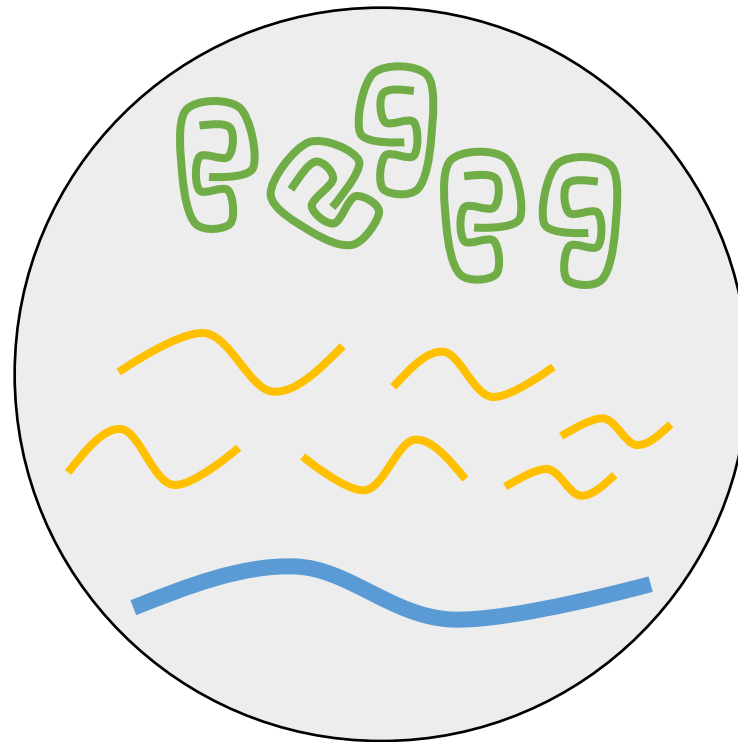


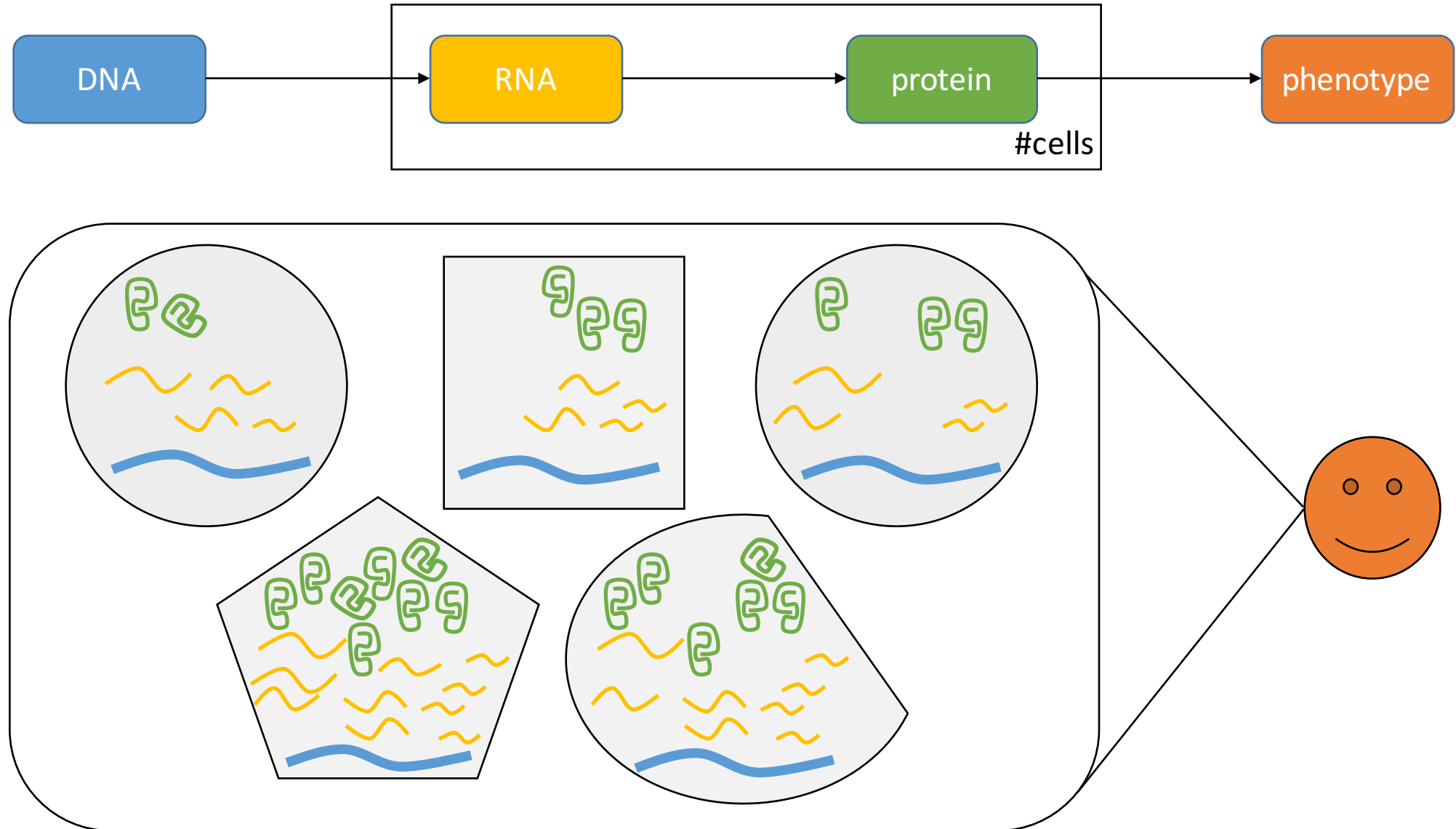
Biological data: a crash course

Yakir Reshef

