

In the name of the most high

Understanding Cell Identity by Deep Neural Networks

Ali Sharifi-Zarchi

Department of Stem Cells, Royan Institute

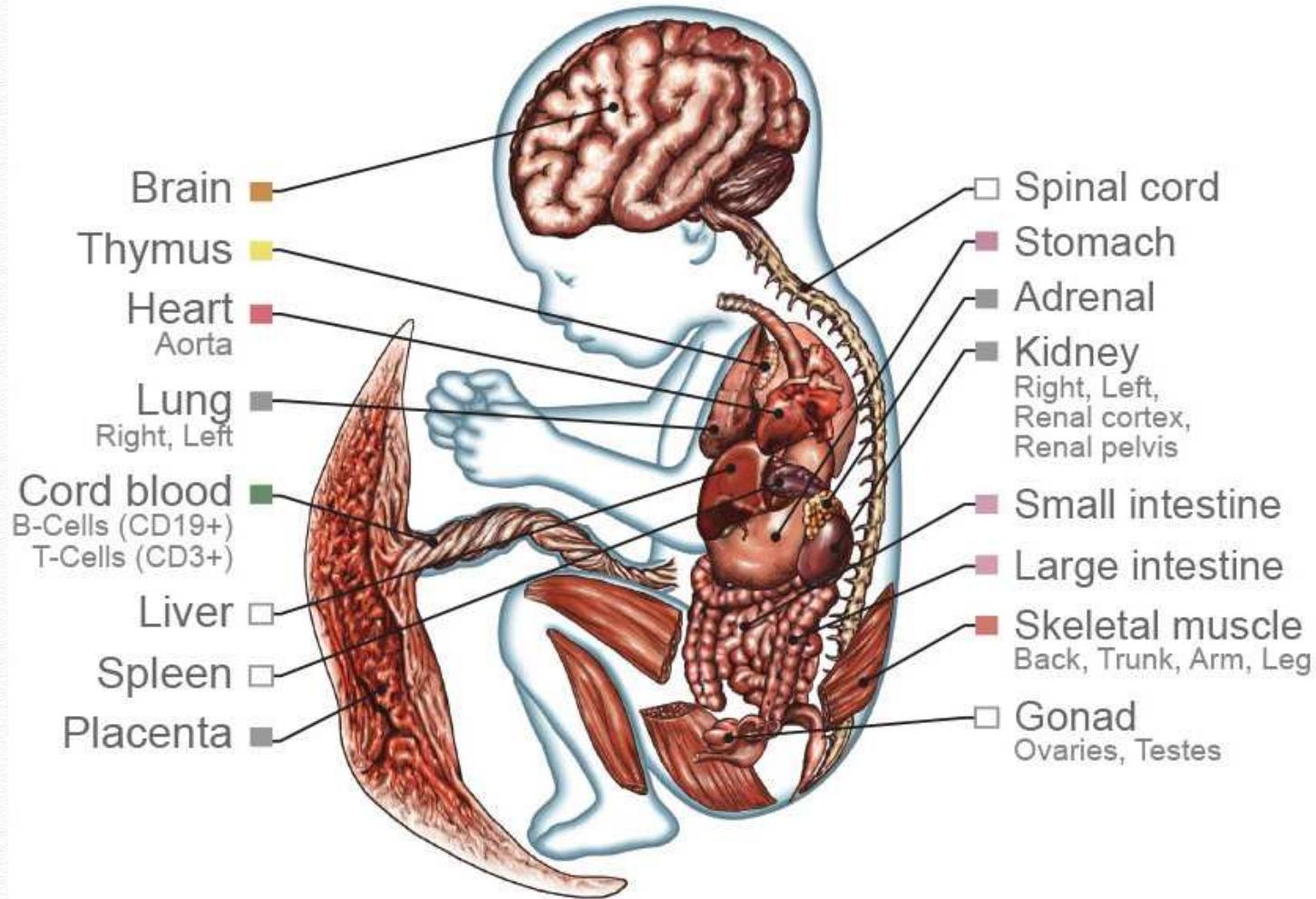
Department of Computer Eng. Sharif Univ. of Tech.

Can algorithmic
and machine learning methods
cure diseases?

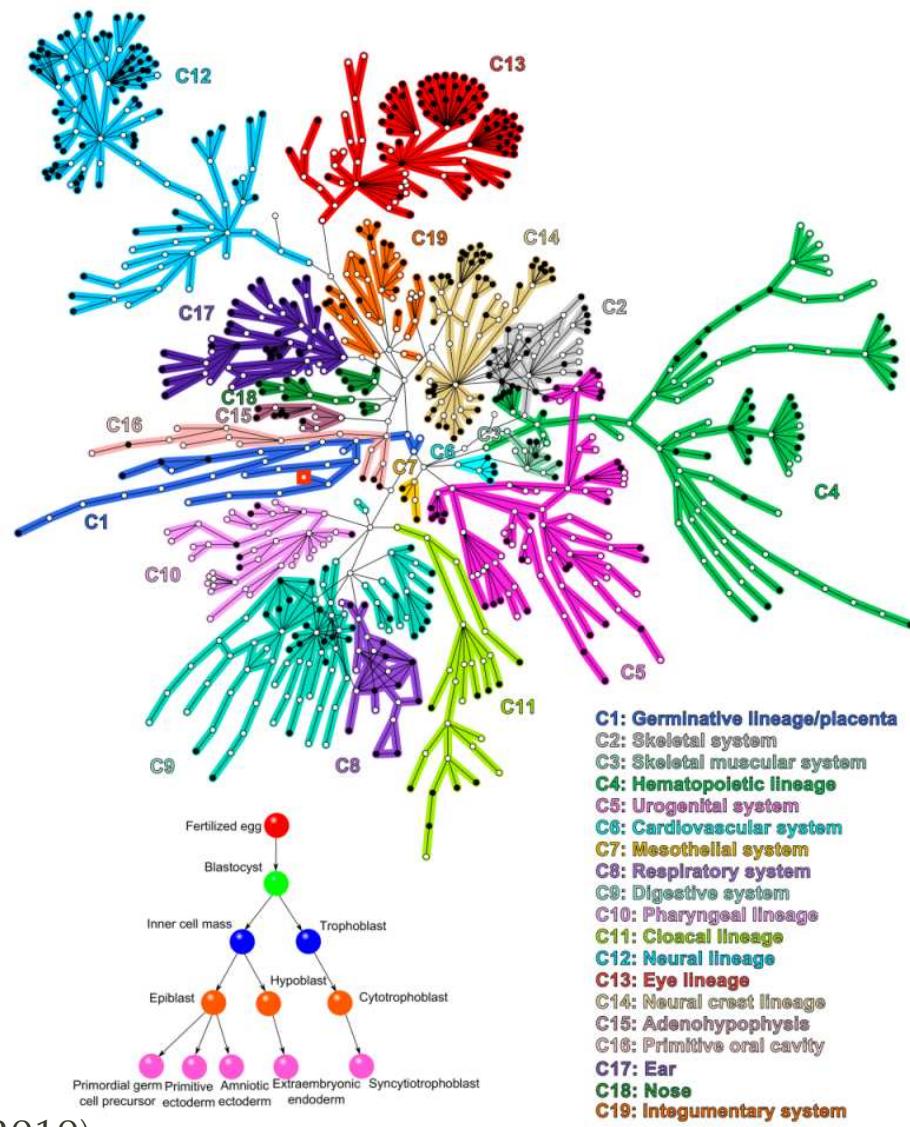
Biological Diversity



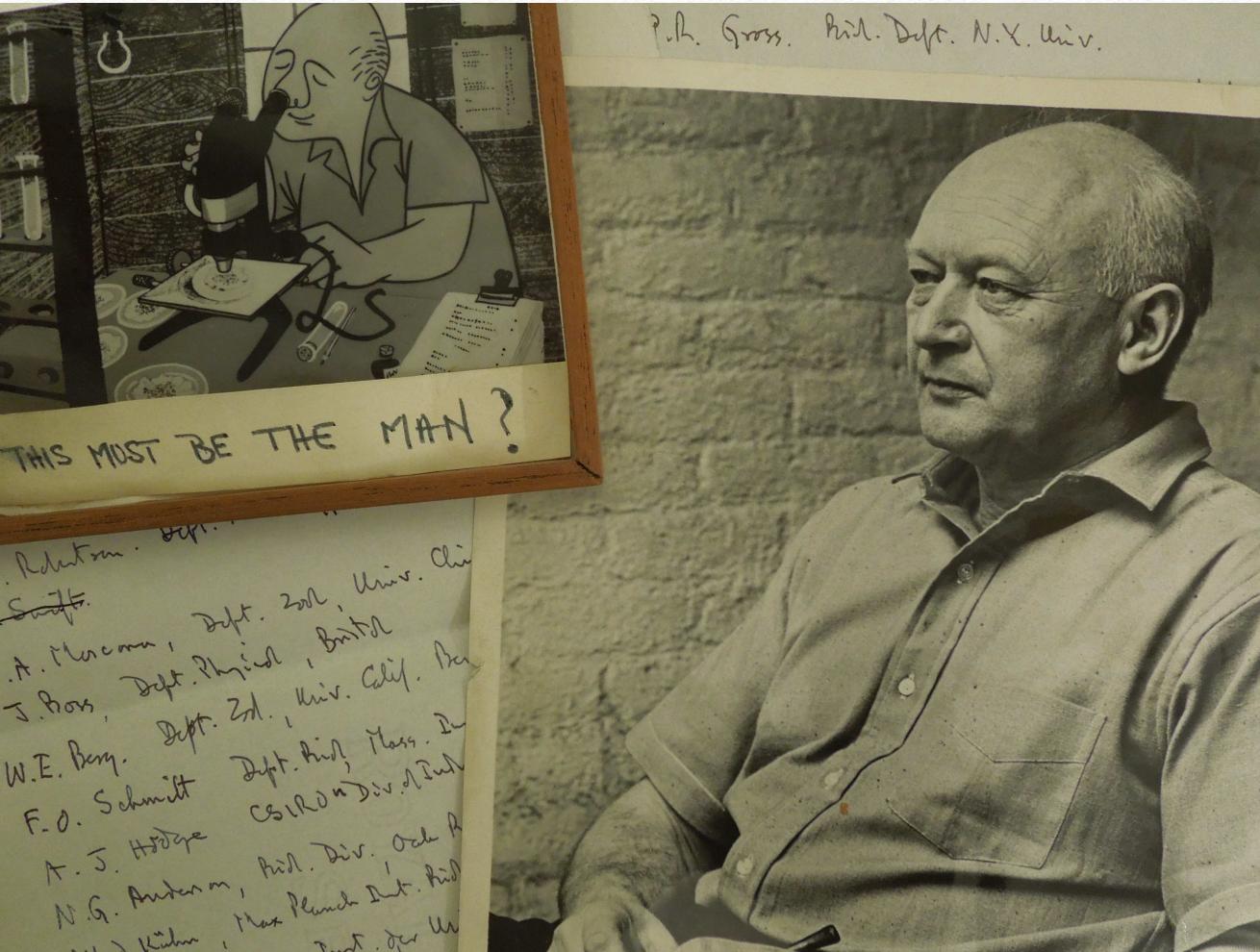
Diversity in Our Body



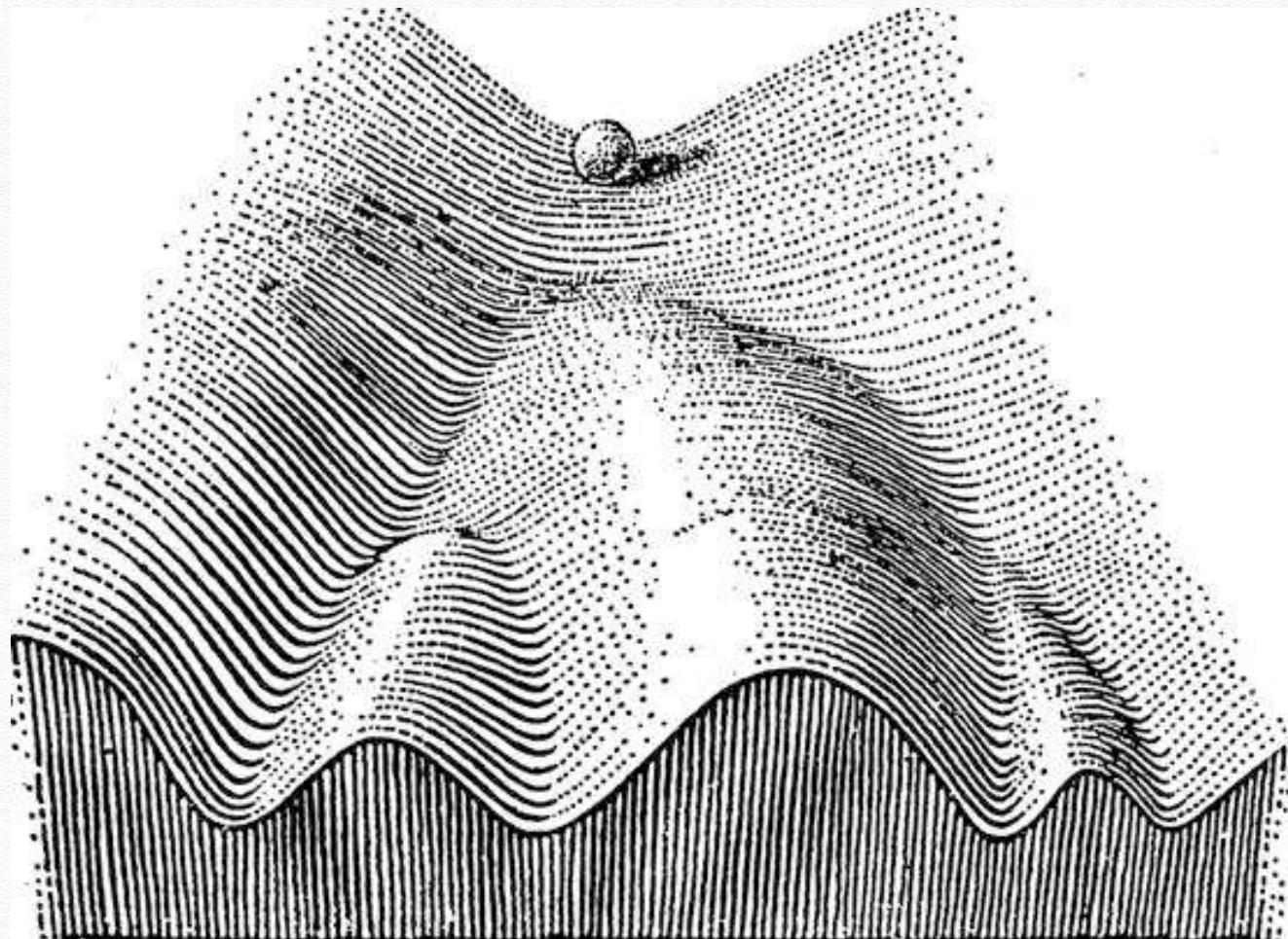
Differentiation Network



Conrad H. Waddington

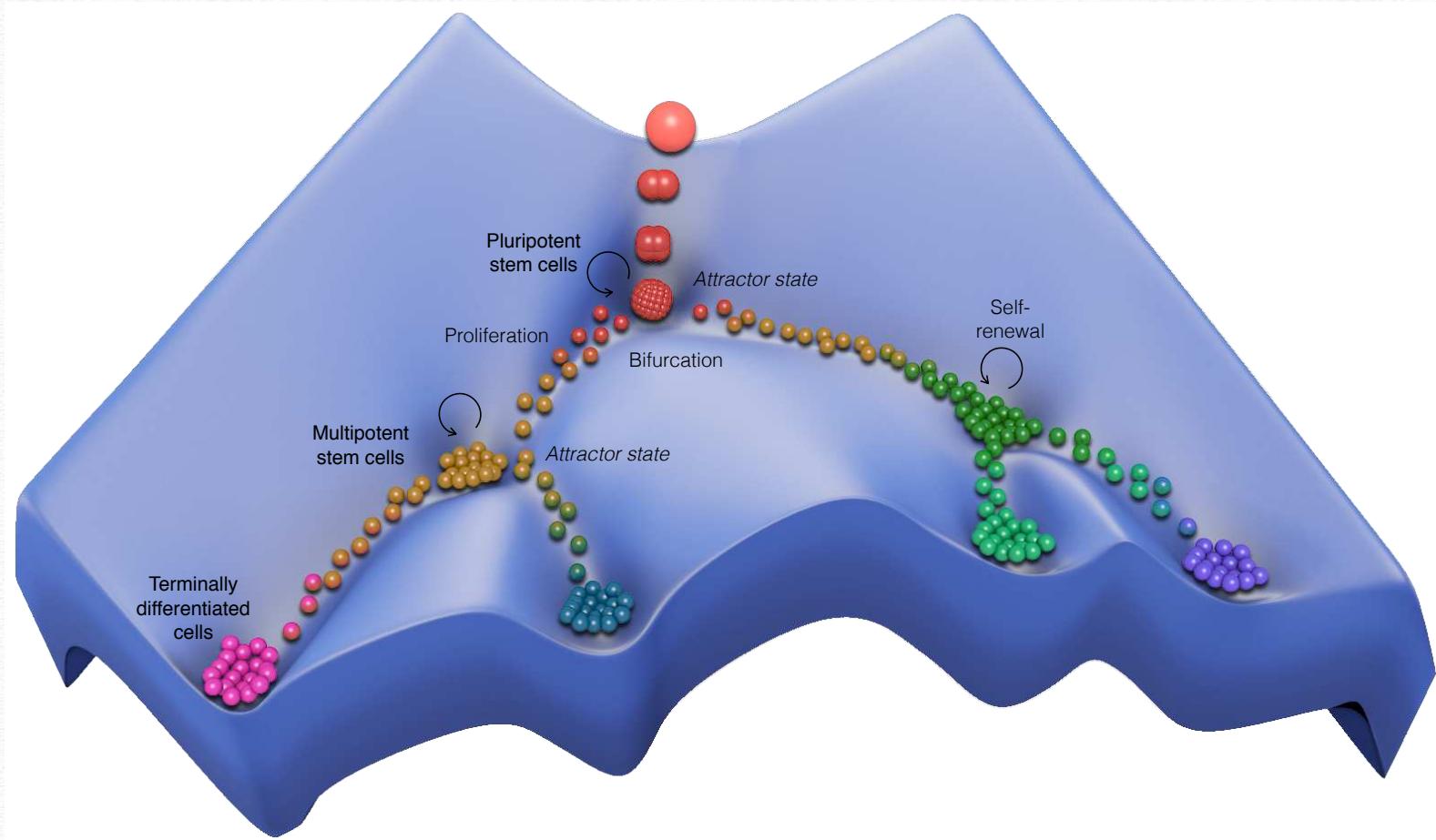


Epigenetic Landscape



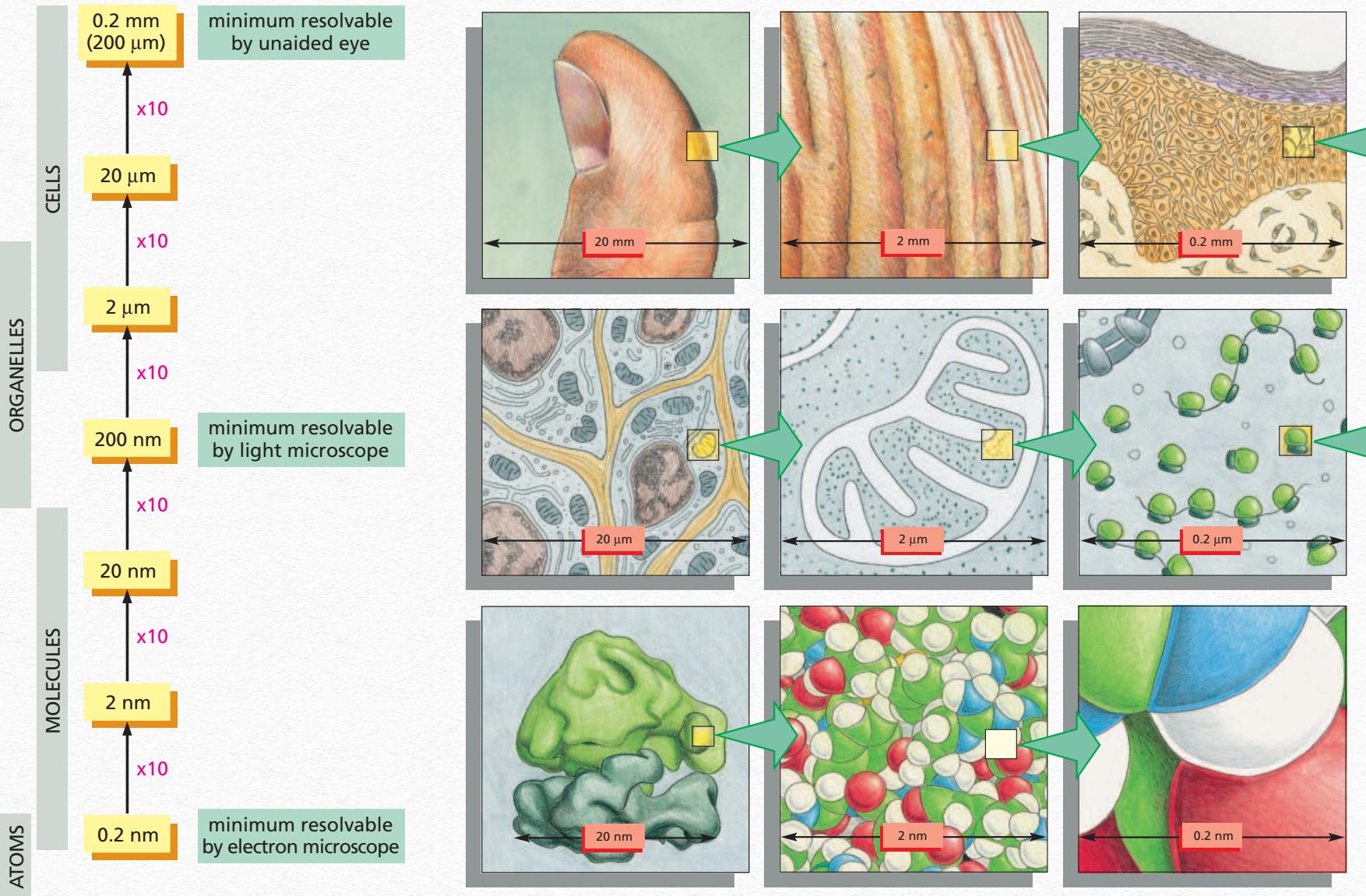
Waddington, C H. *The Strategy of the Genes*. London: George Allen & Unwin, 1957.

