

# Multi-omics

Thomas Stoeger

# Two learning goals:

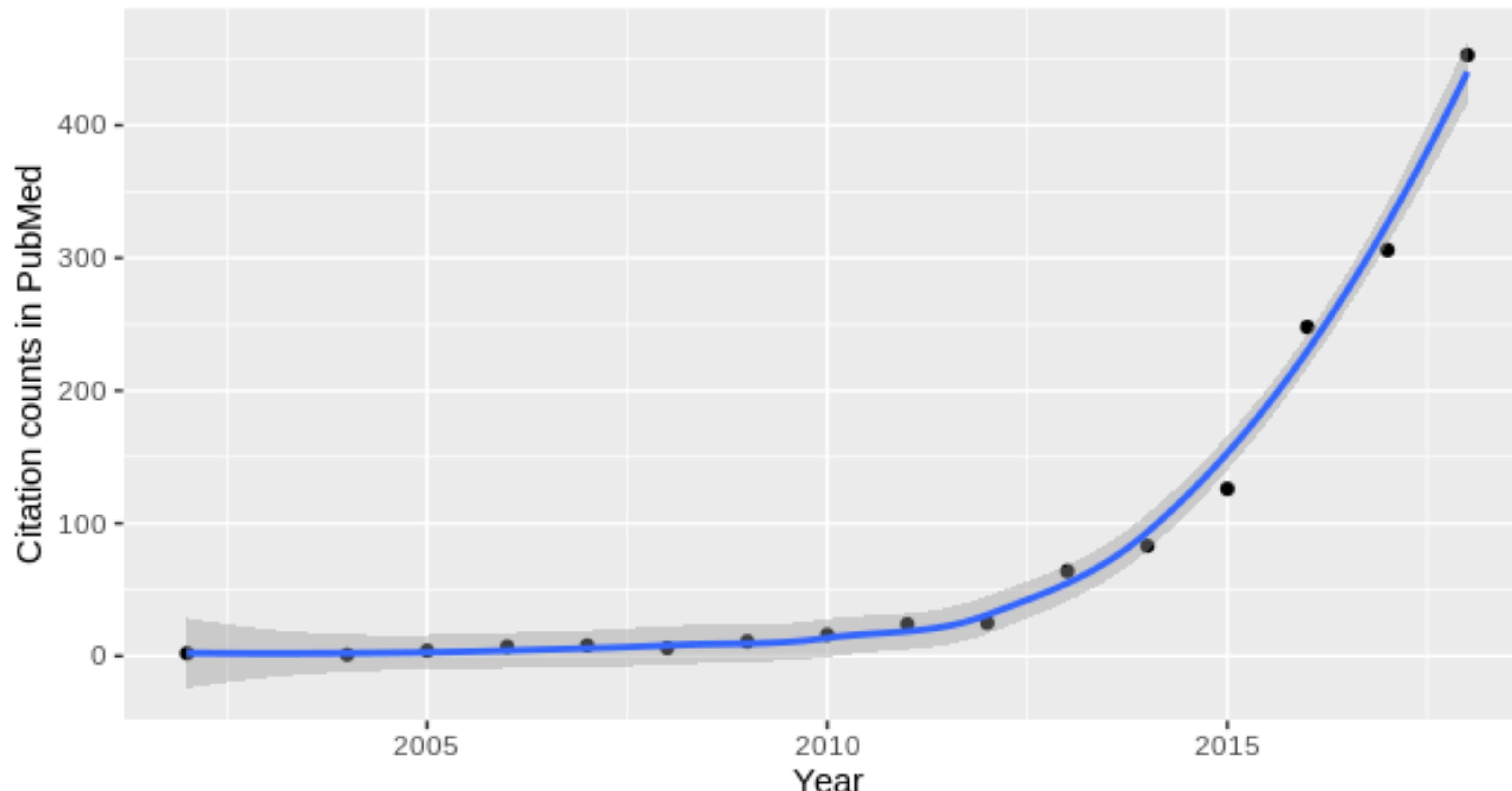
# Two learning goals:

Combining different data modalities (-‘omes’) can answer some interesting scientific questions.

# Two learning goals:

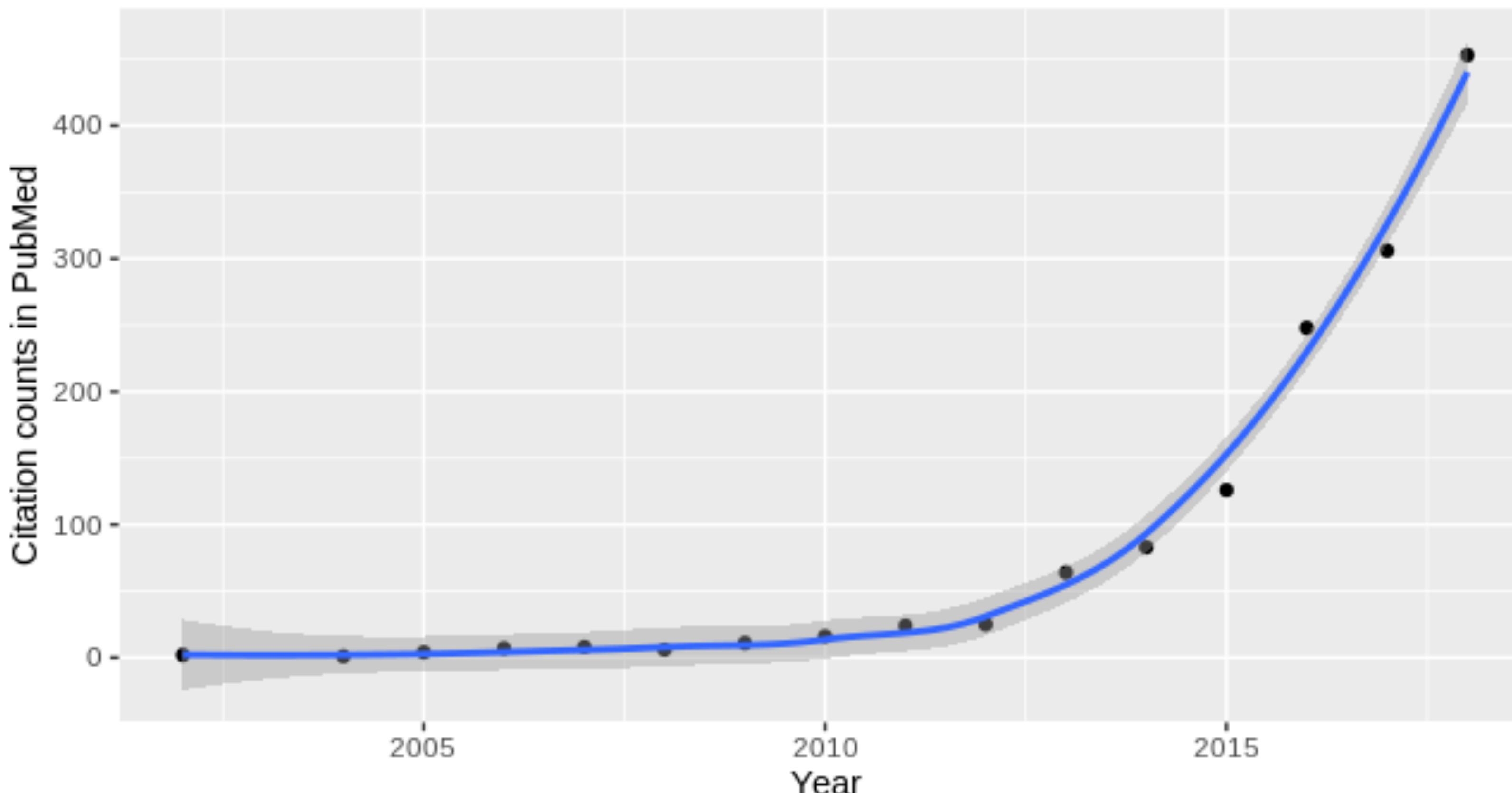
Combining different data modalities (-‘omes’) can answer some interesting scientific questions.

Approach multi-modal data carefully while looking for interesting scientific questions!

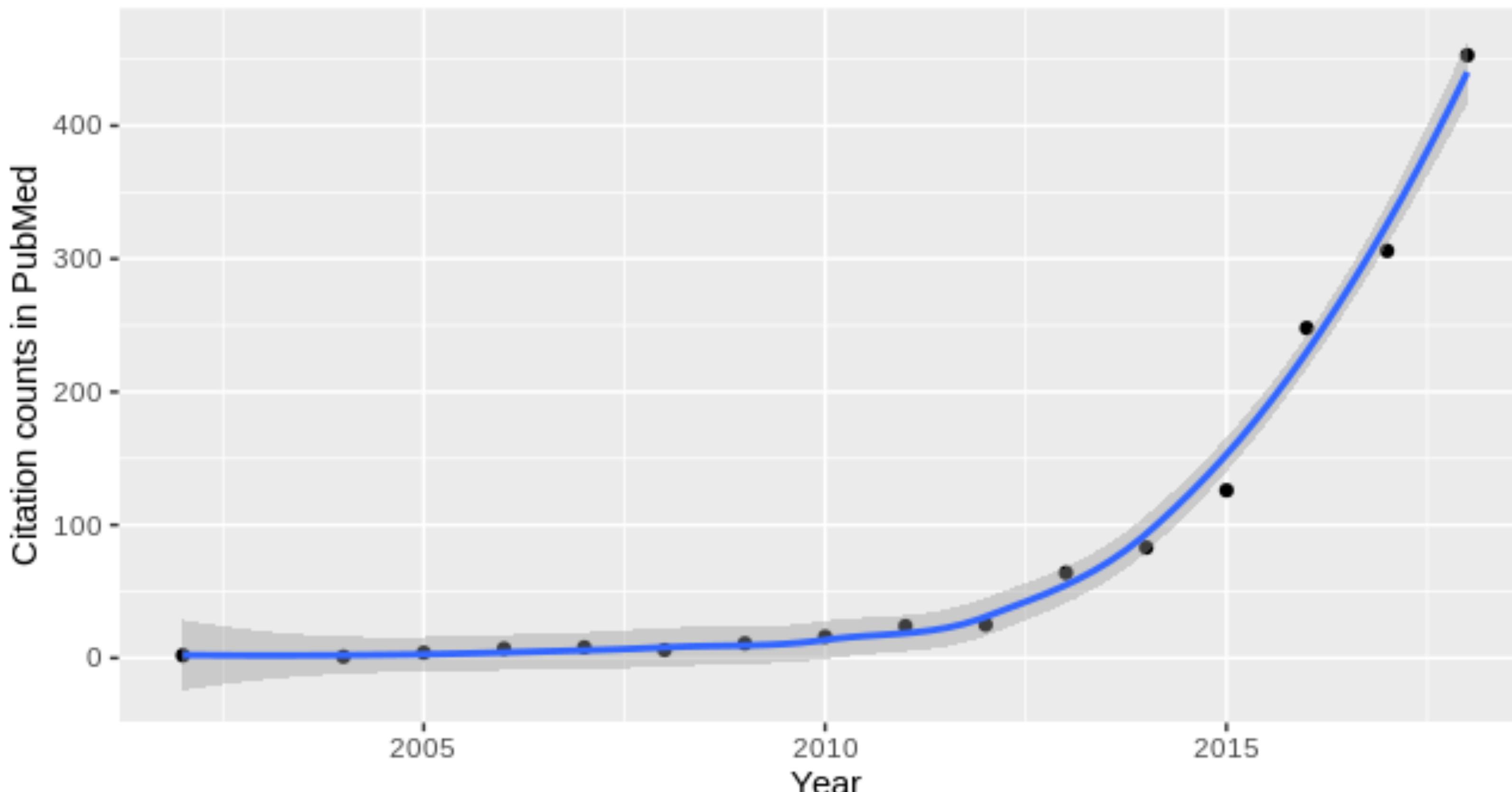


**Wikipedia**

# Multi-omics are growing in popularity.



# Multi-omics are growing in popularity.



# Multi-omics are still a small field.

# Research questions

# Research questions

-> Combine qualitatively  
different data!

# Research questions

# Technology

-> Combine qualitatively  
different data!

# Research questions

-> Combine qualitatively different data!

# Technology

-> How to (start) working with multi-omics!

# Research questions

scales

-> Combine qualitatively  
different data!

# Technology

-> How to (start) working  
with multi-omics!

# Research questions

scales

distorted literature

-> Combine qualitatively  
different data!

# Technology

-> How to (start) working  
with multi-omics!

# Research questions

scales

distorted literature

orientation

-> Combine qualitatively  
different data!

# Technology

-> How to (start) working  
with multi-omics!

# Research questions

scales  
distorted literature  
orientation

-> Combine qualitatively  
different data!

# Technology

organizing

-> How to (start) working  
with multi-omics!

# Research questions

scales  
distorted literature  
orientation

-> Combine qualitatively  
different data!

# Technology

organizing  
analytical approaches

-> How to (start) working  
with multi-omics!

# Research questions

scales  
distorted literature  
orientation

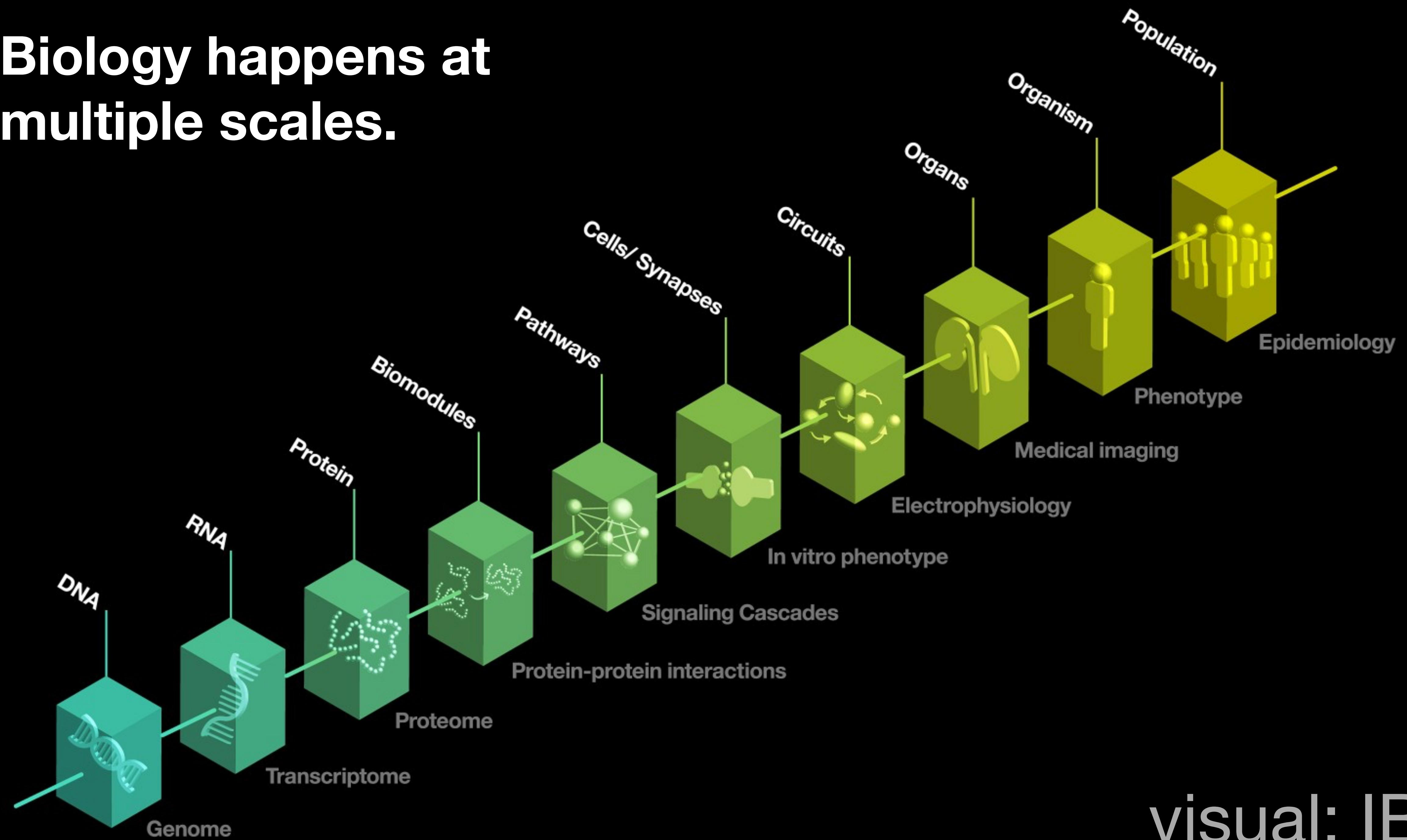
-> Combine qualitatively different data!

# Technology

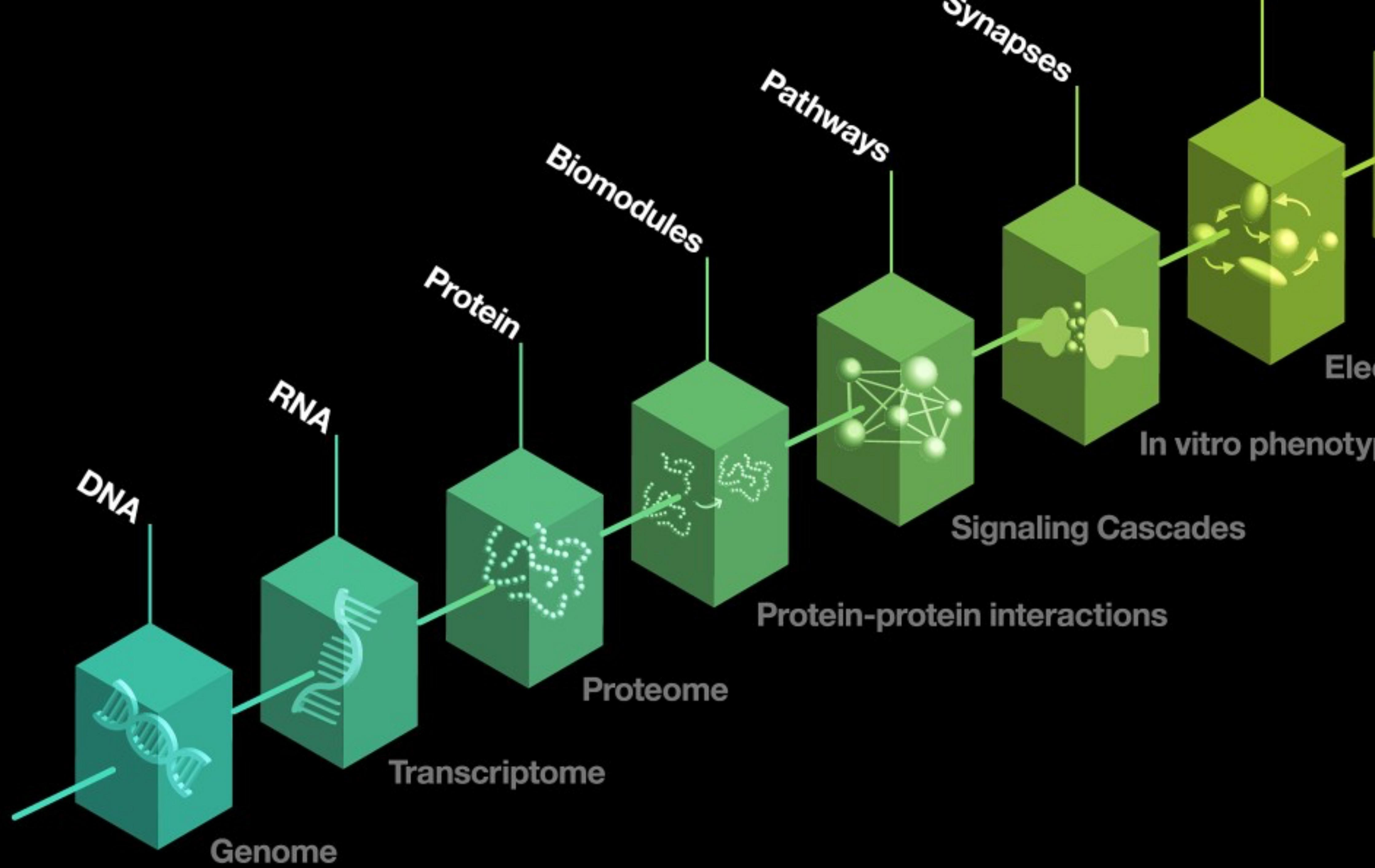
organizing  
analytical approaches  
understanding limits

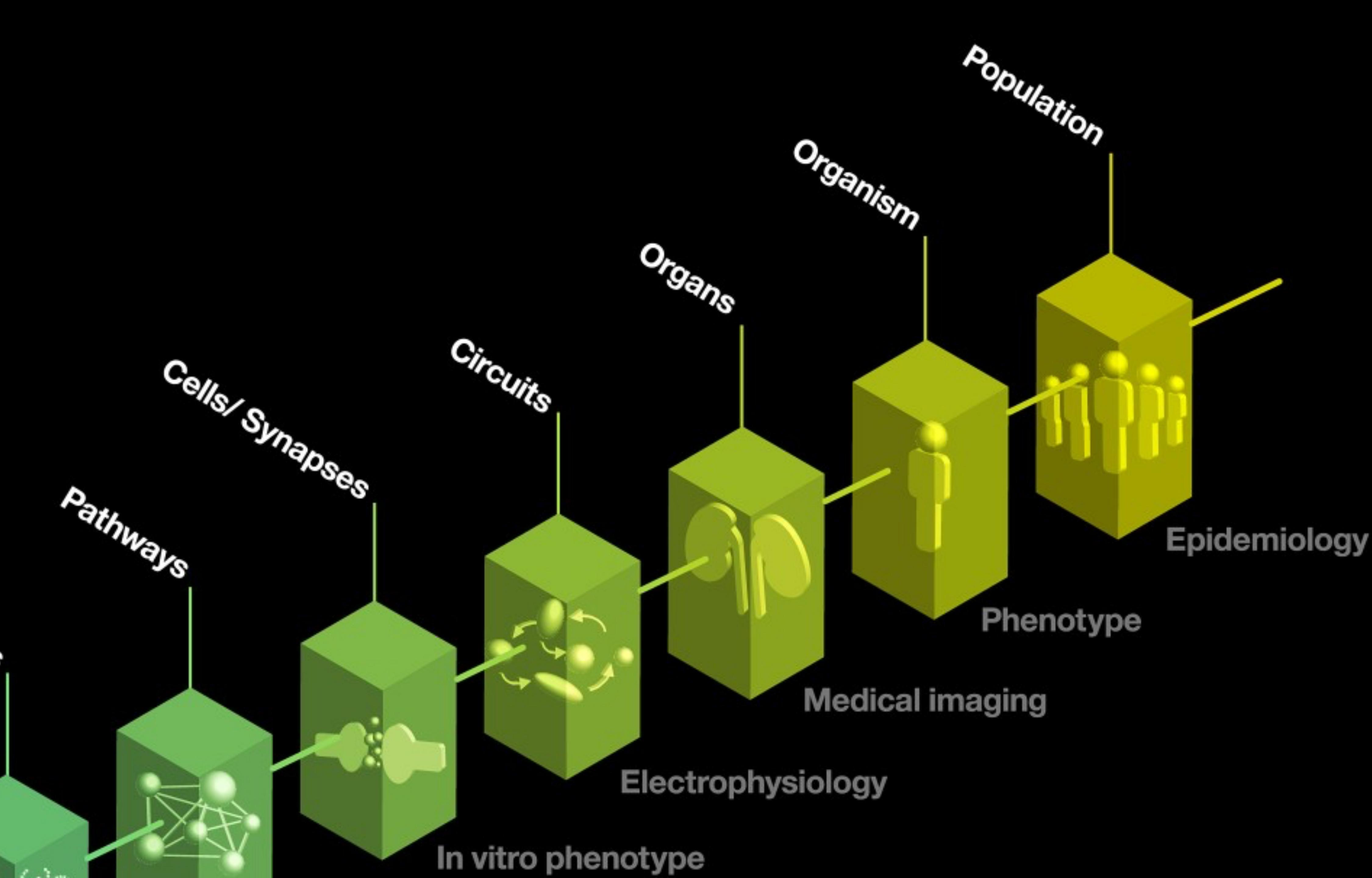
-> How to (start) working with multi-omics!