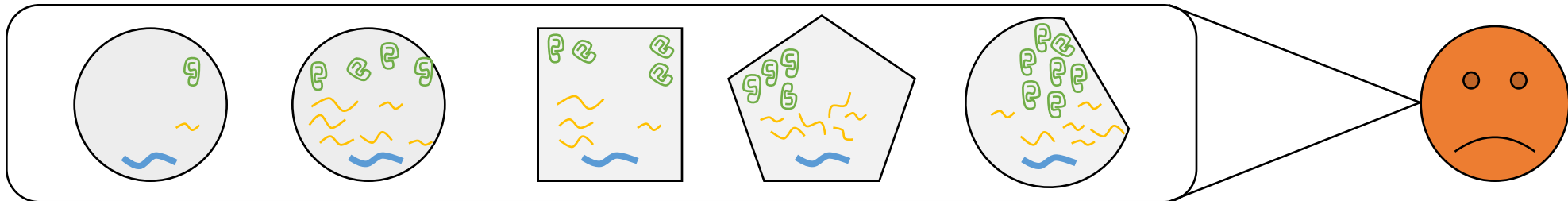
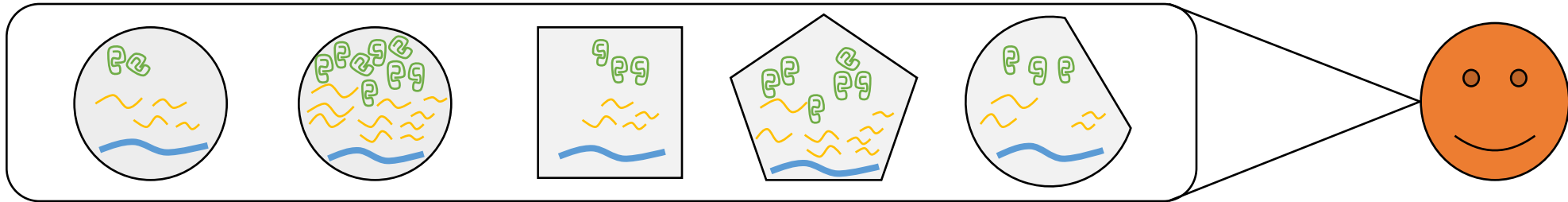
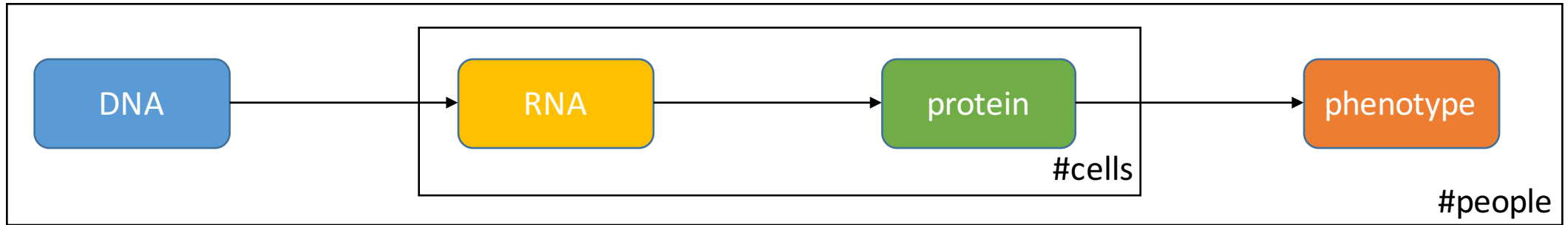
The central dogma