Epigenetic Landscape

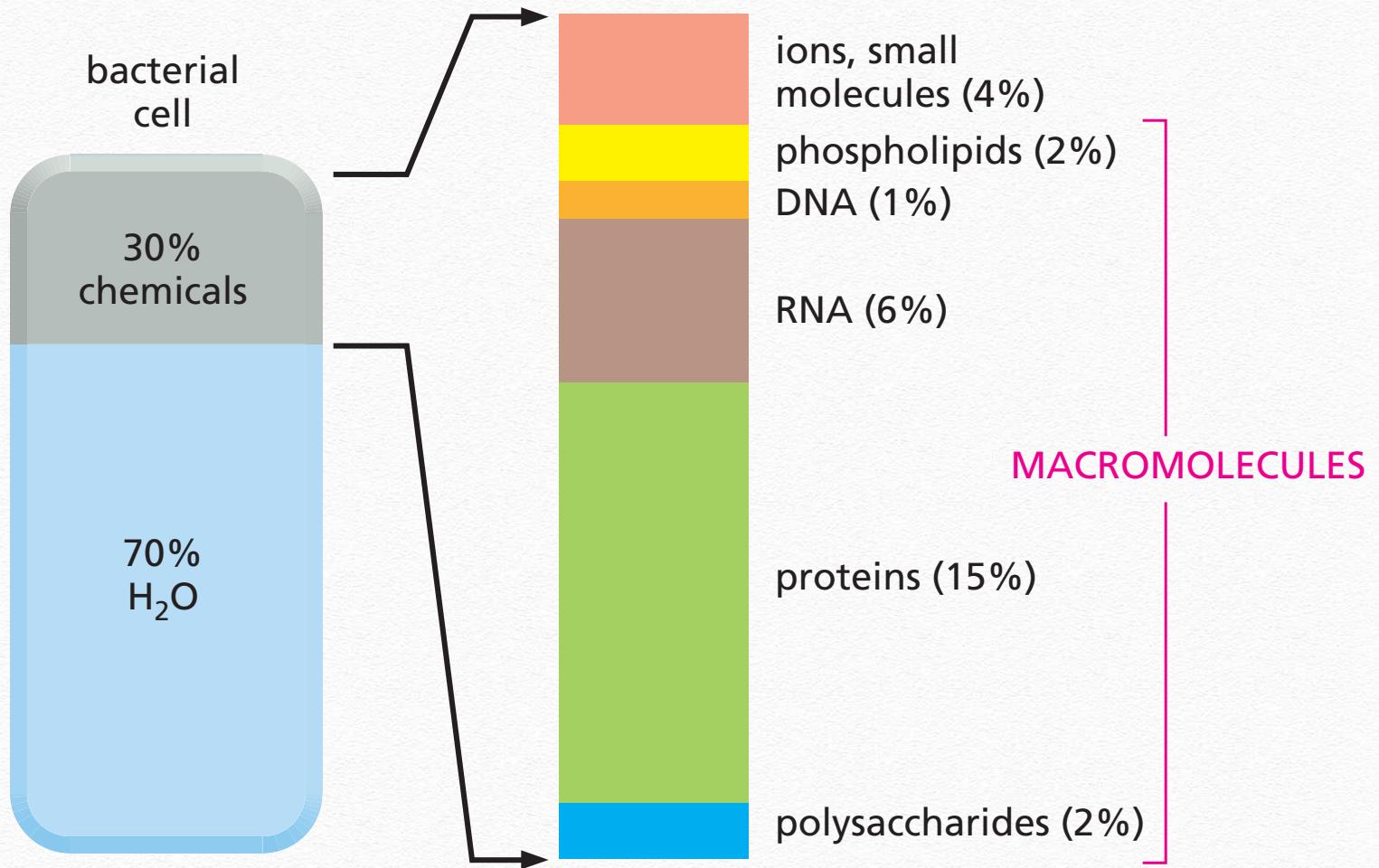


Courtesy of Ms. Razieh Karamzadeh, Royan Institute for Stem Cell Biology & Technology

From Organs to Atoms



Molecules in a Cell



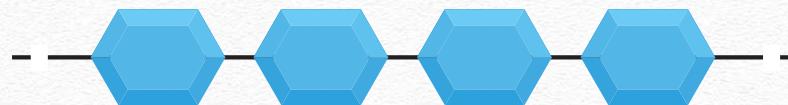
Macromolecules

SUBUNIT



sugar

MACROMOLECULE



polysaccharide



amino
acid



protein

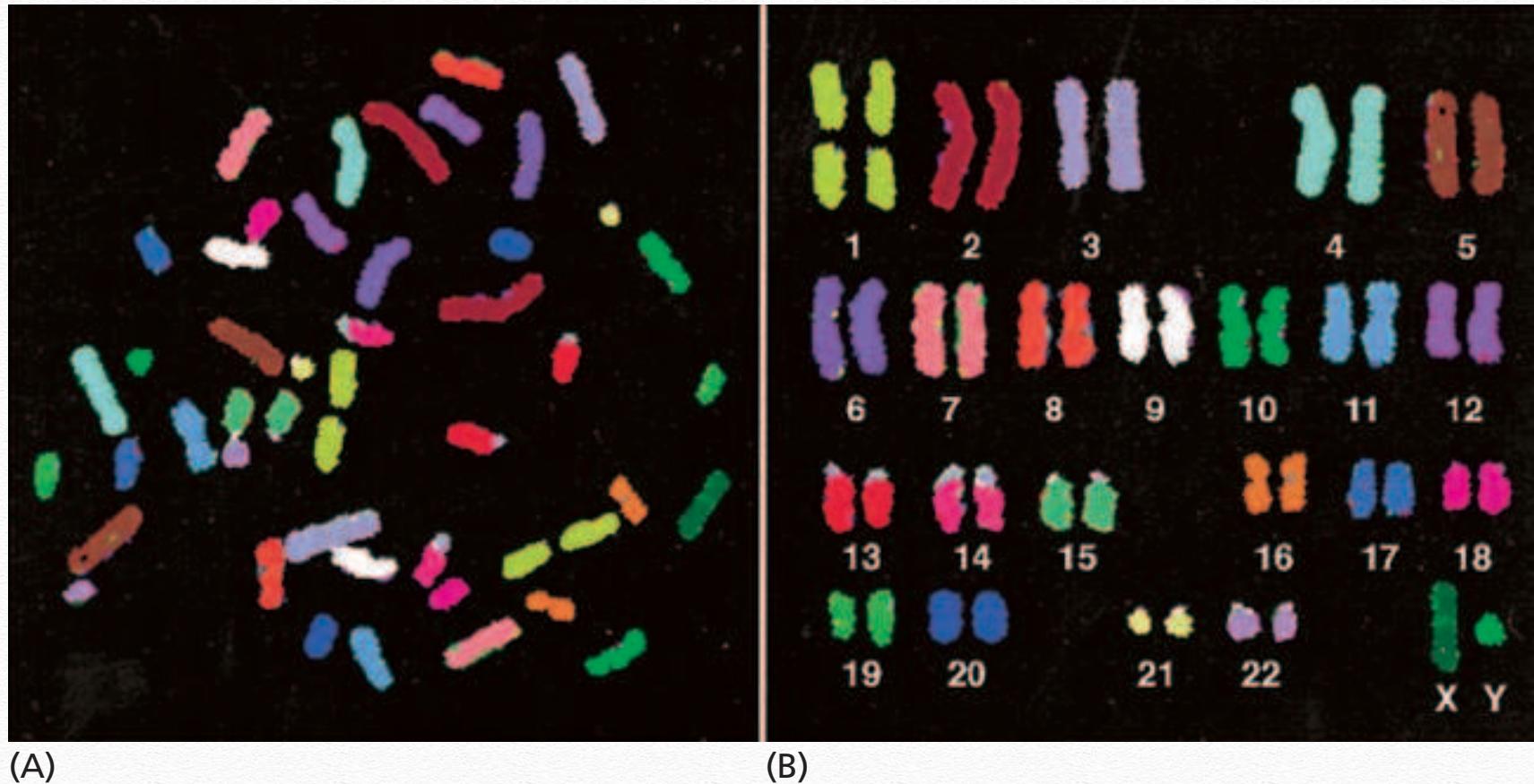


nucleotide

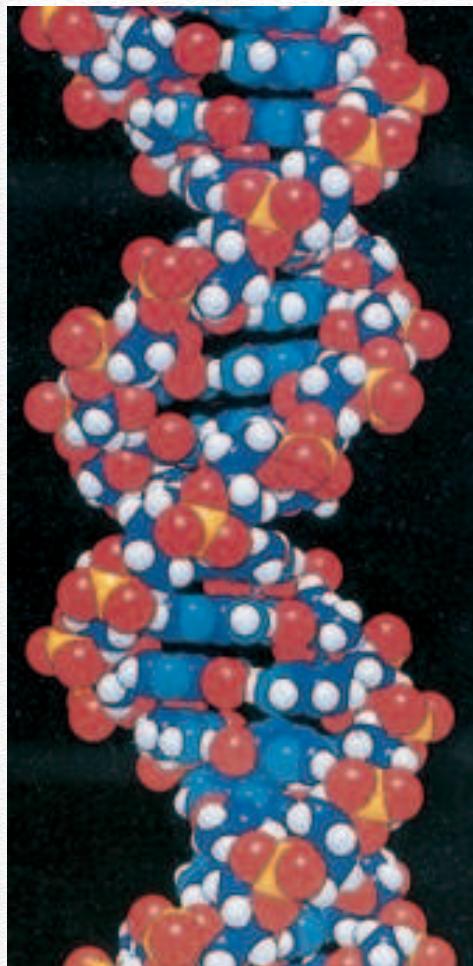


nucleic acid

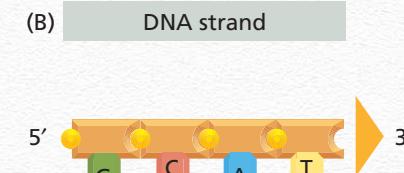
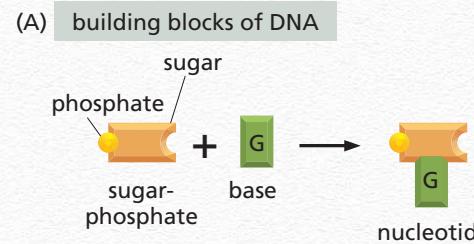
The Genetic Material



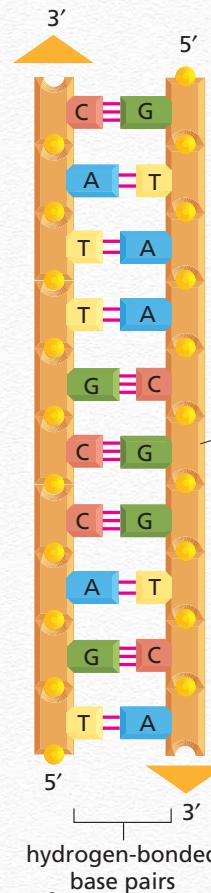
DNA



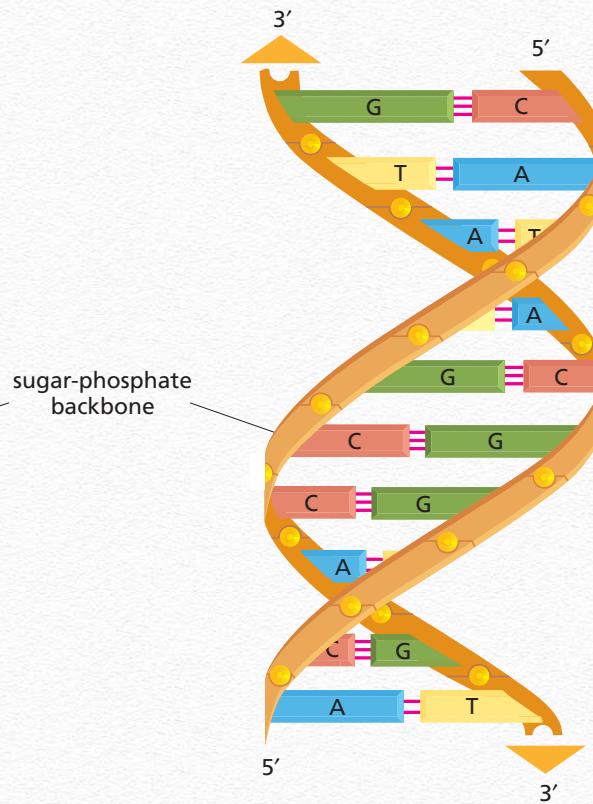
minor
groove
major
groove



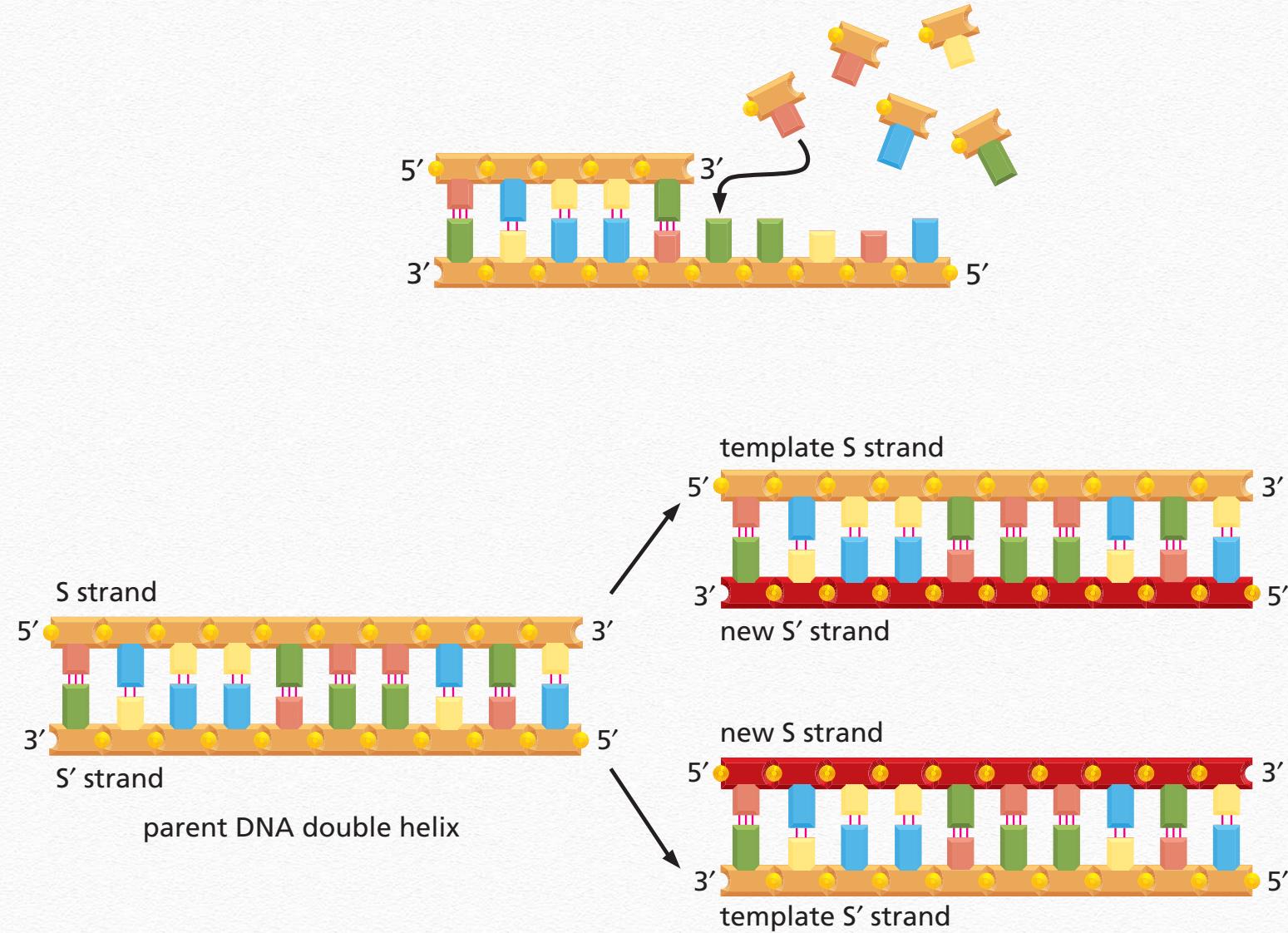
(C) double-stranded DNA



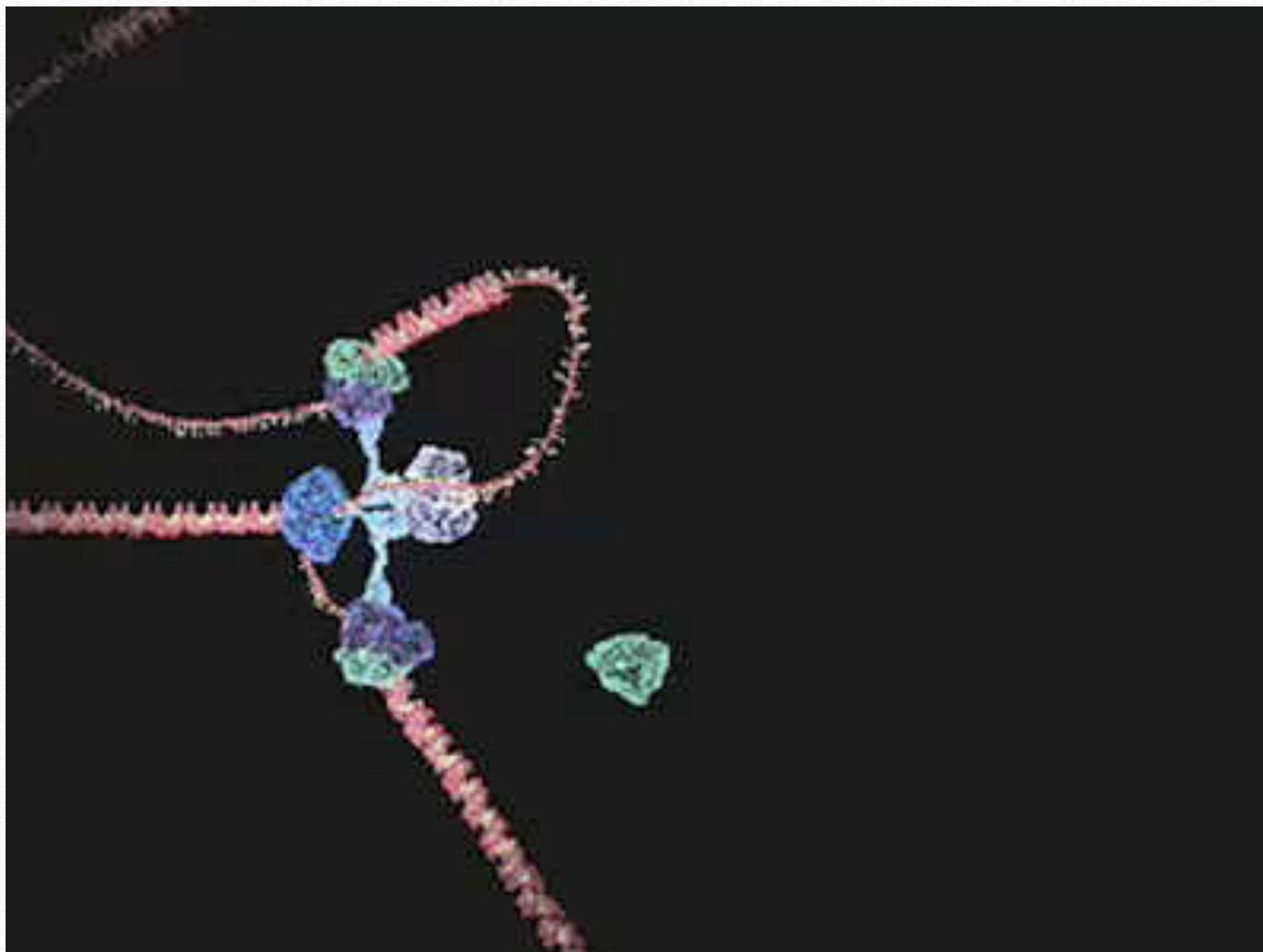
(D) DNA double helix



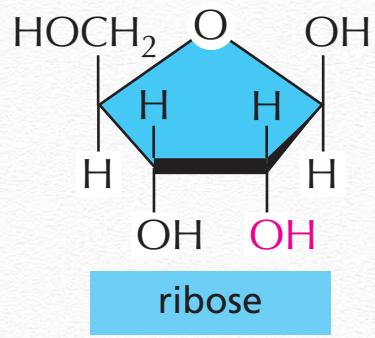
DNA as a Template



DNA Replication Machine

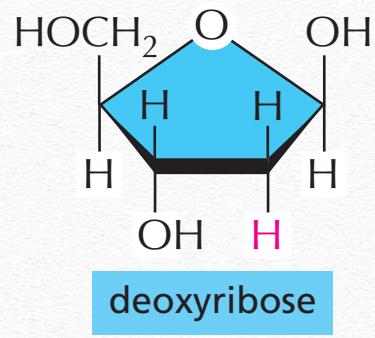


Chemical Structure of RNA



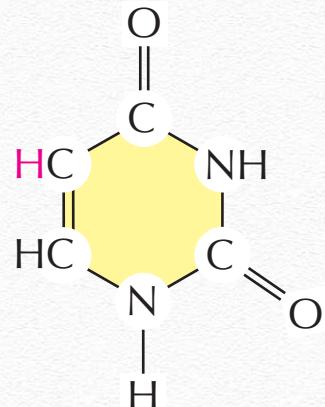
ribose

used in ribonucleic acid (RNA)