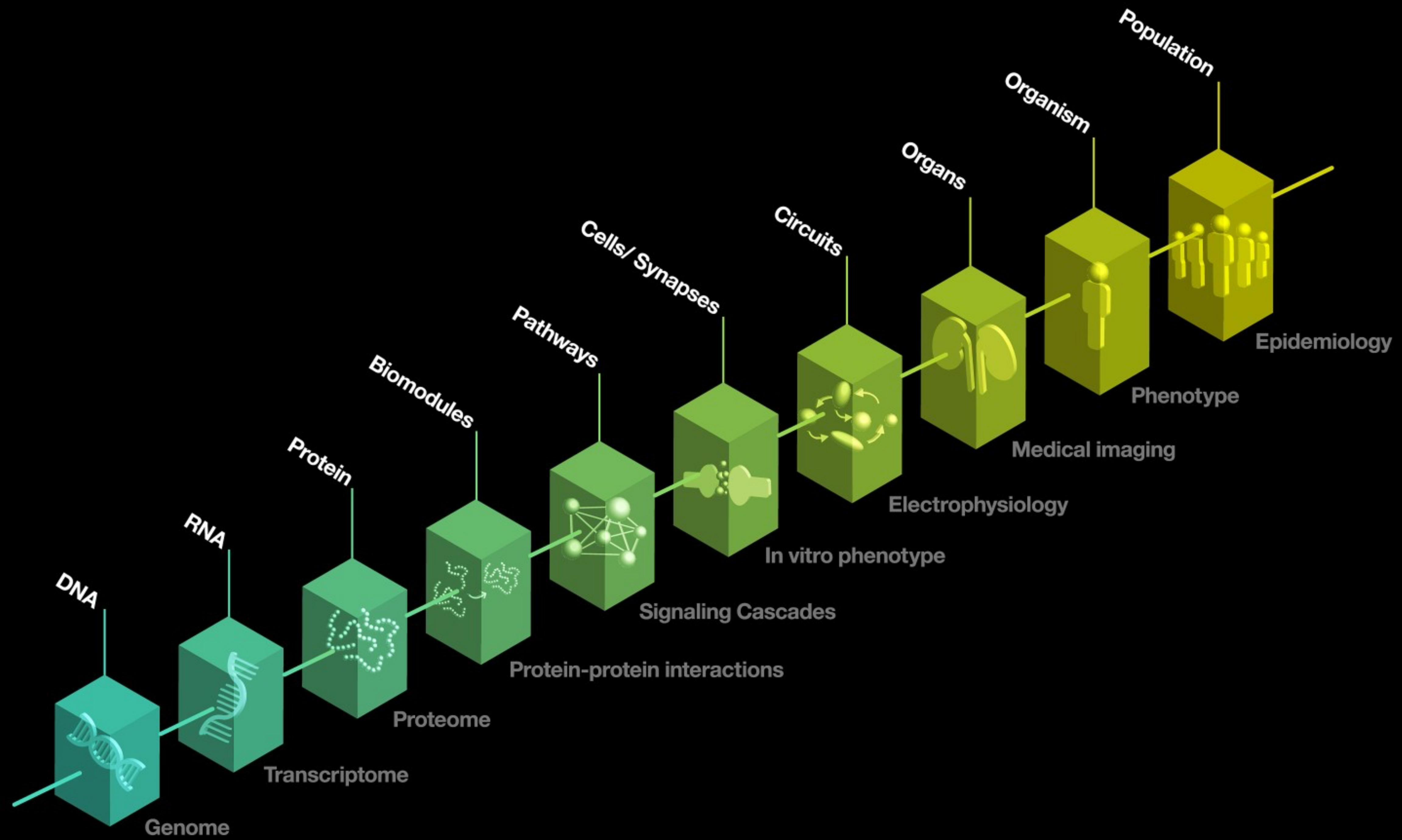
**Biology happens at  
multiple scales.**



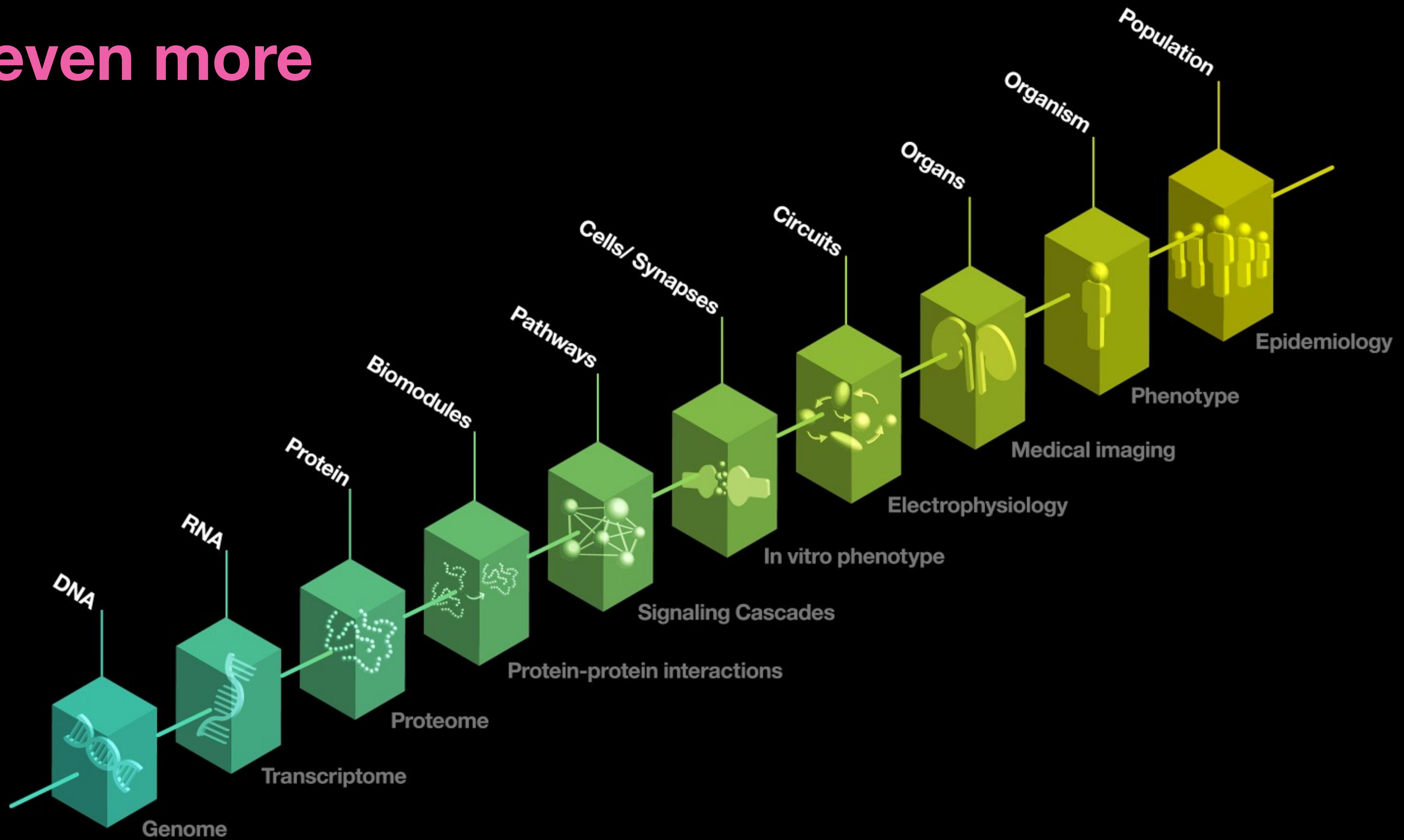
# visual: IBM





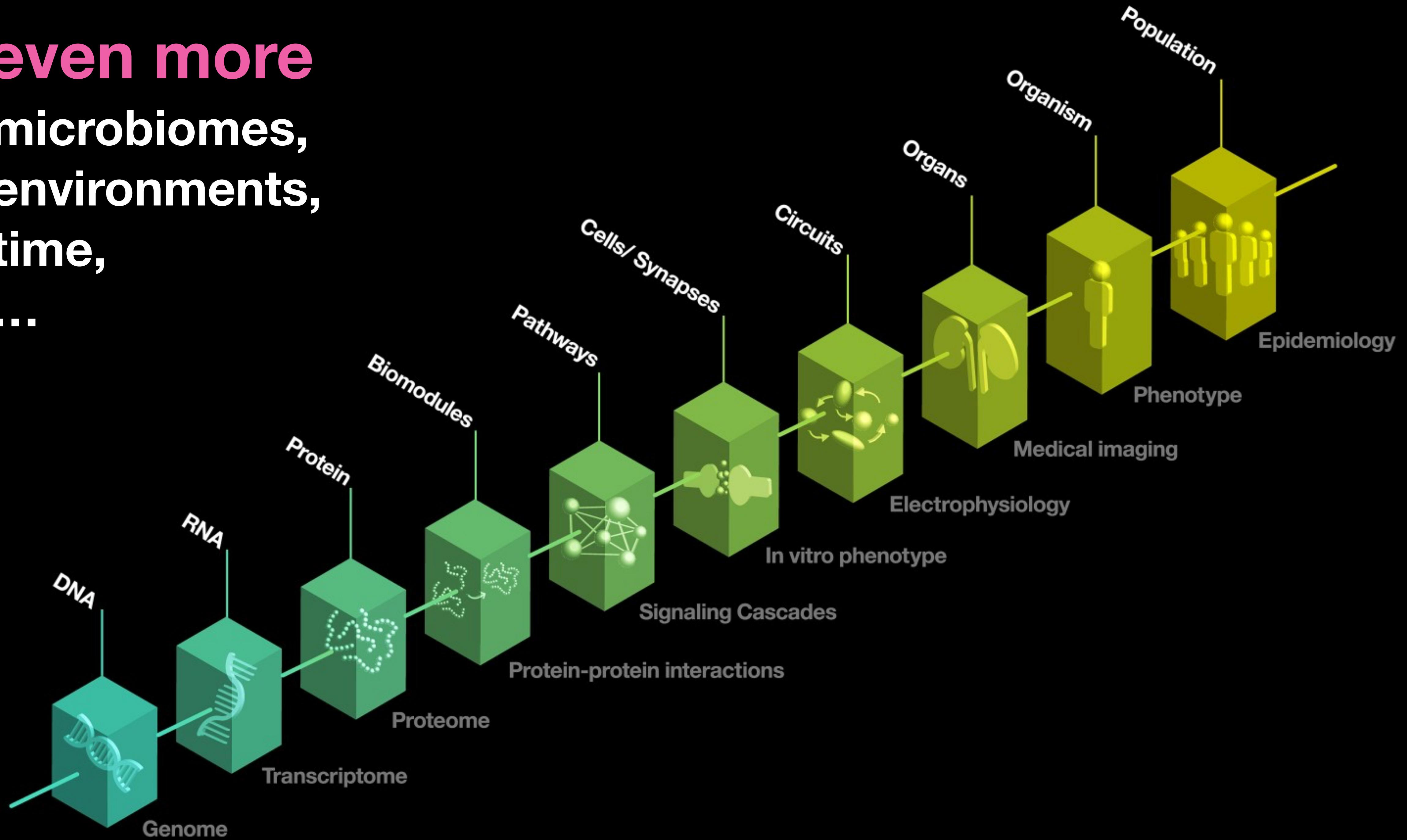


# even more



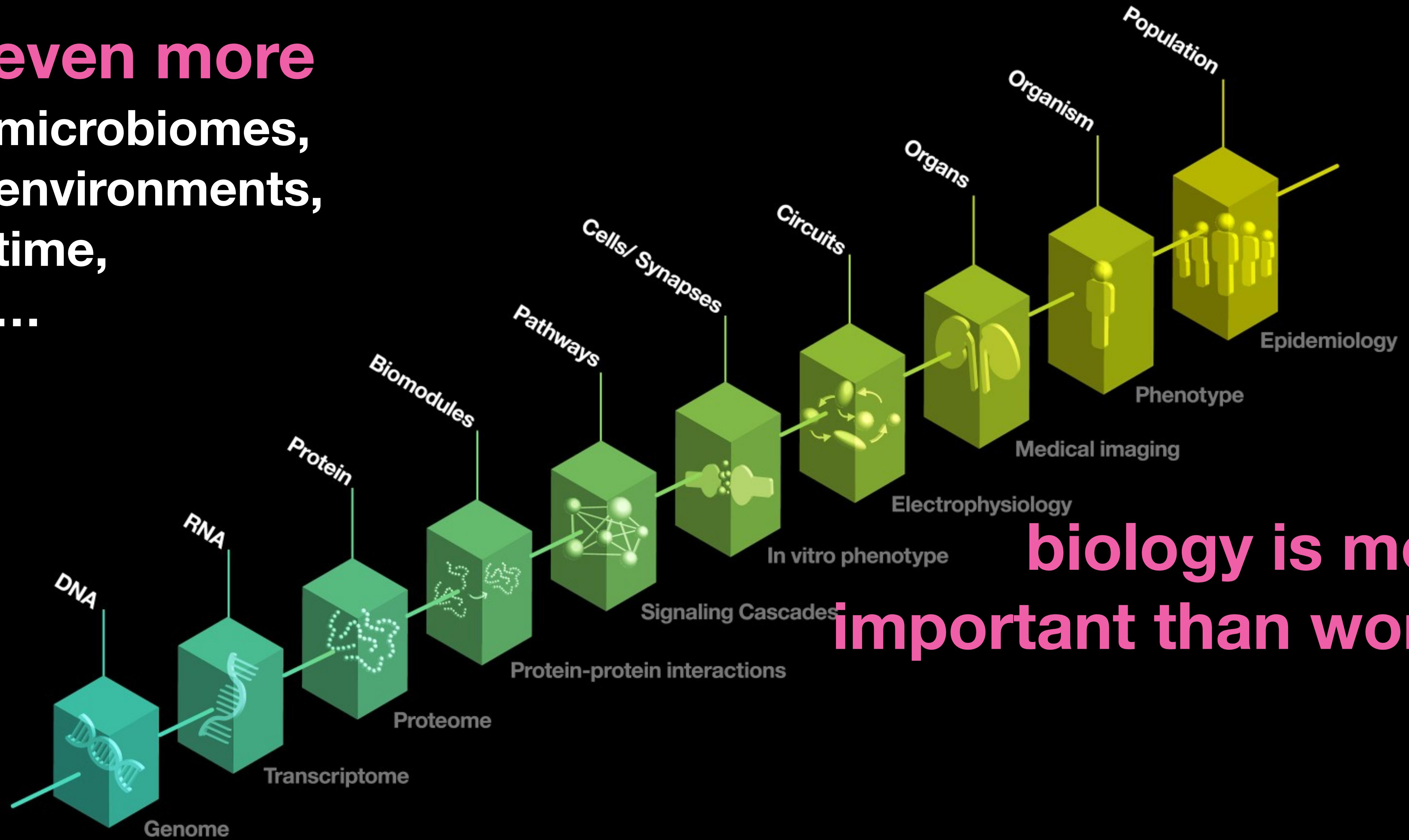
even more  
microbiomes,  
environments,  
time,

...



even more  
microbiomes,  
environments,  
time,

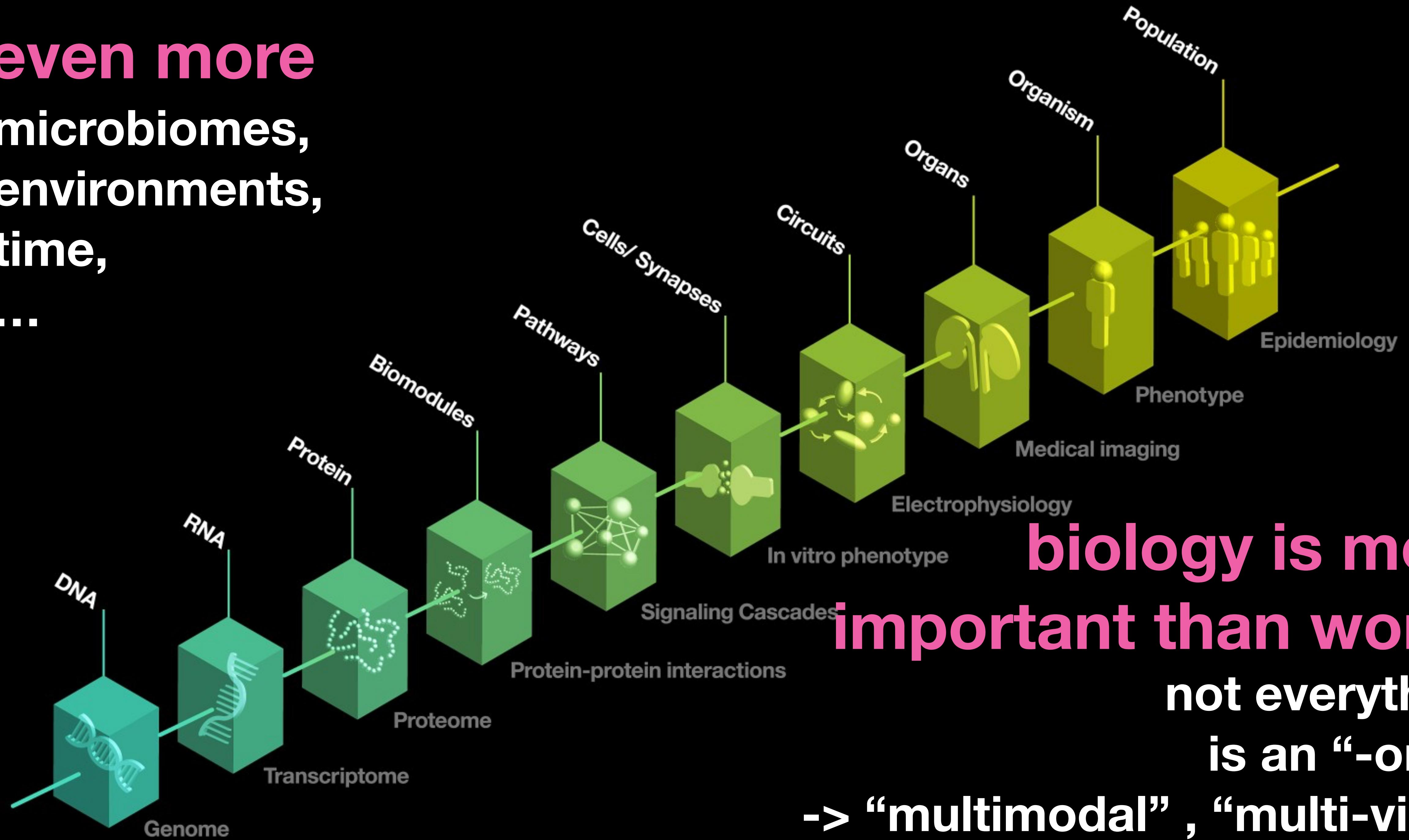
...



biology is more  
important than words

even more  
microbiomes,  
environments,  
time,

...



**biology is more important than words  
not everything is an “-ome”**  
-> “multimodal”, “multi-view”

# Research questions

scales

# Research questions

Example 1: How much does one  
-ome determine another -ome?

scales

# Research questions

scales

Example 1: How much does one -ome determine another -ome?

- sequence features
- technical noise

# Research questions

scales

Example 1: How much does one -ome determine another -ome?

- sequence features
- technical noise

Example 2: Which papilloma virus strains are pathogenic?

# Do transcriptomes explain proteomes?

# Do transcriptomes explain proteomes?



LIKELY

# Do transcriptomes explain proteomes?

~2005



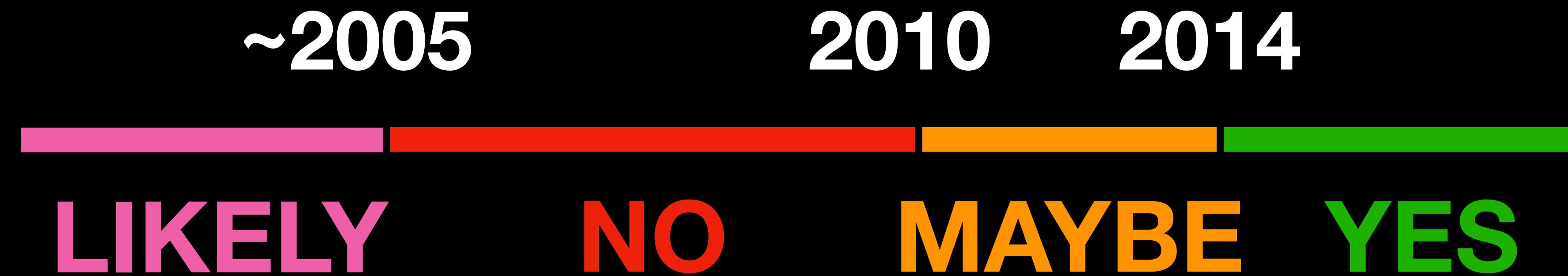
**LIKELY**

**NO**

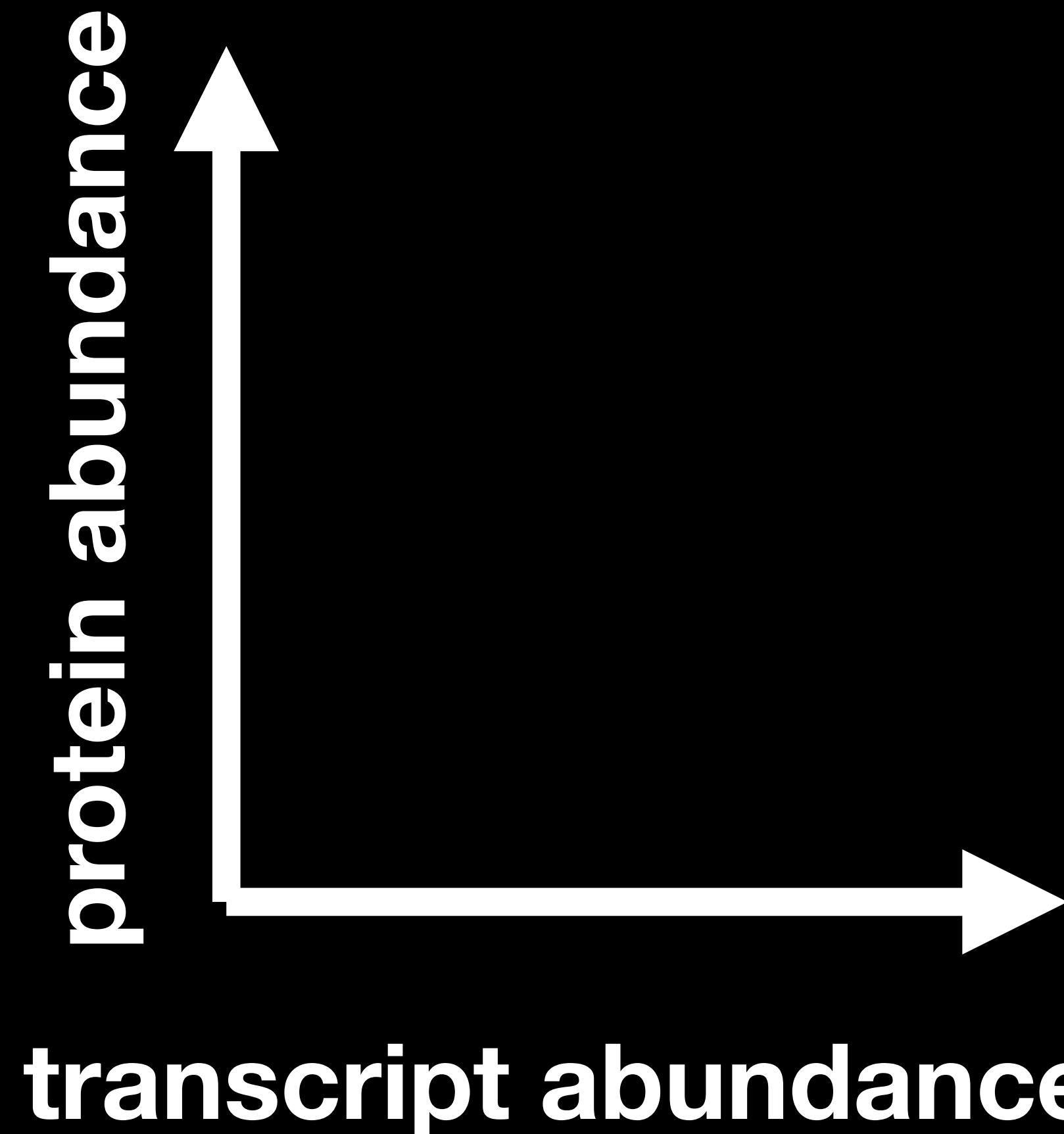
# Do transcriptomes explain proteomes?



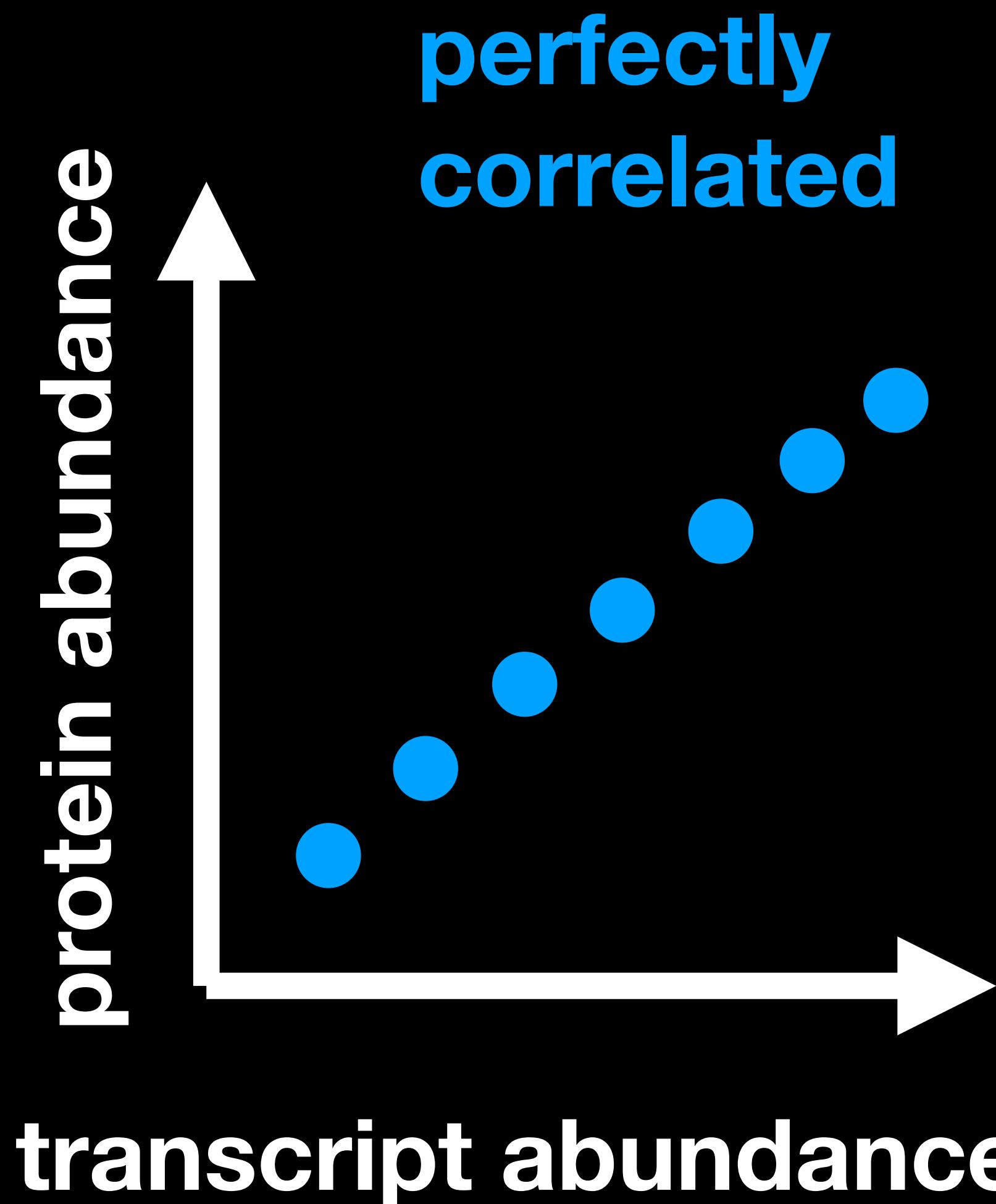
# Do transcriptomes explain proteomes?