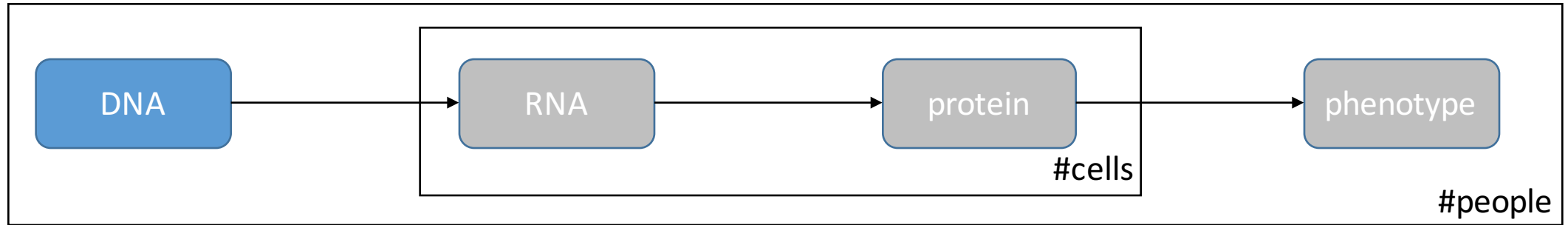
The central dogma



The central dogma



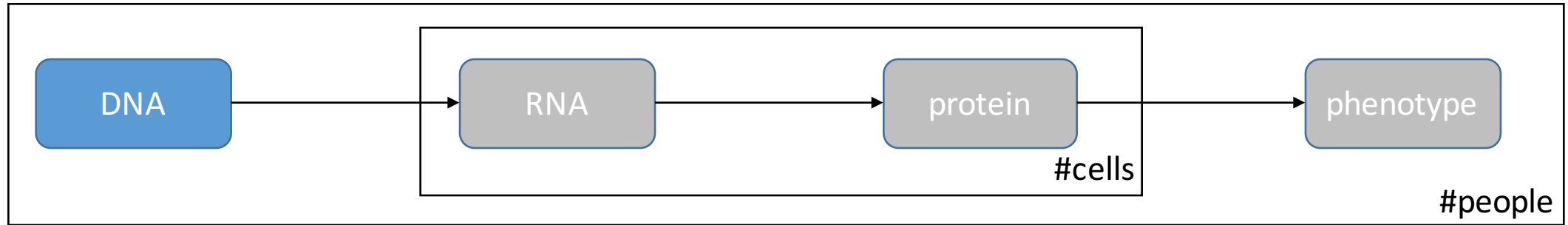
Genotype data



human genome ($\sim 10^9$ base pairs)

people	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A
	A/A	C/C	A/A	G/C	G/G	G/G	T/T	C/C	T/A
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	A/A	C/C	A/A	G/G	G/G	G/G	G/T	C/C	A/A
	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A

Genotype data

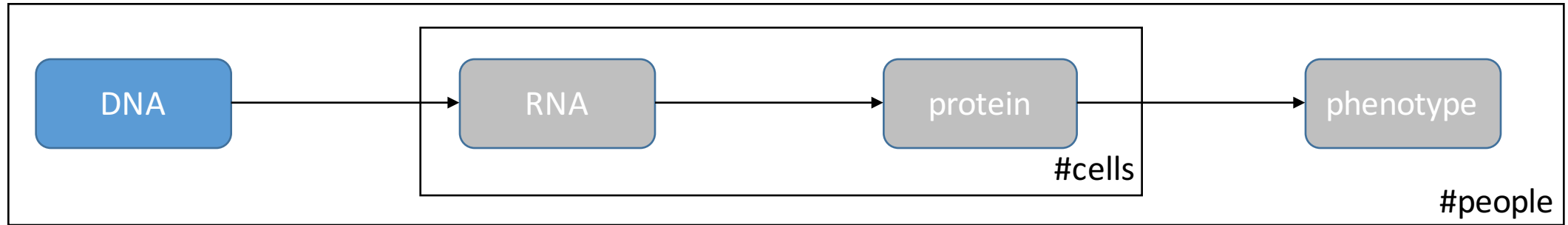


human genome ($\sim 10^9$ base pairs)

Two alleles at each spot

people	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A
	A/A	C/C	A/A	G/C	G/G	G/G	T/T	C/C	T/A
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	A/A	C/C	A/A	G/G	G/G	G/G	G/T	C/C	A/A
	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A

Genotype data



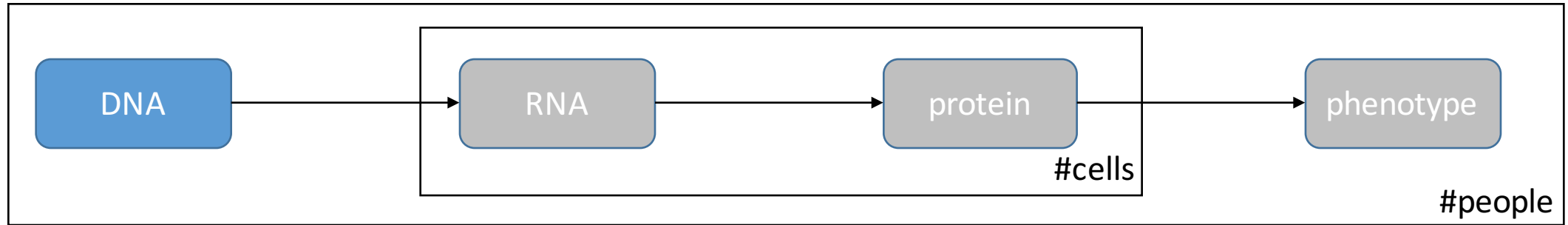
human genome ($\sim 10^9$ base pairs)

Two alleles at each spot

people	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A
	A/A	C/C	A/A	G/C	G/G	G/G	T/T	C/C	T/A
	:	:	:	:	:	:	:	:	:
	A/A	C/C	A/A	G/G	G/G	G/G	G/T	C/C	A/A
	A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A

Most spots:
no variation observed

Genotype data



human genome ($\sim 10^9$ base pairs)