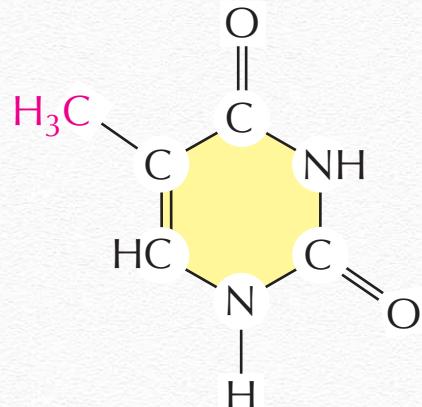
deoxyribose

used in deoxyribonucleic acid (DNA)



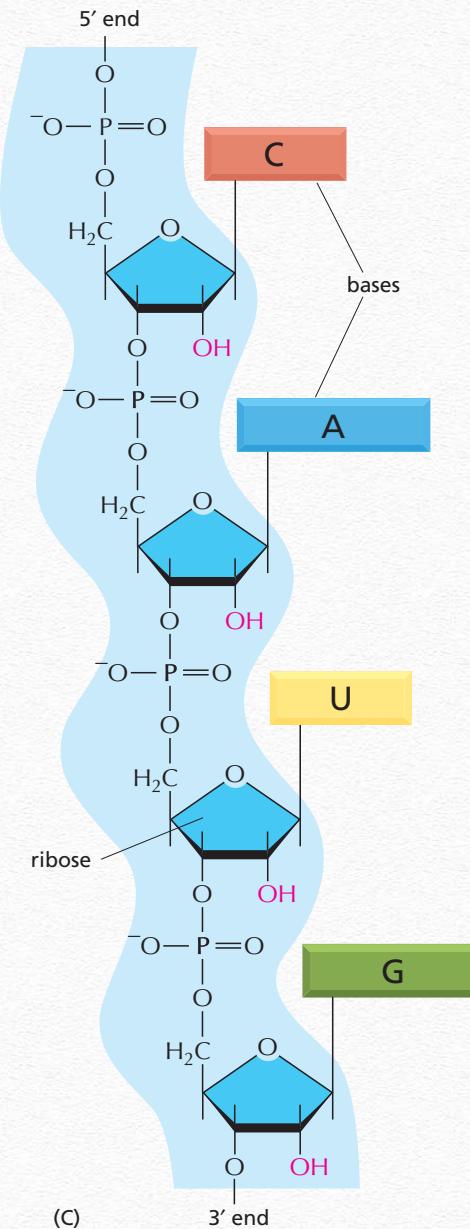
uracil

used in RNA

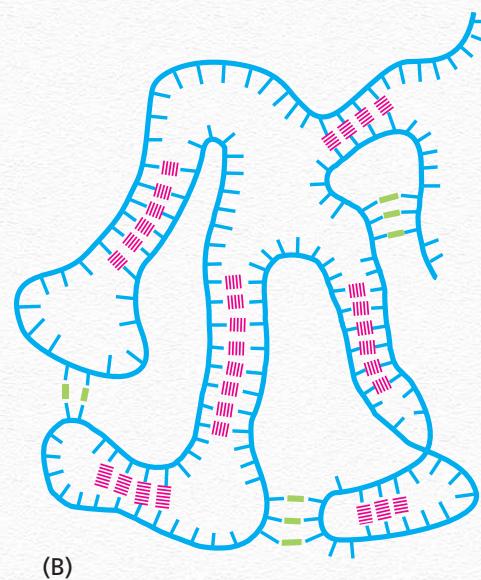
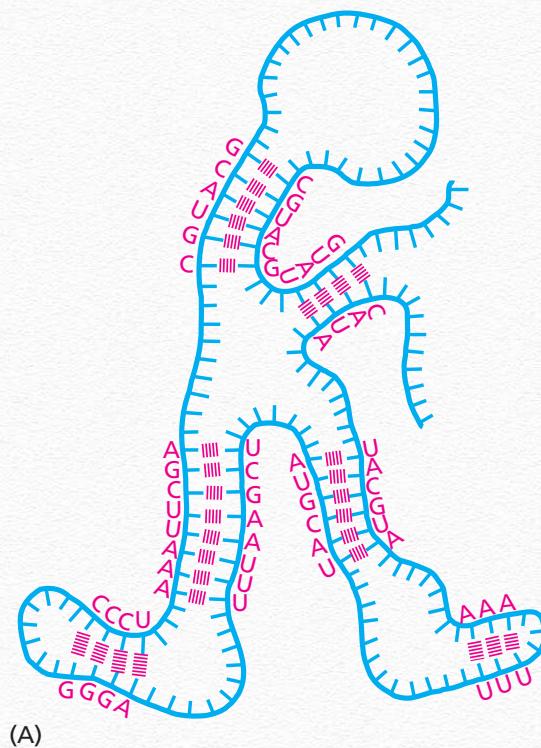


thymine

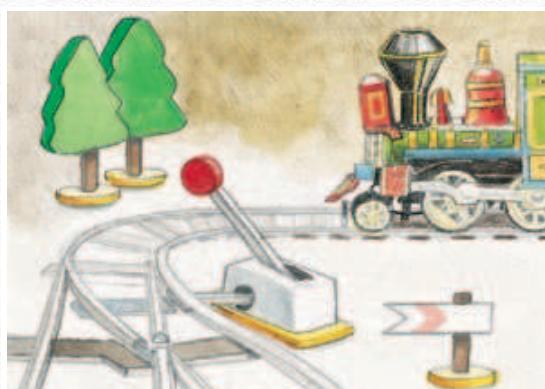
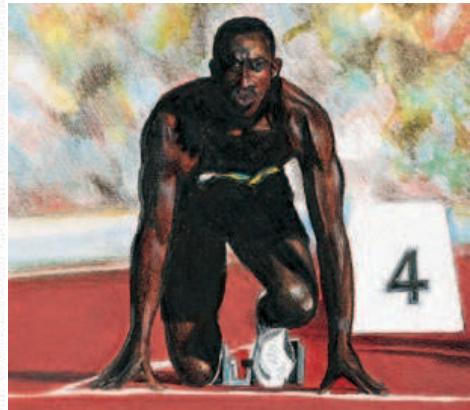
used in DNA



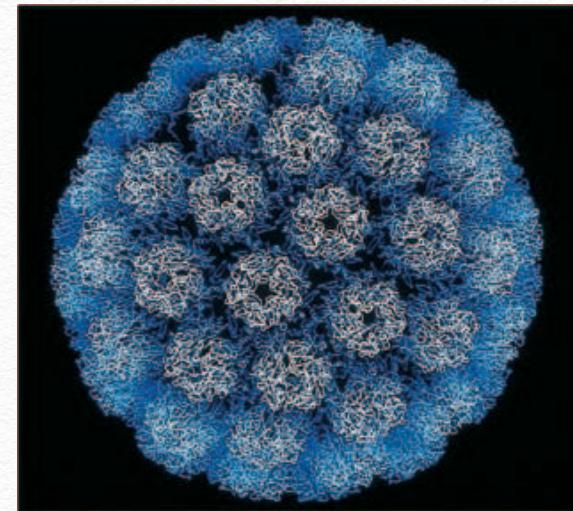
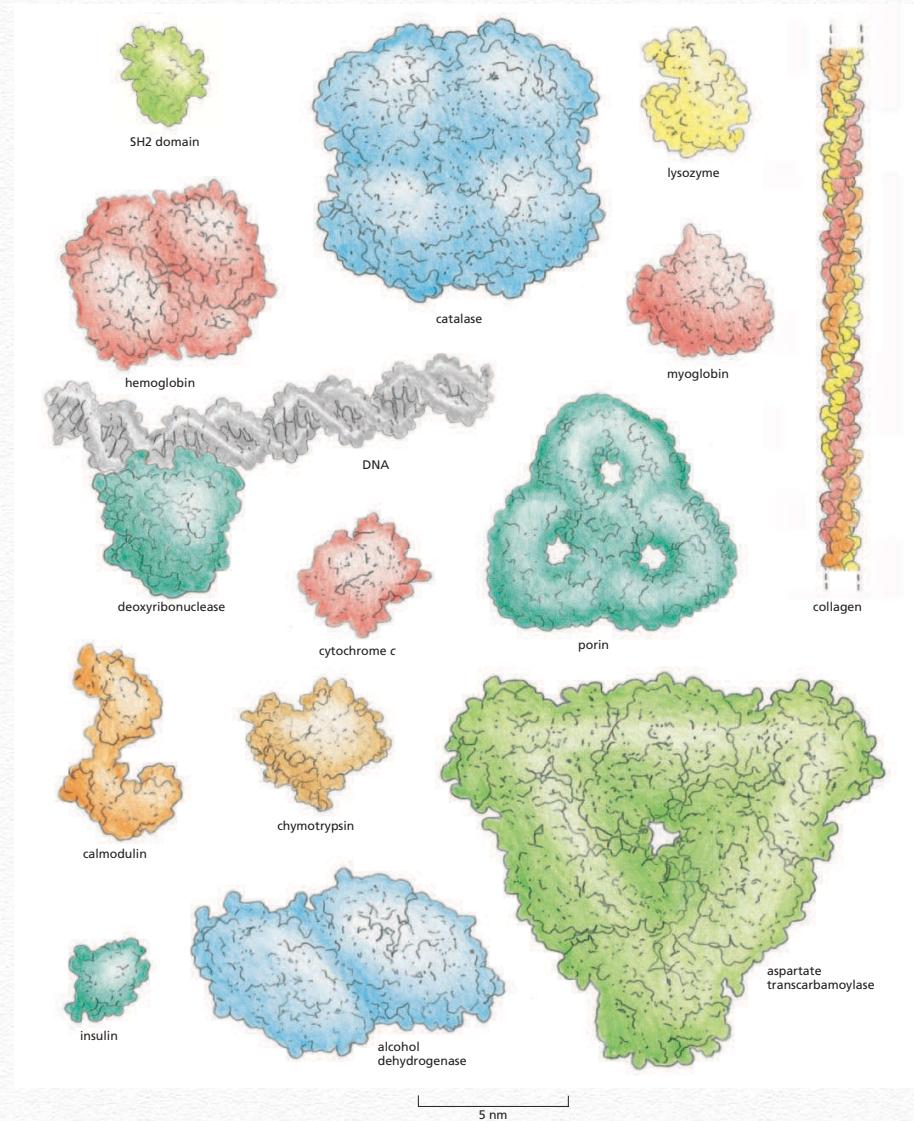
RNA Folding