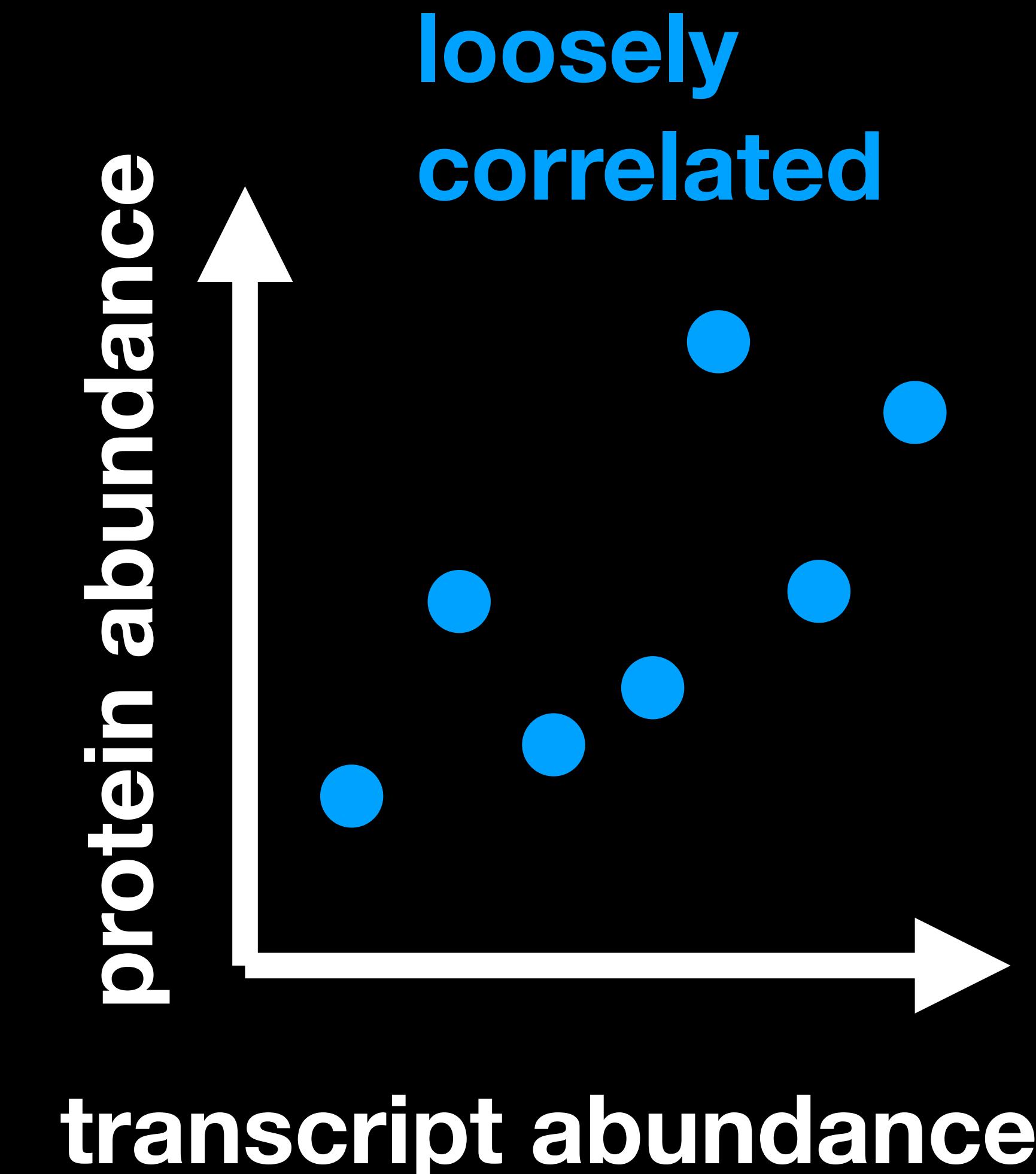
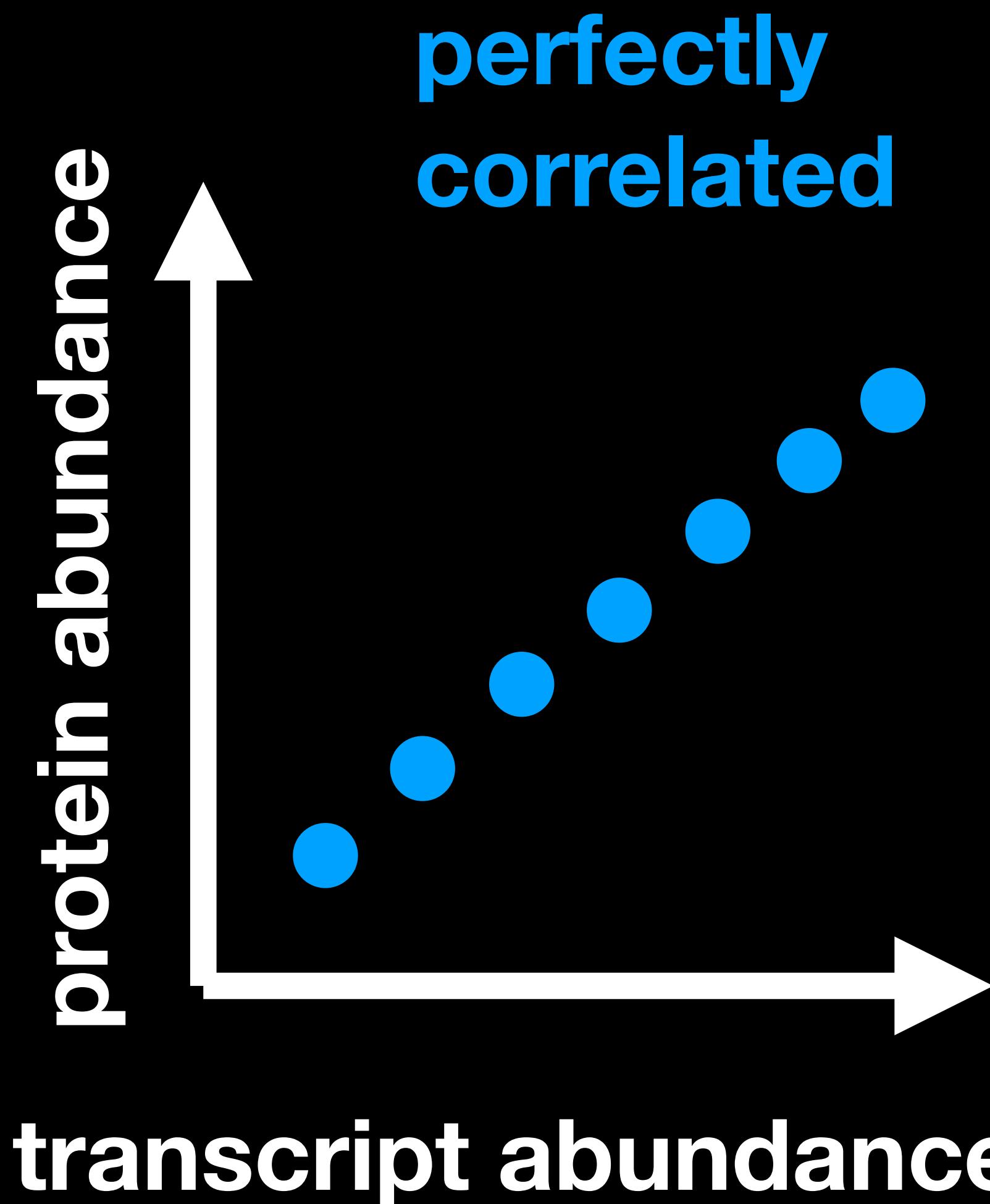
If transcriptomes explain proteomes, there should be a high correlation between them.



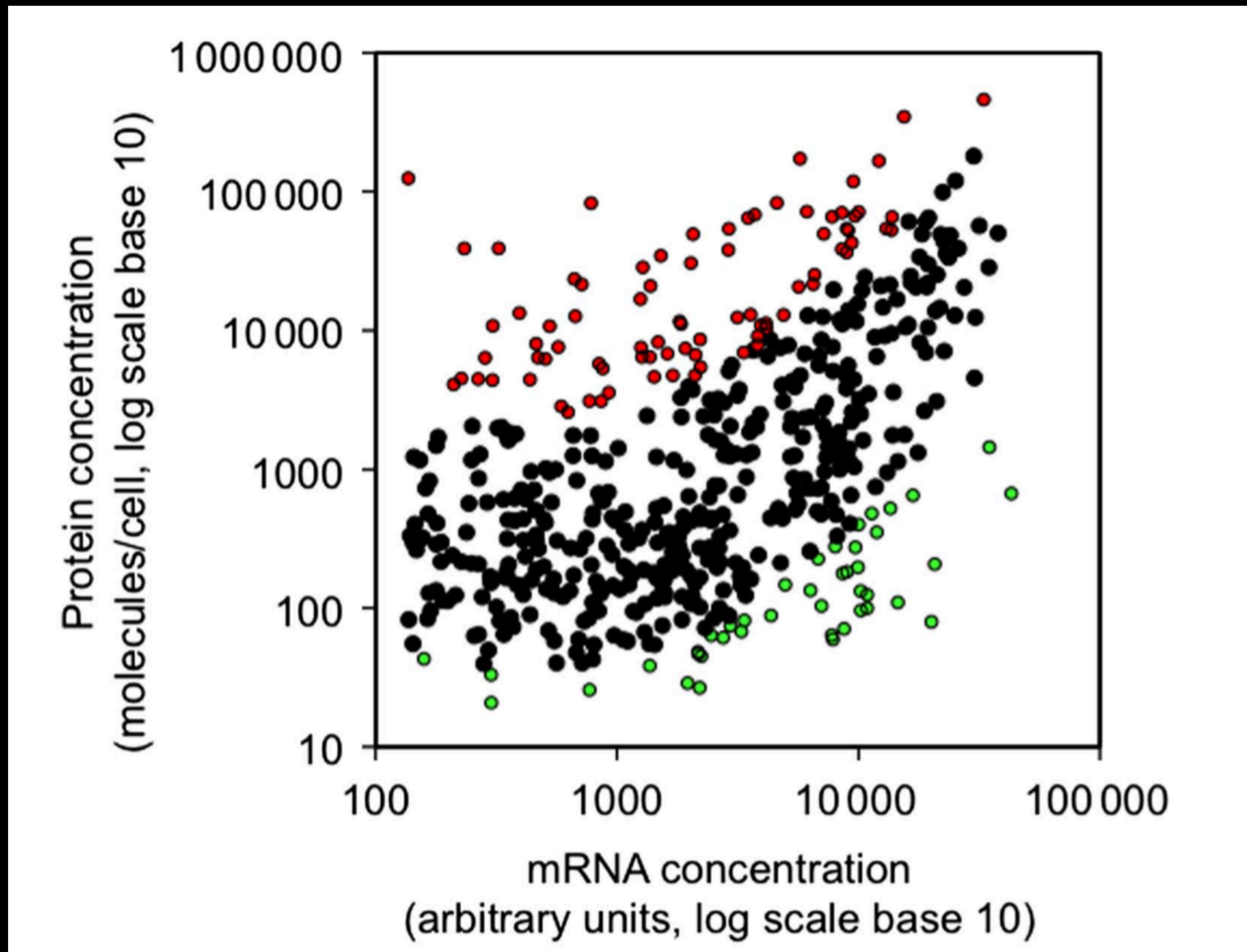
If transcriptomes explain proteomes, there should be a high correlation between them.



If transcriptomes explain proteomes, there should be a high correlation between them.



# Transcriptomes only loosely correlate with proteomes.



Vogel et al., 2010

Brain cancer cell line (soluble fraction)

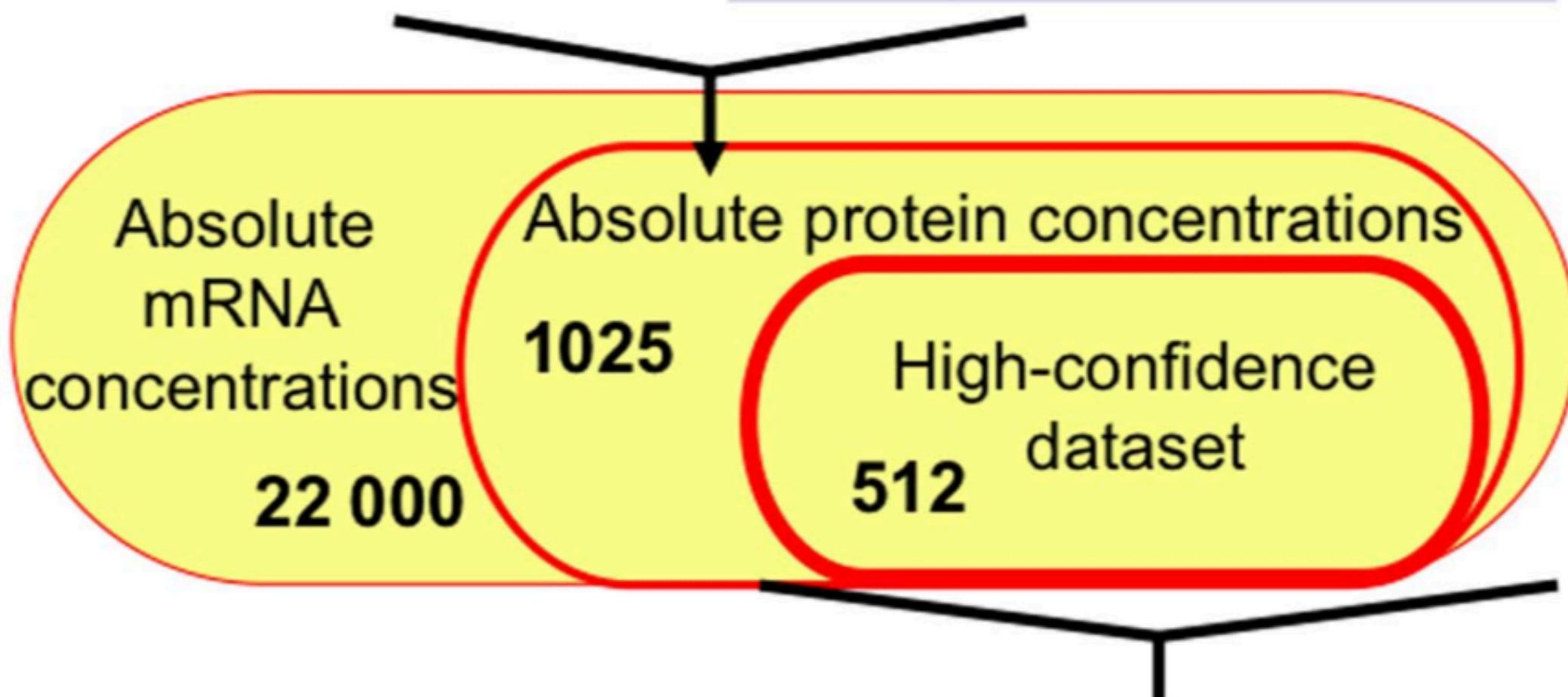
**Microarrays**

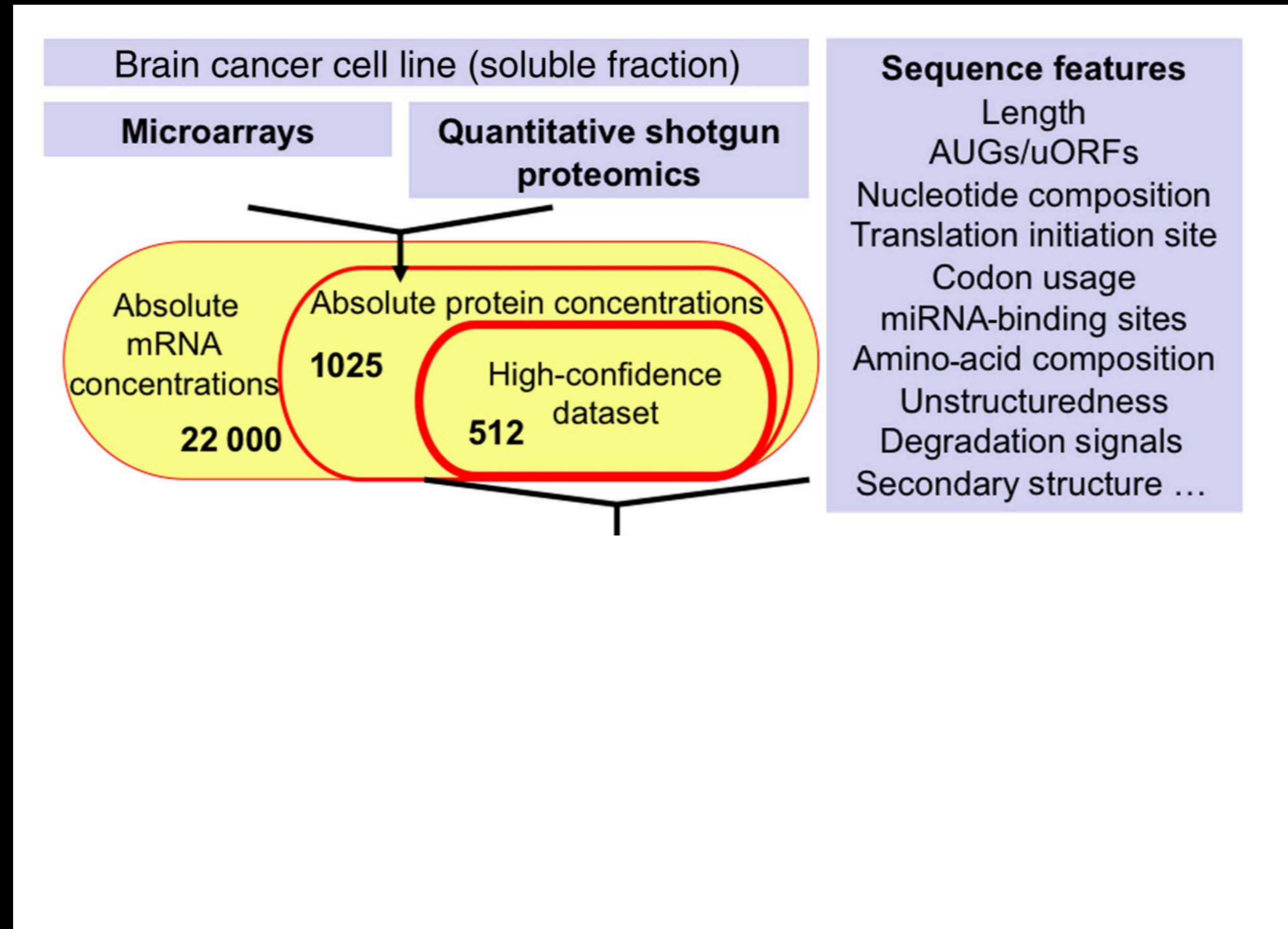
**Quantitative shotgun  
proteomics**

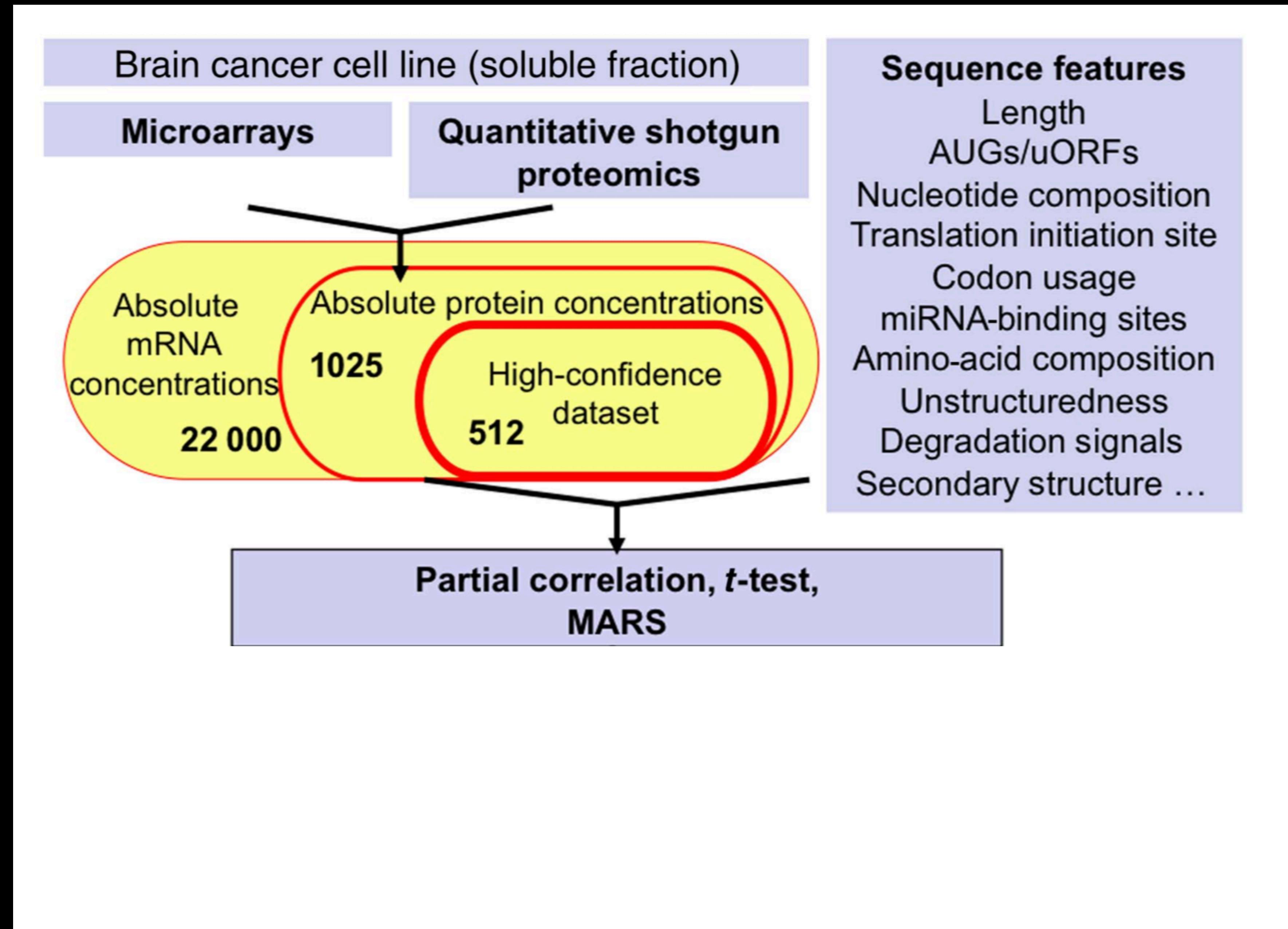
Brain cancer cell line (soluble fraction)