Two alleles at each spot

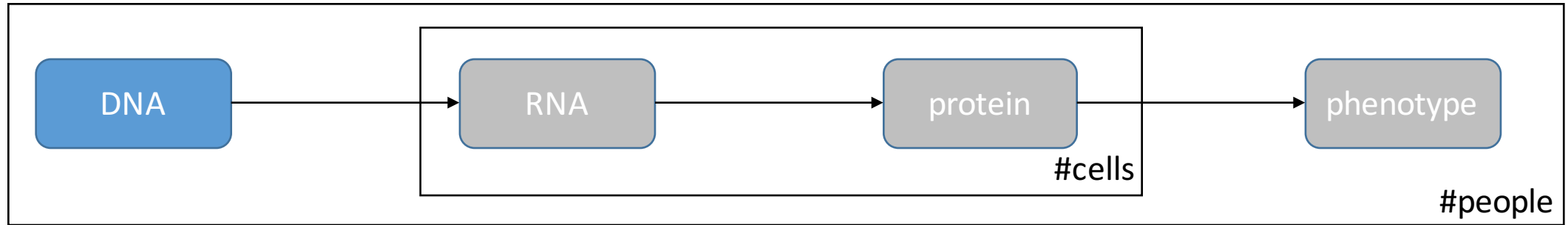
people

A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A
A/A	C/C	A/A	G/C	G/G	G/G	T/T	C/C	T/A
:	:	:	:	:	:	:	:	:
A/A	C/C	A/A	G/G	G/G	G/G	G/T	C/C	A/A
A/A	C/C	A/A	C/C	G/G	G/G	T/T	C/C	A/A

Most spots:
no variation observed

Genetic variants
can be coded as 0/1/2

Genotype data



human genome ($\sim 10^9$ base pairs)

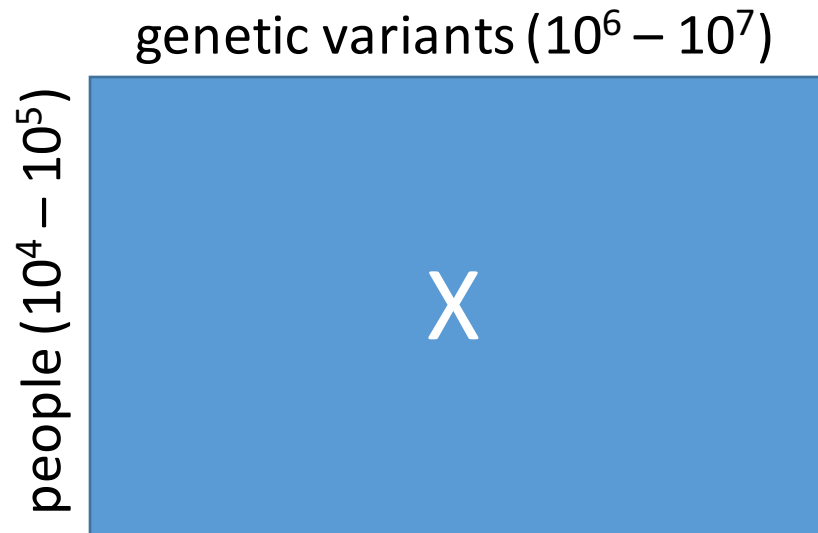
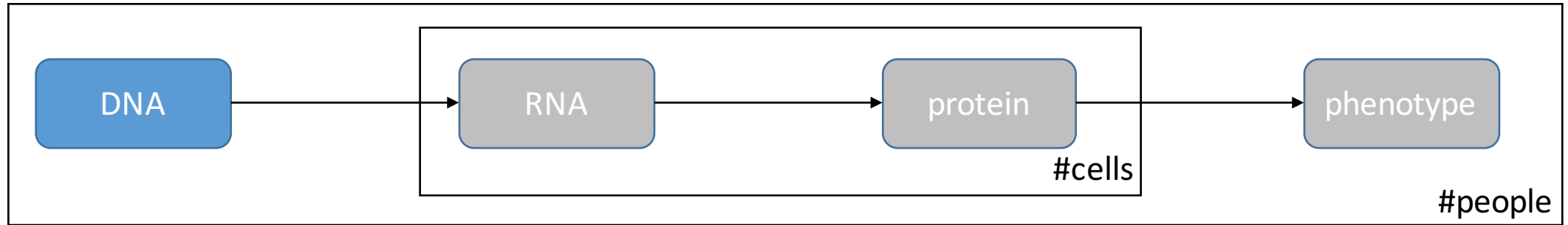
Two alleles at each spot

people	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	1
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	0	0	0	2	0	0	1	0	0
	0	0	0	0	0	0	0	0	0