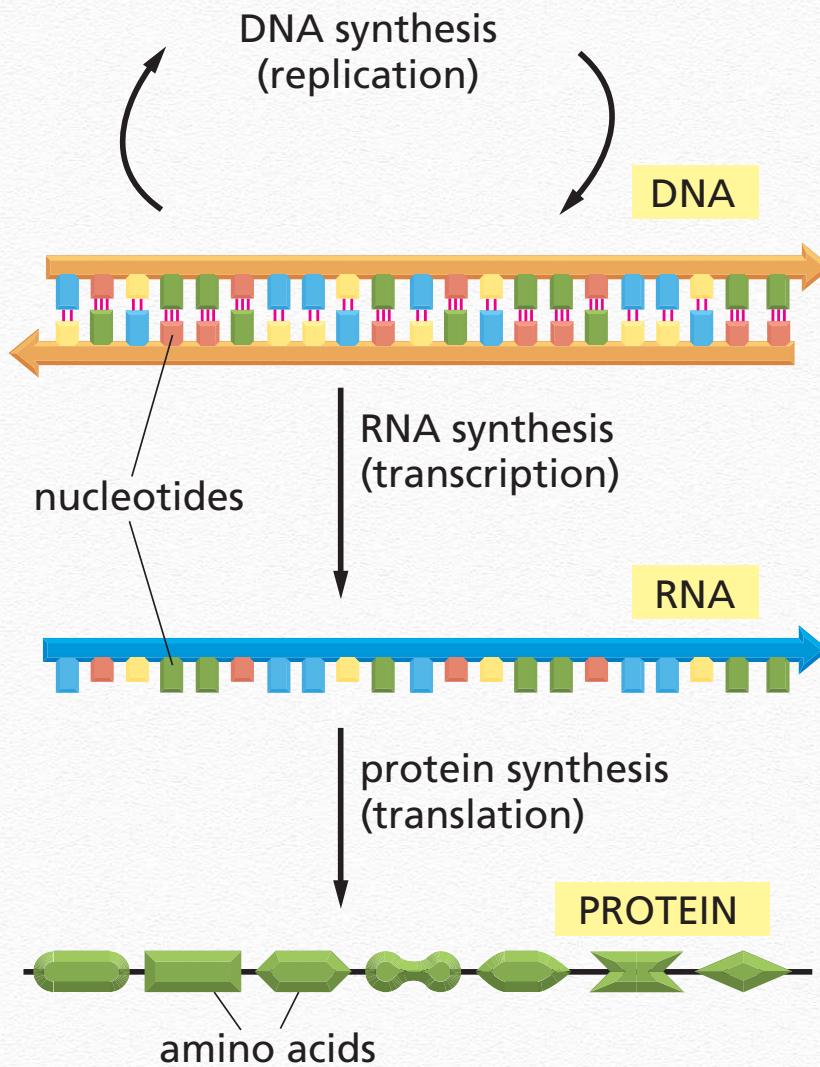
Protein Functions



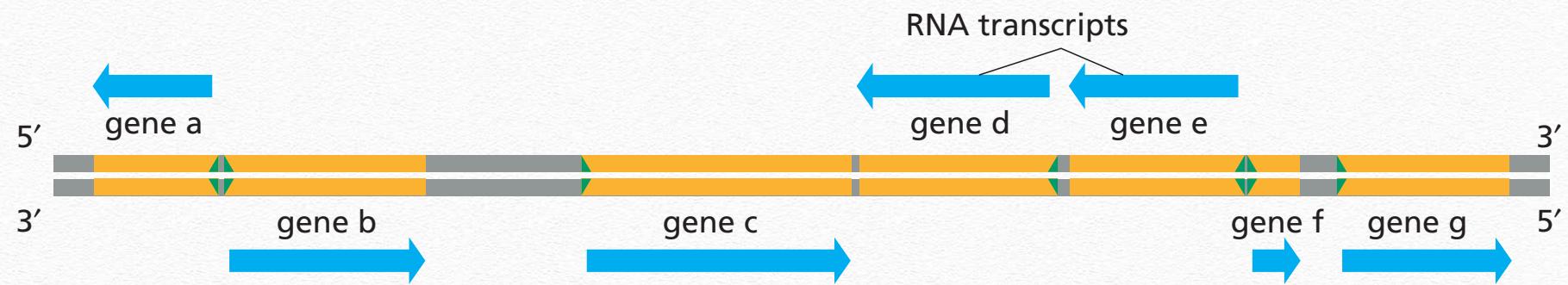
Protein Shapes and Sizes



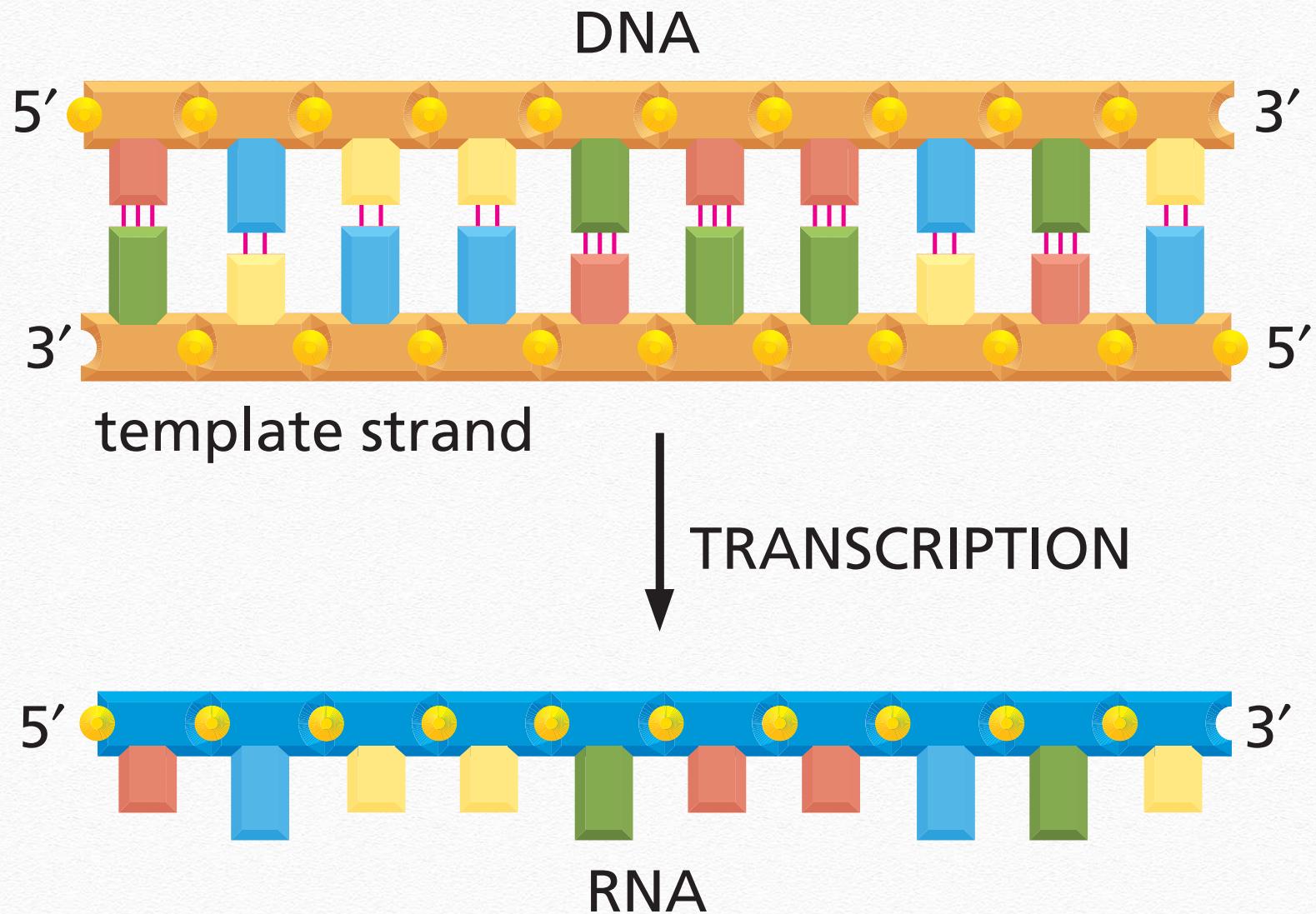
Central Dogma



The Genes



Transcription



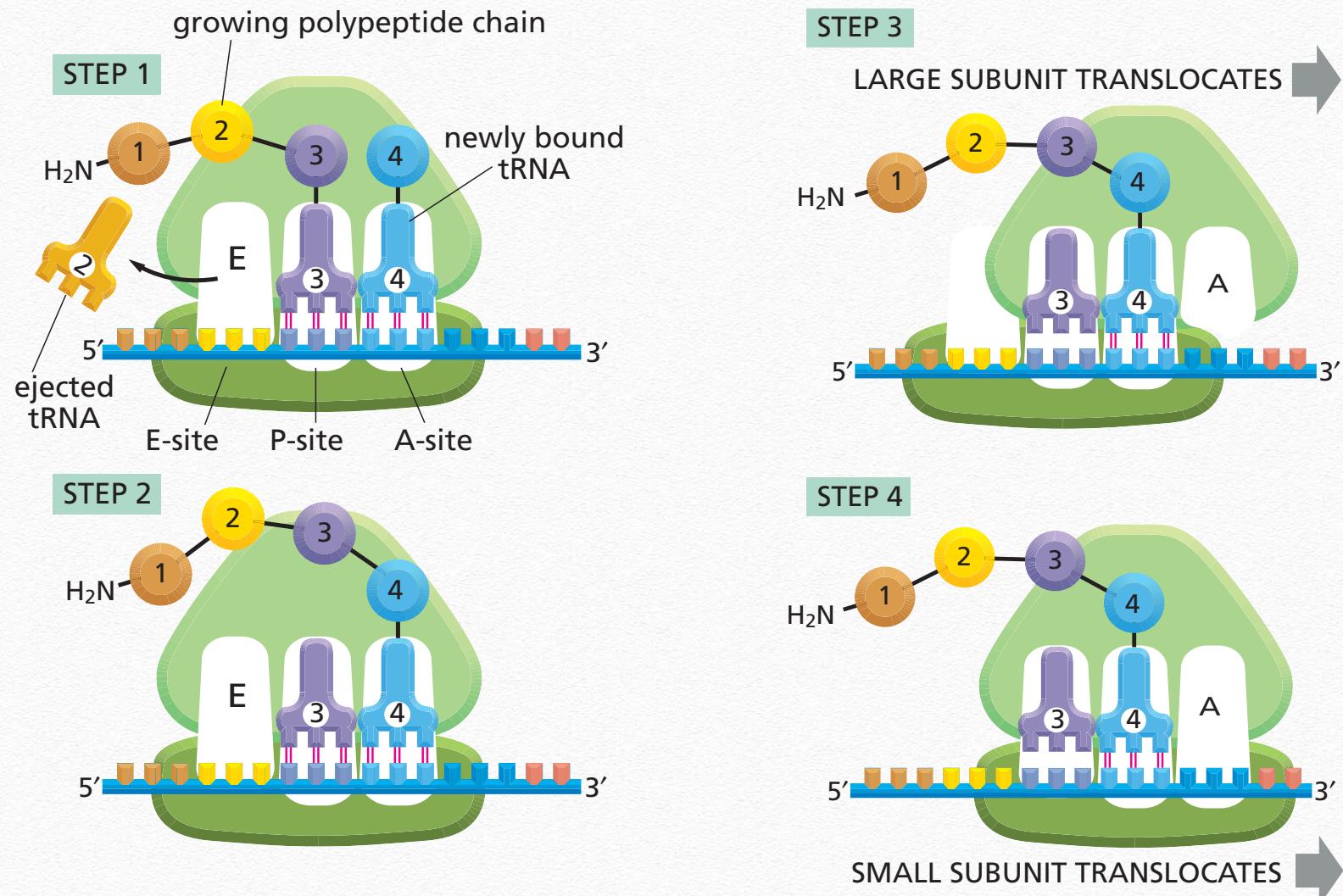
Transcription



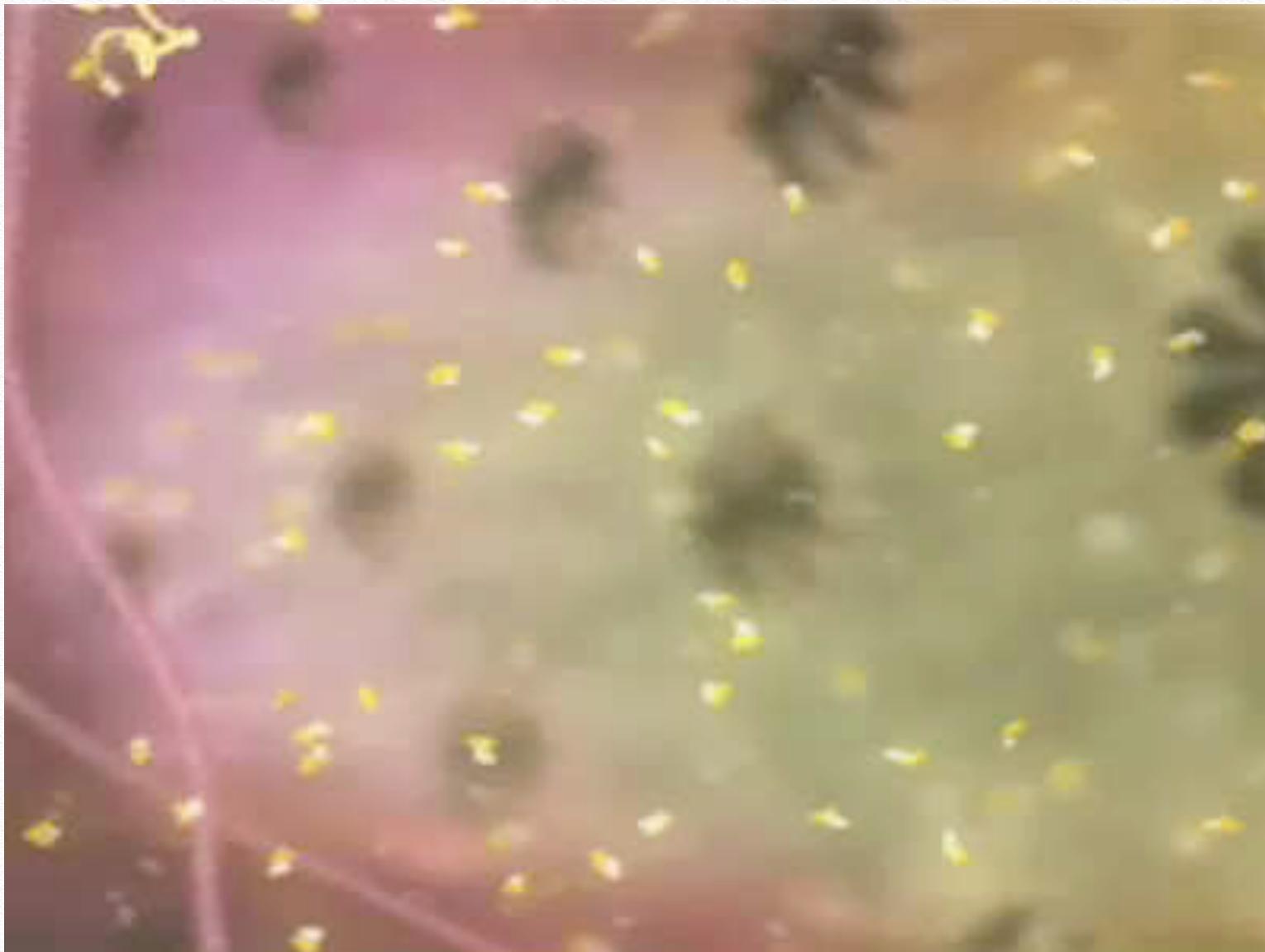
64 -> 20 Code Table

AGA										UUA									AGC				
AGG										UUG									AGU				
GCA	CGA									CUA									UCA			ACA	
GCC	CGC									CUC									CCC			ACC	
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGA	GGC	CAC	AUC	CUG	AAA	UUC	CCG	UCG	ACG		GUA					
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	GGG	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	GUC					
																		UAC					
																		GUU					
																		UAG					
																		UGA					
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val			stop	
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V				

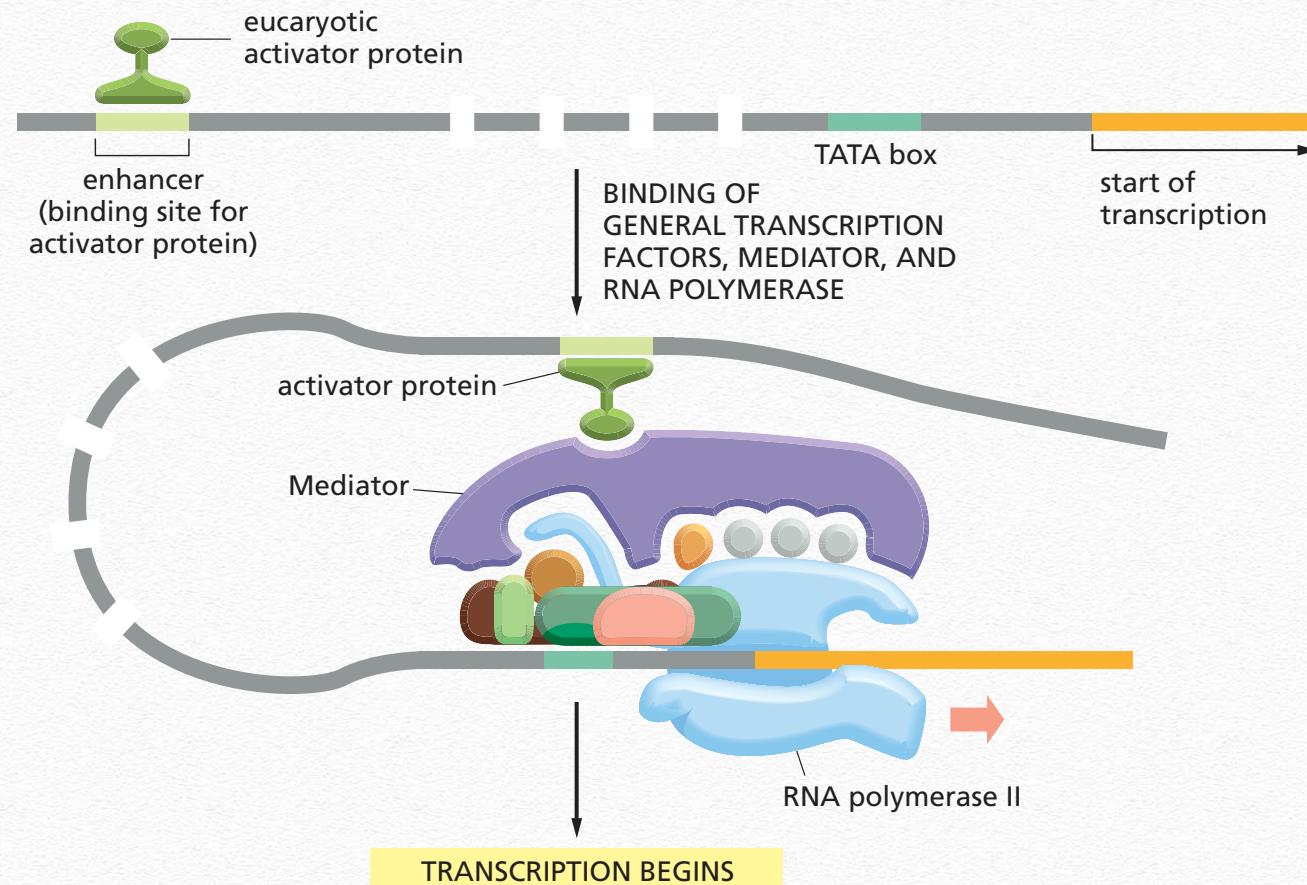
4-Steps in Translation



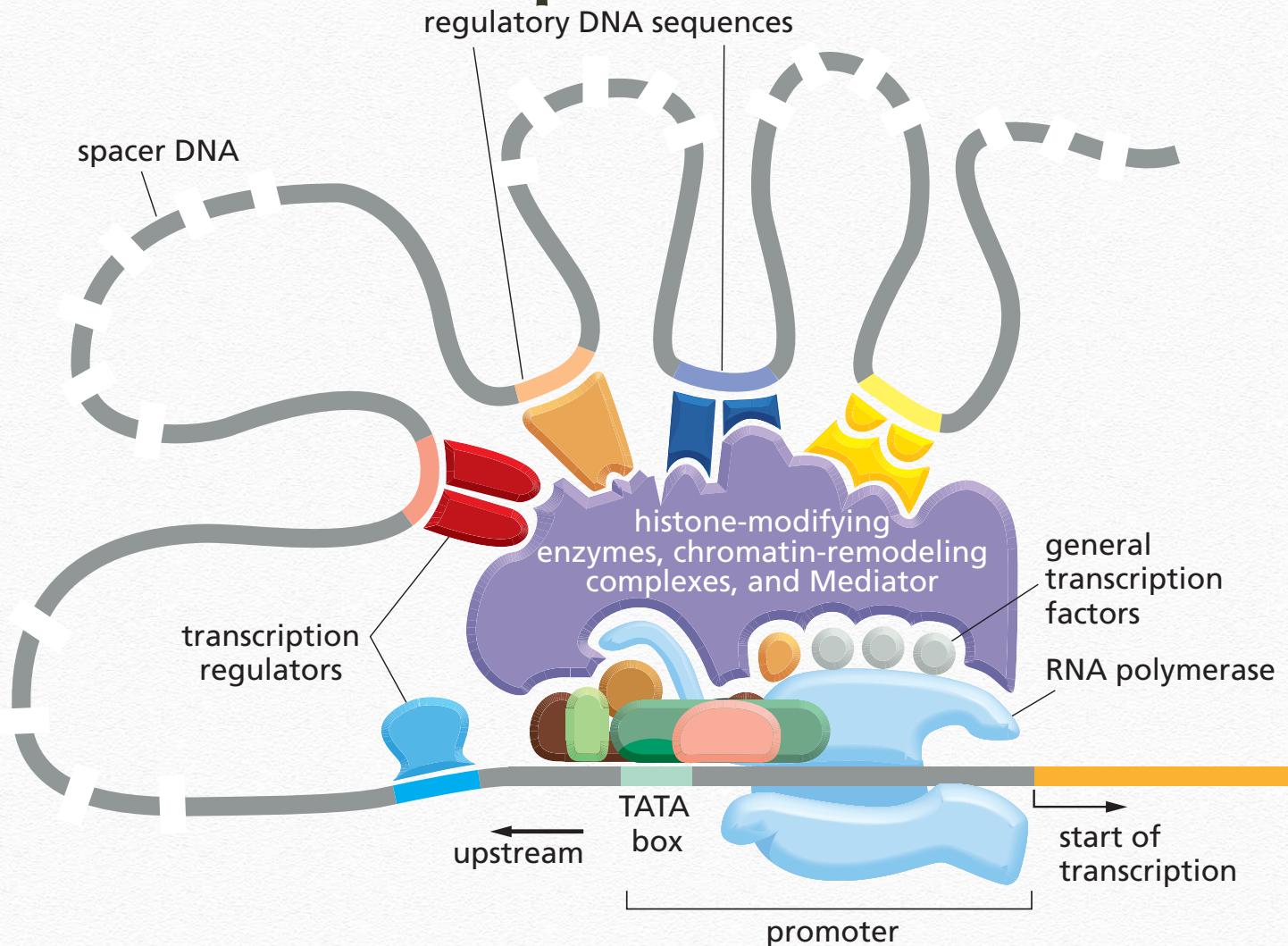
Translation



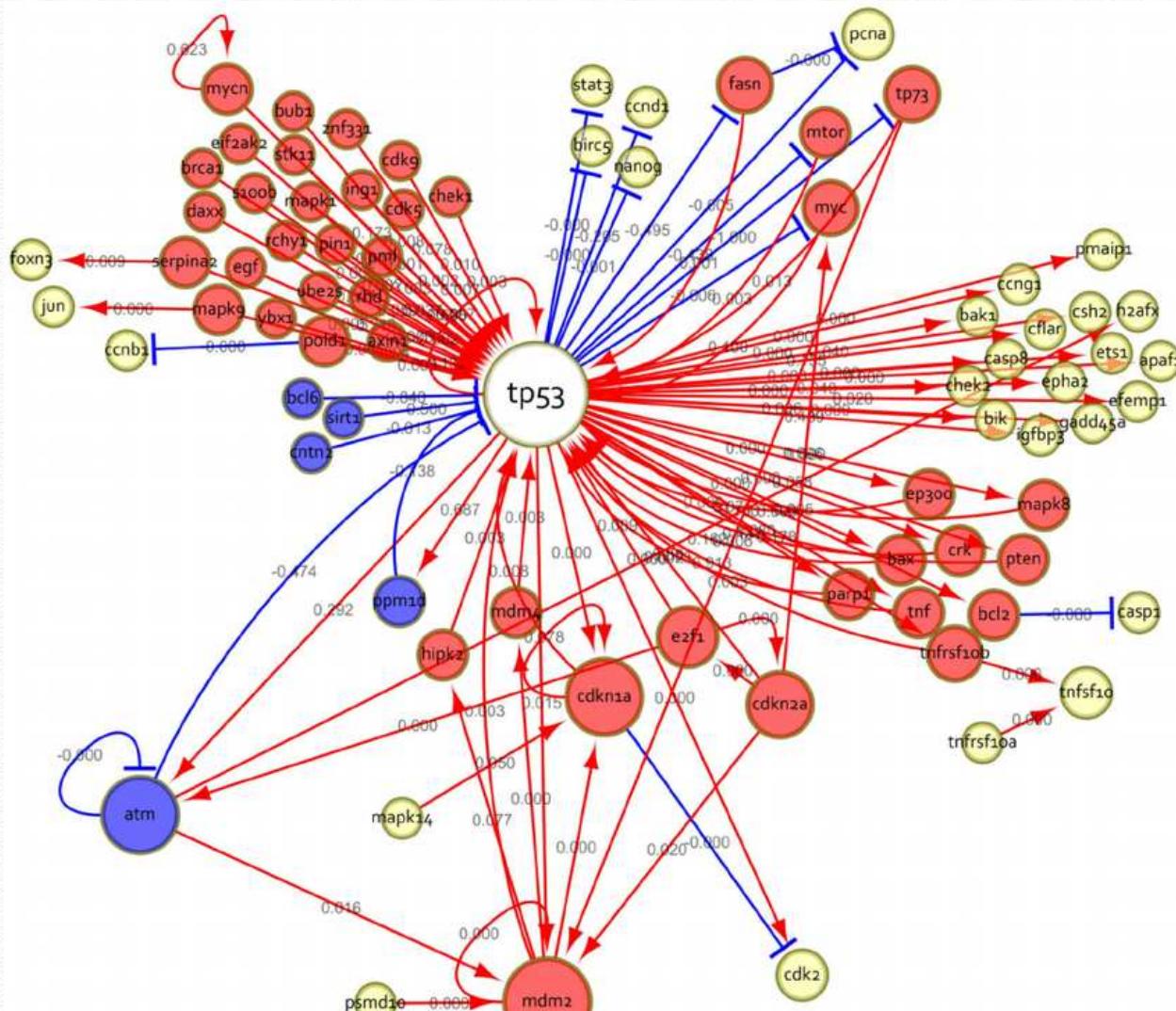
How do Genes Interact?