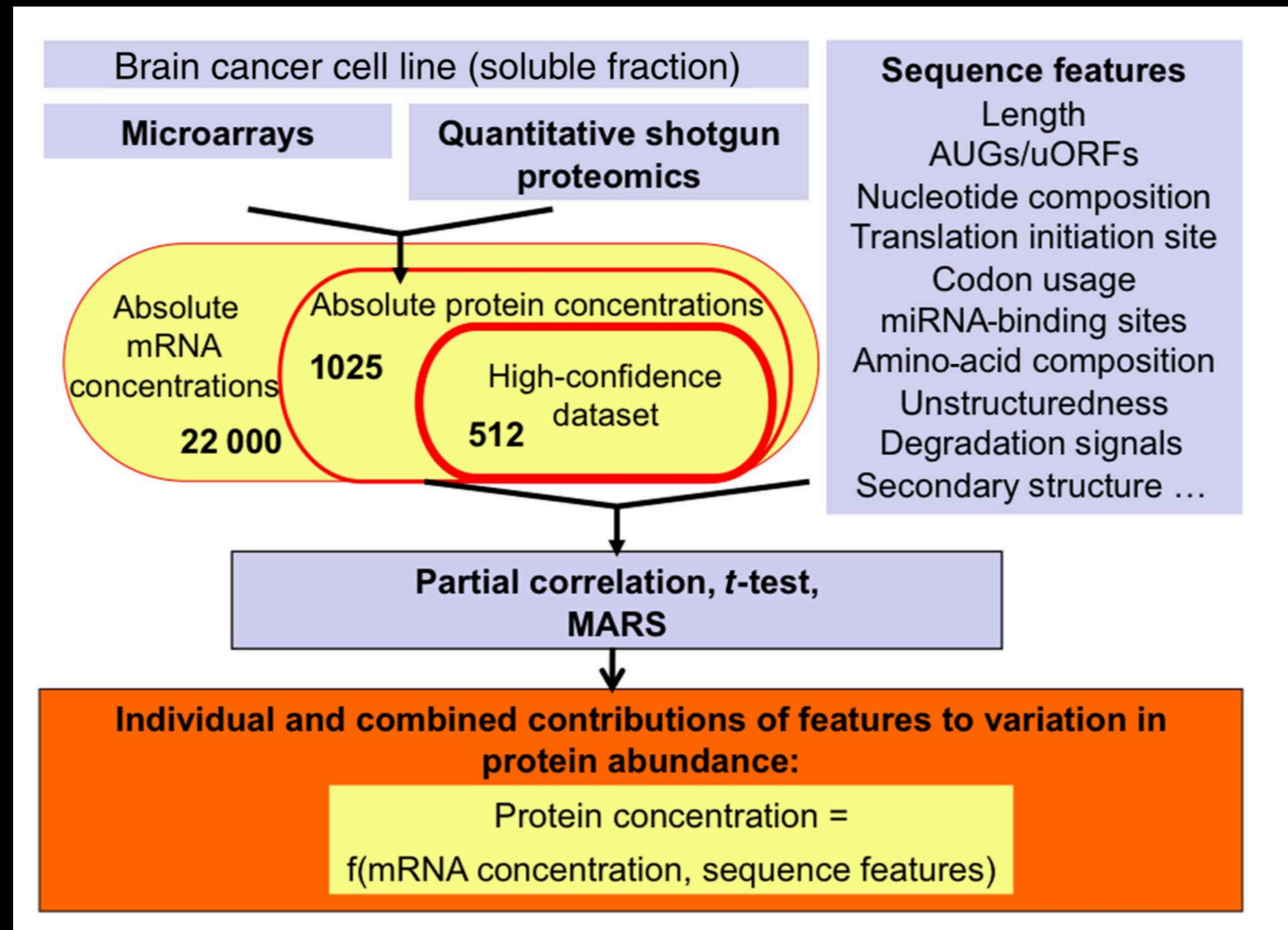
Microarrays

Quantitative shotgun  
proteomics





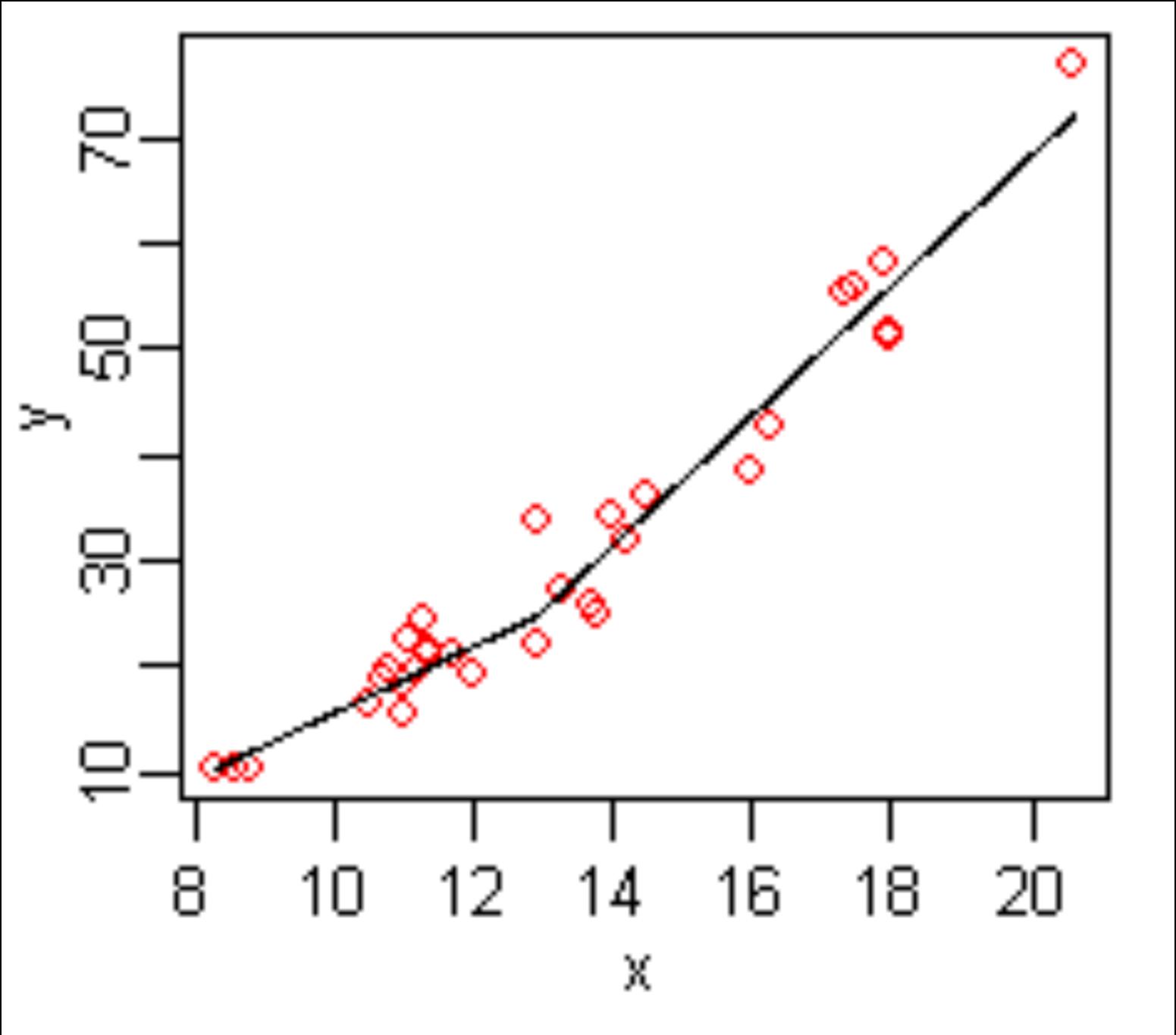




**Partial correlation, *t*-test,  
MARS**

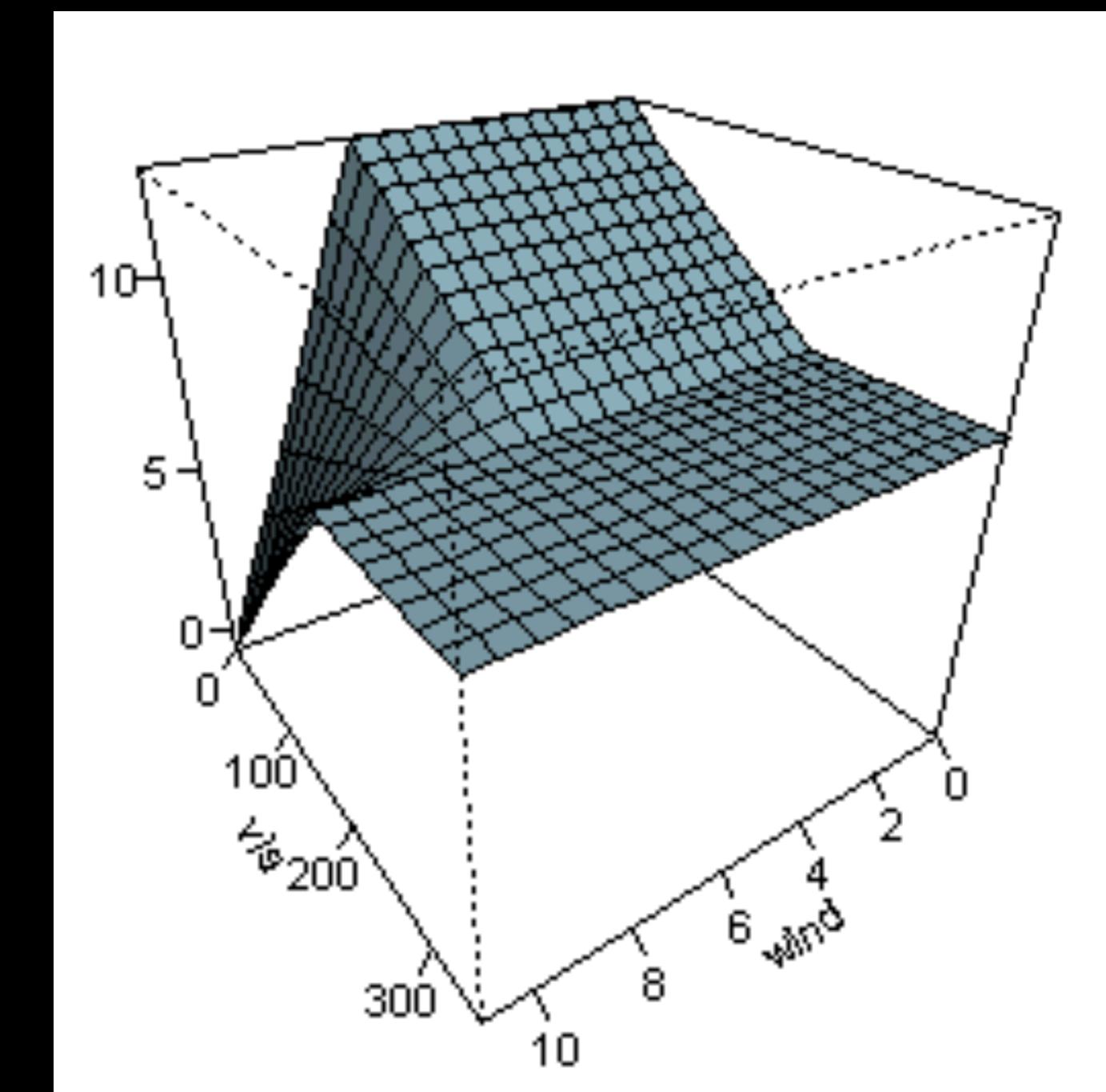
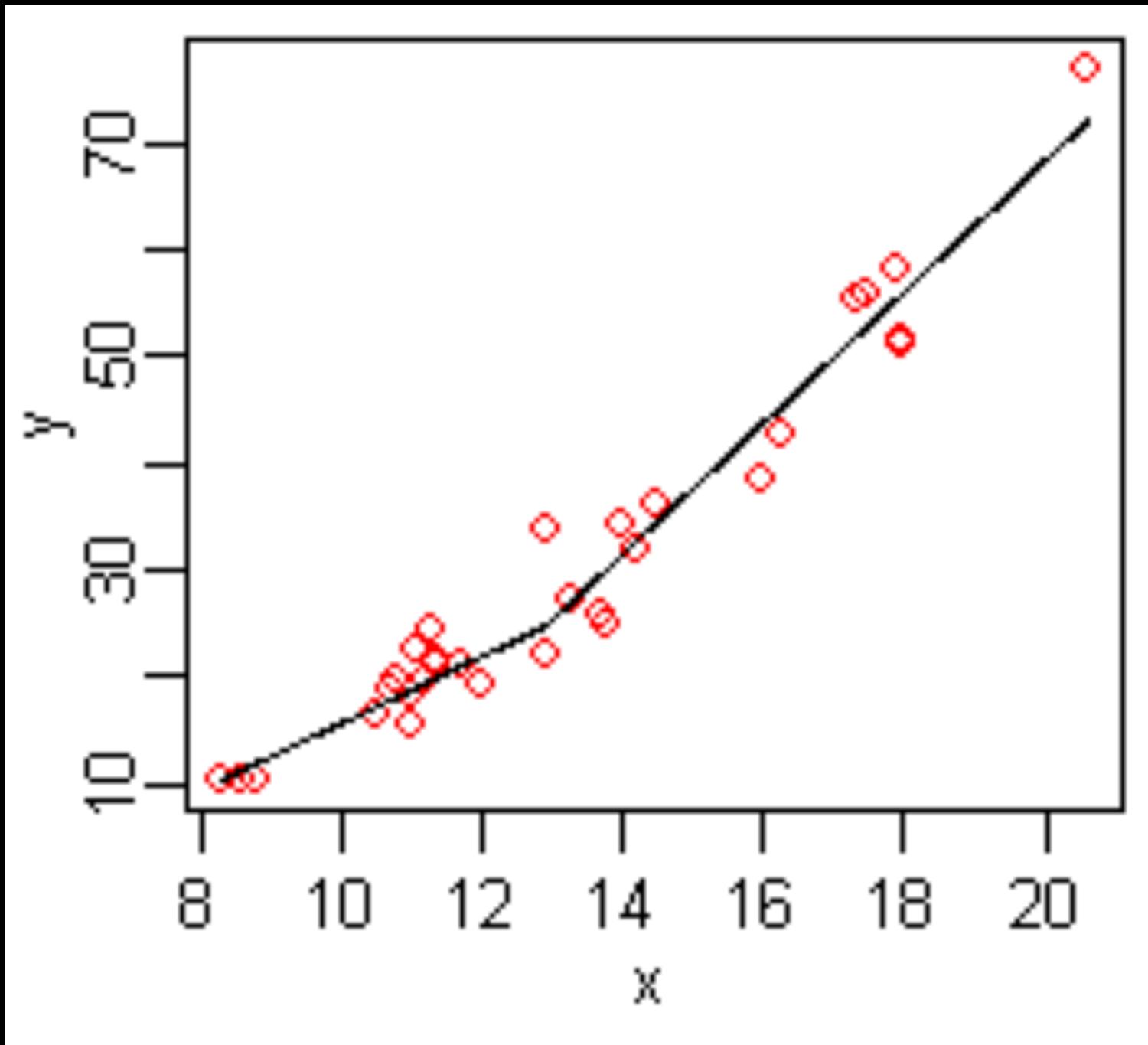
**Partial correlation, *t*-test,  
MARS**

**multi-omics: often non-linear**



Partial correlation,  $t$ -test,  
MARS

multi-omics: often non-linear

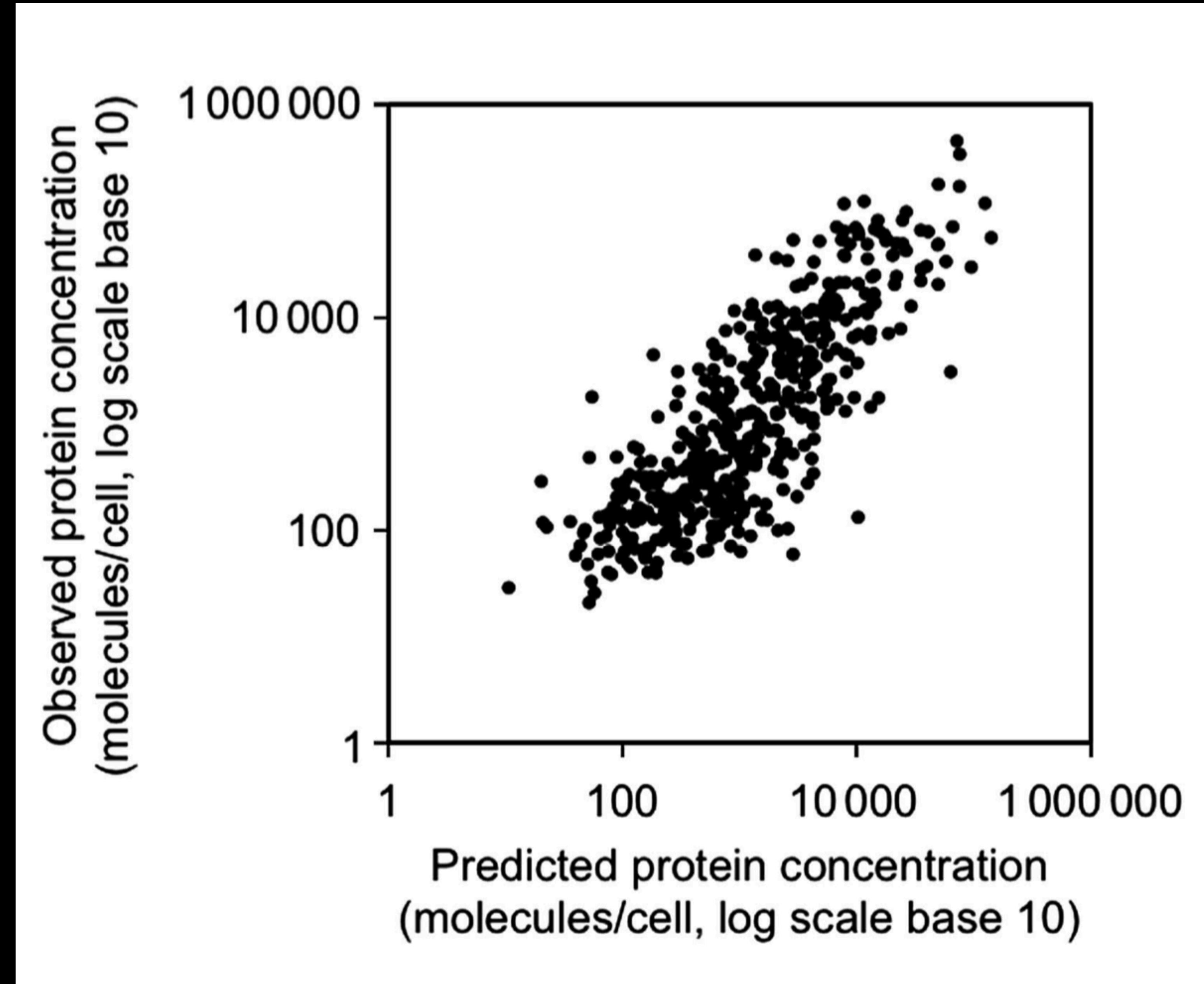


**Partial correlation, *t*-test,  
MARS**

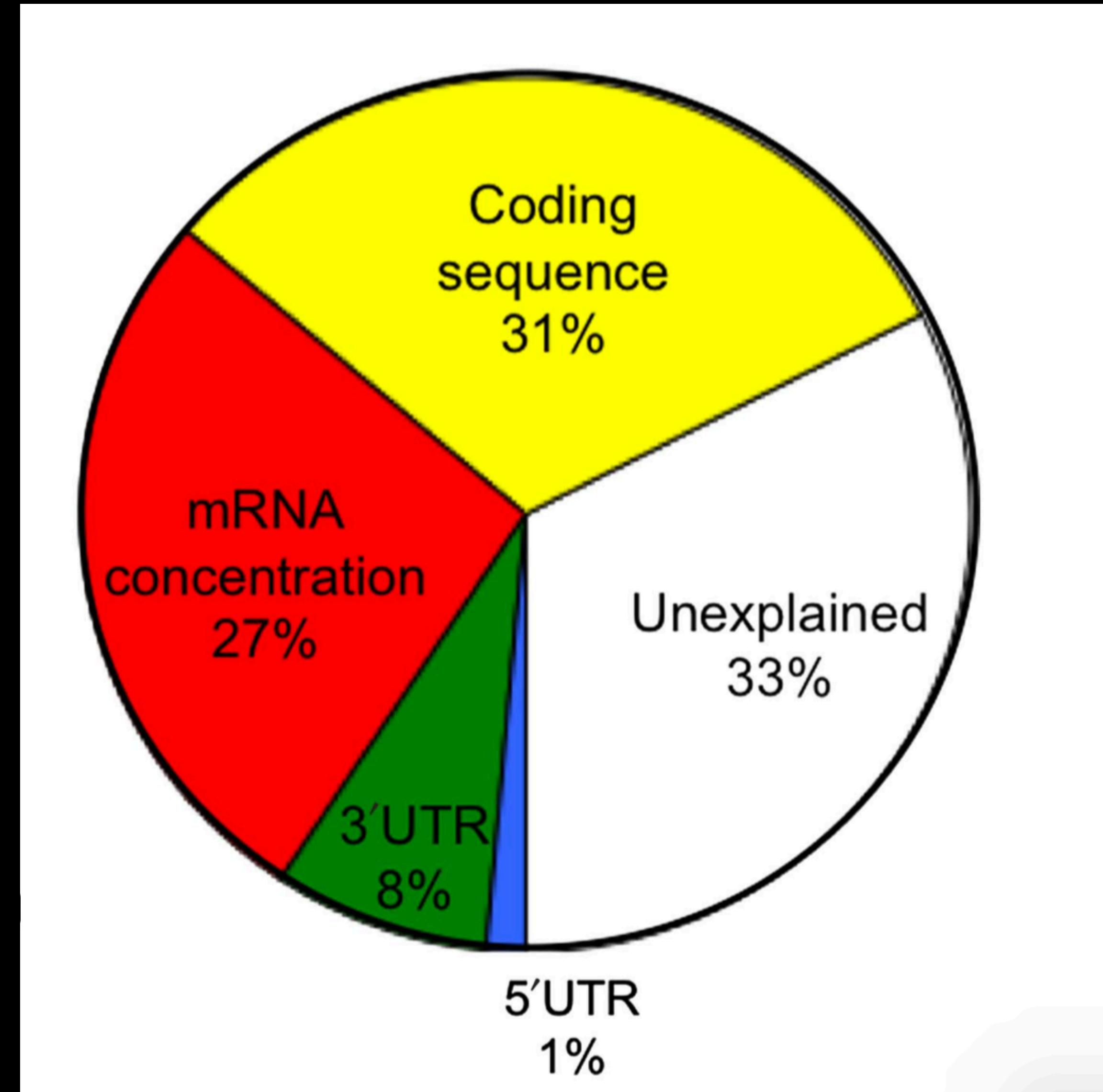
multi-omics: often non-linear

wikipedia

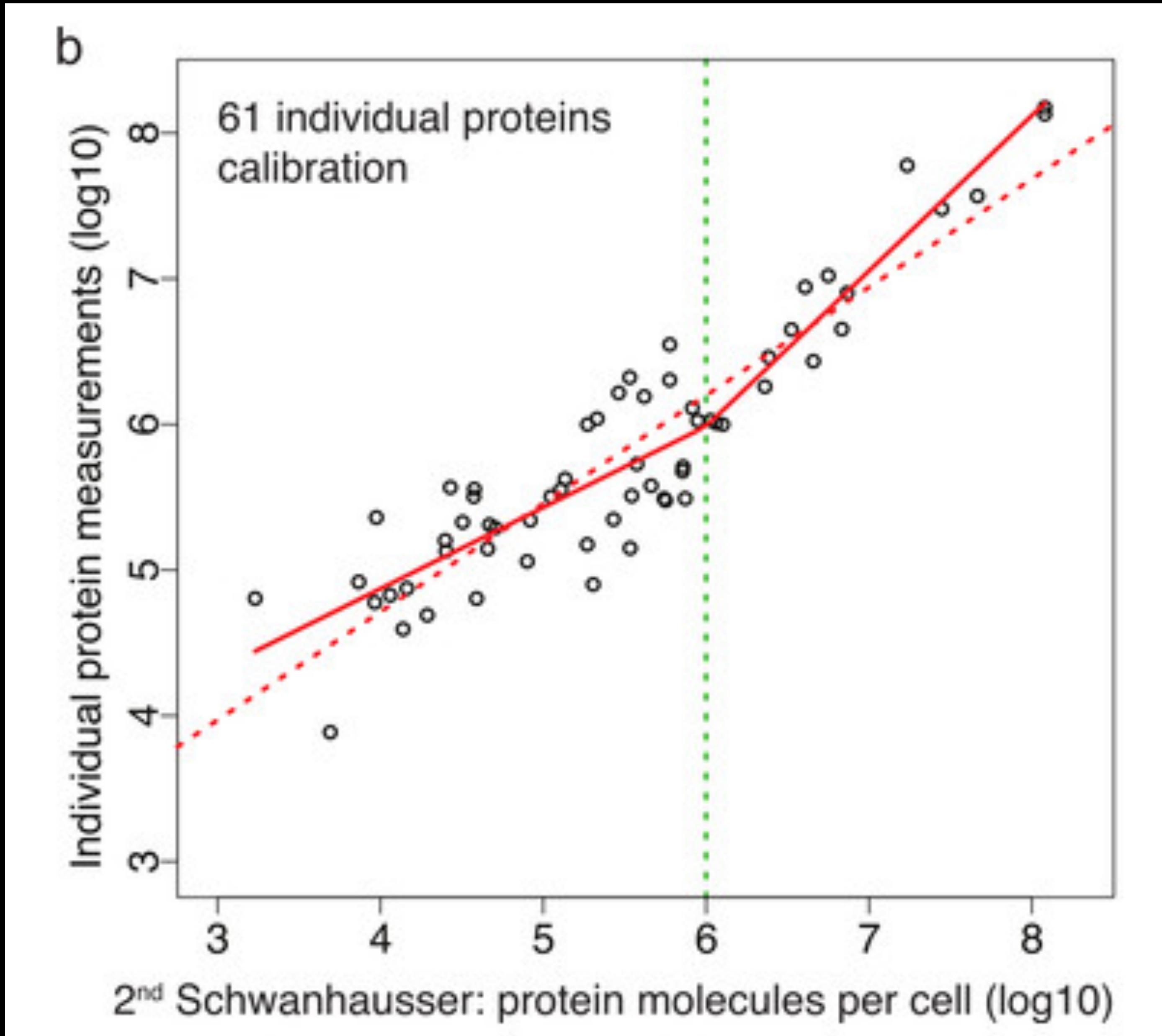
# Transcriptomes and sequence features predict the proteome.



# Transcriptomes and sequence features explain 2/3 of the proteome.

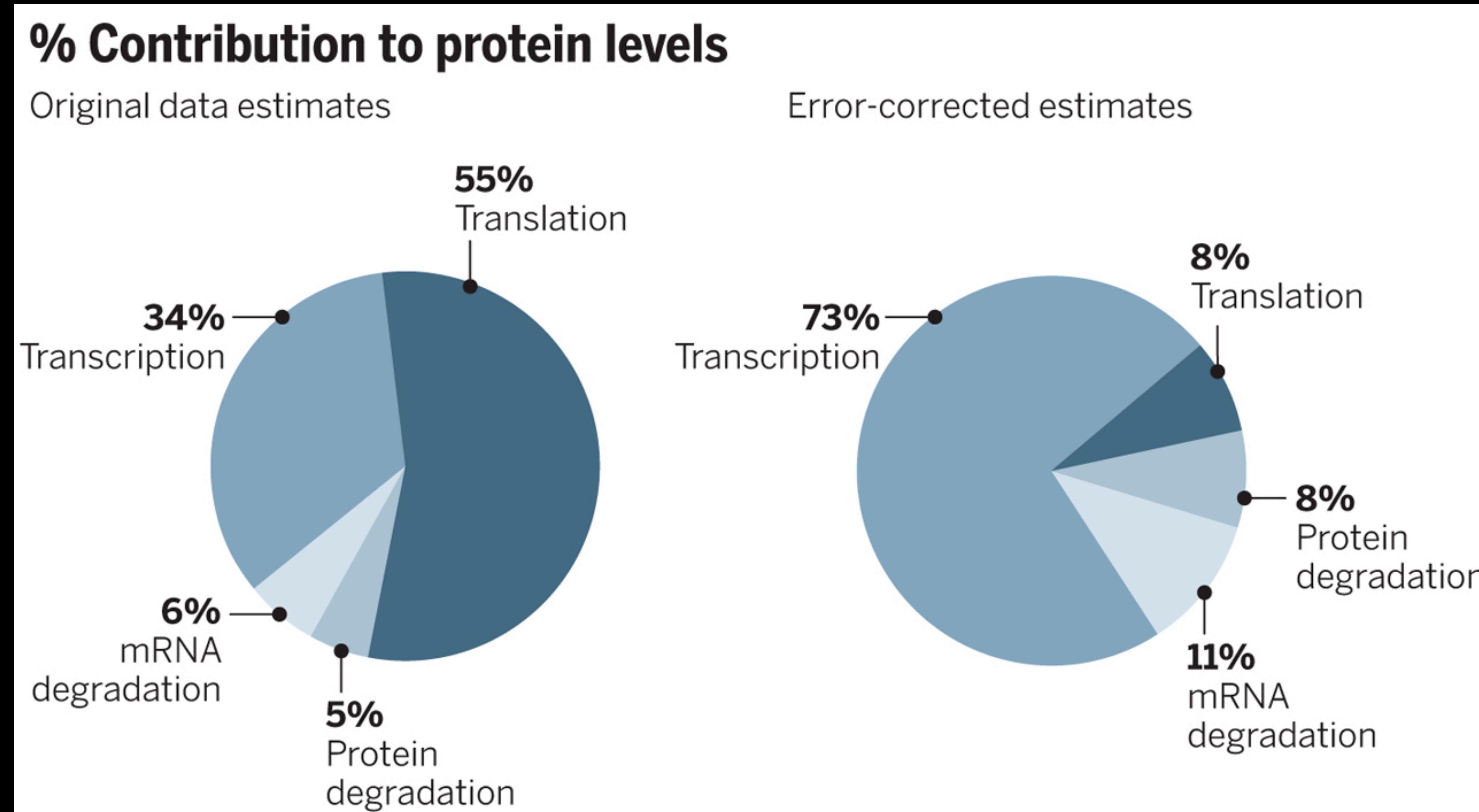


# Datasets have measurement errors!



Li et al. 2014  
Jovanovich et al. 2015  
Battle et al. 2015  
Li et Biggin 2015

# Controlling for measurement errors increases contribution with which transcriptomes explain proteomes.



# Research questions

scales

Example 1: How much does one -ome determine another -ome?