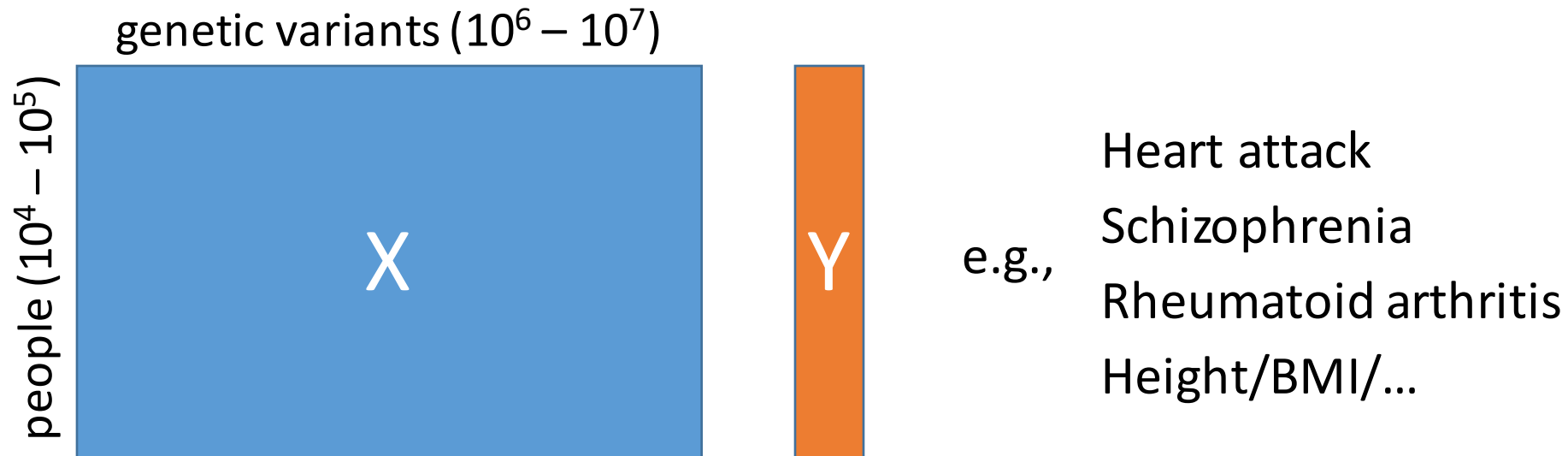
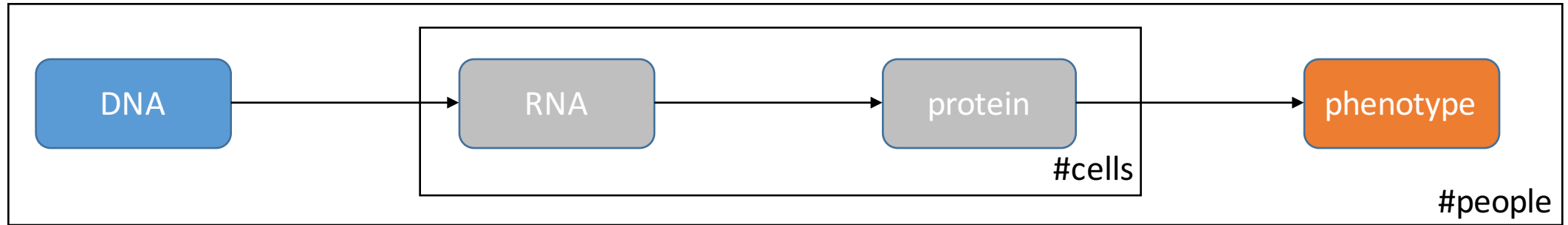
Most spots:
no variation observed

Genetic variants
can be coded as 0/1/2

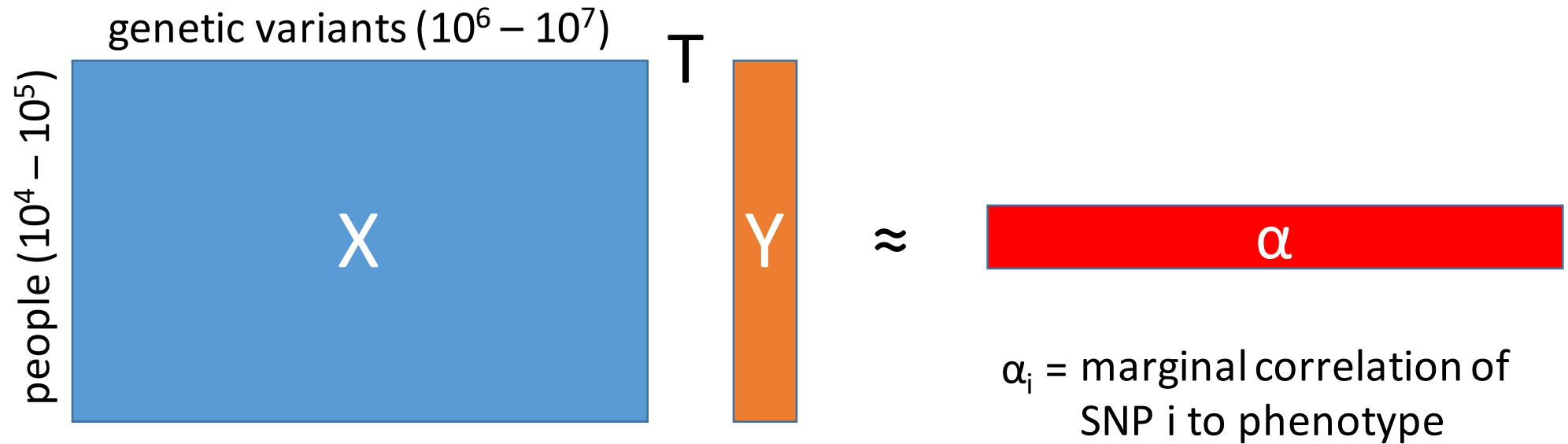
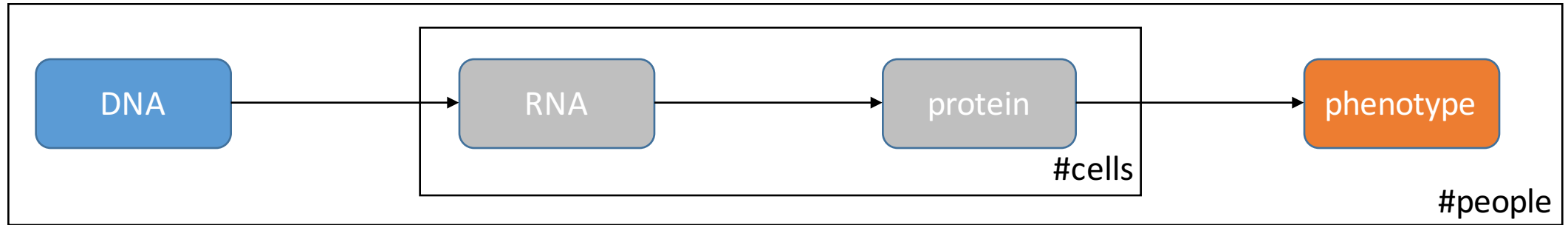
Genotype data