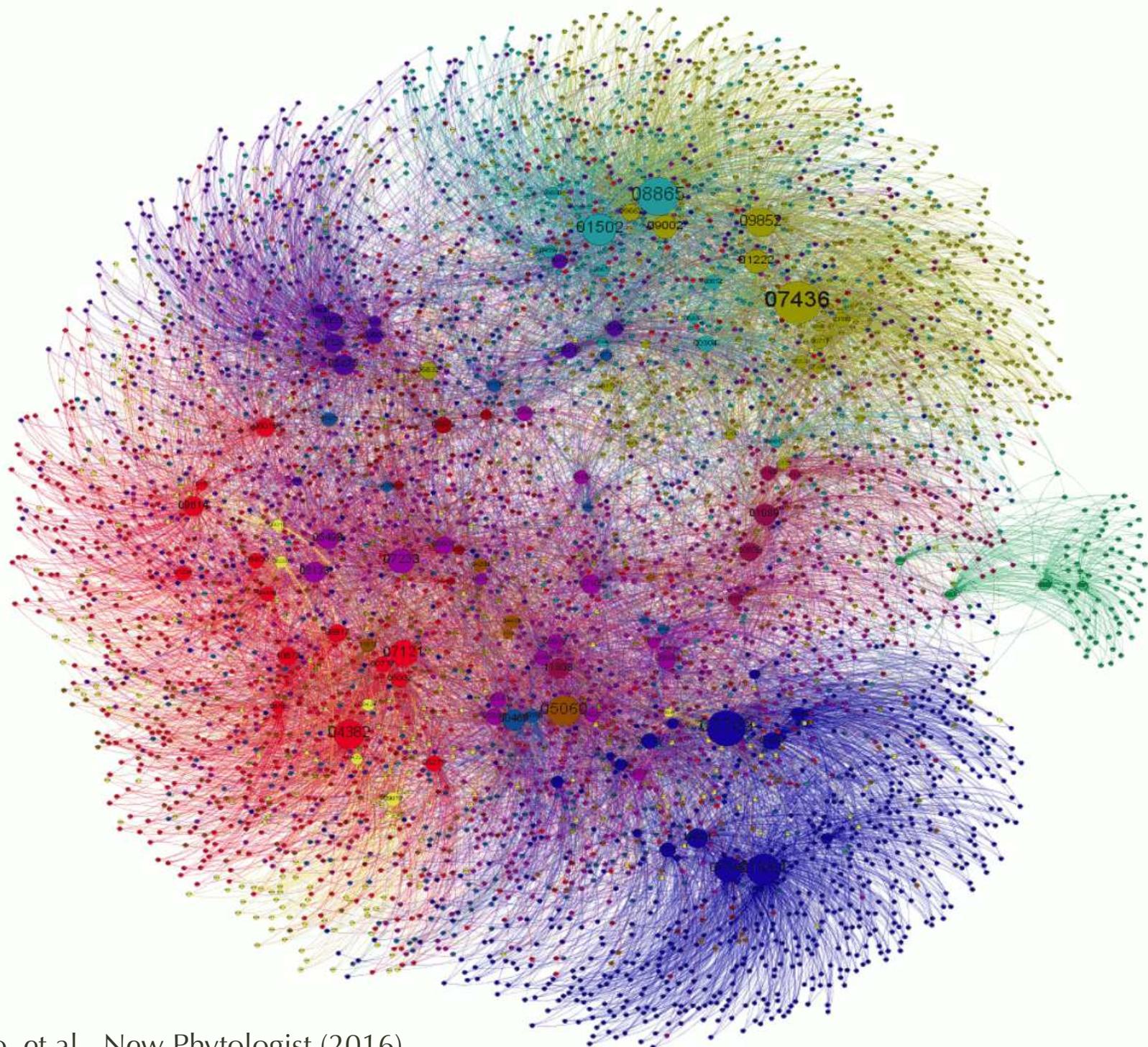


Can be More Complicated

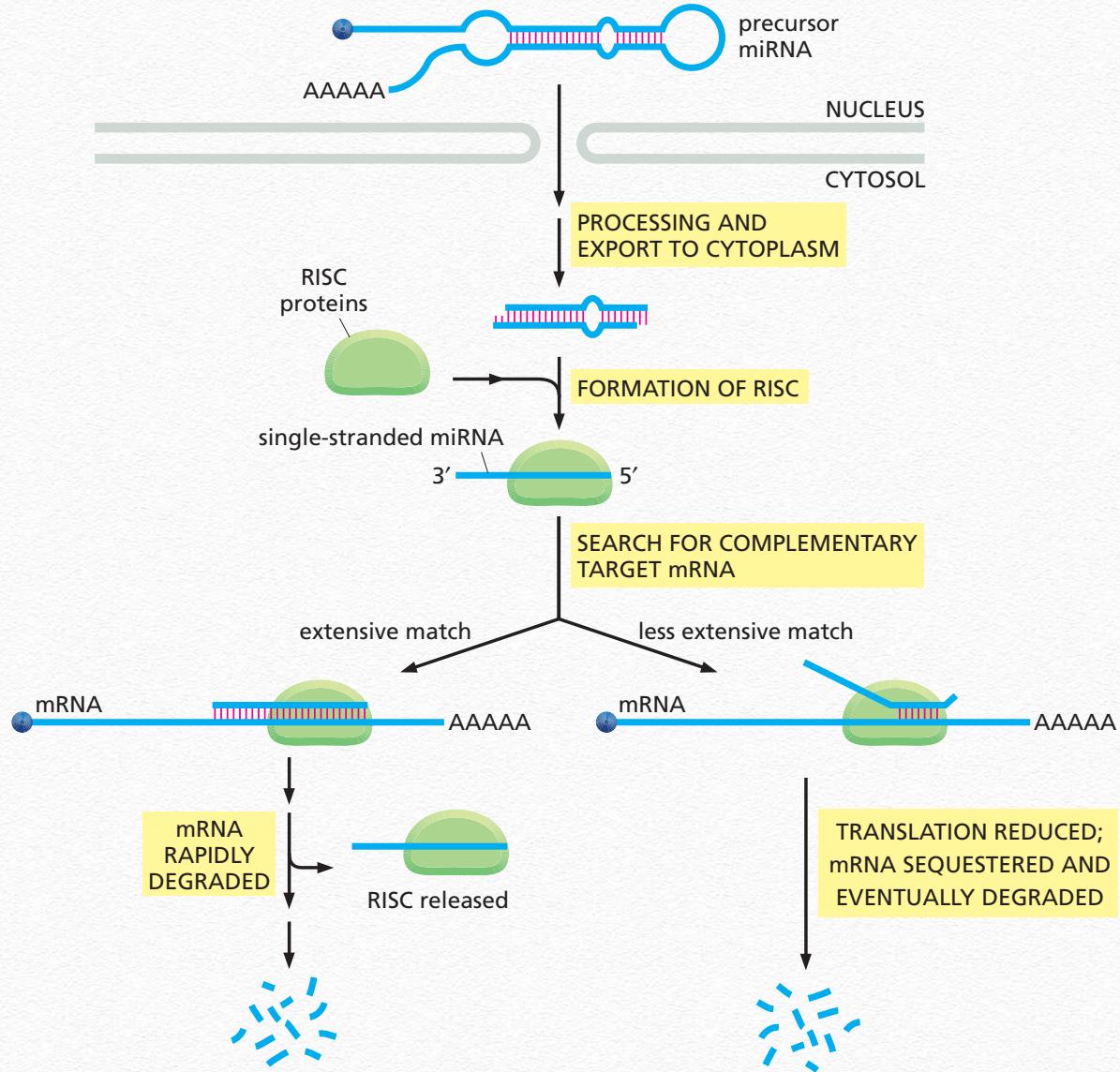


Gene Regulatory Networks (GRN)

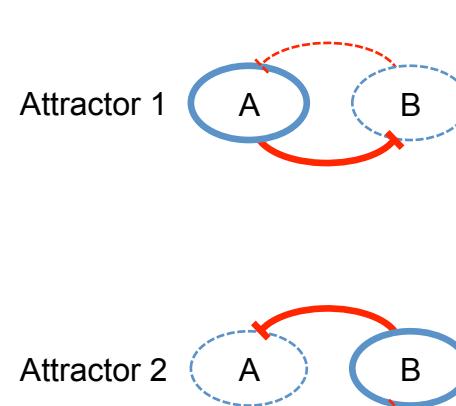
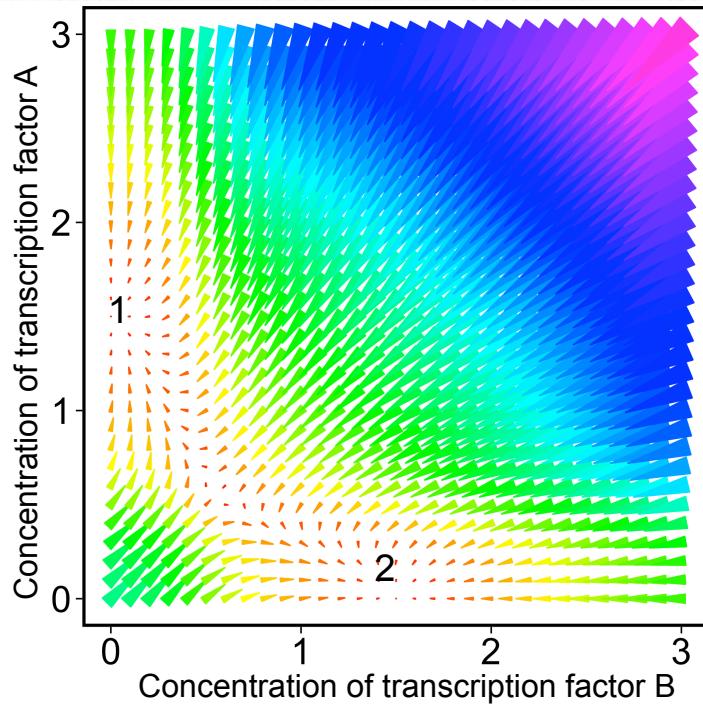
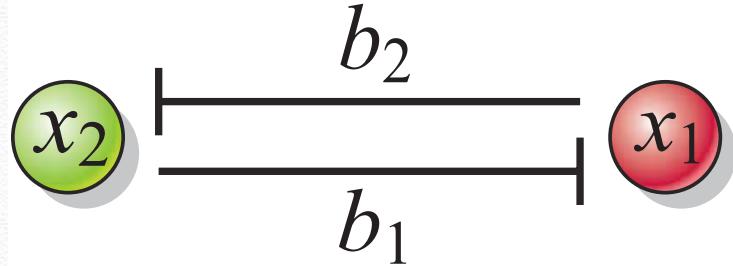




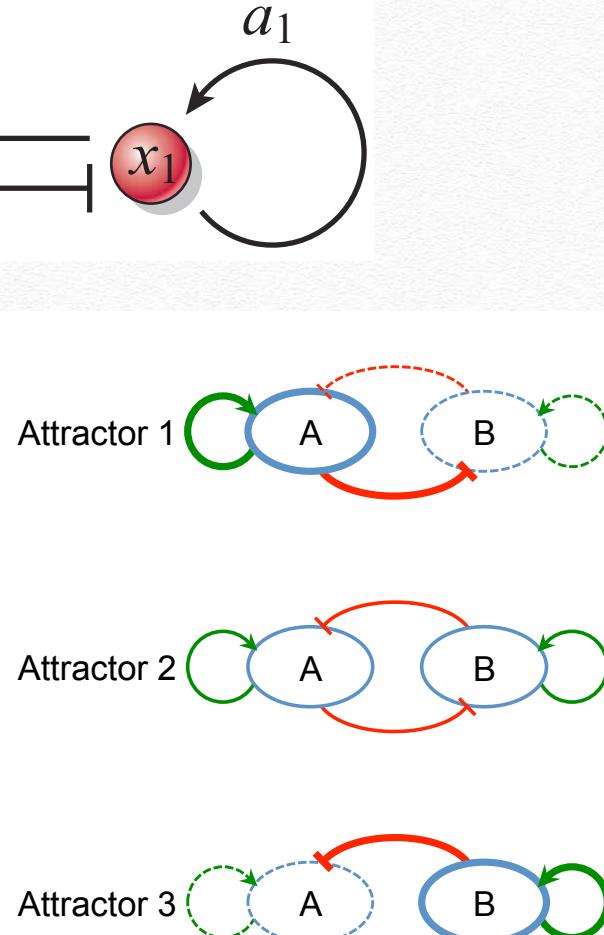
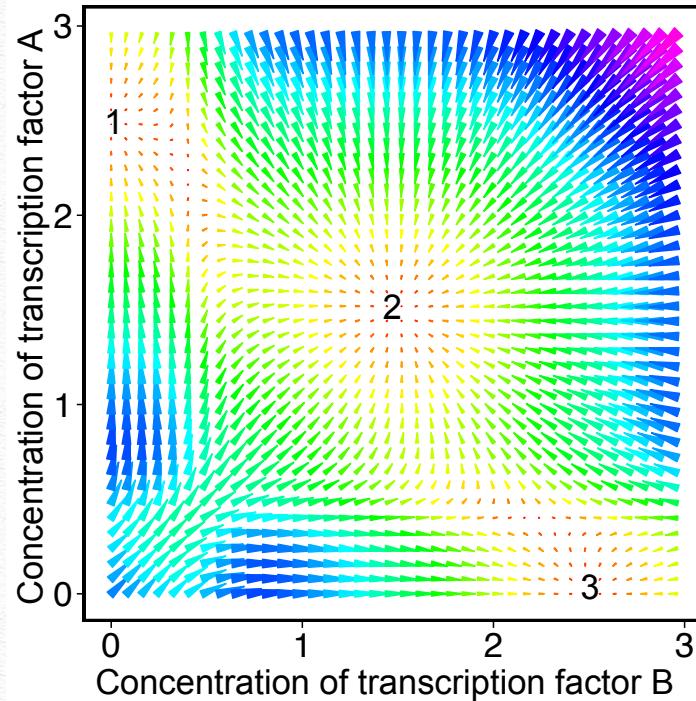
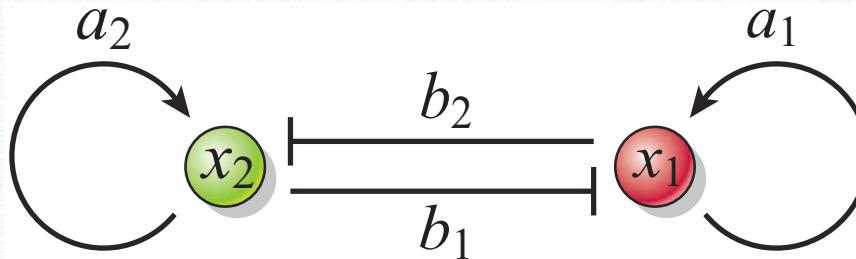
microRNAs



Bistable Switch



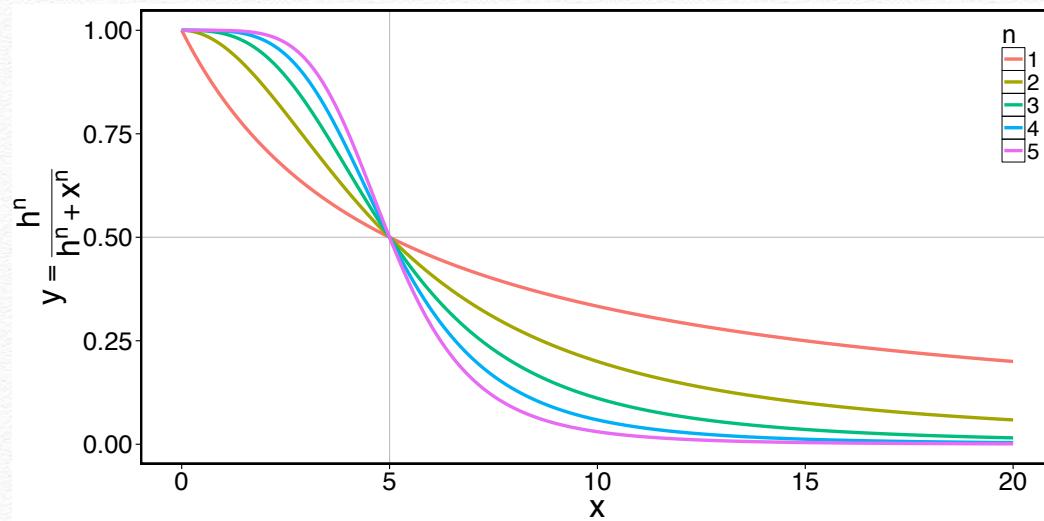
Tristable Switch



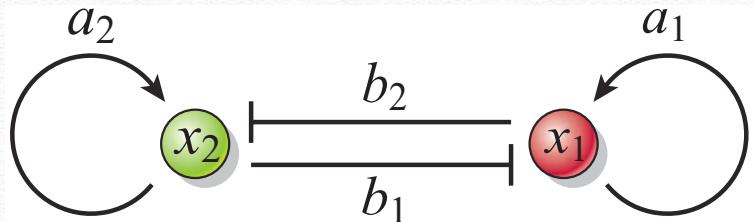
Mutual Inhibition



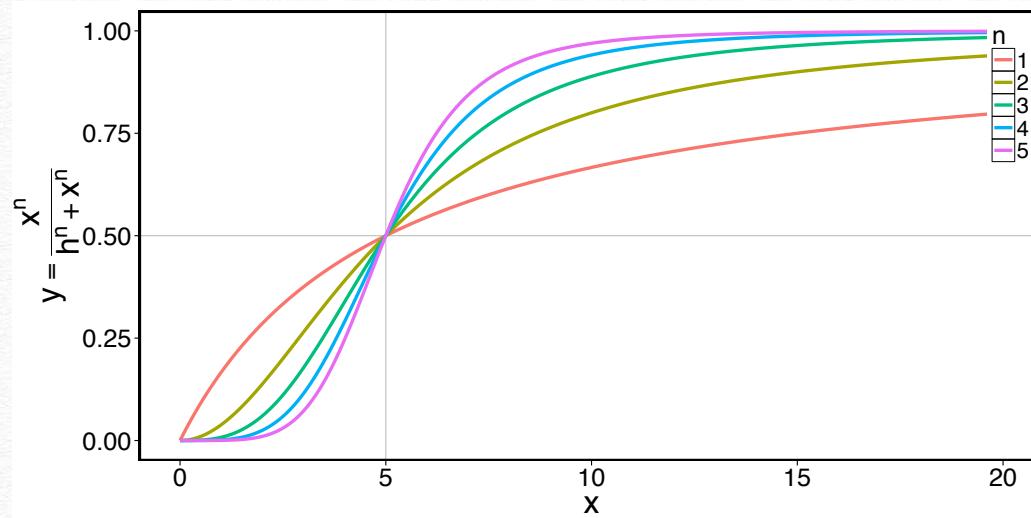
$$y = \frac{h^n}{h^n + x^n}$$



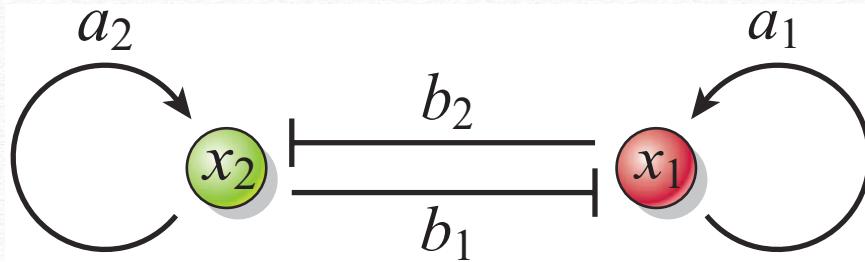
Self Activation



$$y = \frac{h^n}{h^n + x^n} \rightarrow y = \frac{x^n}{h^n + x^n}$$



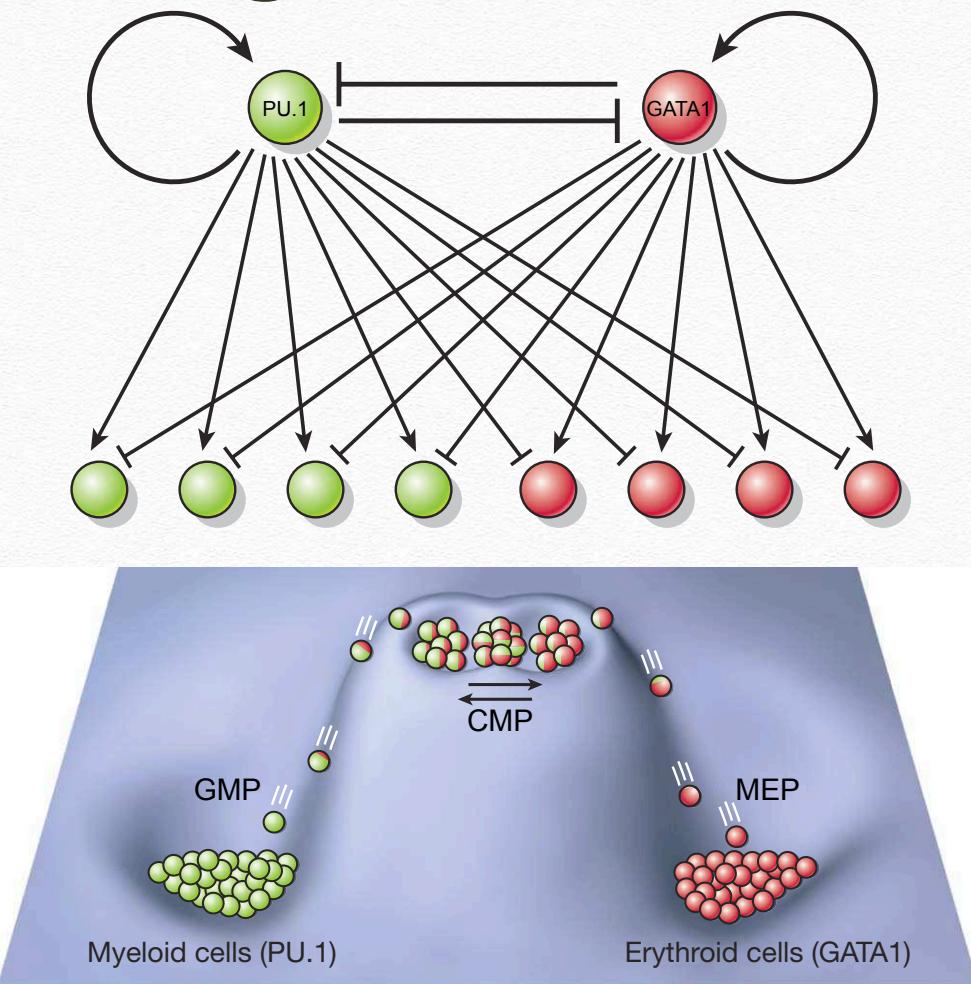
Together



$$\frac{dx_1}{dt} = a_1 \frac{x_1^n}{h^n + x_1^n} + b_1 \frac{h^n}{h^n + x_2^n} - cx_1$$

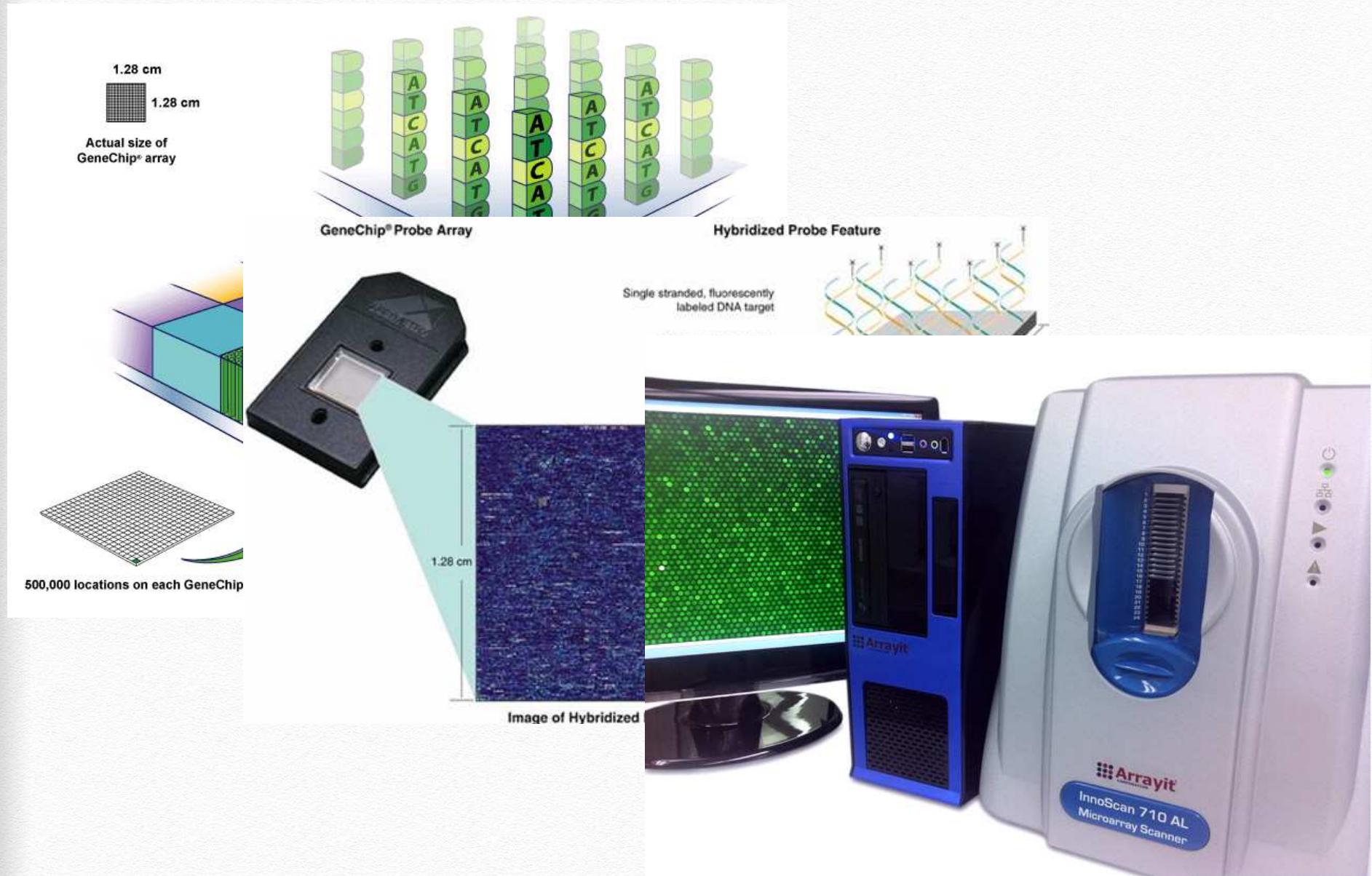
$$\frac{dx_2}{dt} = a_2 \frac{x_2^n}{h^n + x_2^n} + b_2 \frac{h^n}{h^n + x_1^n} - cx_2$$

Lineage Bifurcation



Graf, Thomas and Enver, Tariq. Nature, 462(7273):587– 594 (2009).