- sequence features
- technical noise

# Research questions

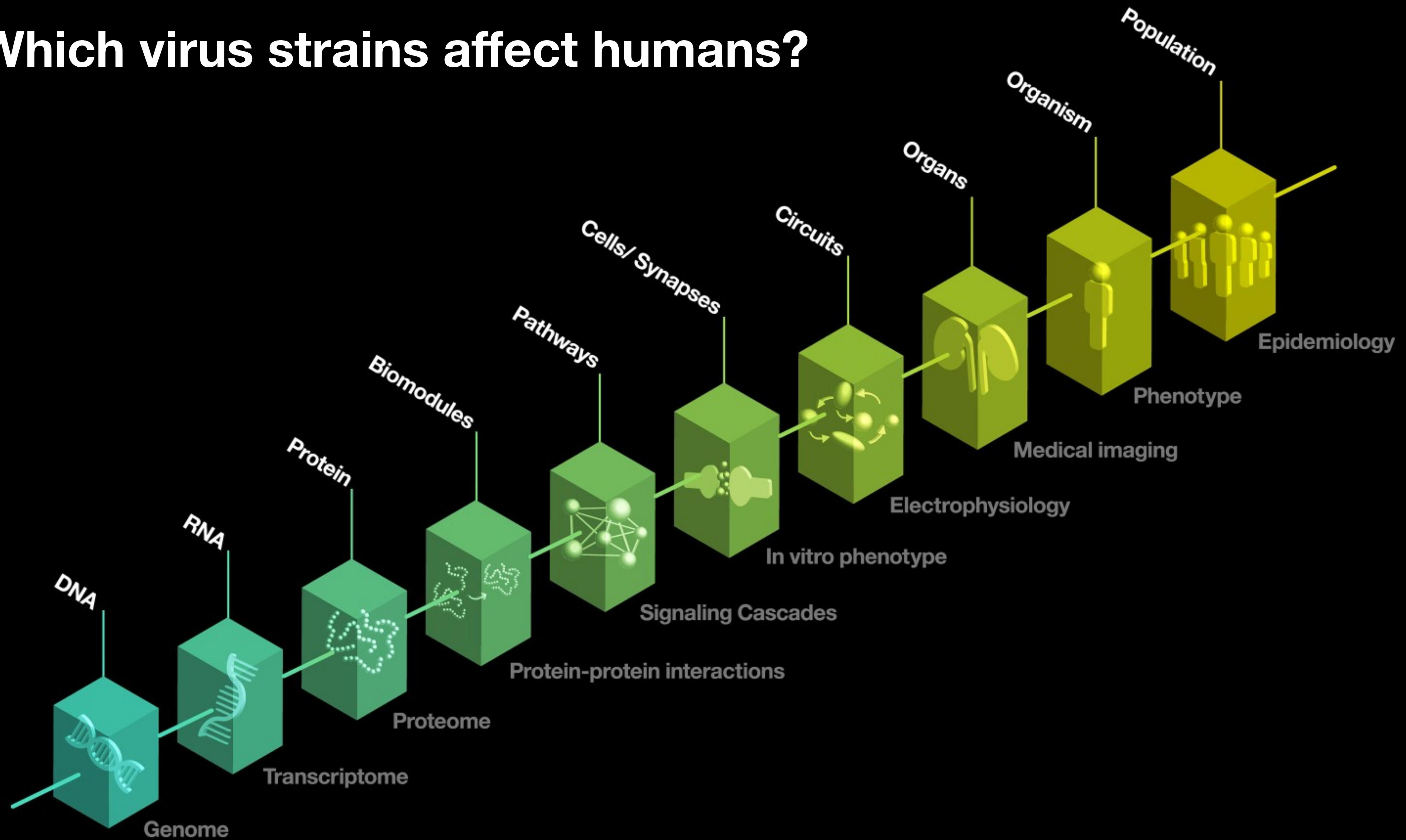
scales

Example 1: How much does one -ome determine another -ome?

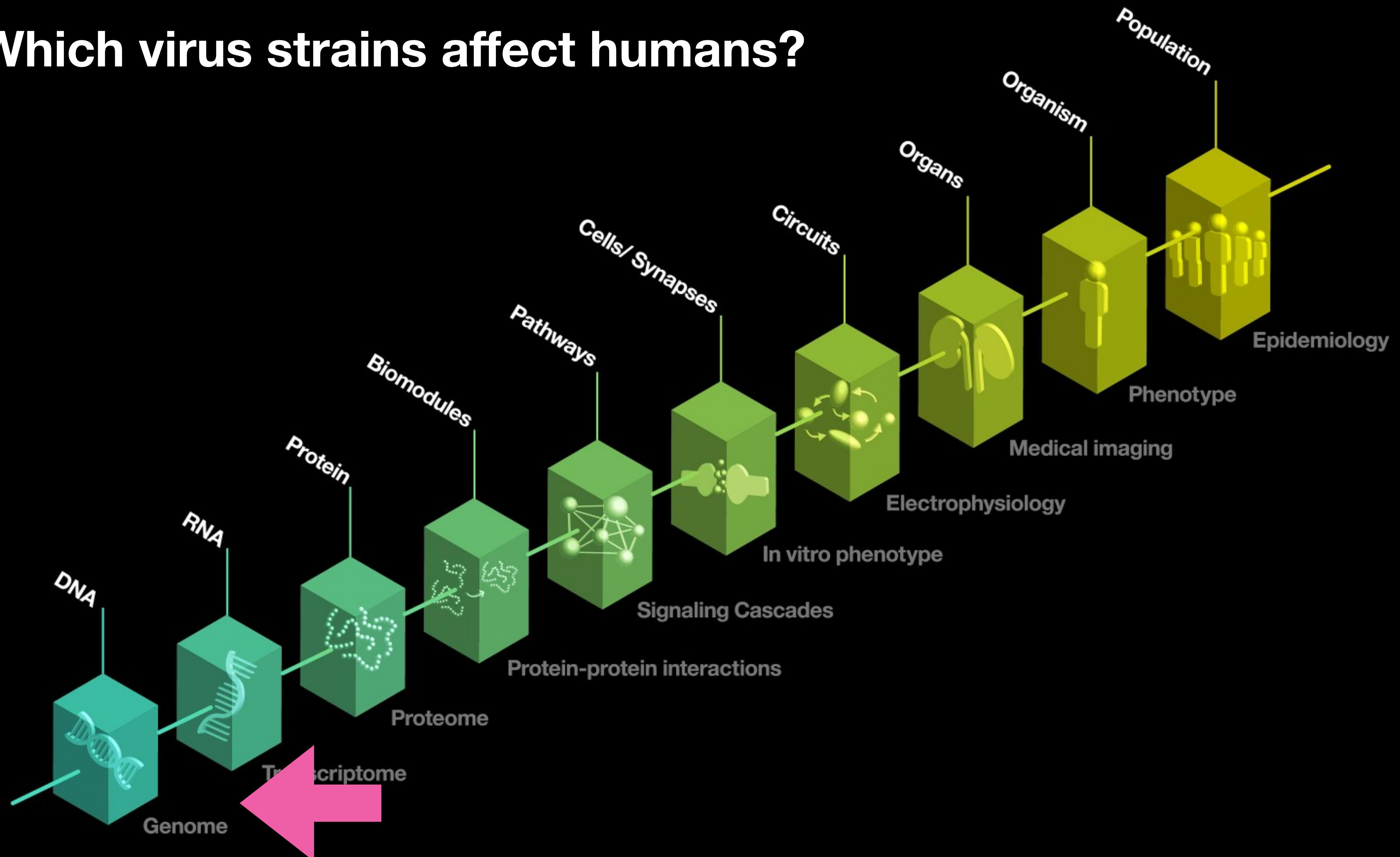
- sequence features
- technical noise

Example 2: Which papilloma virus strains are pathogenic?

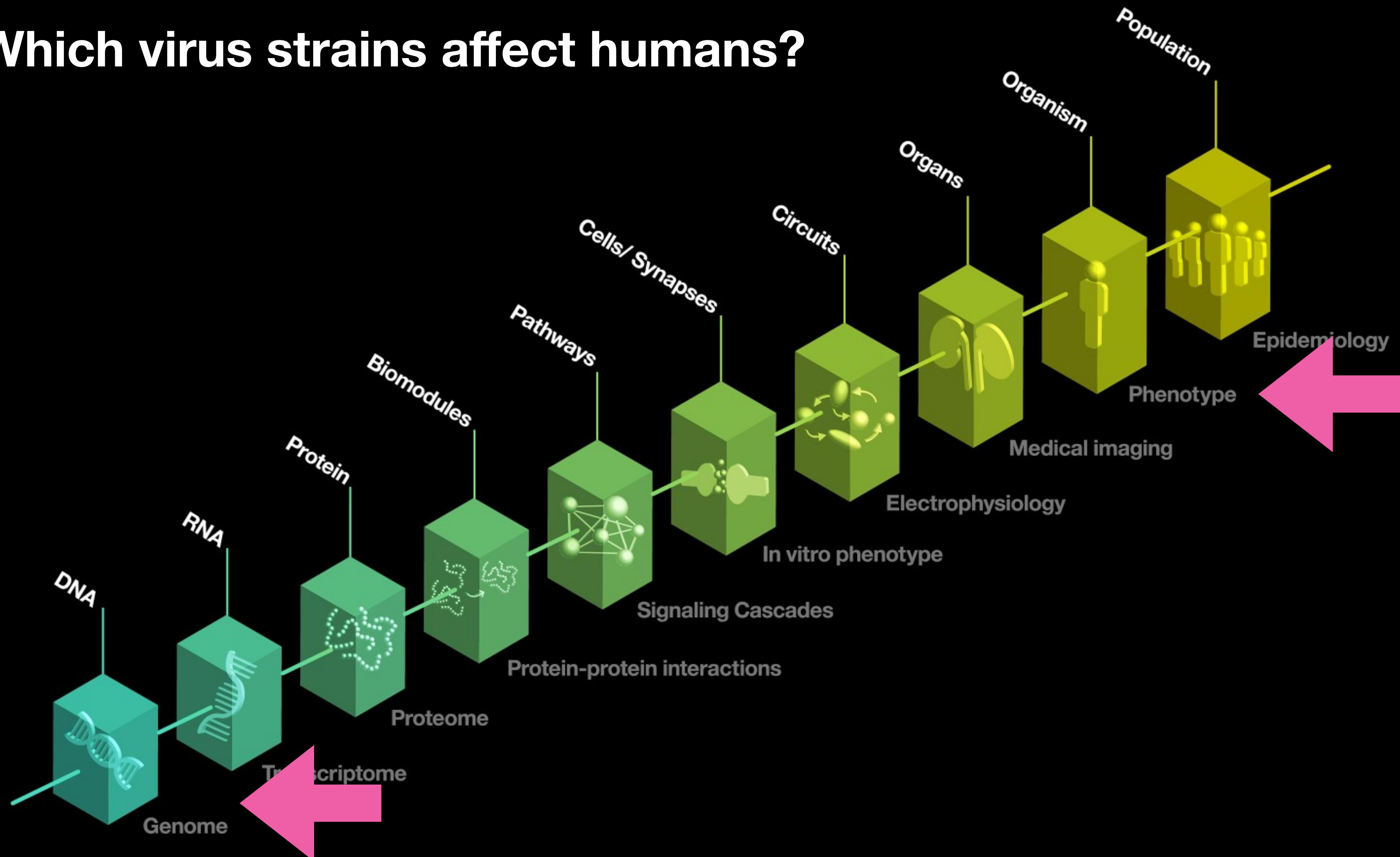
# Which virus strains affect humans?



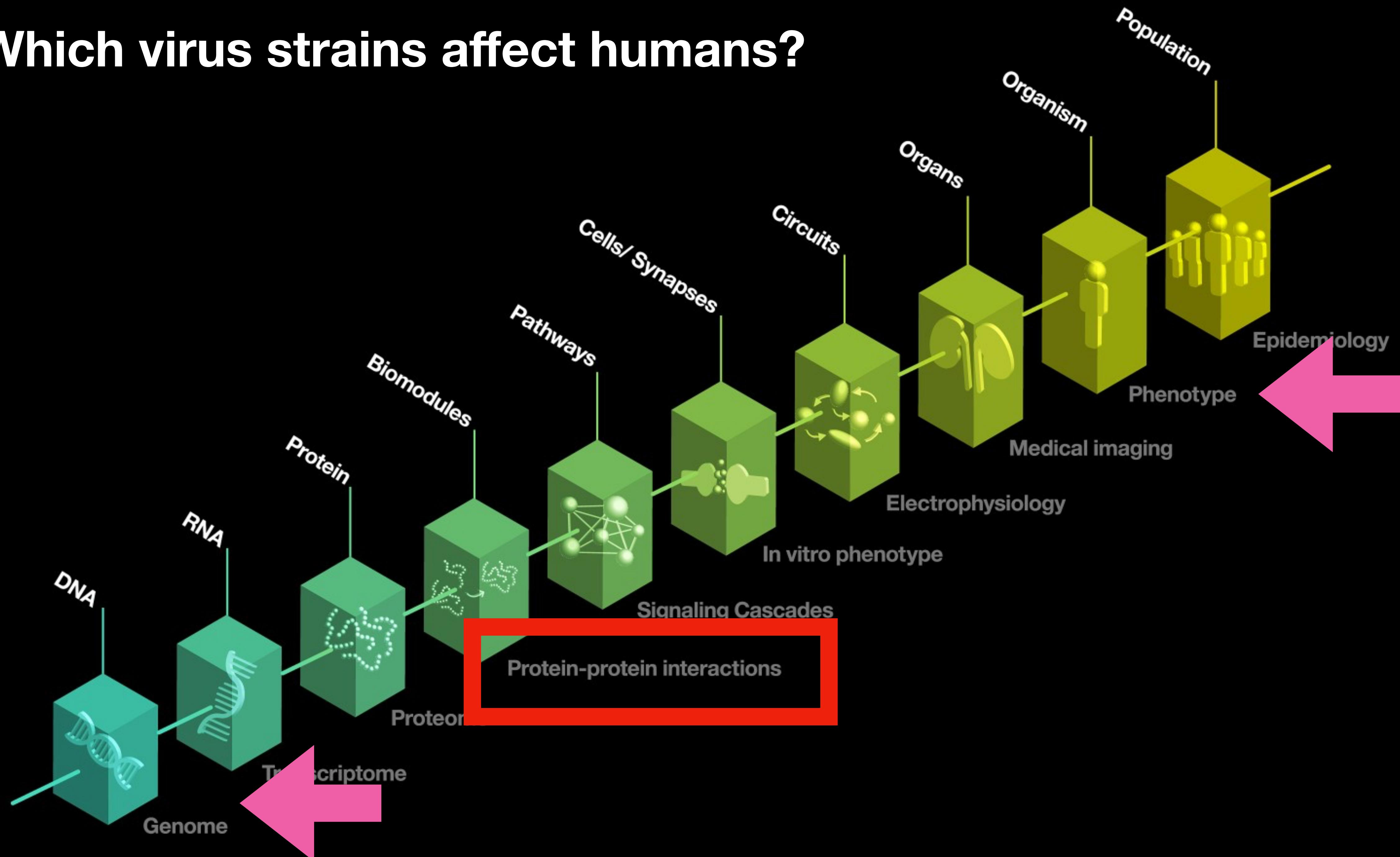
# Which virus strains affect humans?



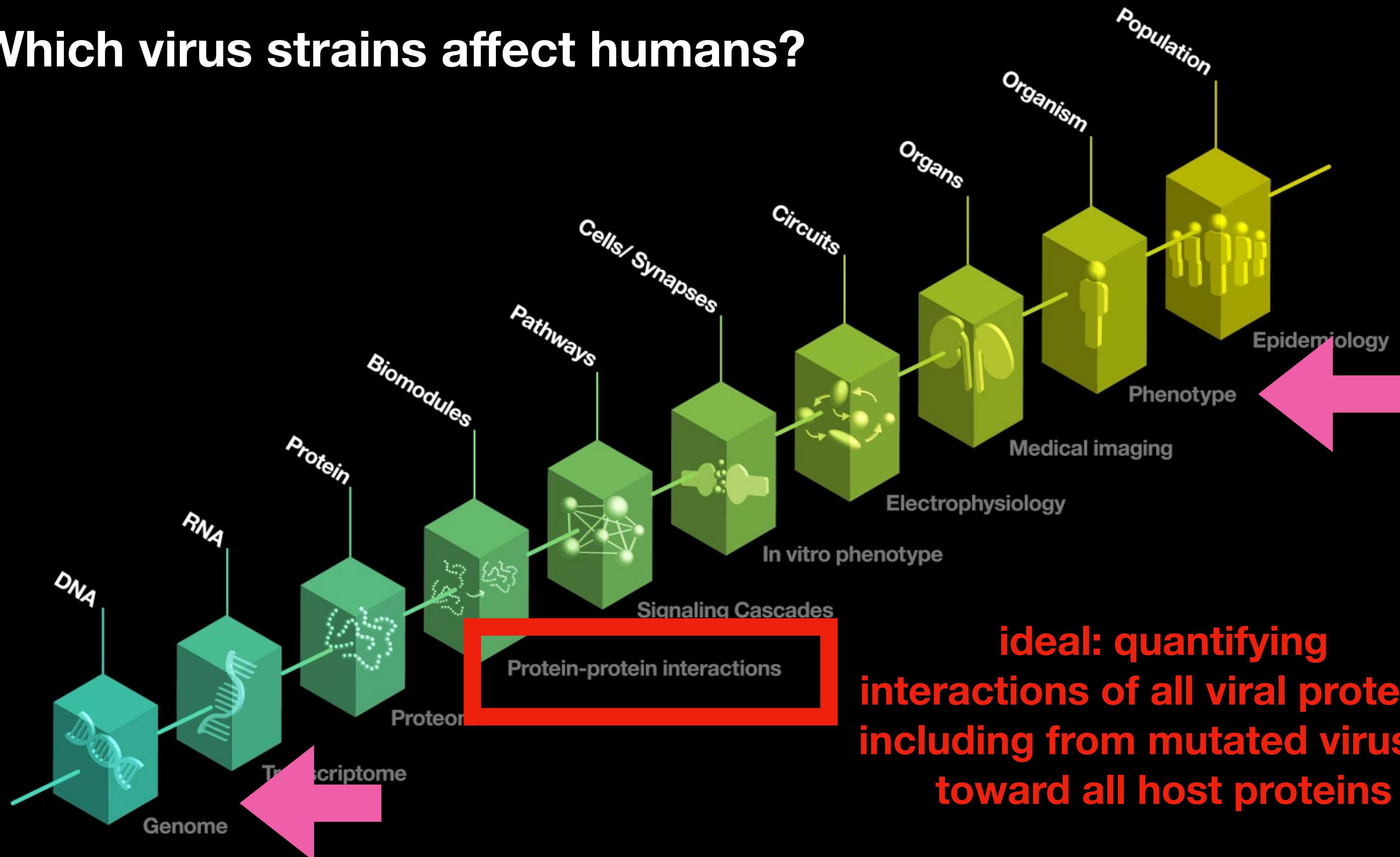
# Which virus strains affect humans?



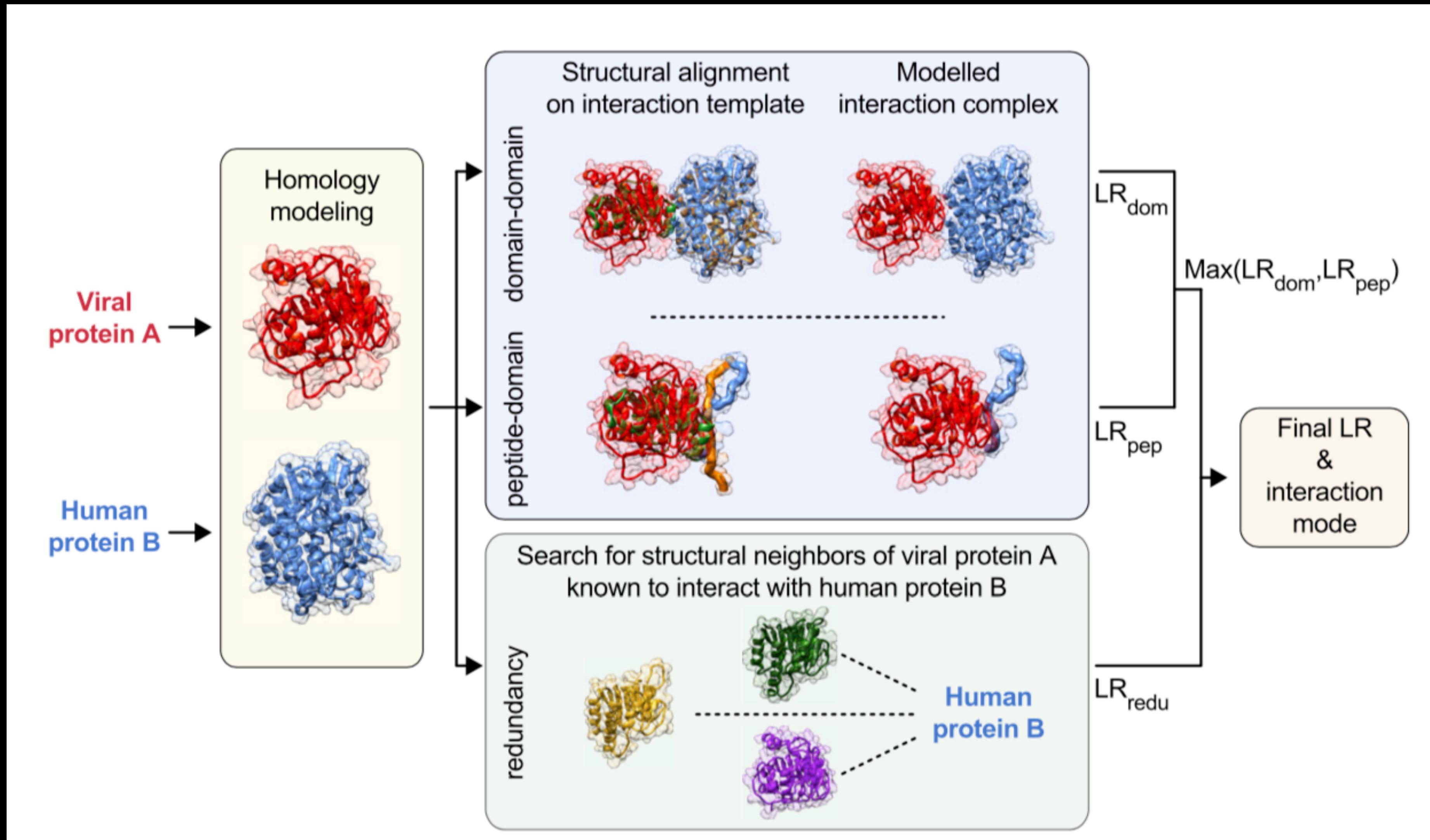
# Which virus strains affect humans?



