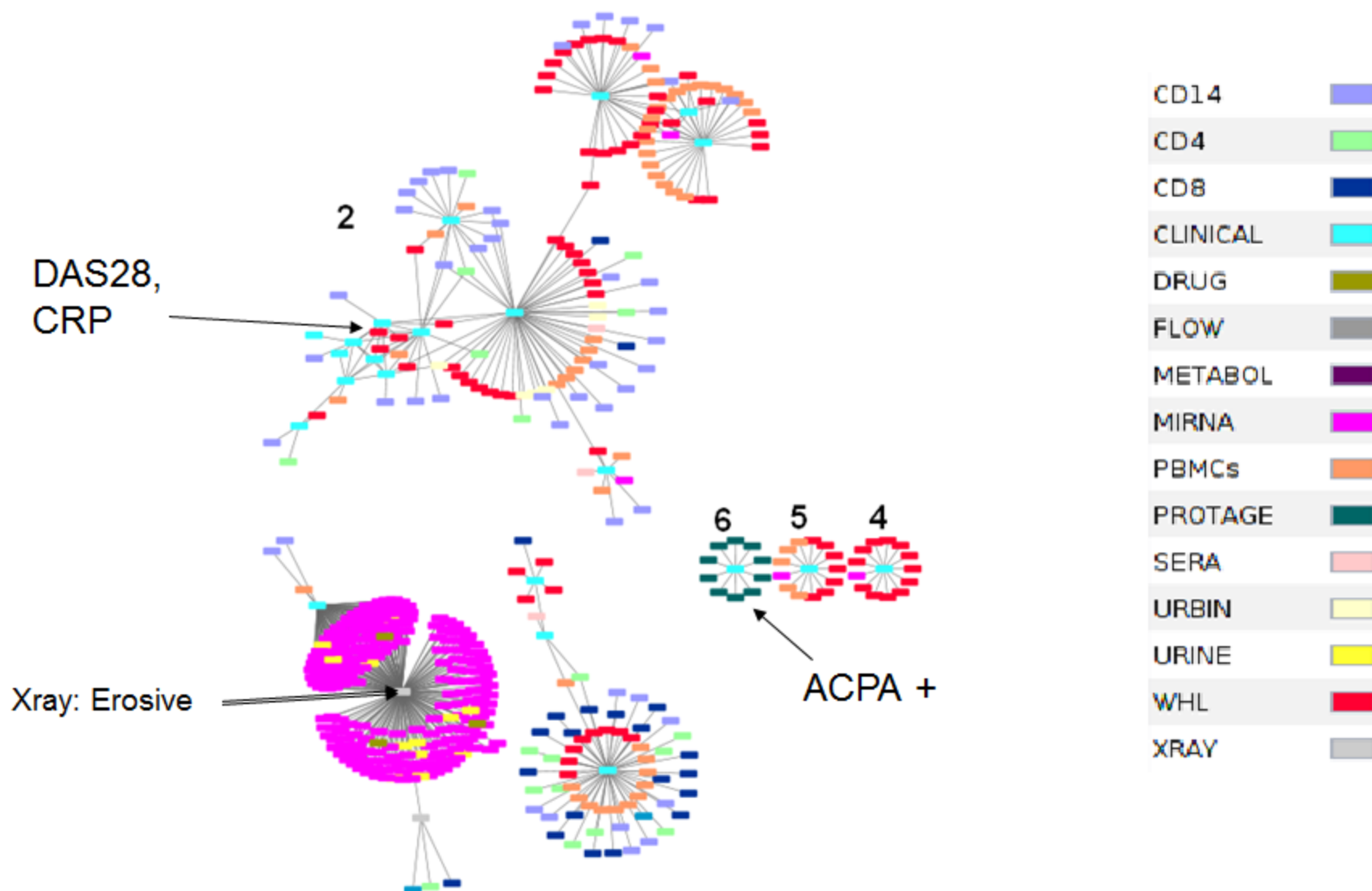
# ARACNE: Multi-Omics Network



# ARACNE: 6M Multi-Omics Network



# ARACNE: Baseline Clinical & X RAY





# ARACNE: 6M Clinical Nodes