Microarrays



Next Generation Sequencing (NGS)



Roche GS-FLX



Life Technologies SOLiD



Illumina HiSeq



Life Technologies Ion Torrent/Proton

NCBI Resources How To

GEO Home Documentation Query & Browse Email GEO

asharifz@gmail.com My NCBI Sign Out

My GEO Submissions

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.



Keyword or GEO Accession

ArrayExpress

Home Browse Submit Help About ArrayExpress

Search Examples: E-MEXP-31, cancer, p53, Geuvadis

Contact Us

ArrayExpress – functional genomics data

ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

Data Content

Updated today at 03:00

- 69642 experiments
- 2198358 assays
- 44.50 TB of archived data



EVENTS | NEWS | PUBLICATIONS | PRESENTATIONS | [Twitter](#) | [LinkedIn](#) | CONTACT

WHO WE ARE | RESEARCH | CHALLENGES | PLATFORMS | MOBILE HEALTH | GOVERNANCE | JOIN OUR TEAM



NCBI Resources How To

SRA Advanced

asharifz@gmail.com My NCBI Sign Out



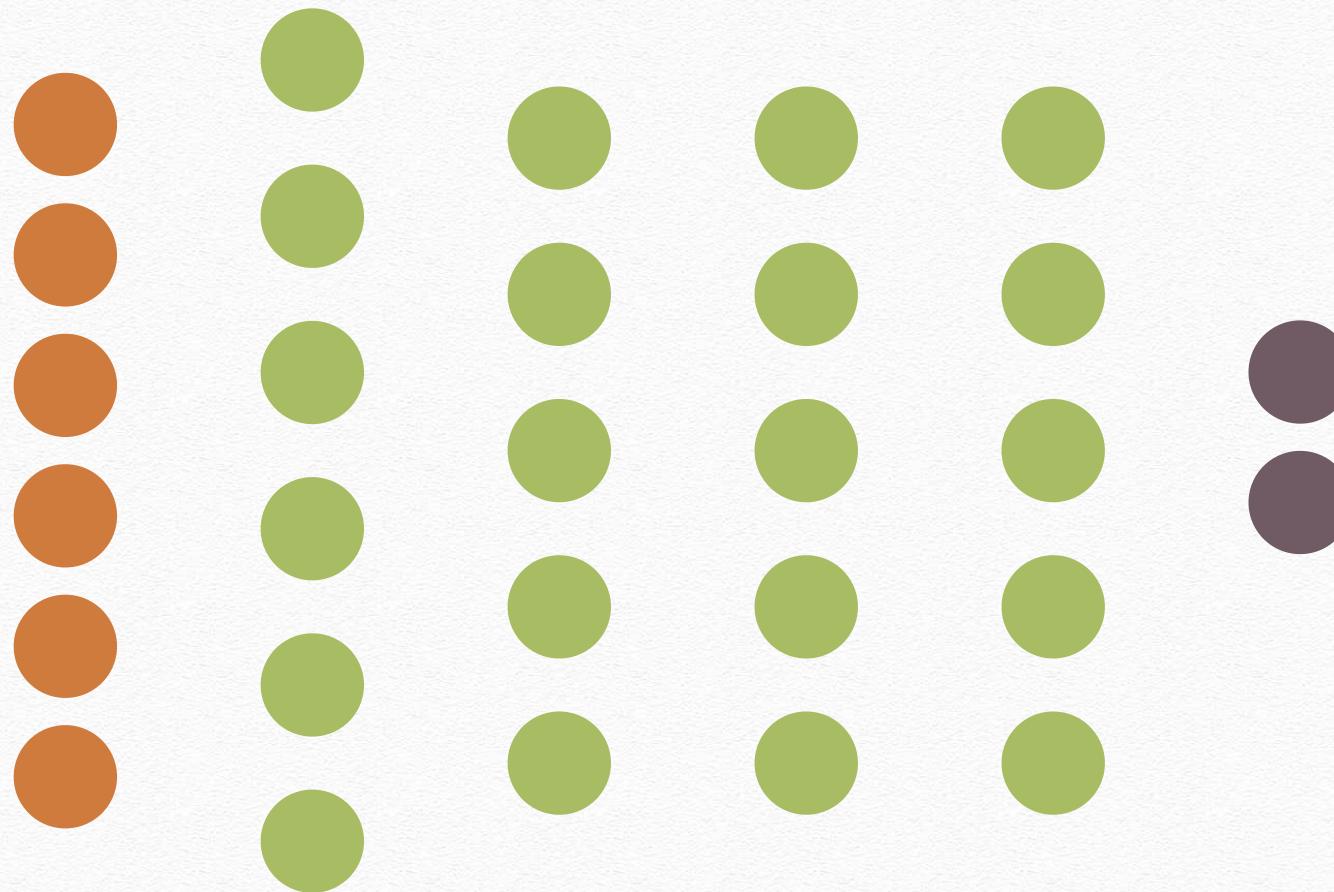
SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

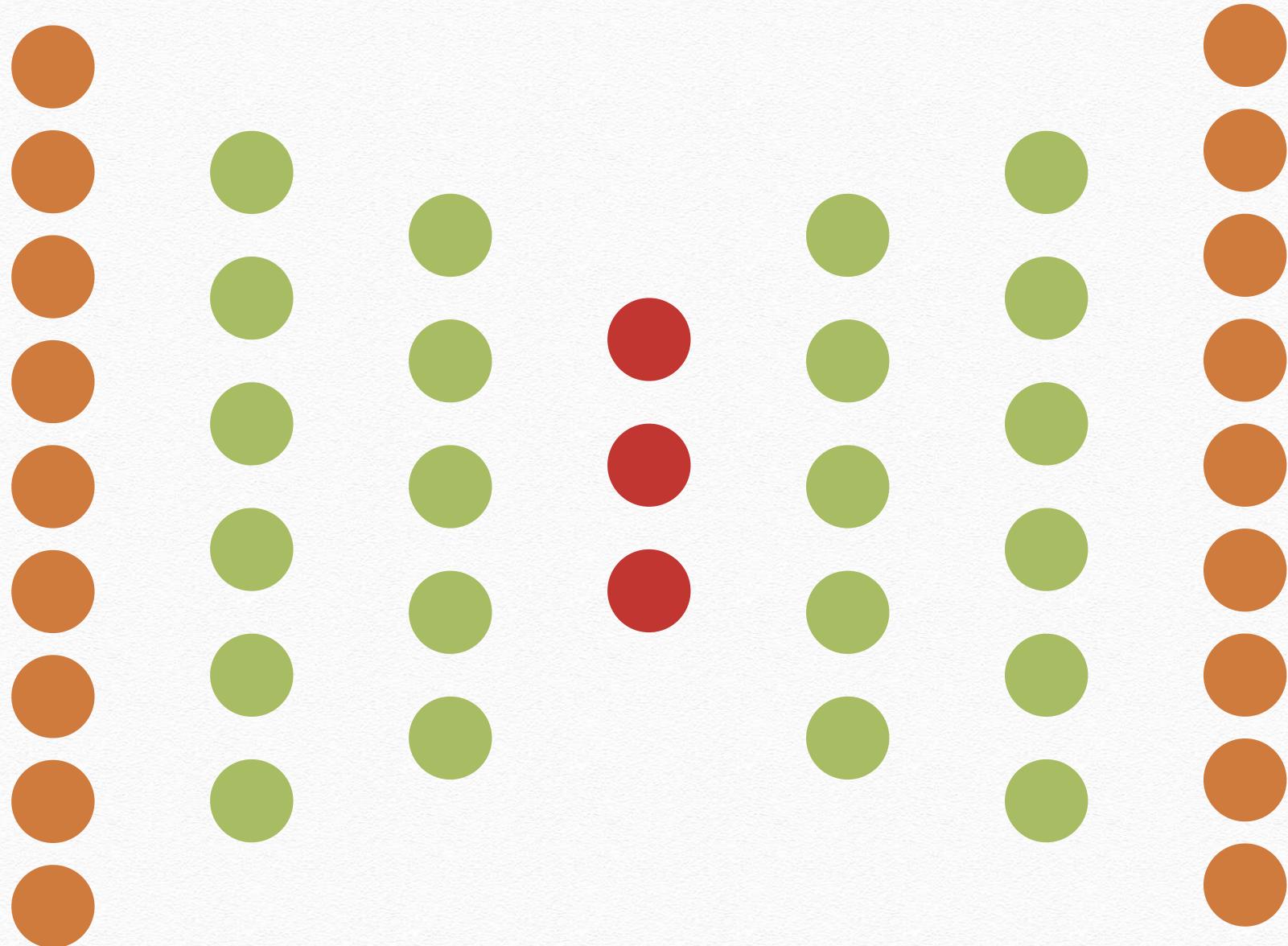
Take-home Messages

- ❖ DNA(Info.) → RNA (Info. + Func.) → Protein (Func.)
- ❖ The genes (parts of the genome that produce proteins) can regulate each other
- ❖ A (simplistic) view of the cell identity is its state in Gene Expression Space