# Which virus strains affect humans?

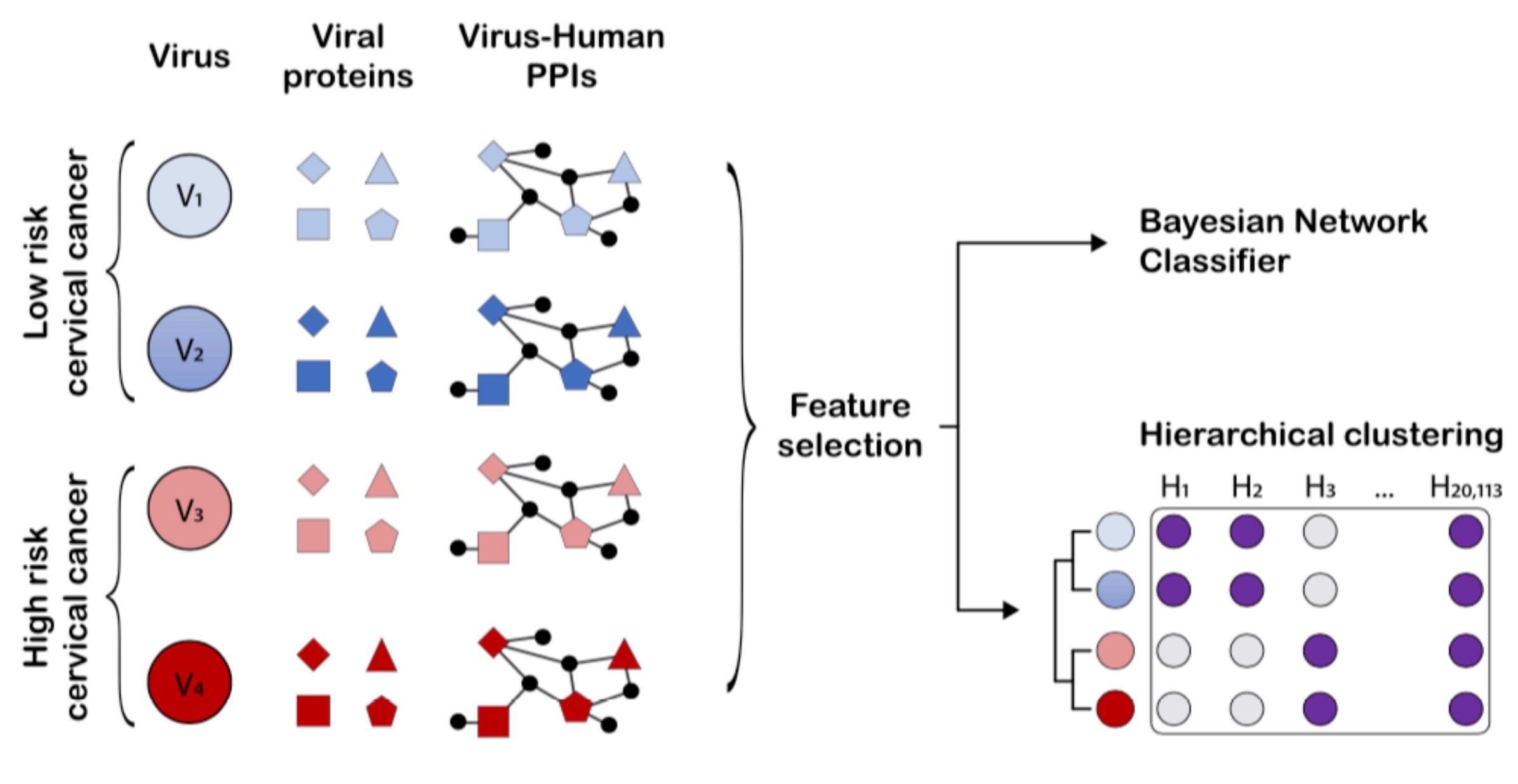


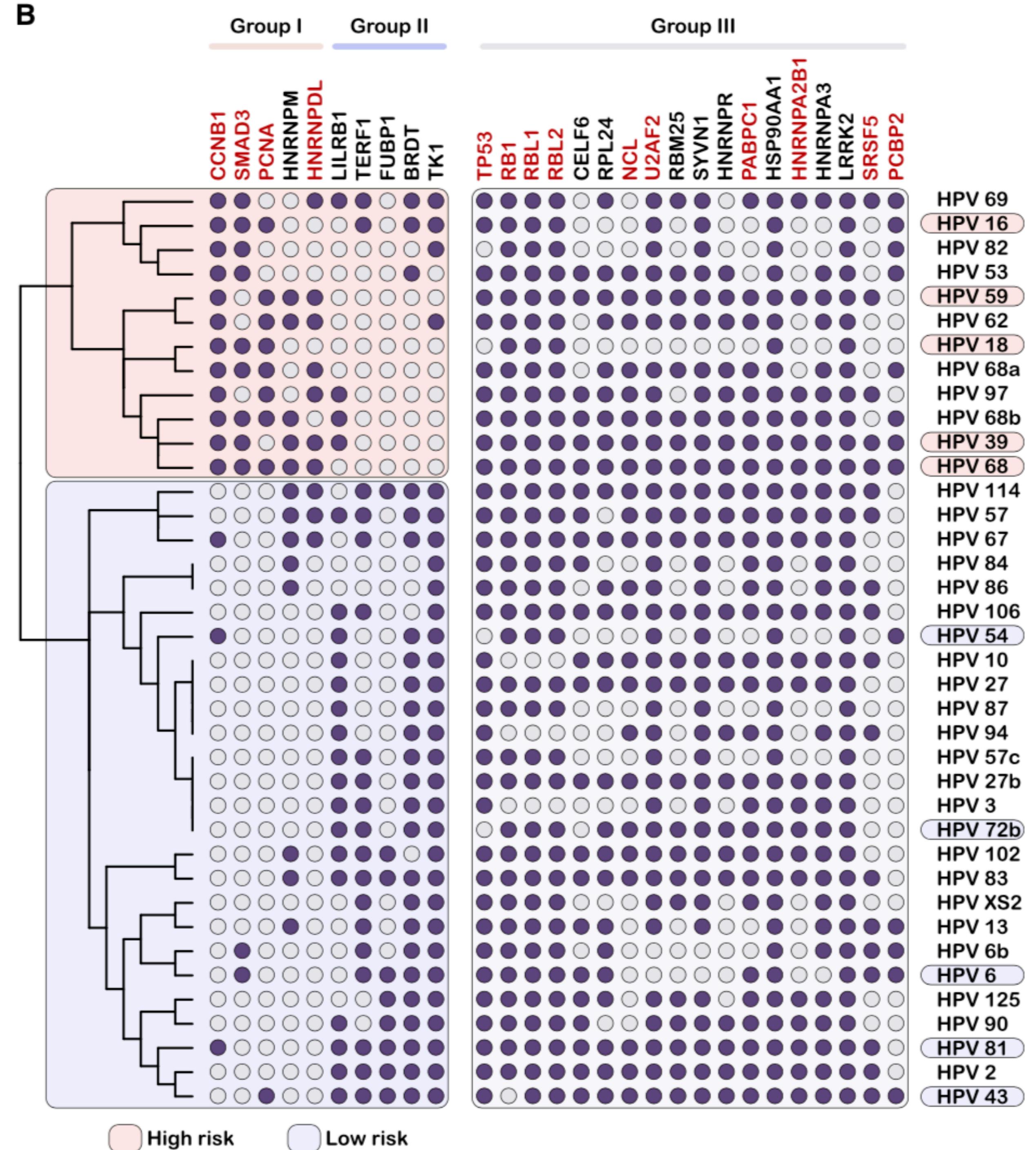
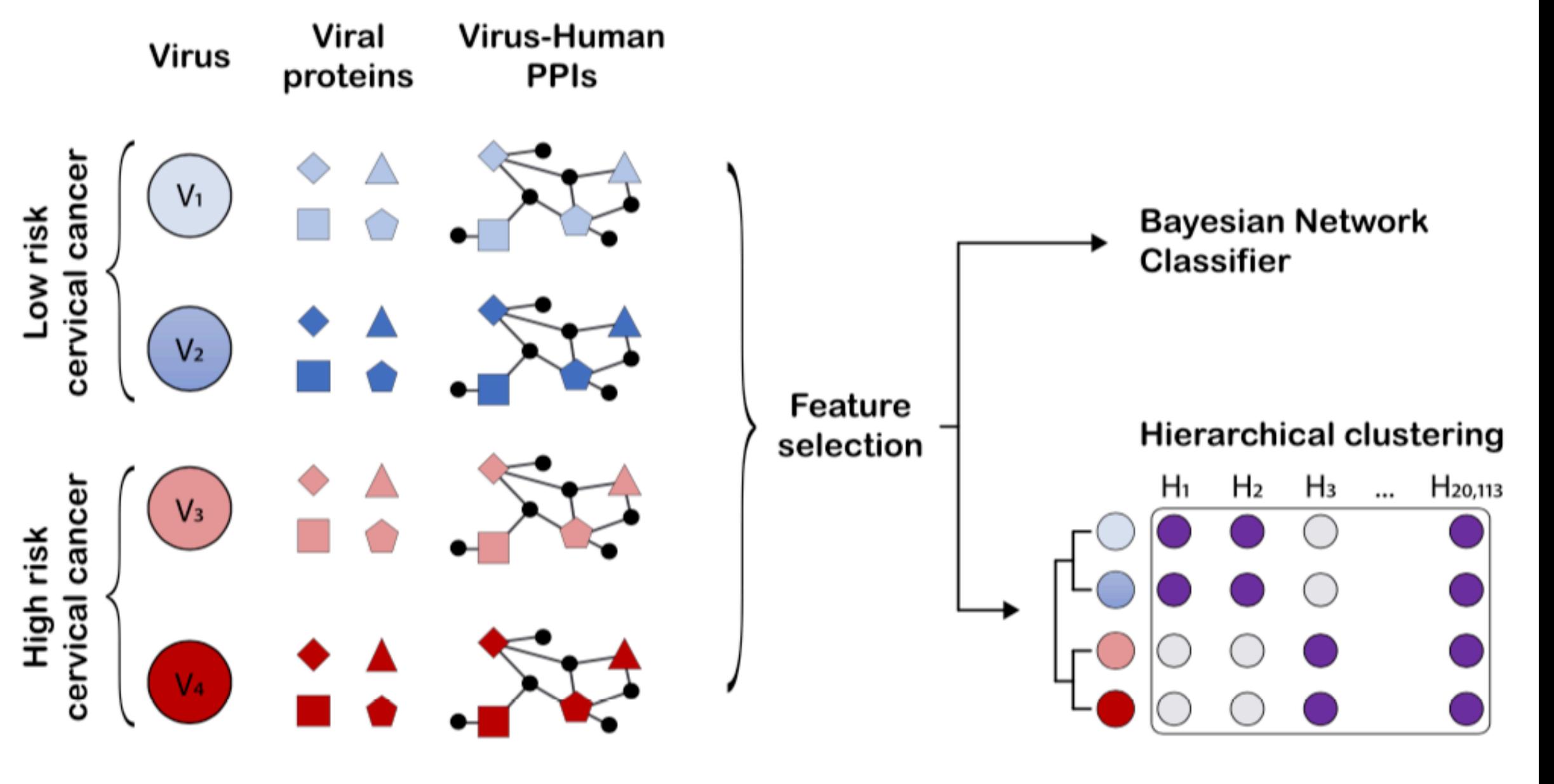
# Some '-omes' can be predicted sufficiently well.



Genome -> similar protein fold -> interaction

Lasso et al, 2019





# Research questions

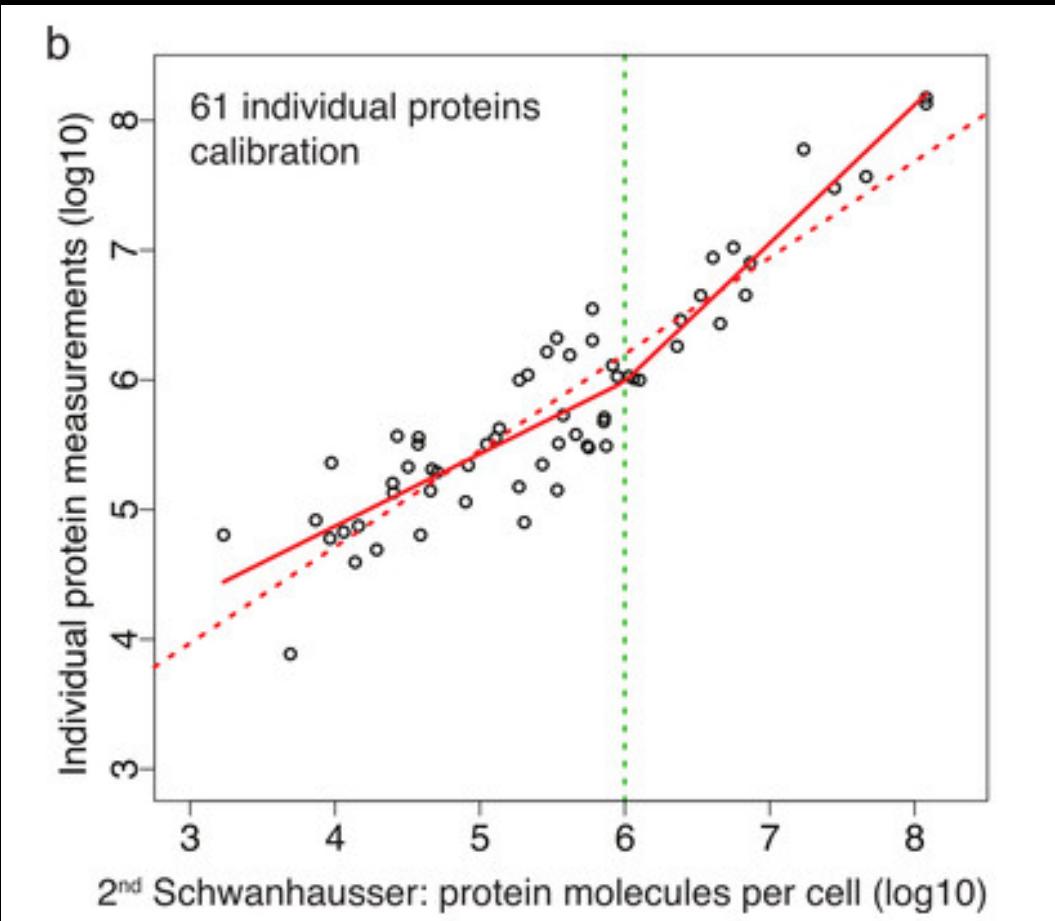
scales

Example 1: How much does one -ome determine another -ome?

- sequence features
- technical noise

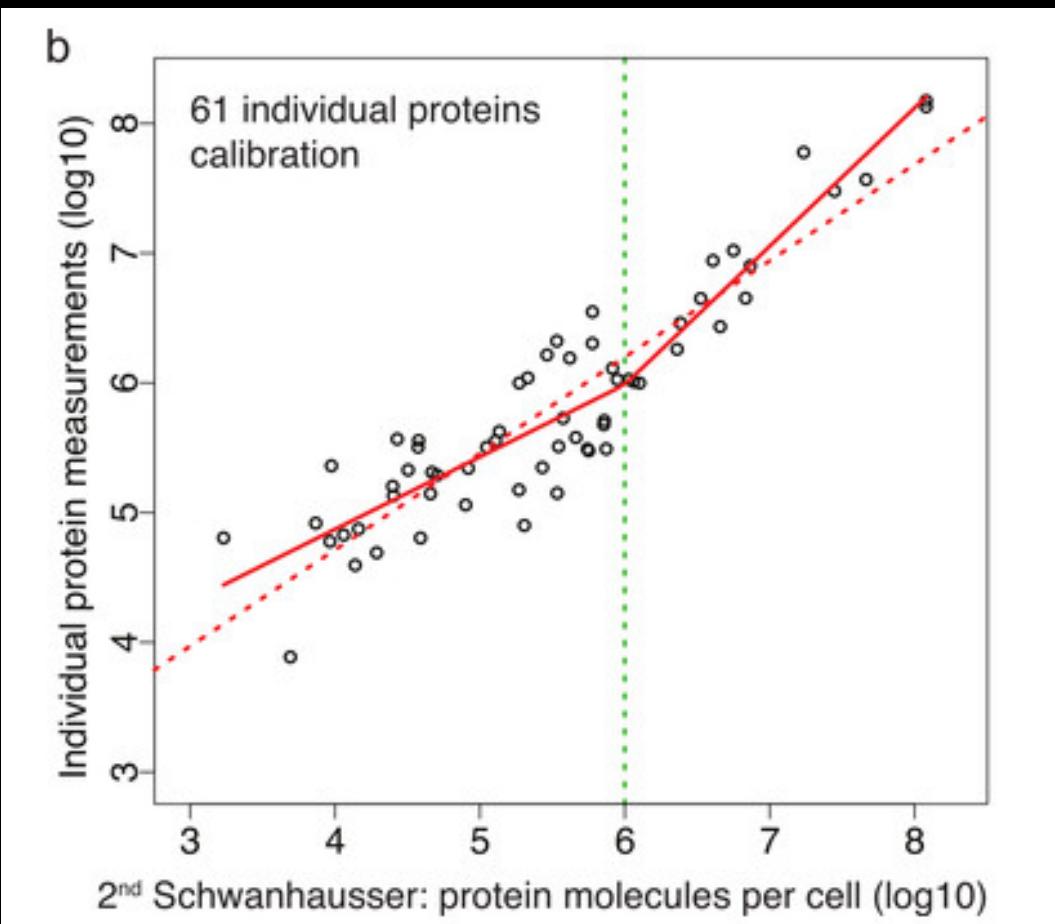
Example 2: Which papilloma virus strains are pathogenic?