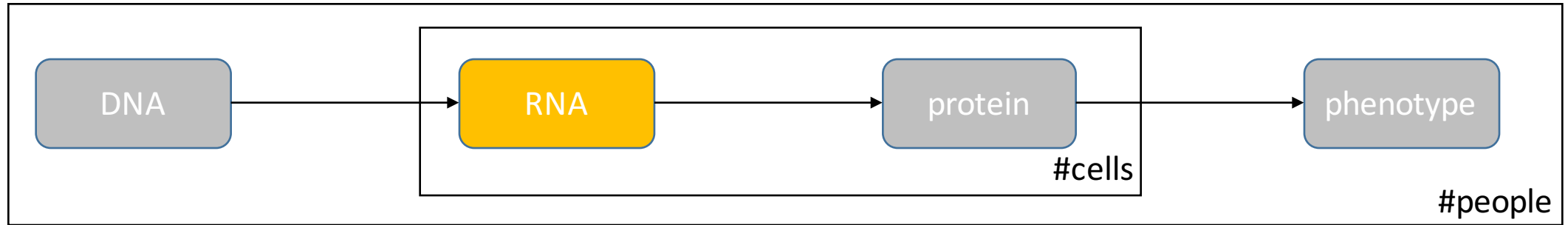
Genome-wide association study (GWAS)



Genome-wide association study (GWAS)