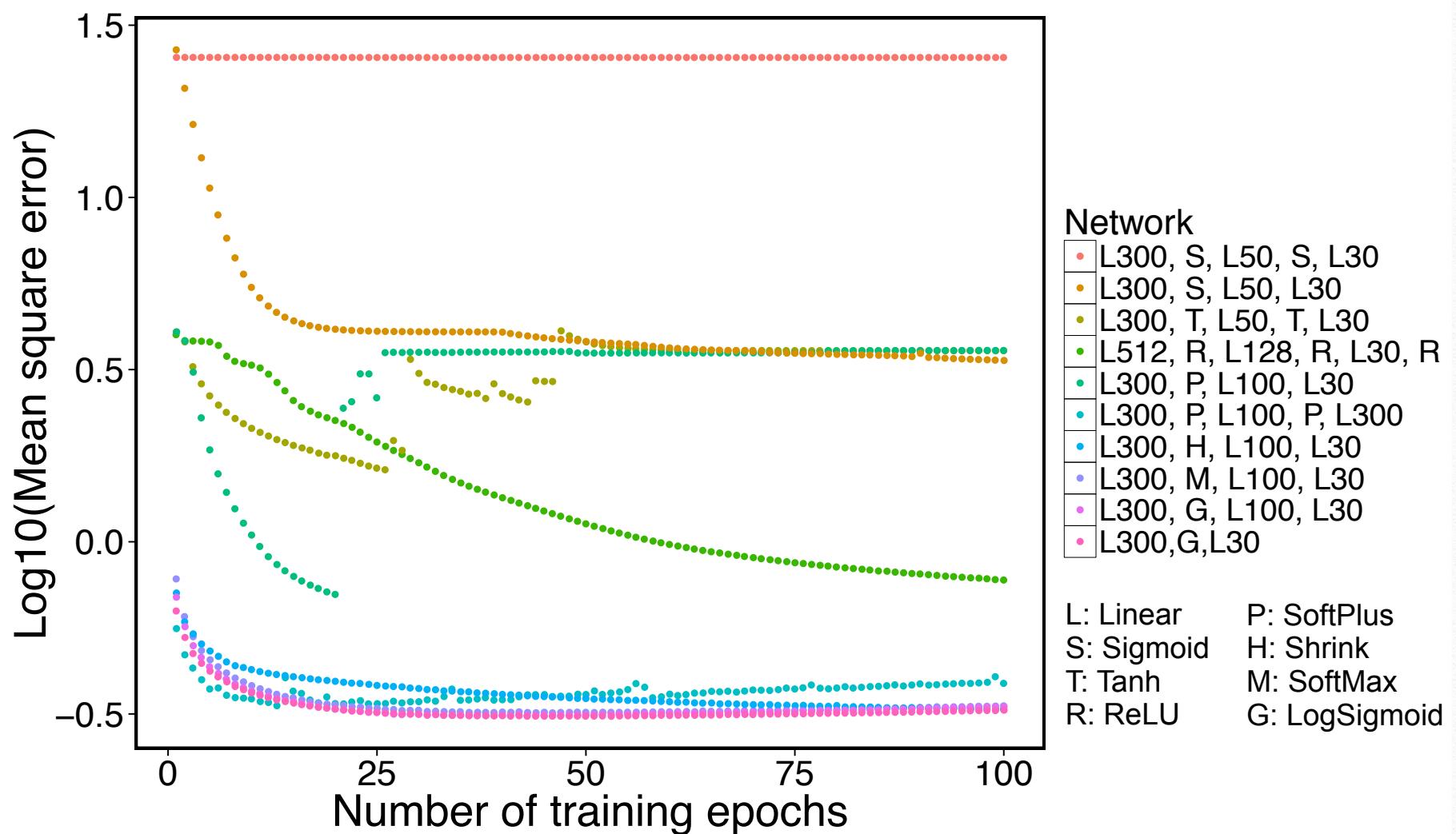
Classifiers



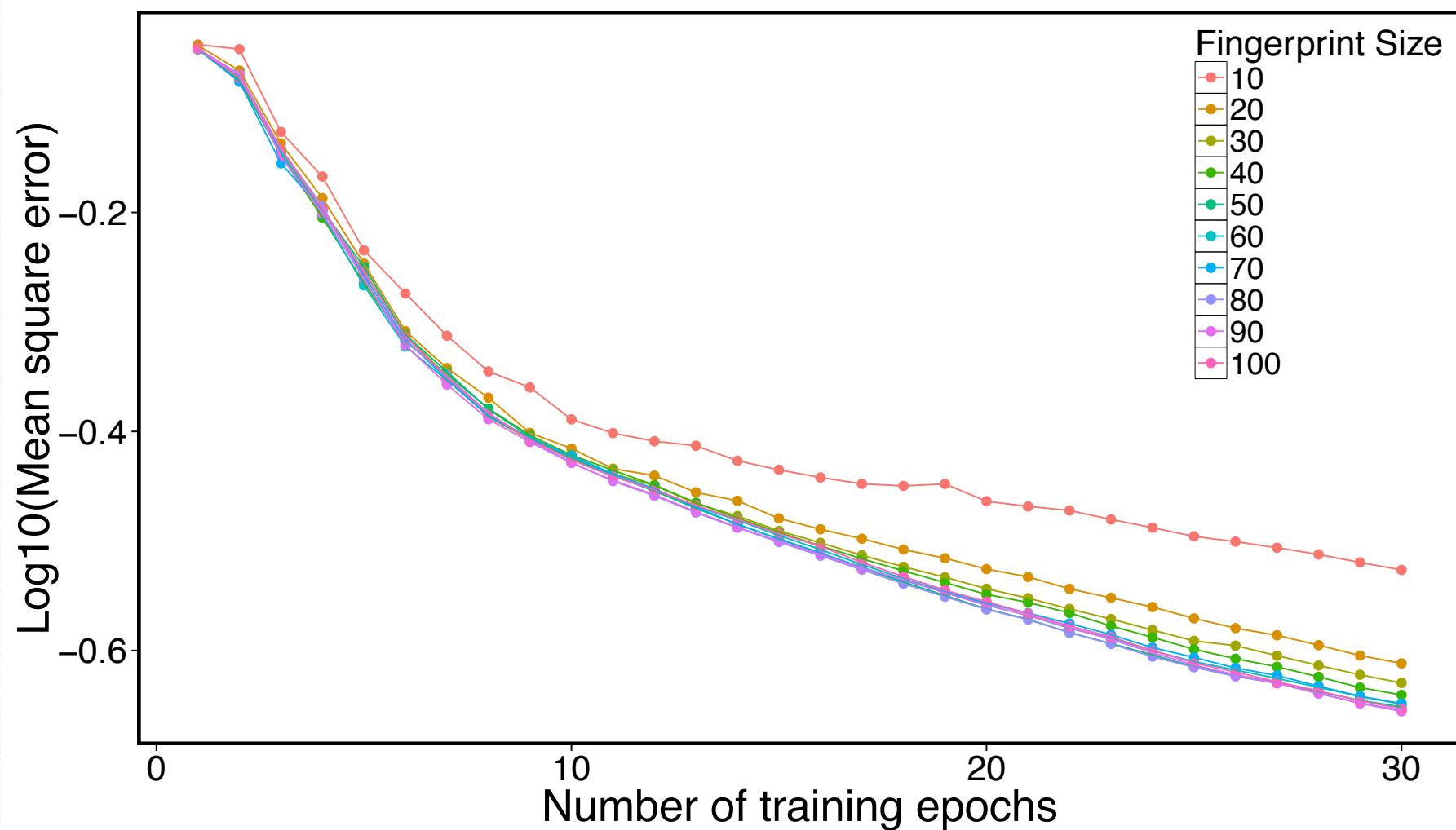
Autoencoders



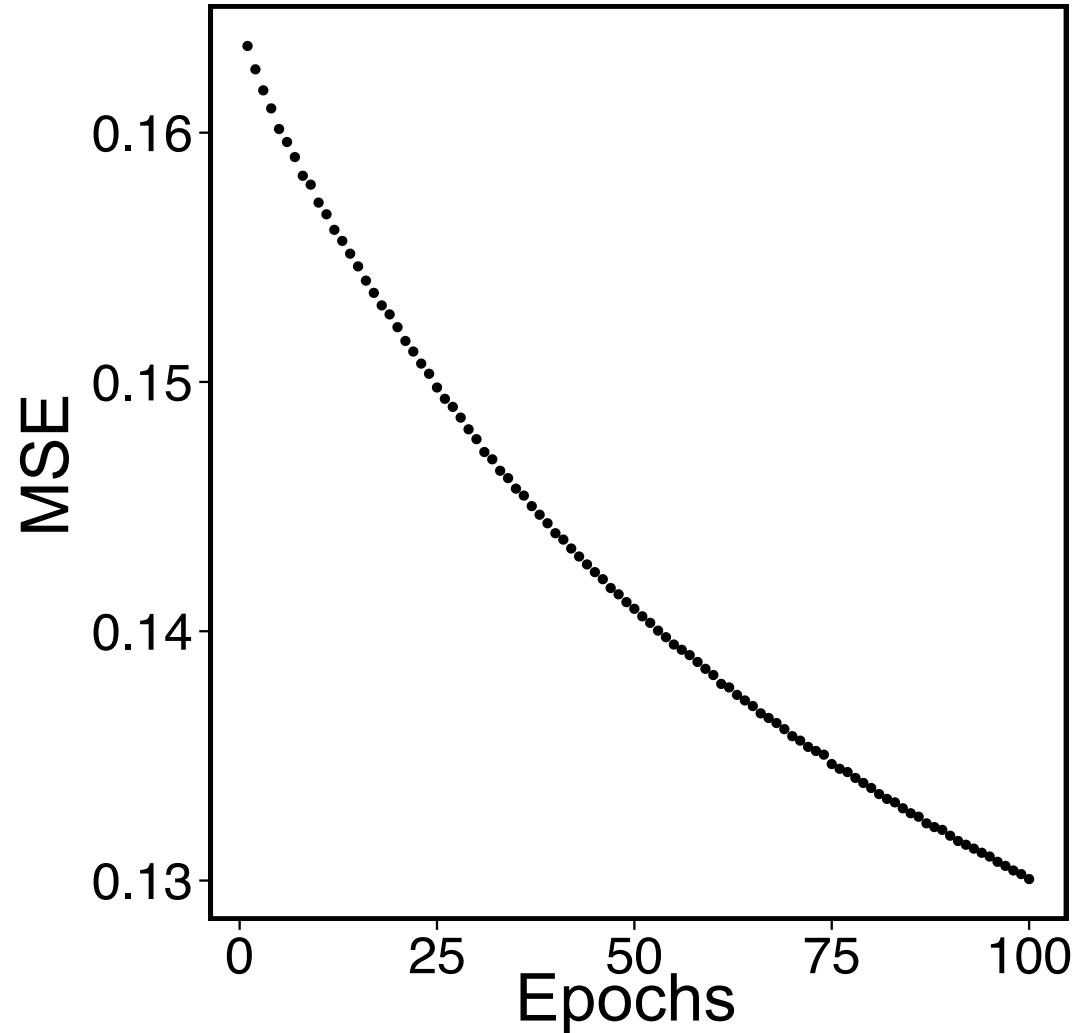
Architecture



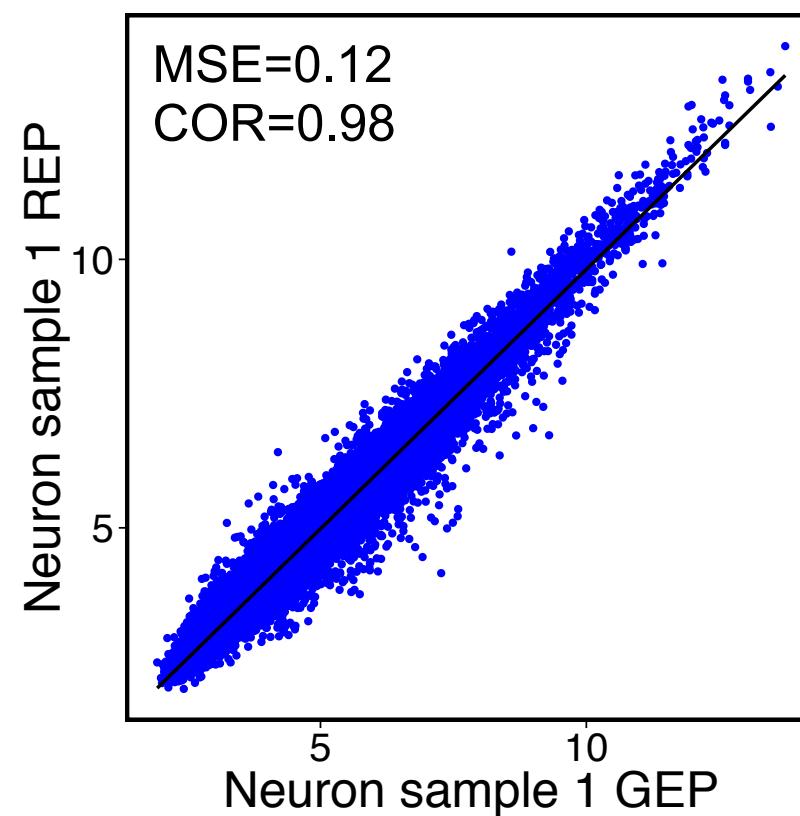
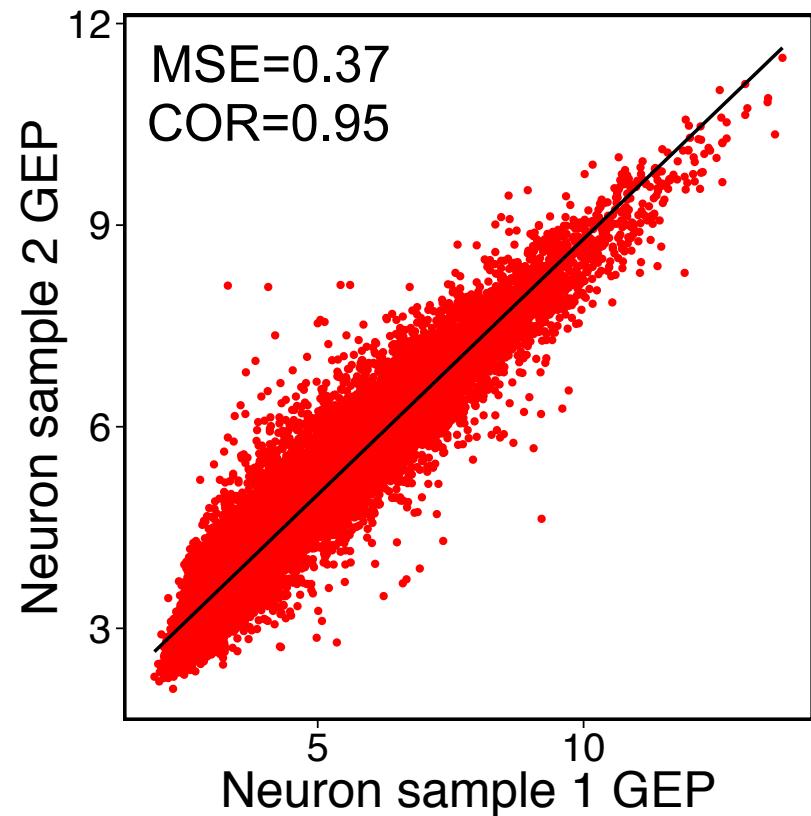
Size of Fingerprints



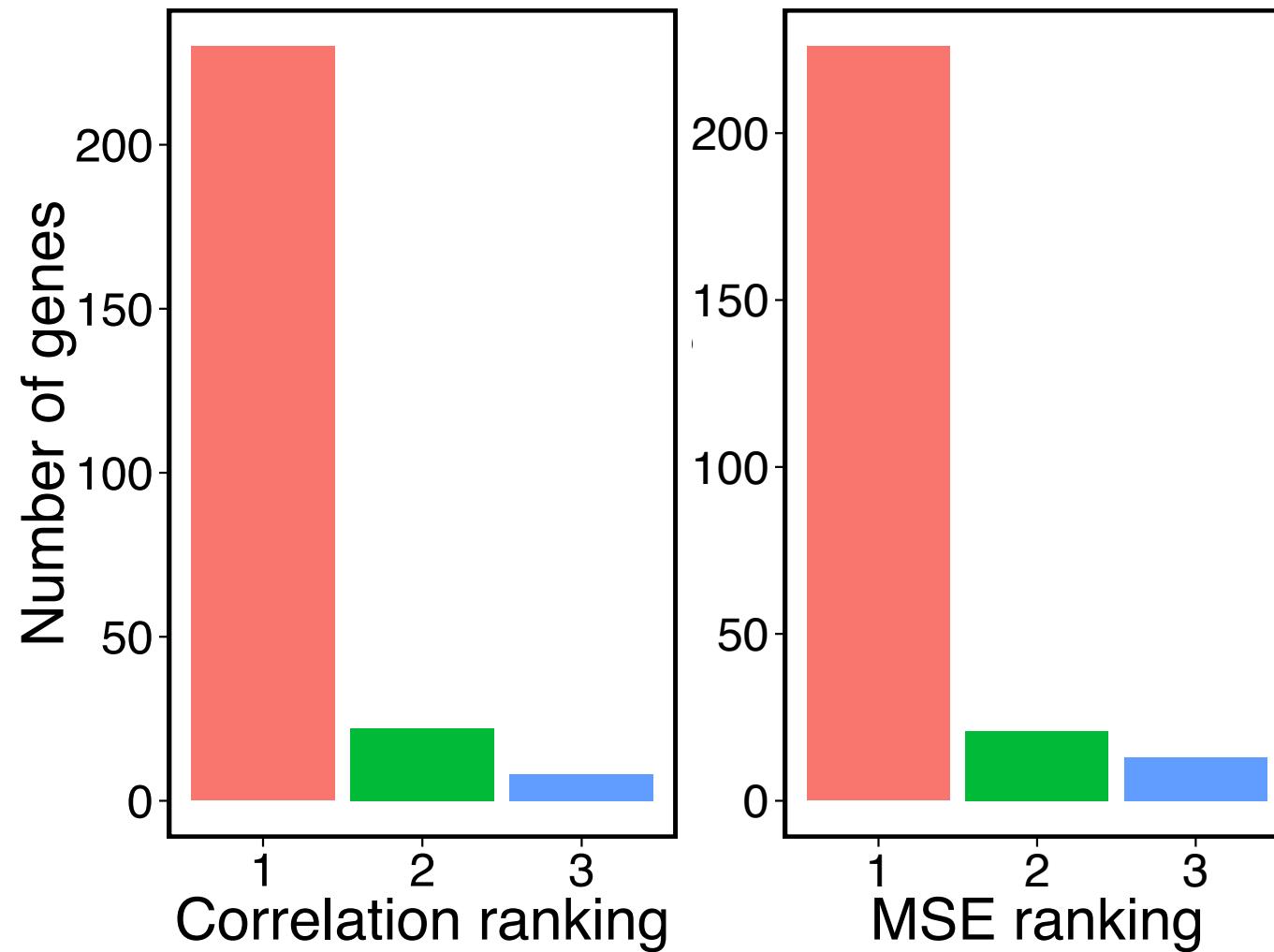
The Actual Data



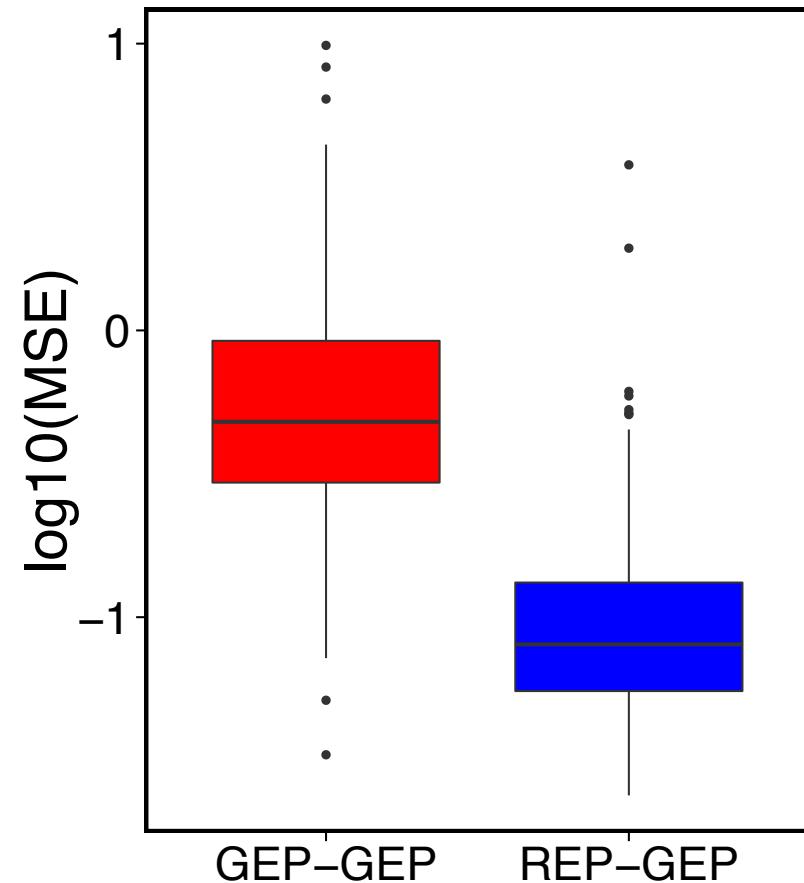
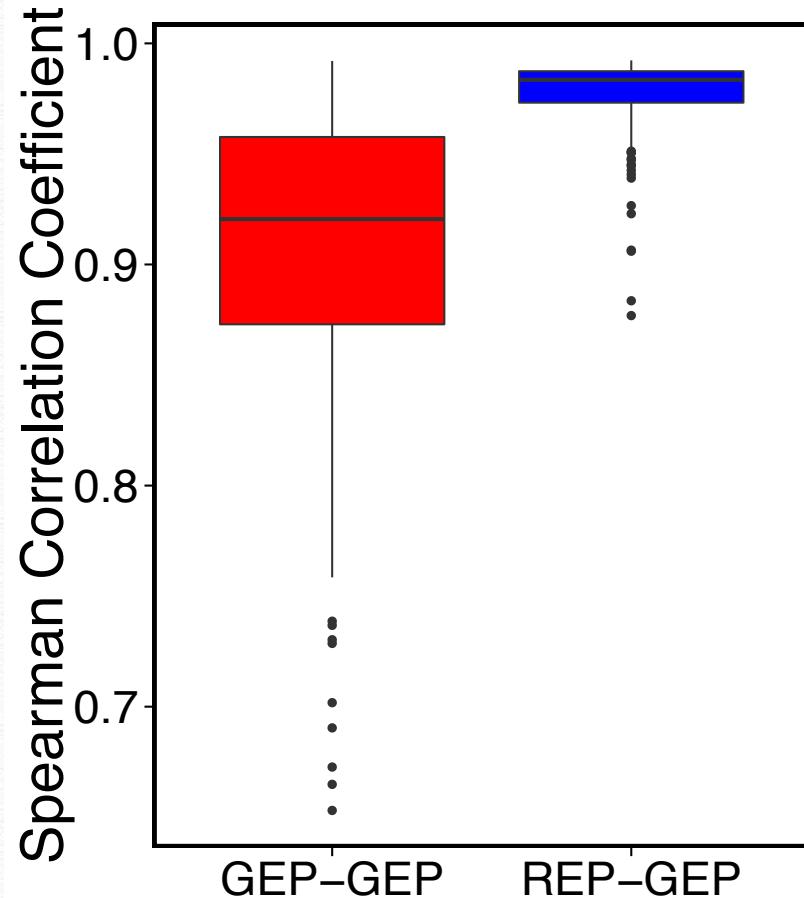
Better than Replicates



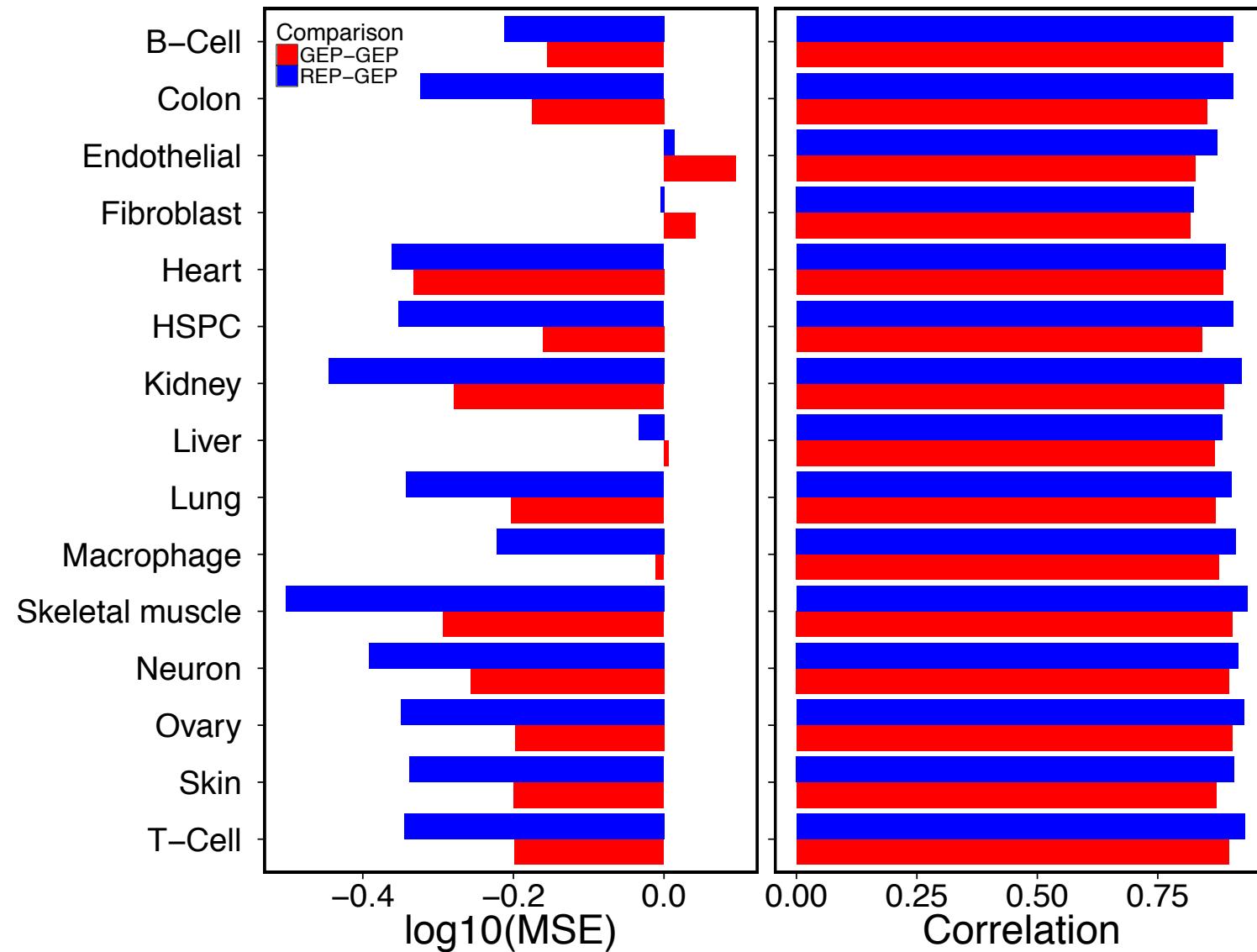
Better than Replicates



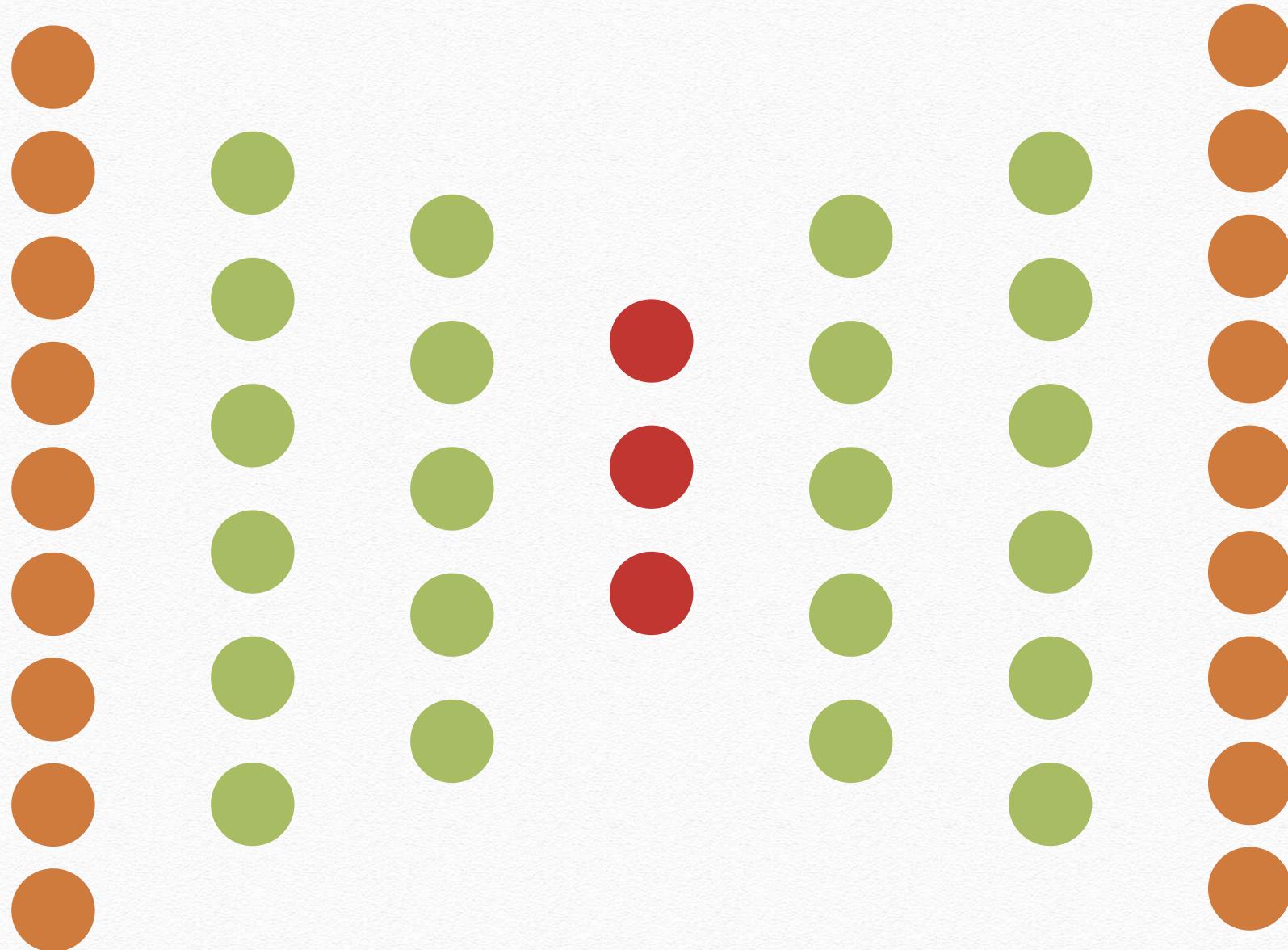
Better than Replicates



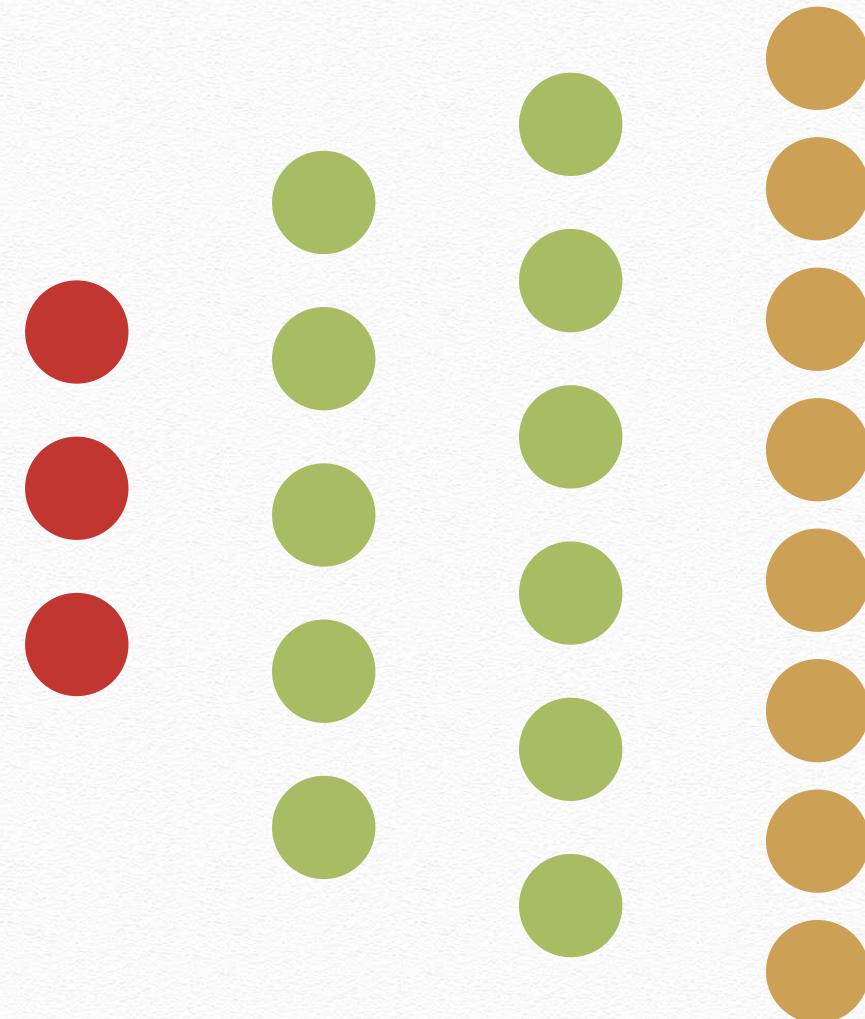
Generalizable!