## Relation between transcriptome and proteome.



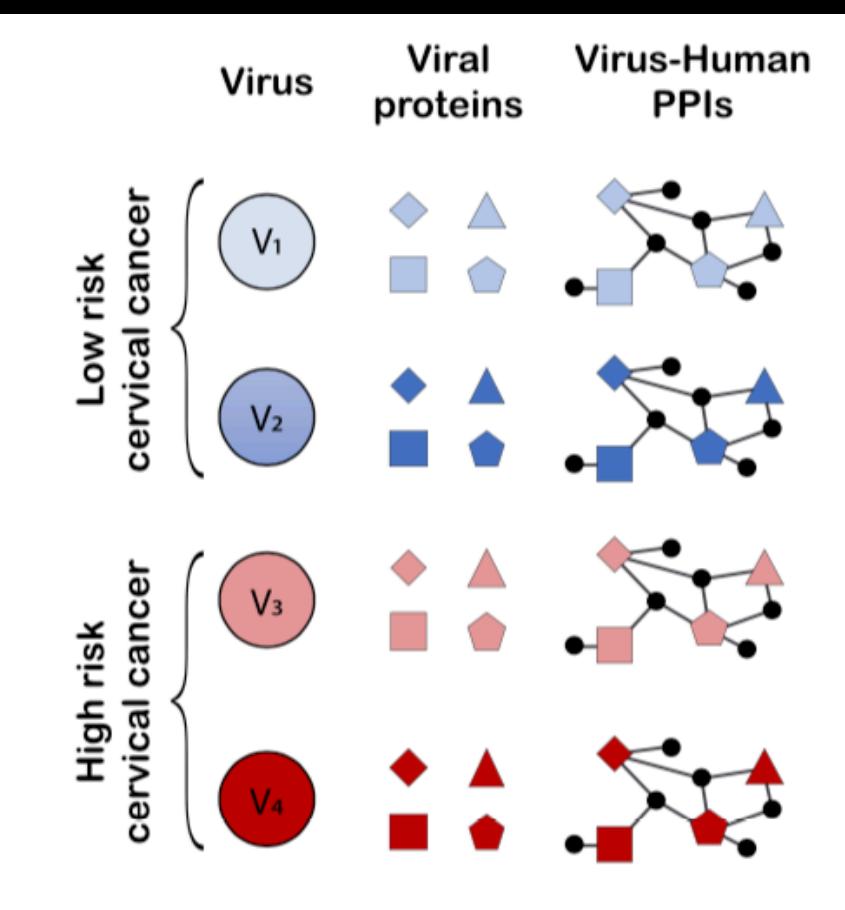
Vogel et al. 2010; Li et al. 2014f

## Relation between transcriptome and proteome.



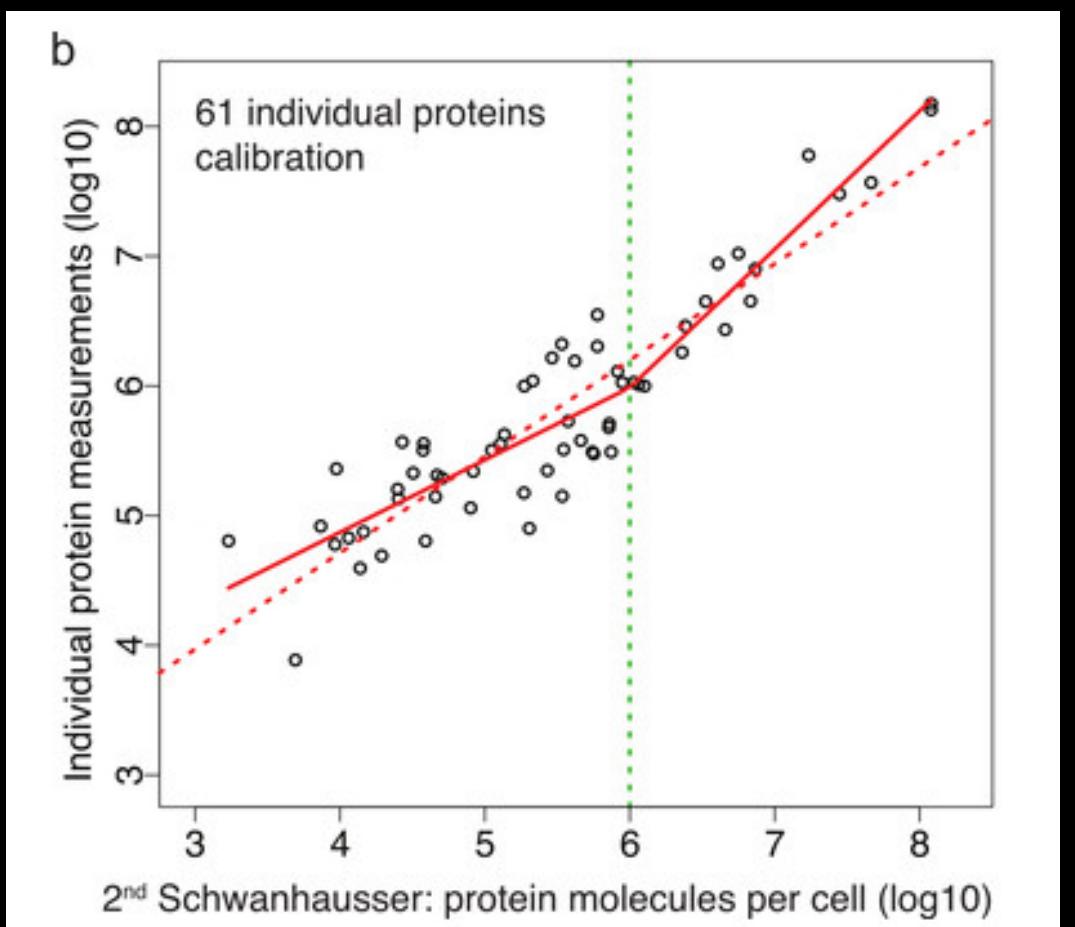
Vogel et al. 2010; Li et al. 2014f

## Disease-causing strains.



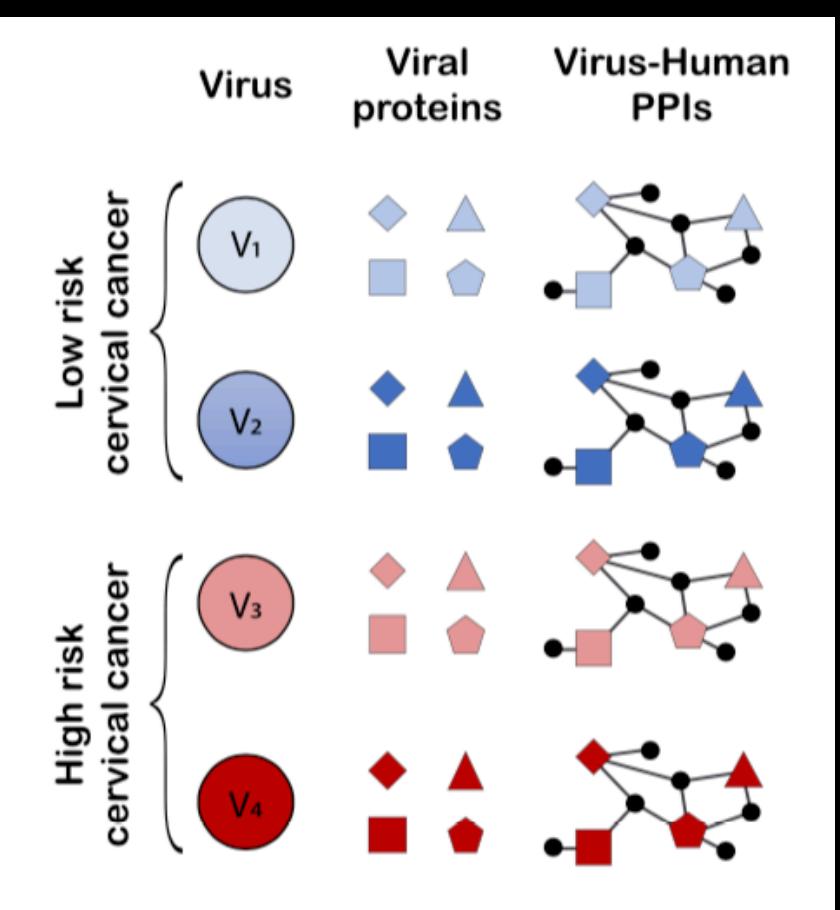
Lasso et al. 2019

## Relation between transcriptome and proteome.



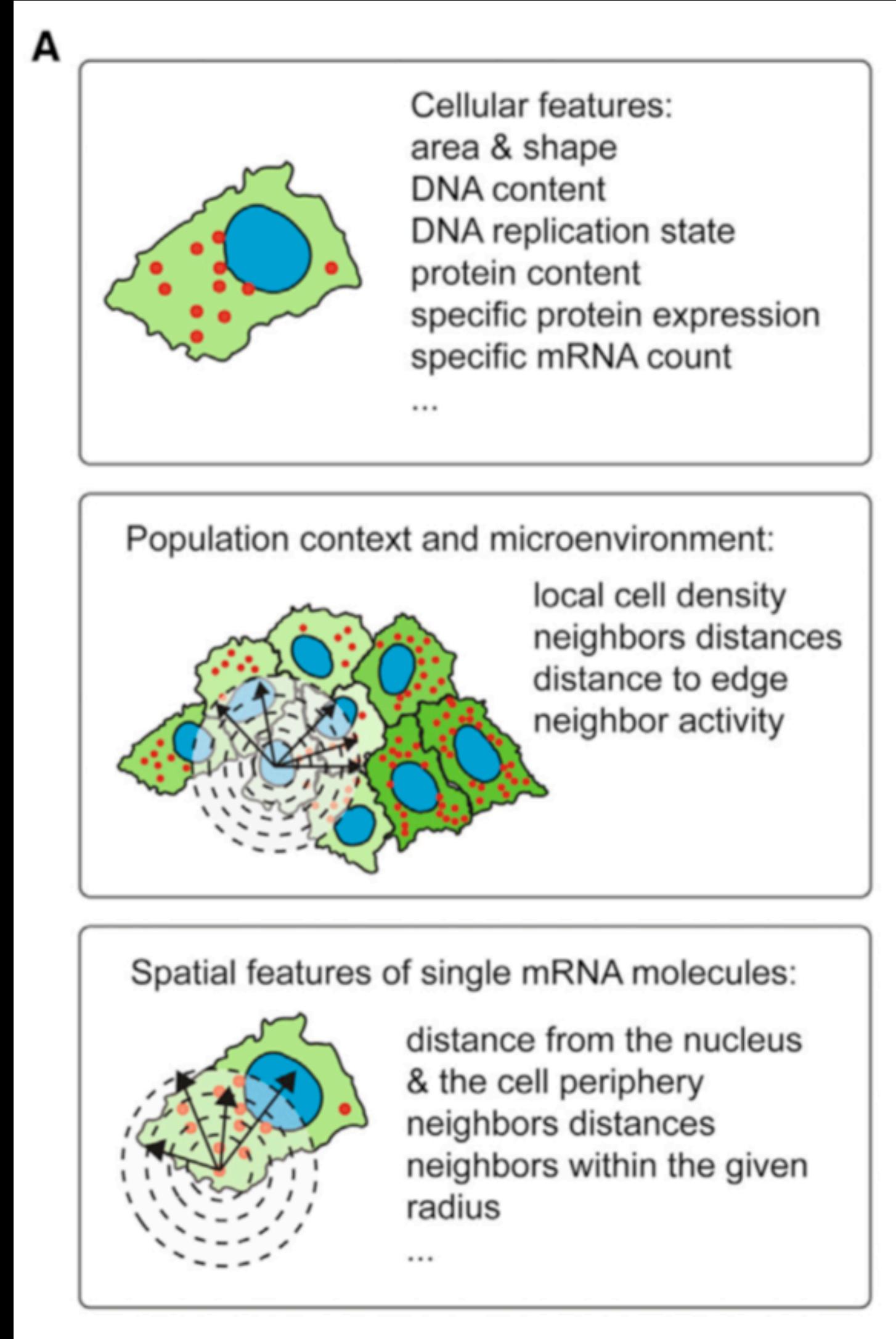
Vogel et al. 2010; Li et al. 2014f

## Disease-causing strains.



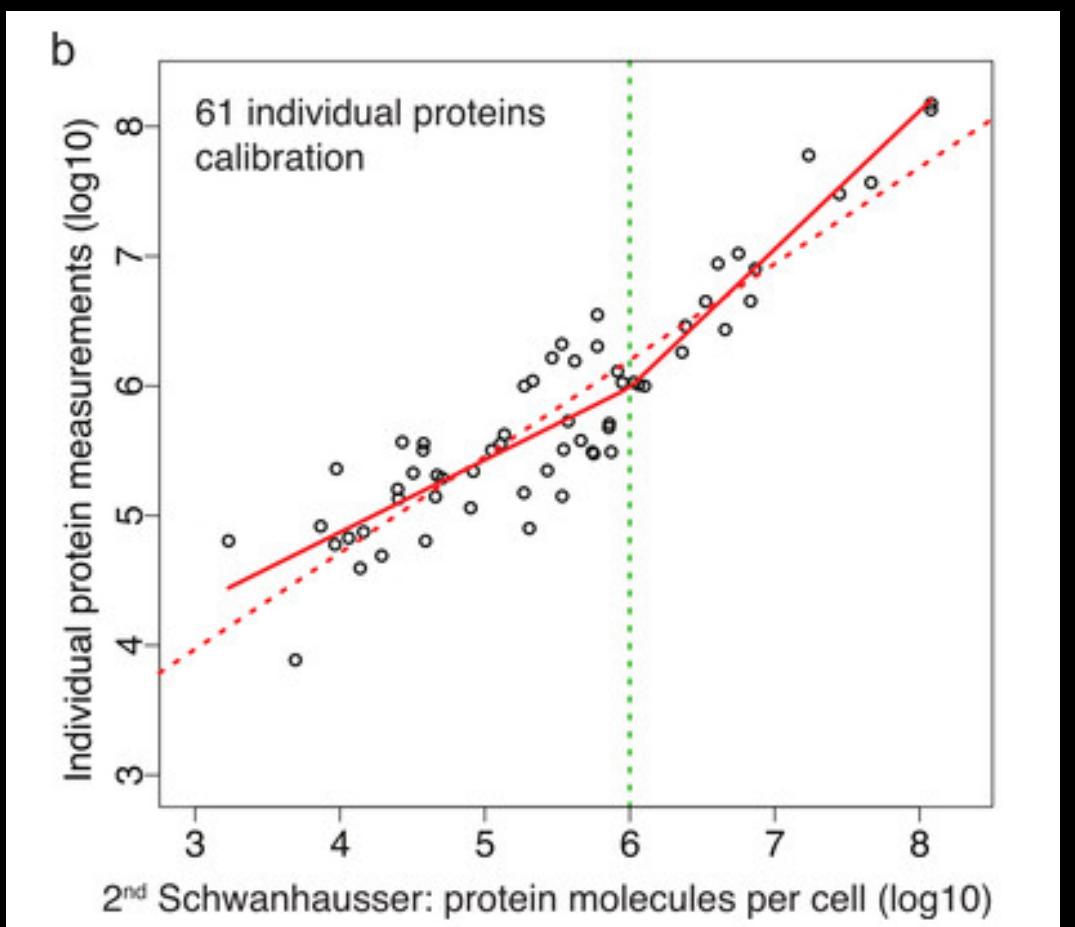
Lasso et al. 2019

## Adding spatial information of molecules and cells.



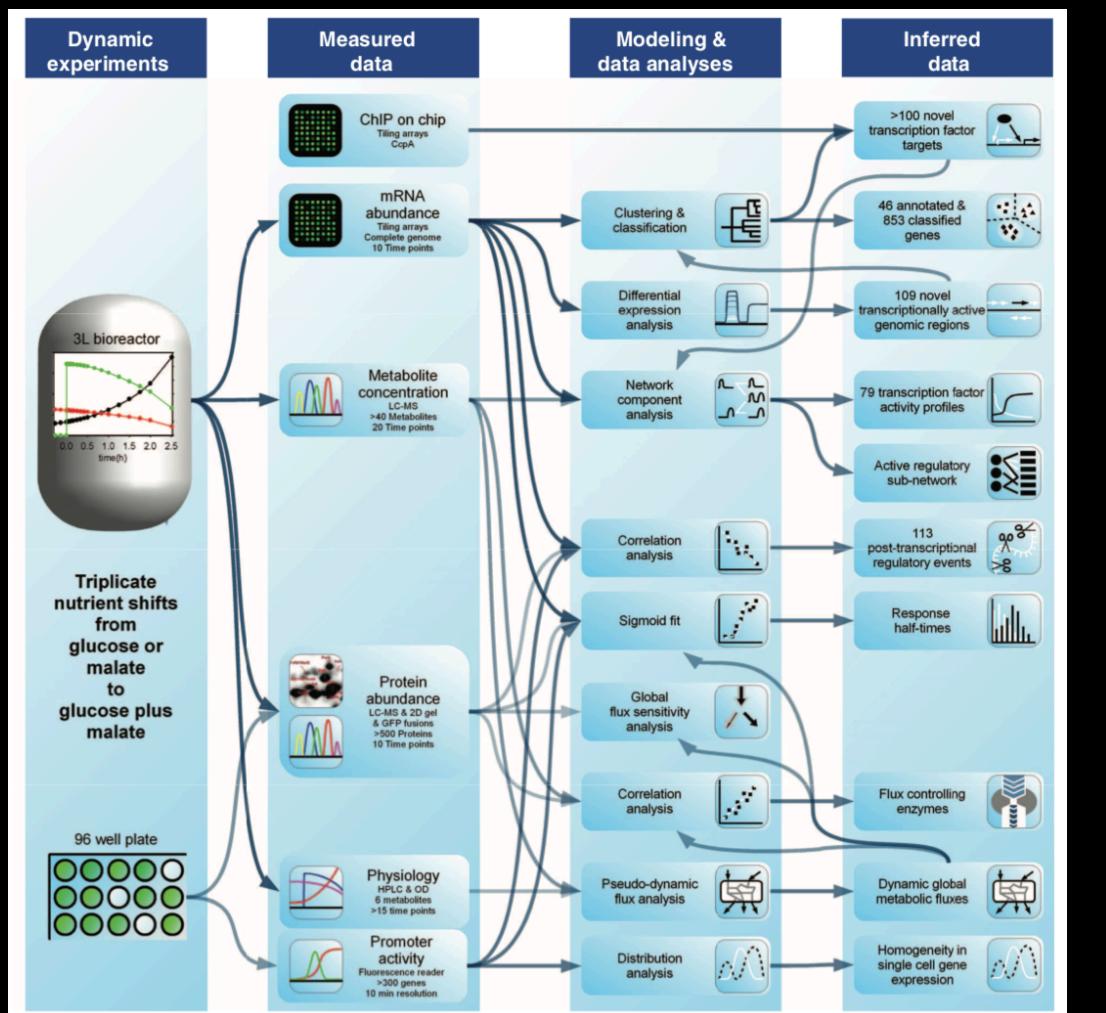
Popovich et al. 2019

## Relation between transcriptome and proteome.



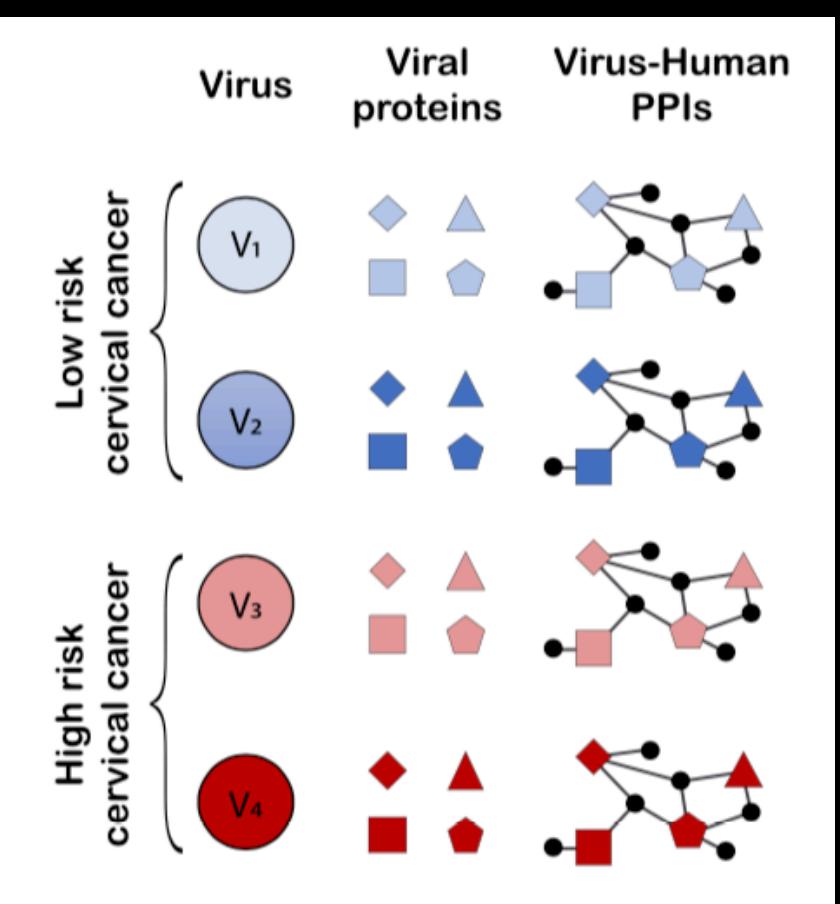
Vogel et al. 2010; Li et al. 2014f

## Transcriptomics, protein -modifications, fluxes



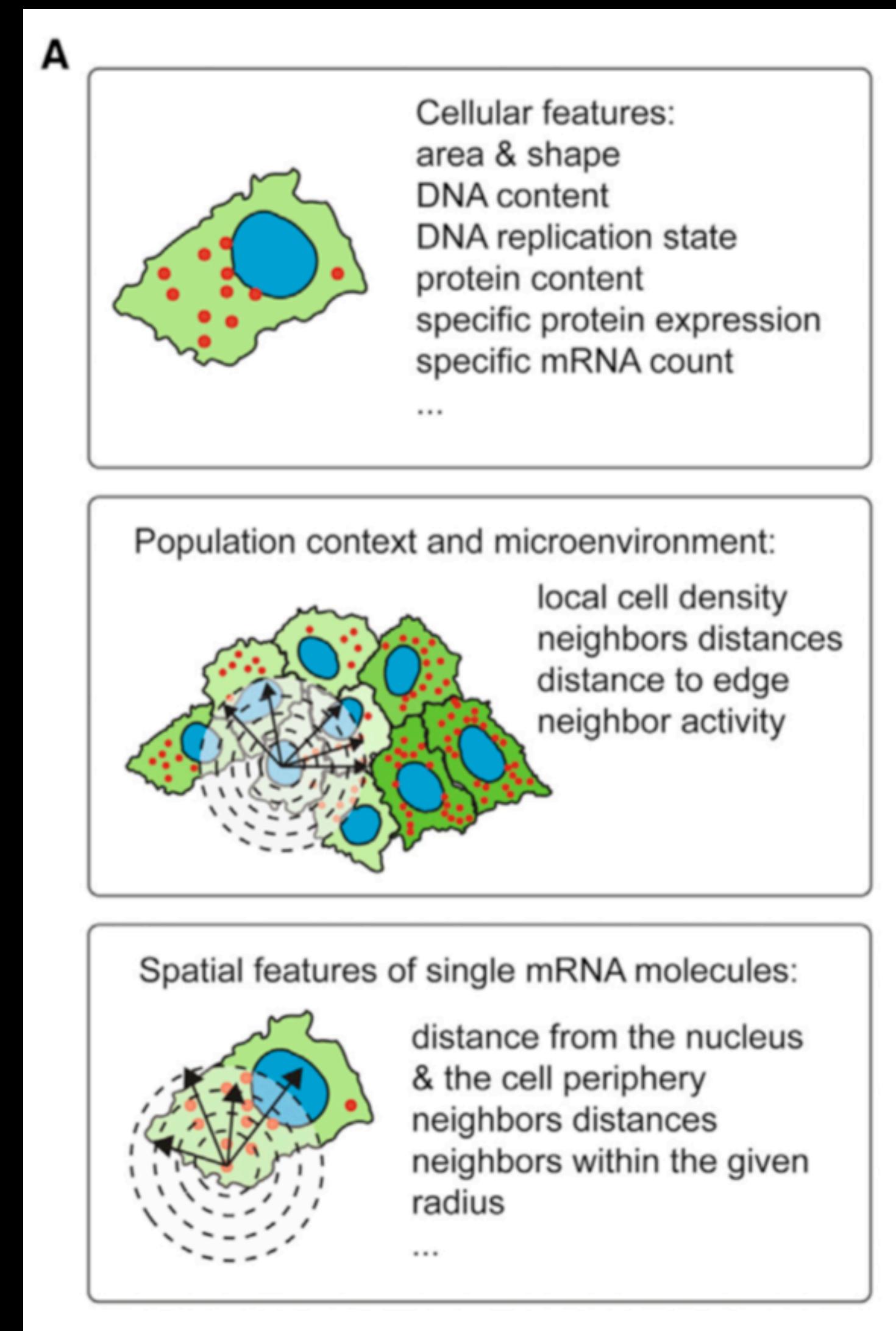
Buescher et al. 2012

## Disease-causing strains.



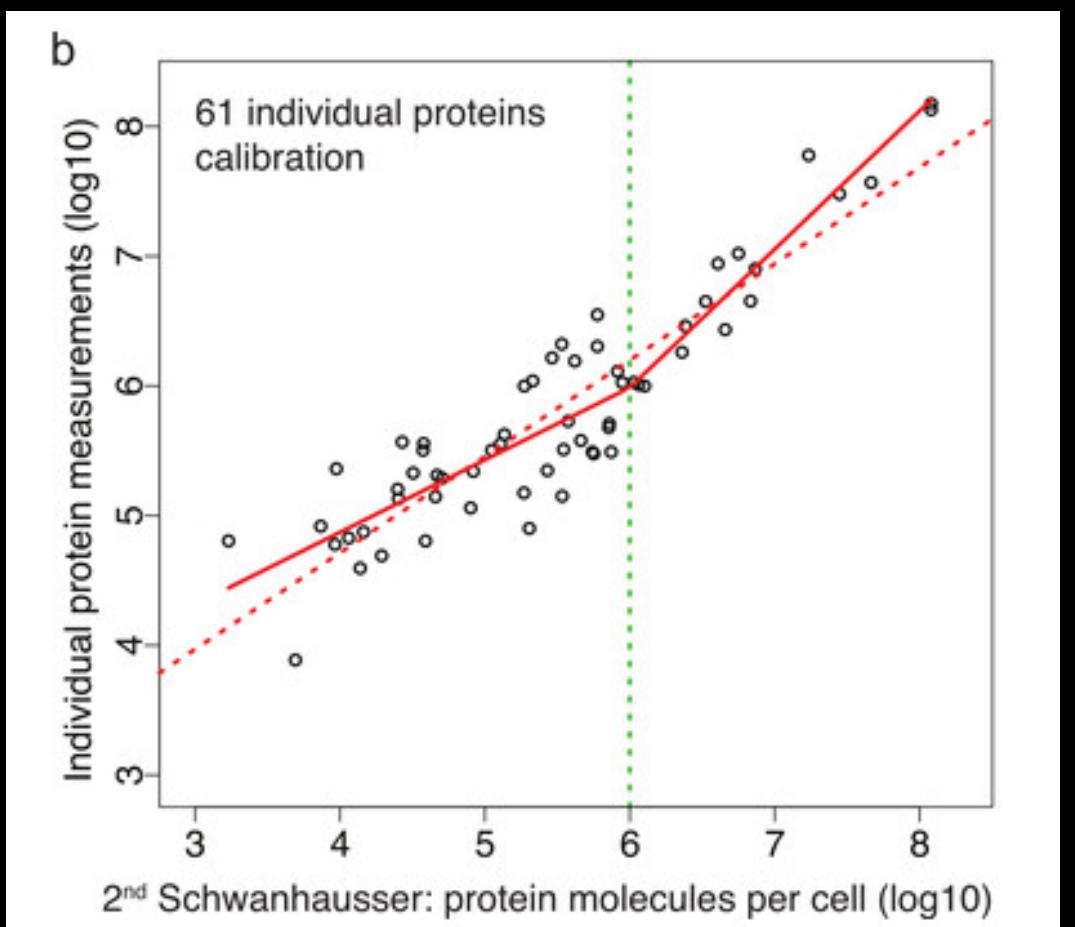
Lasso et al. 2019

## Adding spatial information of molecules and cells.



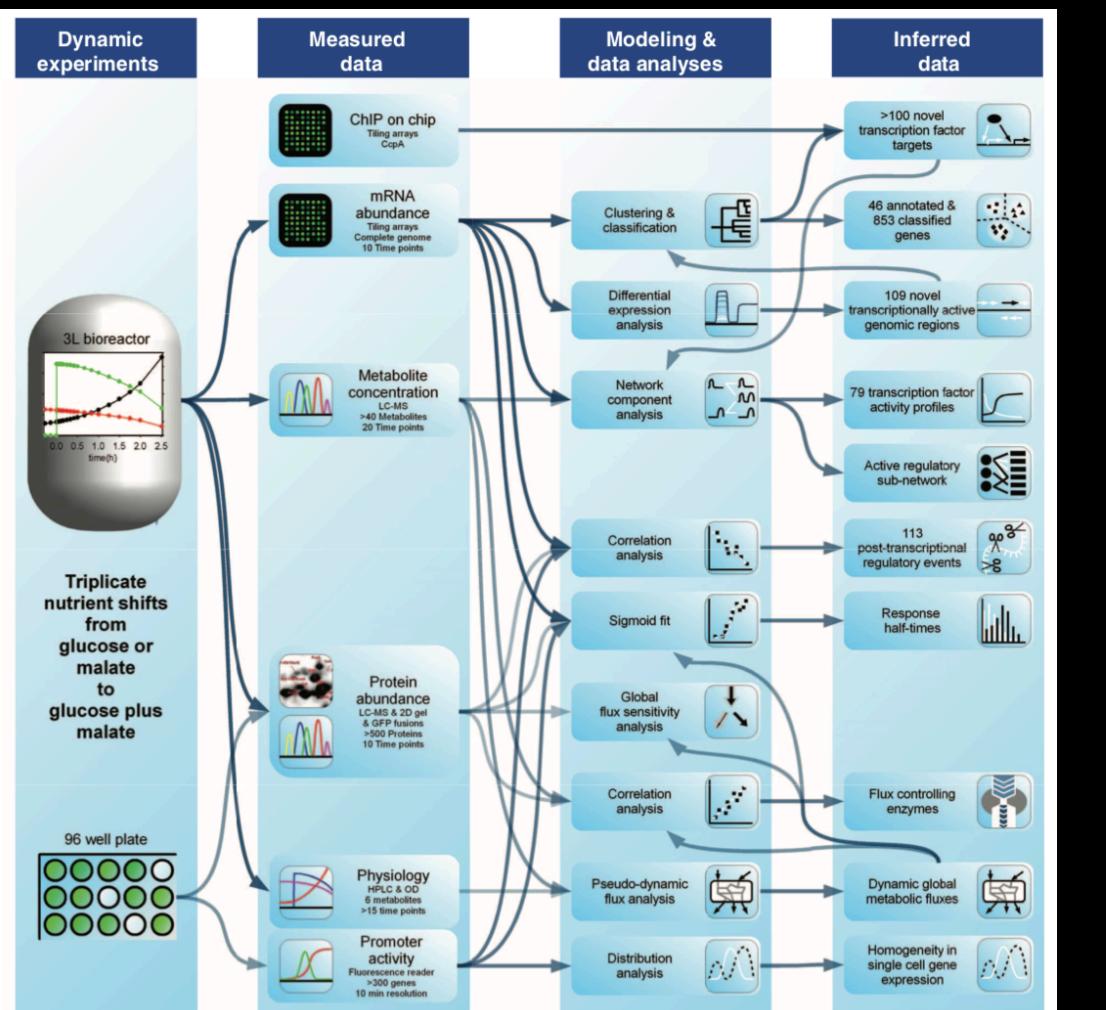
Popovich et al. 2019

## Relation between transcriptome and proteome.



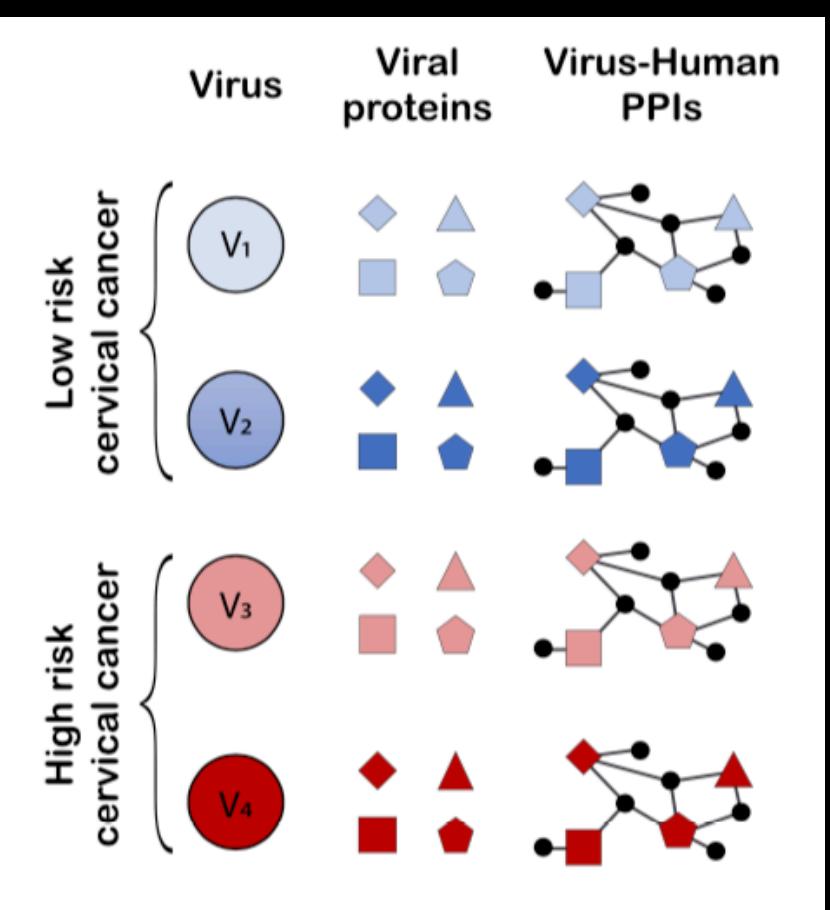
Vogel et al. 2010; Li et al. 2014f

## Transcriptomics, protein -modifications, fluxes



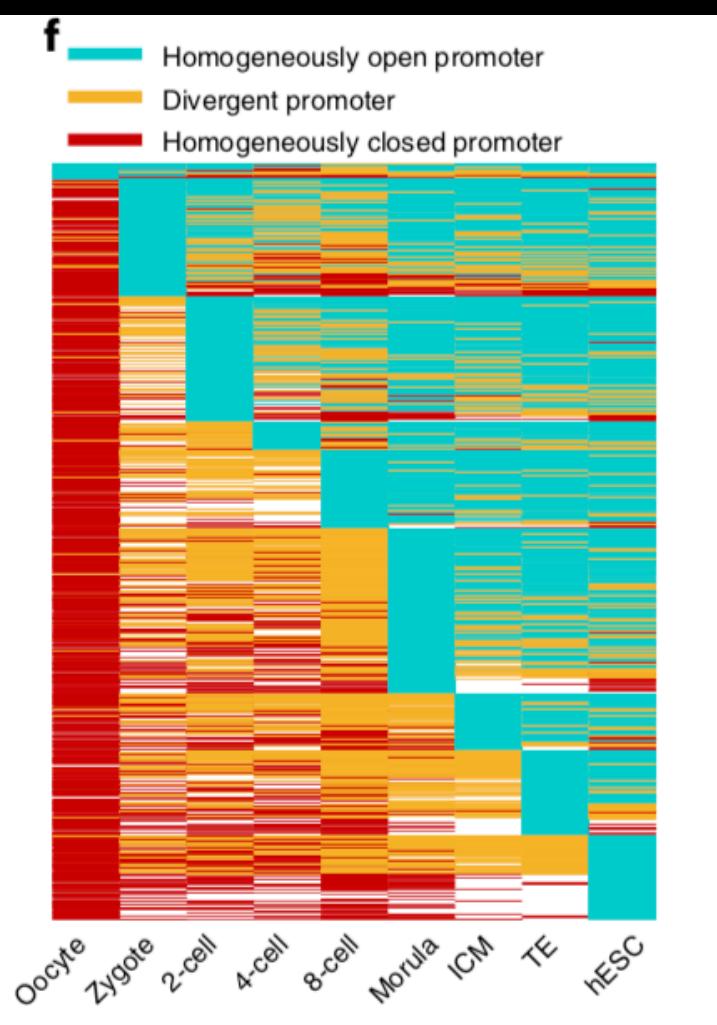
Buescher et al. 2012

## Disease-causing strains.



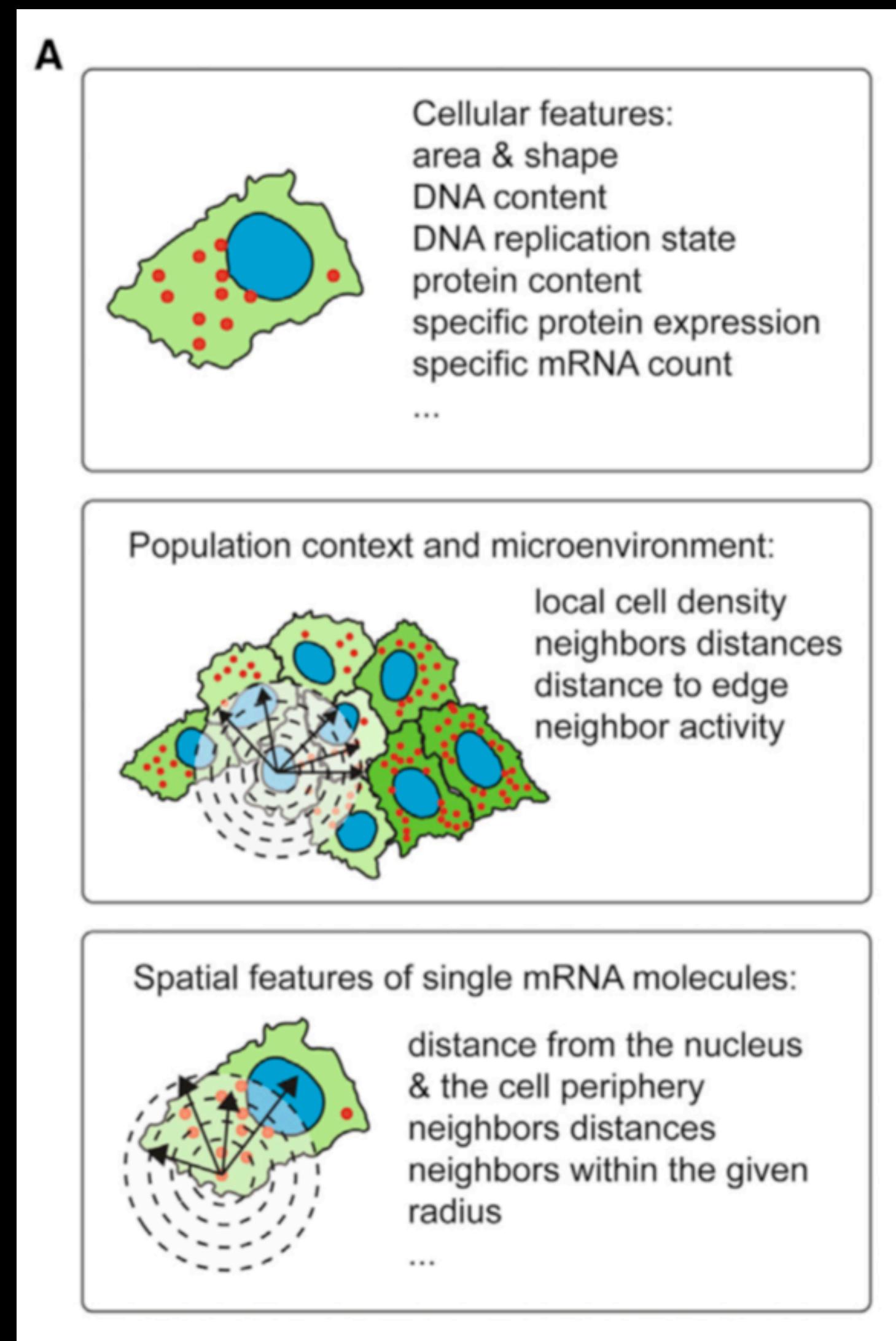
Lasso et al. 2019

## DNA methylation and nucleosomes in single cells.



Li et al. 2018

## Adding spatial information of molecules and cells.



Popovich et al. 2019

# Research questions

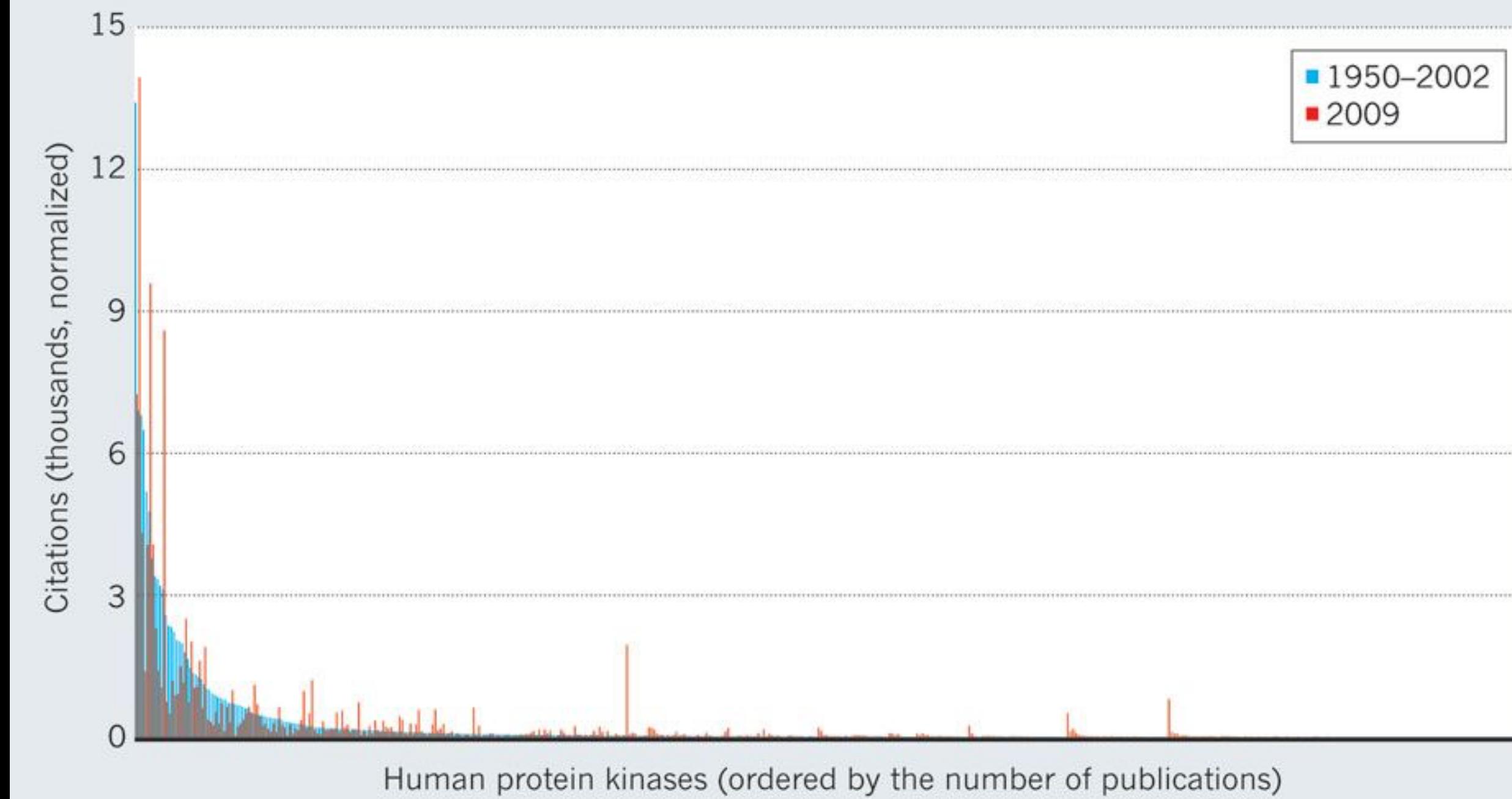
distorted literature

# Research questions

distorted literature

## FONDLING OUR PROBLEMS

Researchers' 'favourite kinases' have remained the same for decades with a few exceptions (kinases linked to diseases of great interest to industry).



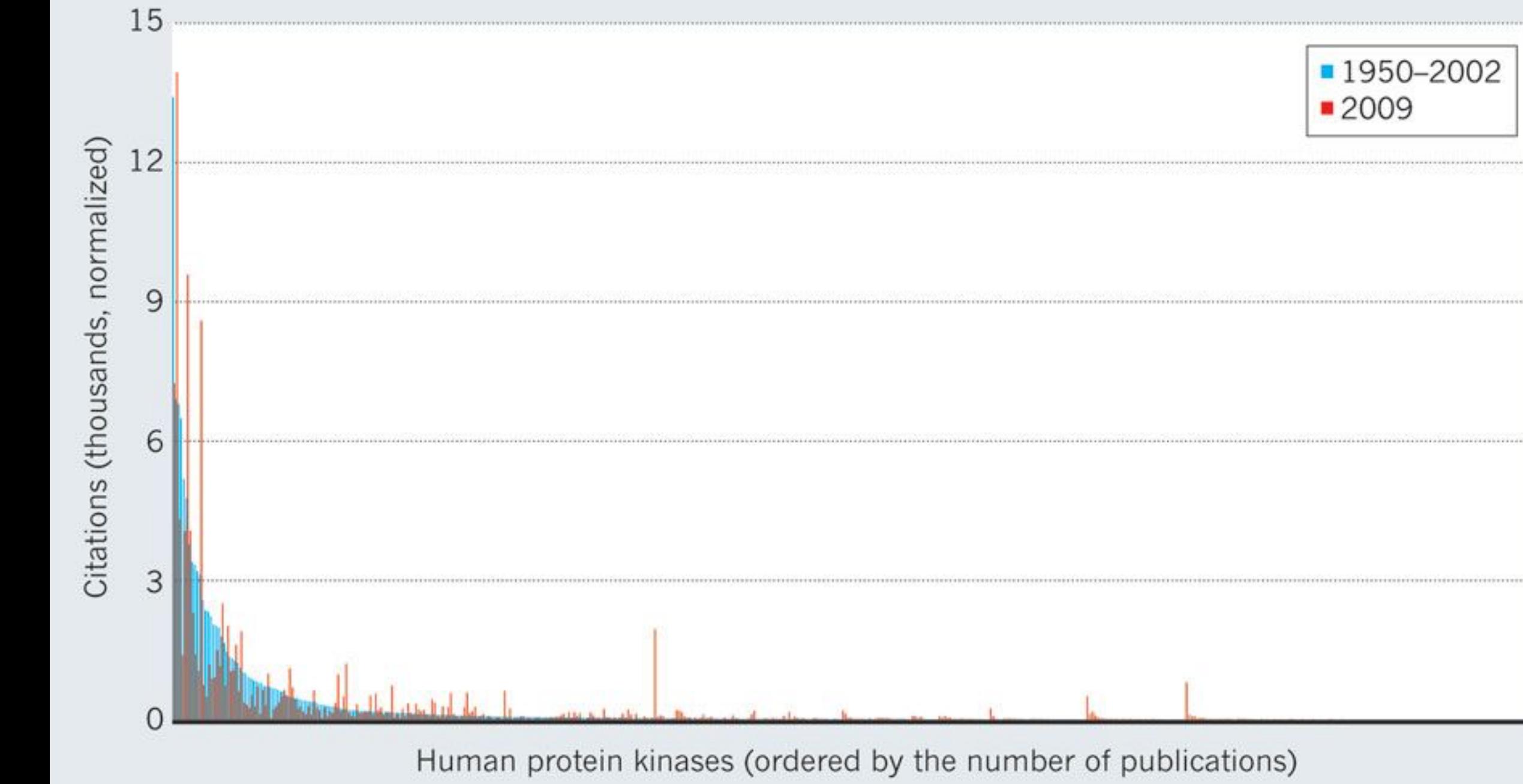
Edwards et al. 2012

# Research questions

distorted literature

## FONDLING OUR PROBLEMS

Researchers' 'favourite kinases' have remained the same for decades with a few exceptions (kinases linked to diseases of great interest to industry).



Edwards et al. 2012

Example: Can the success of drug trials be predicted, if ignoring small-scale studies?