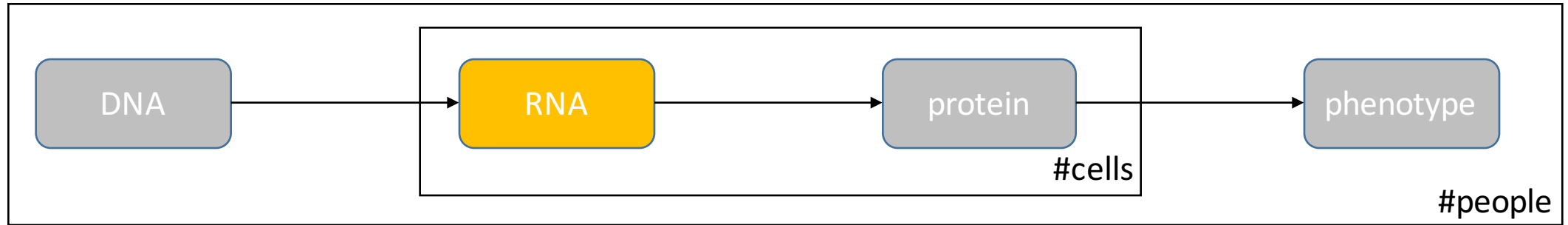


Gene expression



human genome

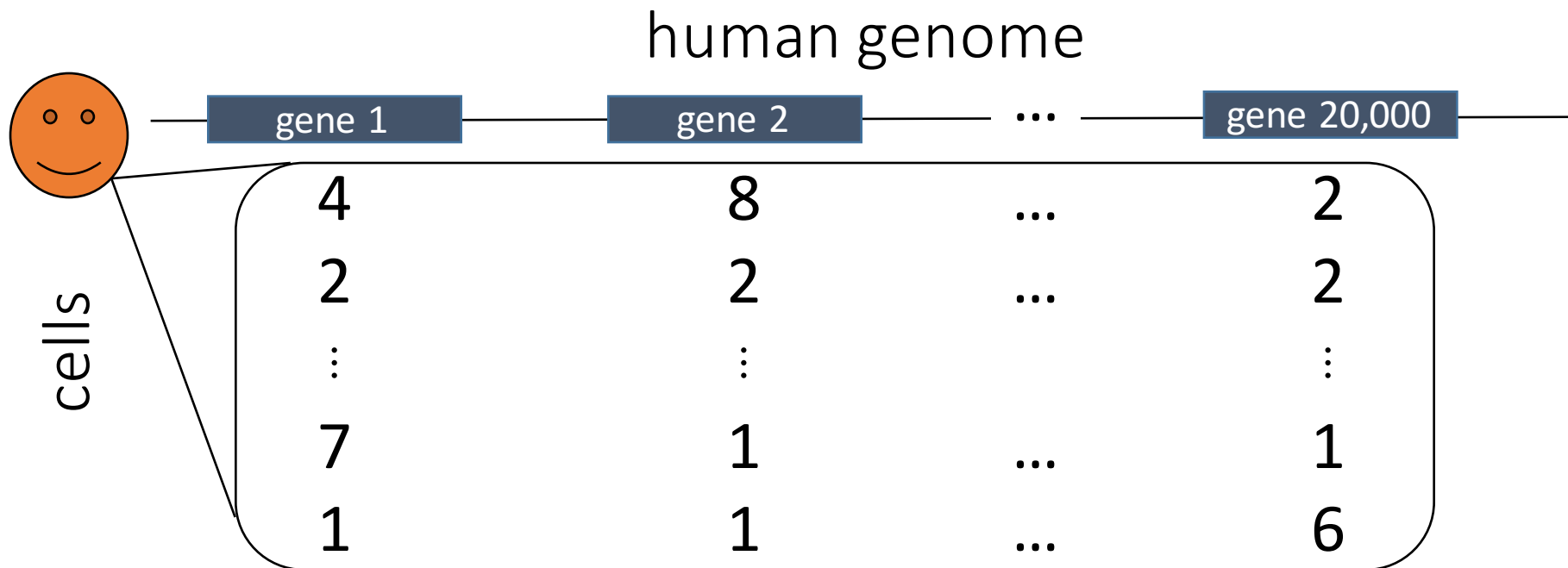
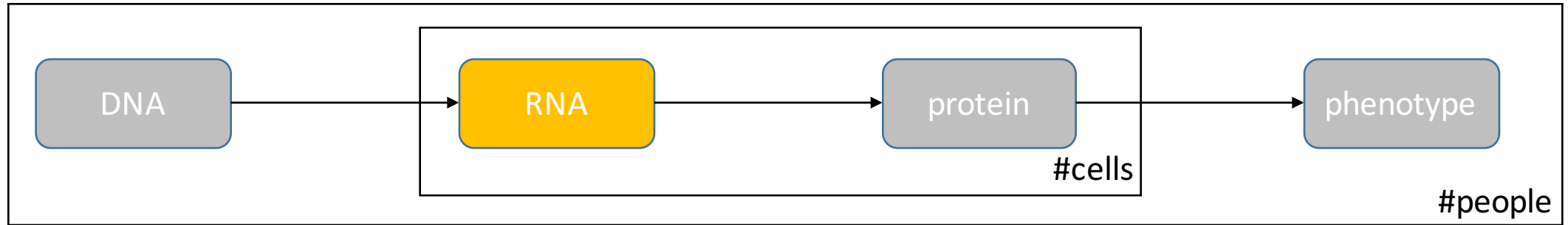
Gene expression