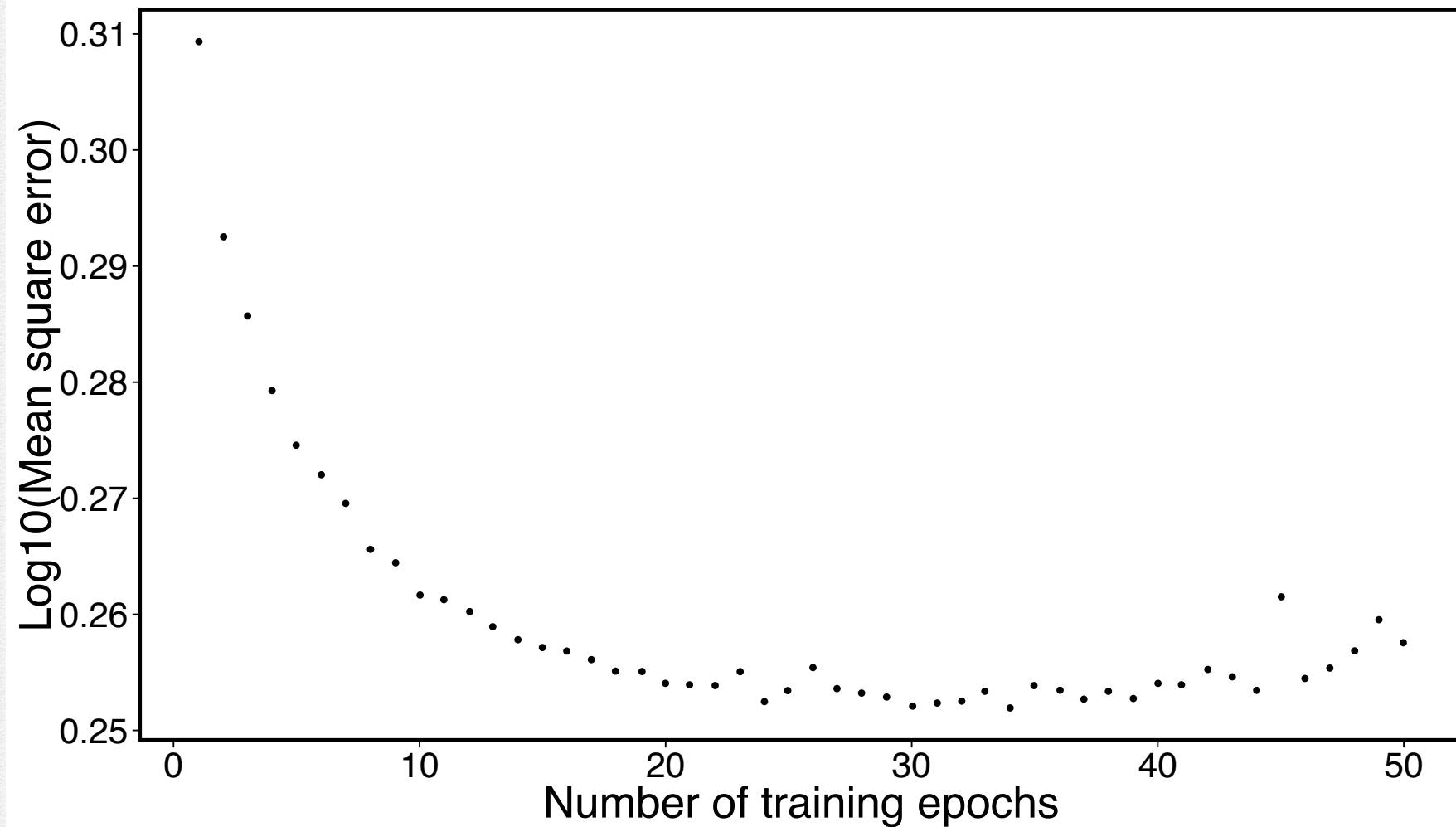
Learning microRNAs



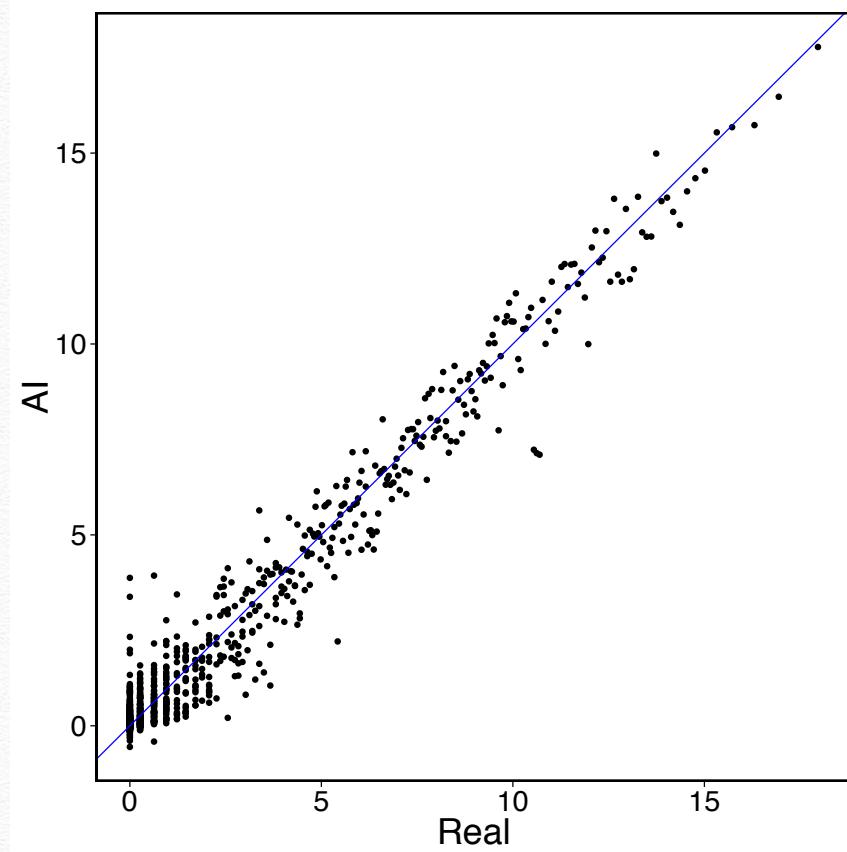
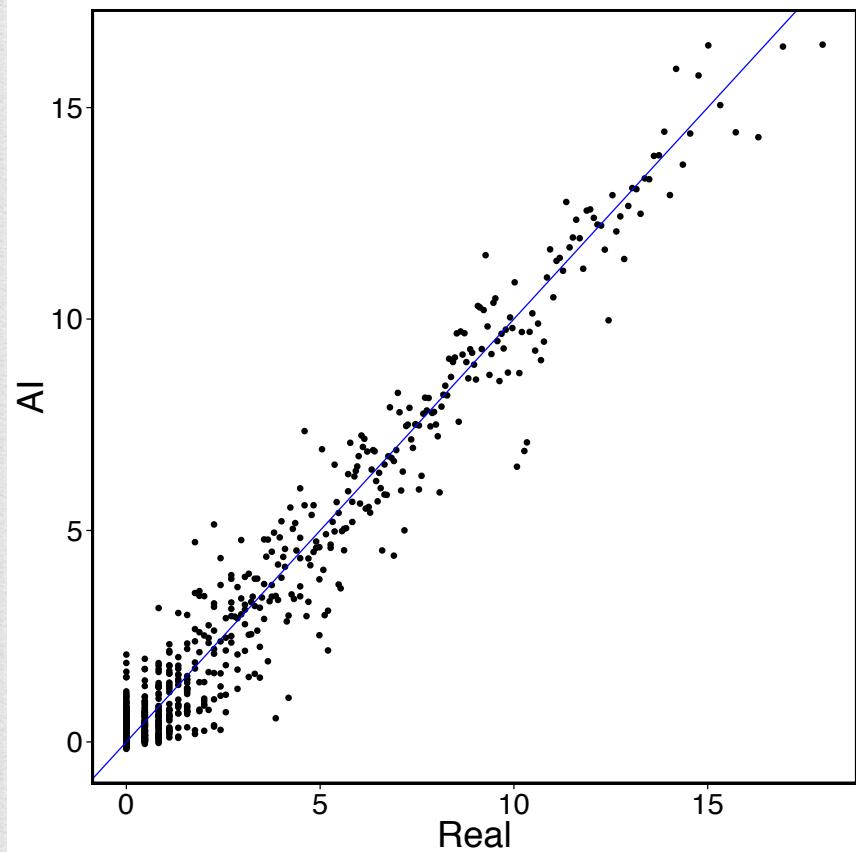
Learning microRNAs



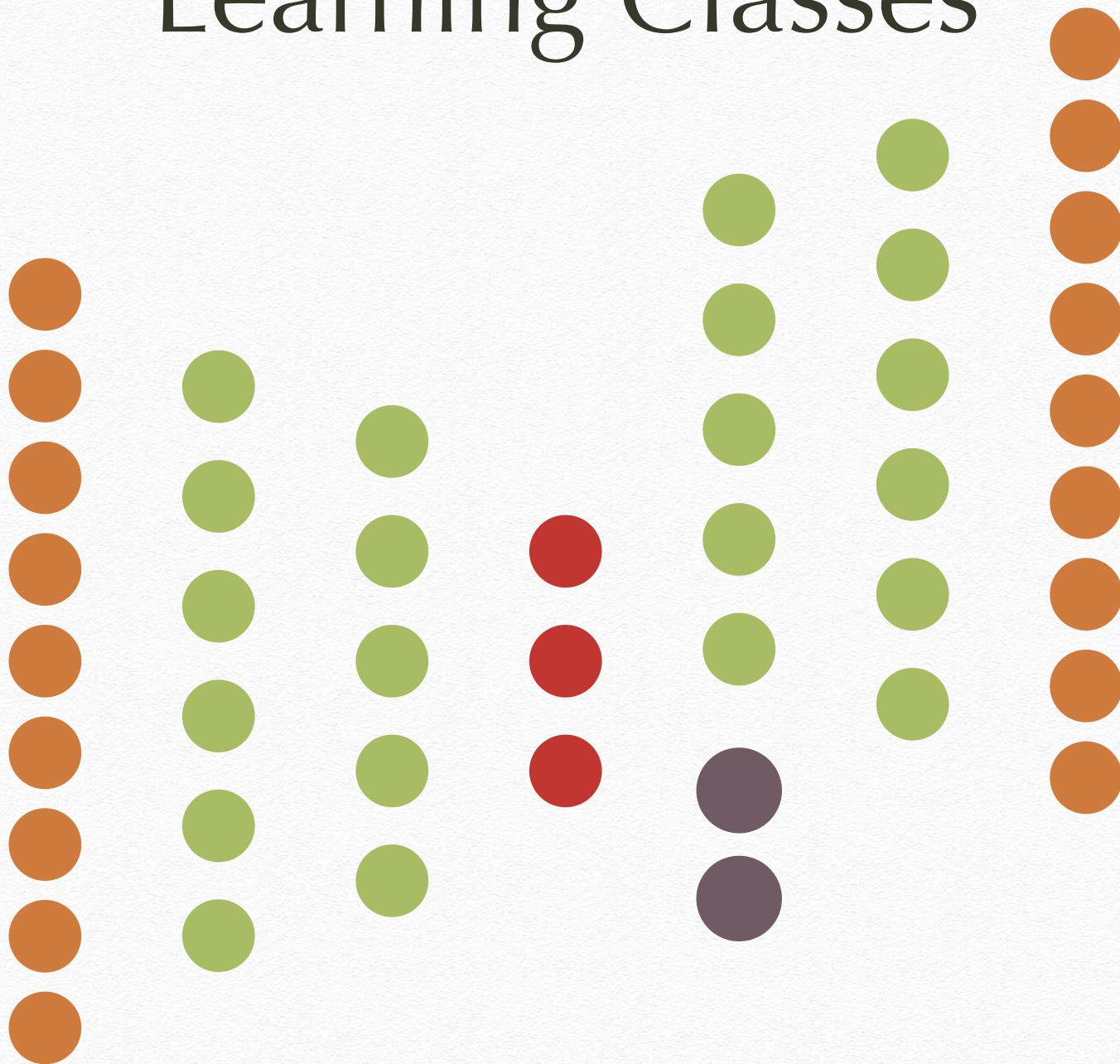
Learning microRNAs



Learning microRNAs



Learning Classes



Learning Classes

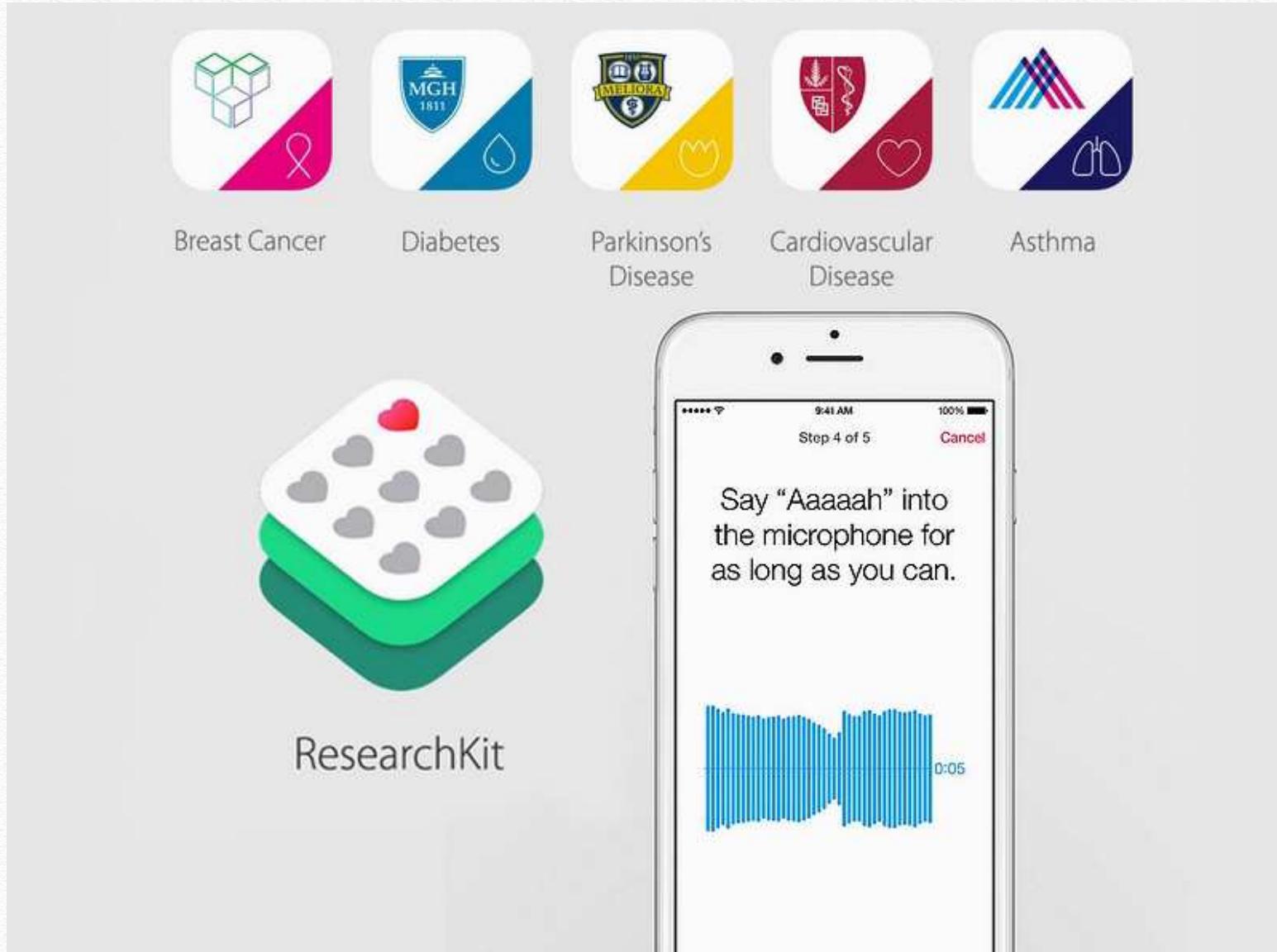
- ❖ There were 10 misclassifications (3.8% error) by both SVM and RF methods
- ❖ 60% error improvement in classifying autoencoder, in comparison with PCA



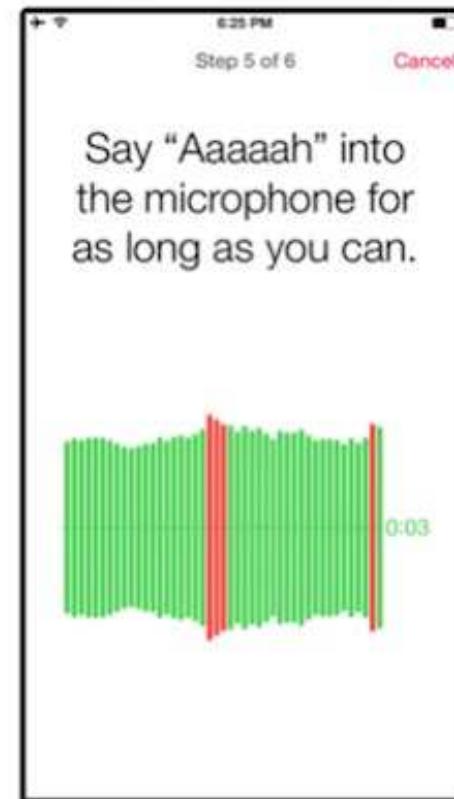
کم/نا شنوایی



ResearchKit



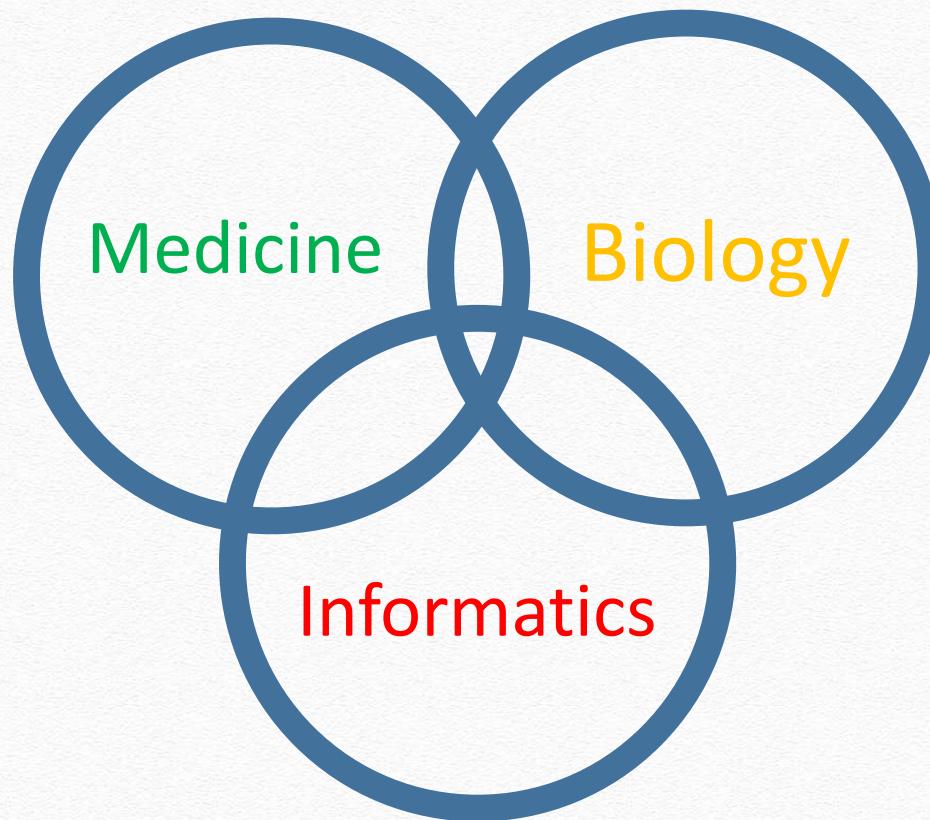
mPower



Can algorithmic
and machine learning methods
cure diseases?

No!

Collaboration



Special Thanks

Dr. Hamidreza Chitsaz (Colorado State University)
Farzad Abdolhosseini (Sharif Univ. of Tech)
Aryan Kamal (Sharif Univ. of Tech)
Abbas Ma'azallahi (University of Tehran)

هُوَ الَّذِي
يُصَوِّرُكُمْ فِي الْأَرْحَامِ
كَيْفَ يَشَاءُ
لَا إِلَهَ إِلَّا هُوَ
الْعَزِيزُ الْحَكِيمُ
آلَ عمران٦