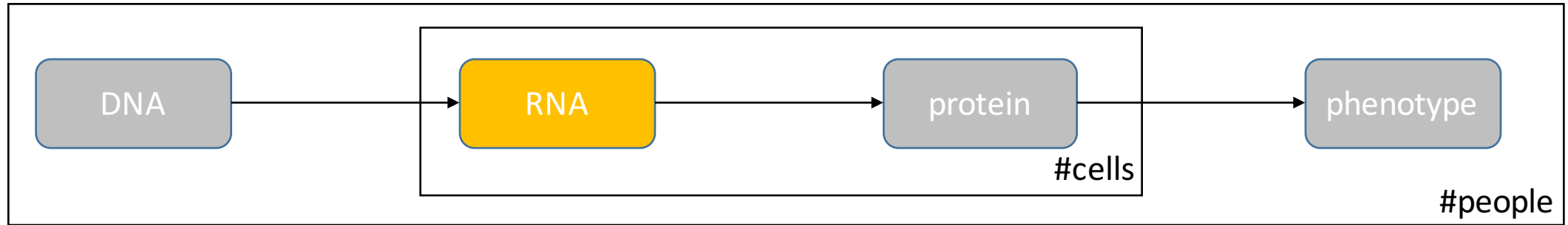
human genome



Single-cell RNA sequencing



Single-cell RNA sequencing



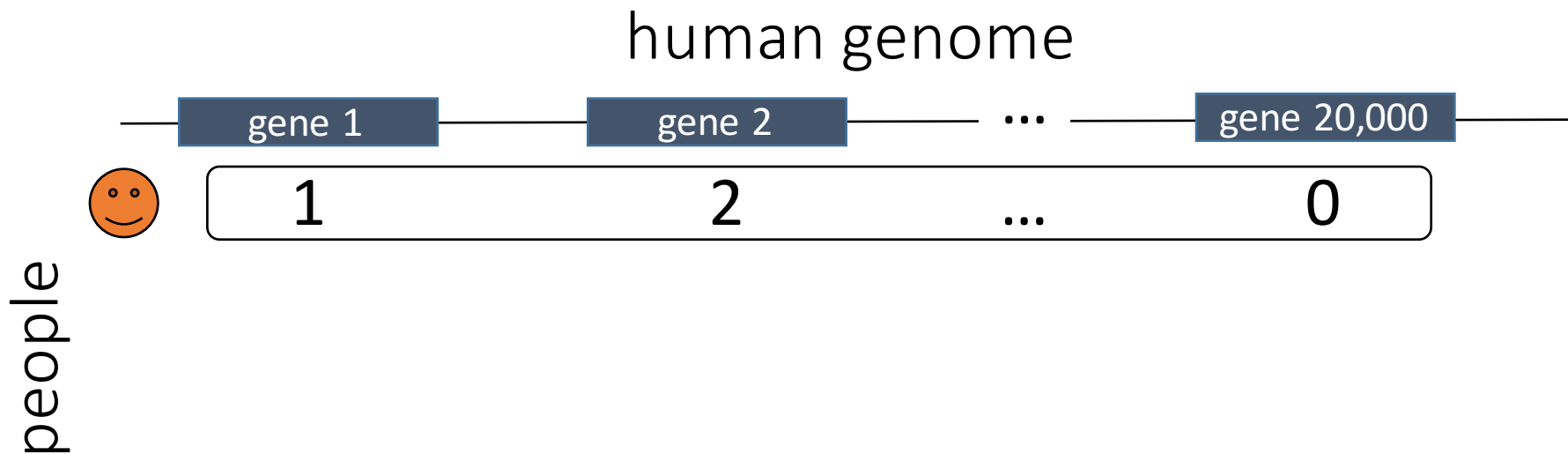
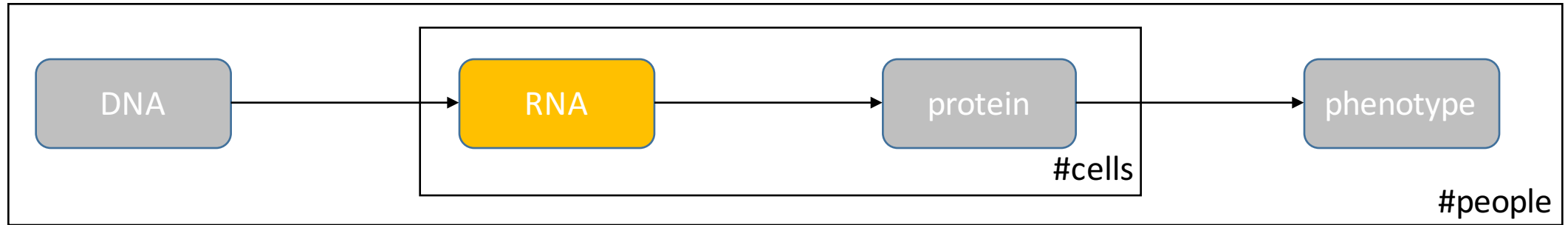
human genome

	gene 1	gene 2	...	gene 20,000
cells	4	8	...	2
	2	2	...	2
	⋮	⋮		⋮
	7	1	...	1
	1	1	...	6

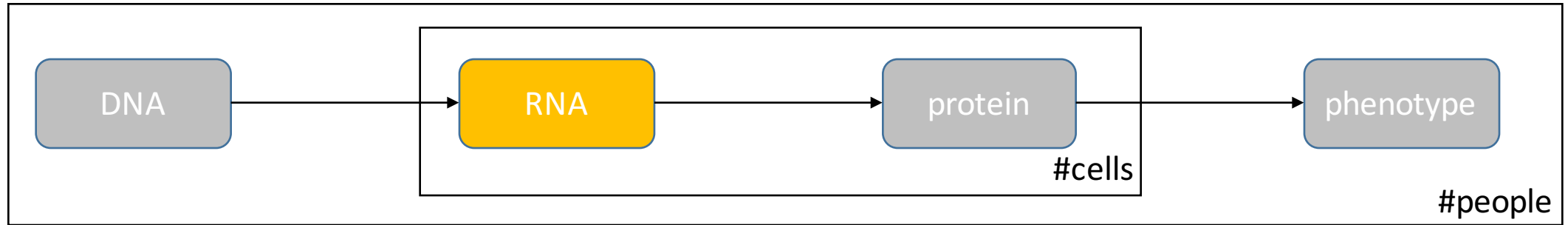


expensive
few individuals
but: single-cell resolution

Bulk RNA sequencing



Bulk RNA sequencing

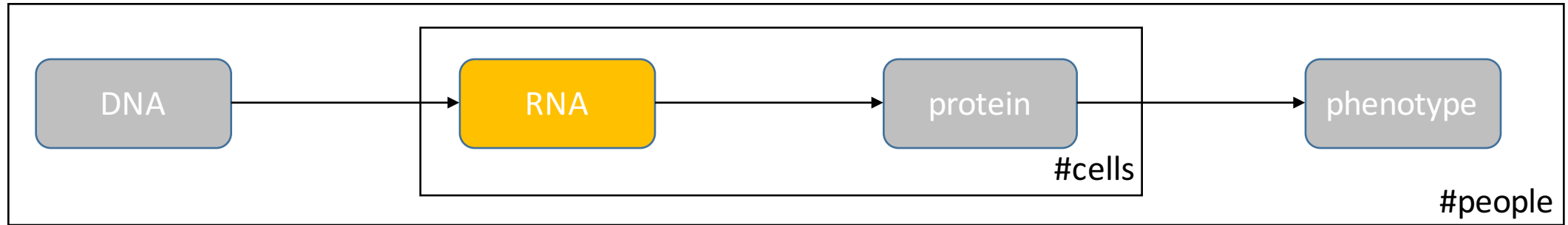


human genome

	gene 1	gene 2	...	gene 20,000
people	1	2	...	0
😊	3	6	...	2
😞	⋮	⋮		⋮
😊	6	7	...	6
😞	3	2	...	9



Bulk RNA sequencing



human genome

	gene 1	gene 2	...	gene 20,000
people	1	2	...	0
😊	3	6	...	2
😞	⋮	⋮		⋮
😊	6	7	...	6
😞	3	2	...	9



cheap
many individuals
but: only averages

Summary