# Some DNA polymorphisms associate with disease.

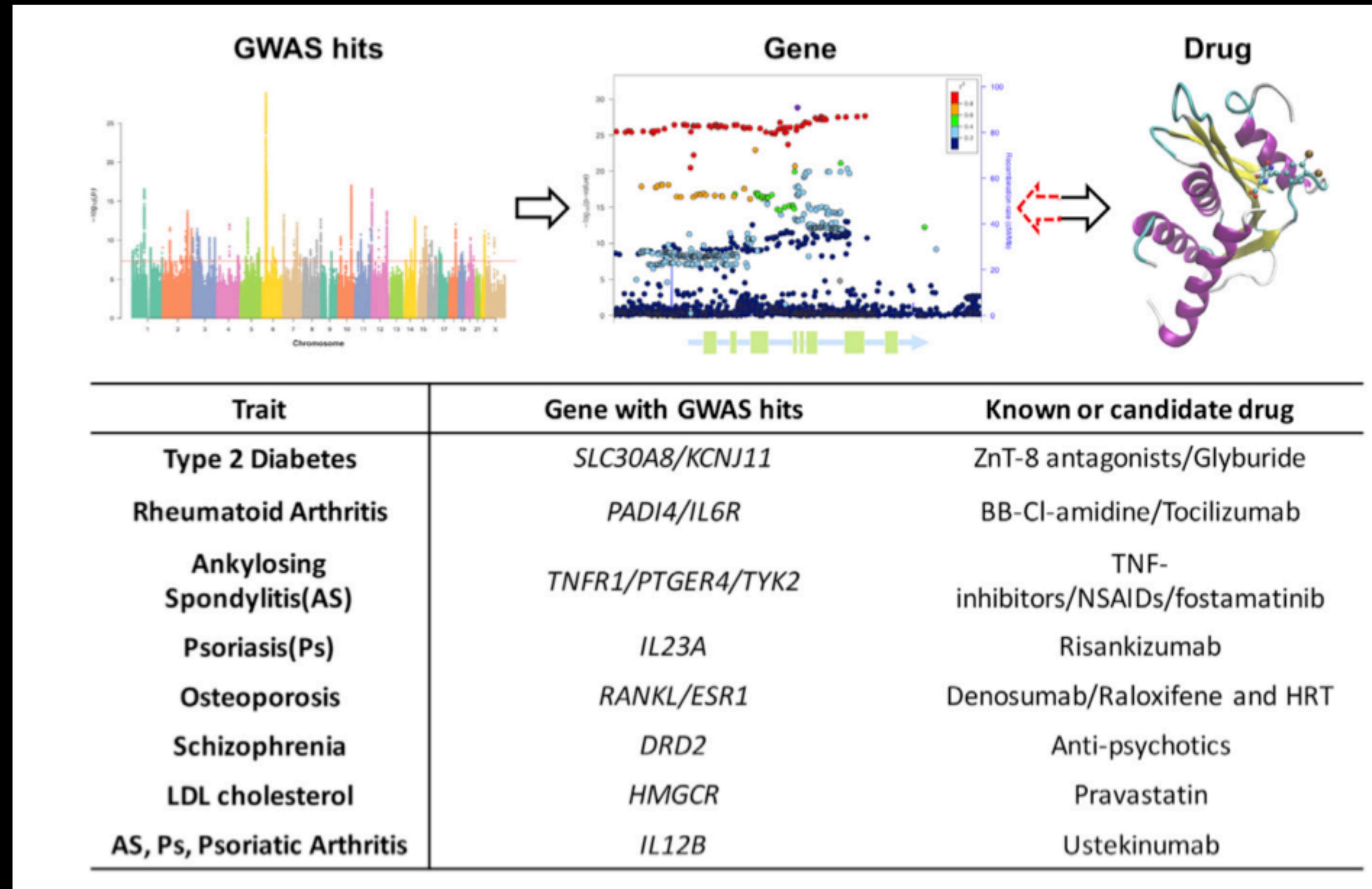
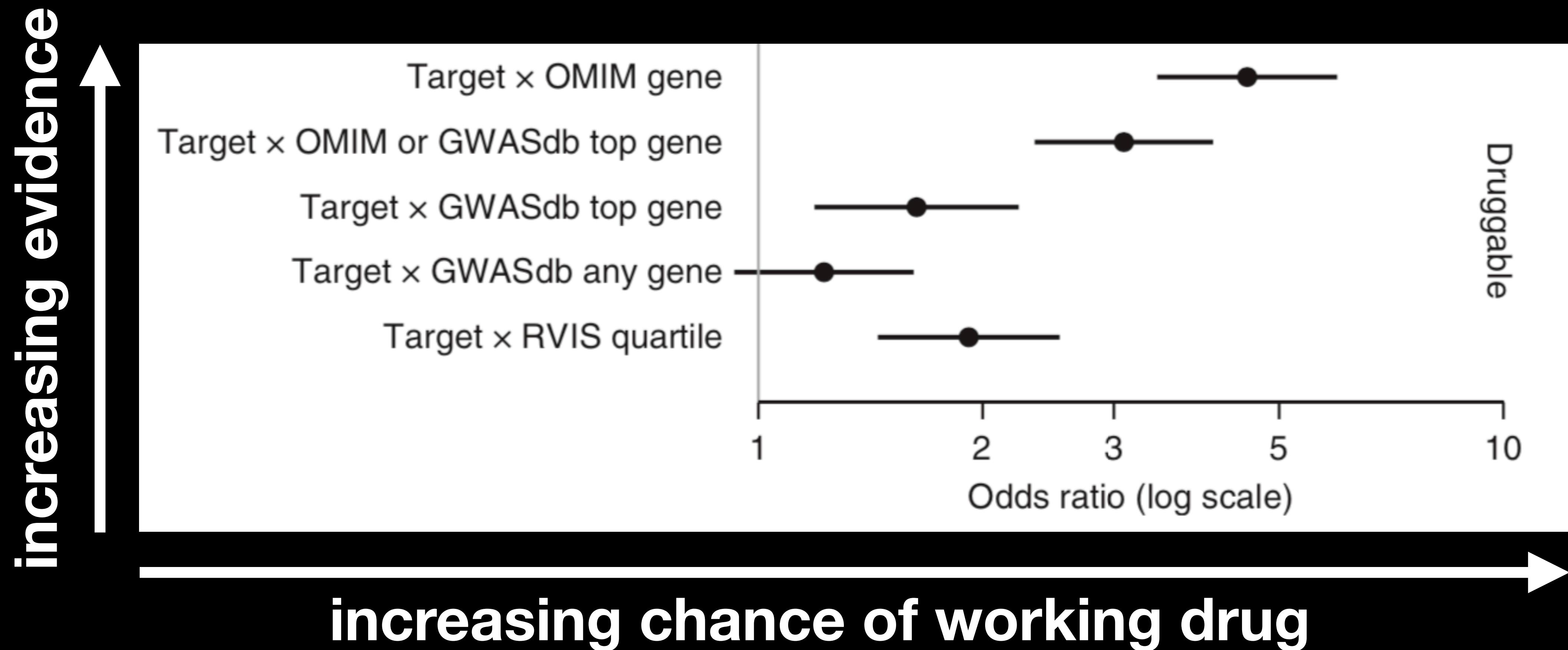
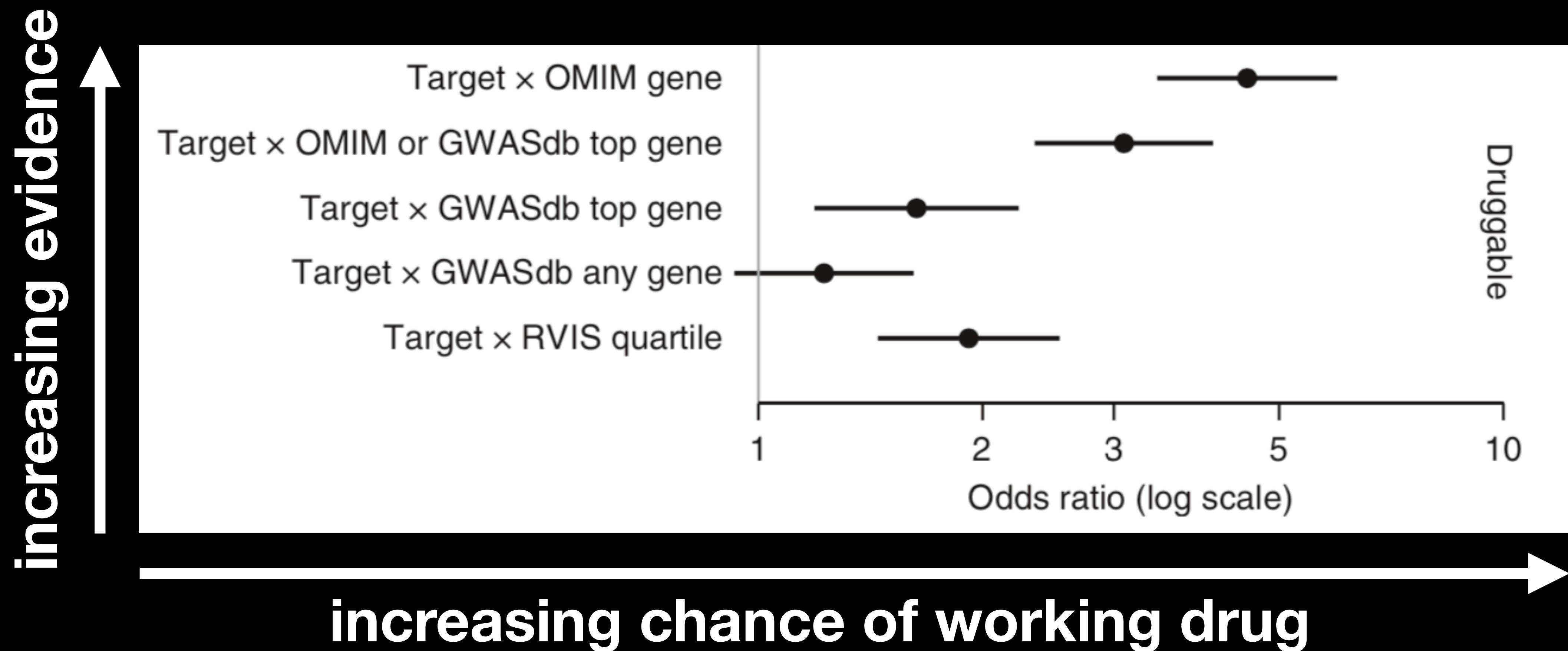


Figure 3. Examples of Links between GWAS Discoveries and Drugs

If there is evidence that the targeted gene is involved in the targeted disease, the drug is more likely to work.



If there is evidence that the targeted gene is involved in the targeted disease, the drug is more likely to work.



## Step 1: Authors excluded low-quality literature.

We excluded all data from PharmGKB and the Genetic Association Database. Genetic associations reported from these two sources contained no supporting statistical association evidence (with most  $P$  values equal to zero) to accompany the entries, and the new associations included were largely drawn from candidate gene association studies that lacked rigorous criteria for reporting a statistical association. In particular, we found that there were a large number of candidate gene associations in PharmGKB for drug target genes, which would result in an upward bias in the number of drug targets with supposed genetic associations.

## Step 1: Authors excluded low-quality literature.

We excluded all data from PharmGKB and the Genetic Association Database. Genetic associations reported from these two sources contained no supporting statistical association evidence (with most  $P$  values equal to zero) to accompany the entries, and the new associations included were largely drawn from candidate gene association studies that lacked rigorous criteria for reporting a statistical association. In particular, we found that there were a large number of candidate gene associations in PharmGKB for drug target genes, which would result in an upward bias in the number of drug targets with supposed genetic associations.

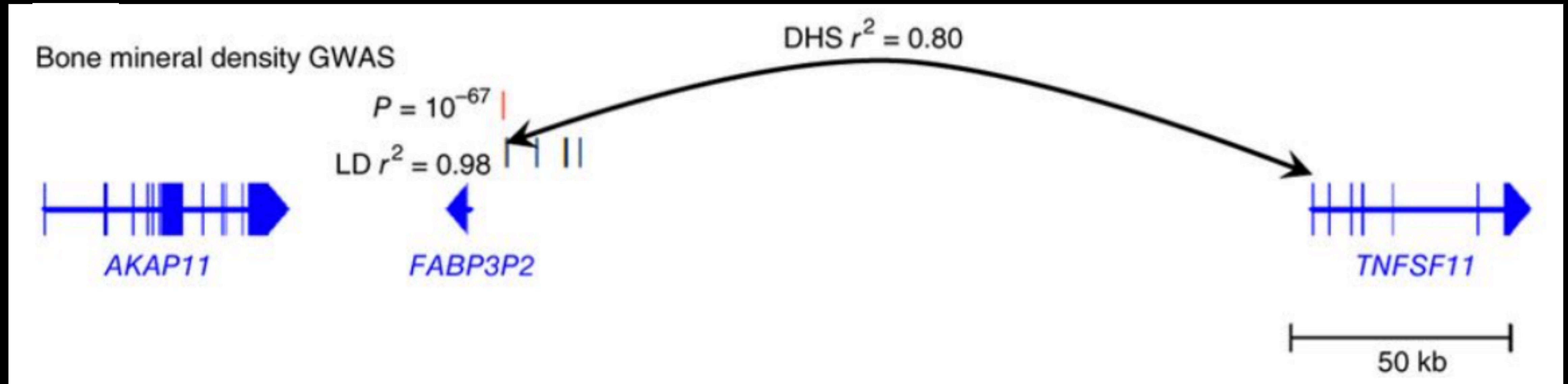
We also excluded a few large metabolomic studies with numerous traits screened that had very large numbers of associations reported. Finally, we identified one study<sup>15</sup> where a supplementary table was misinterpreted, leading to many falsely identified associations that were also excluded.

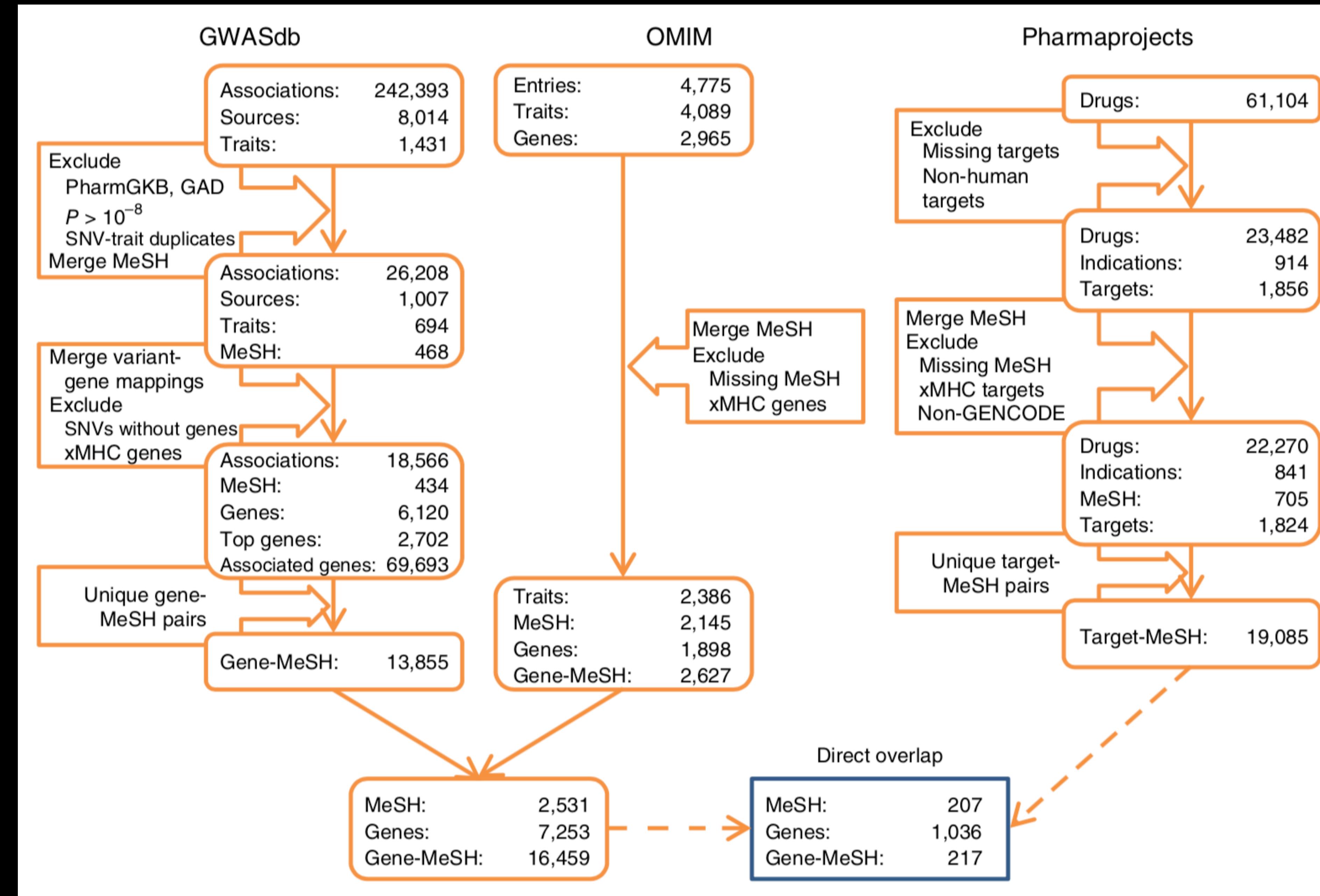
## **Step 2: Authors curate words used to describe phenotypes.**

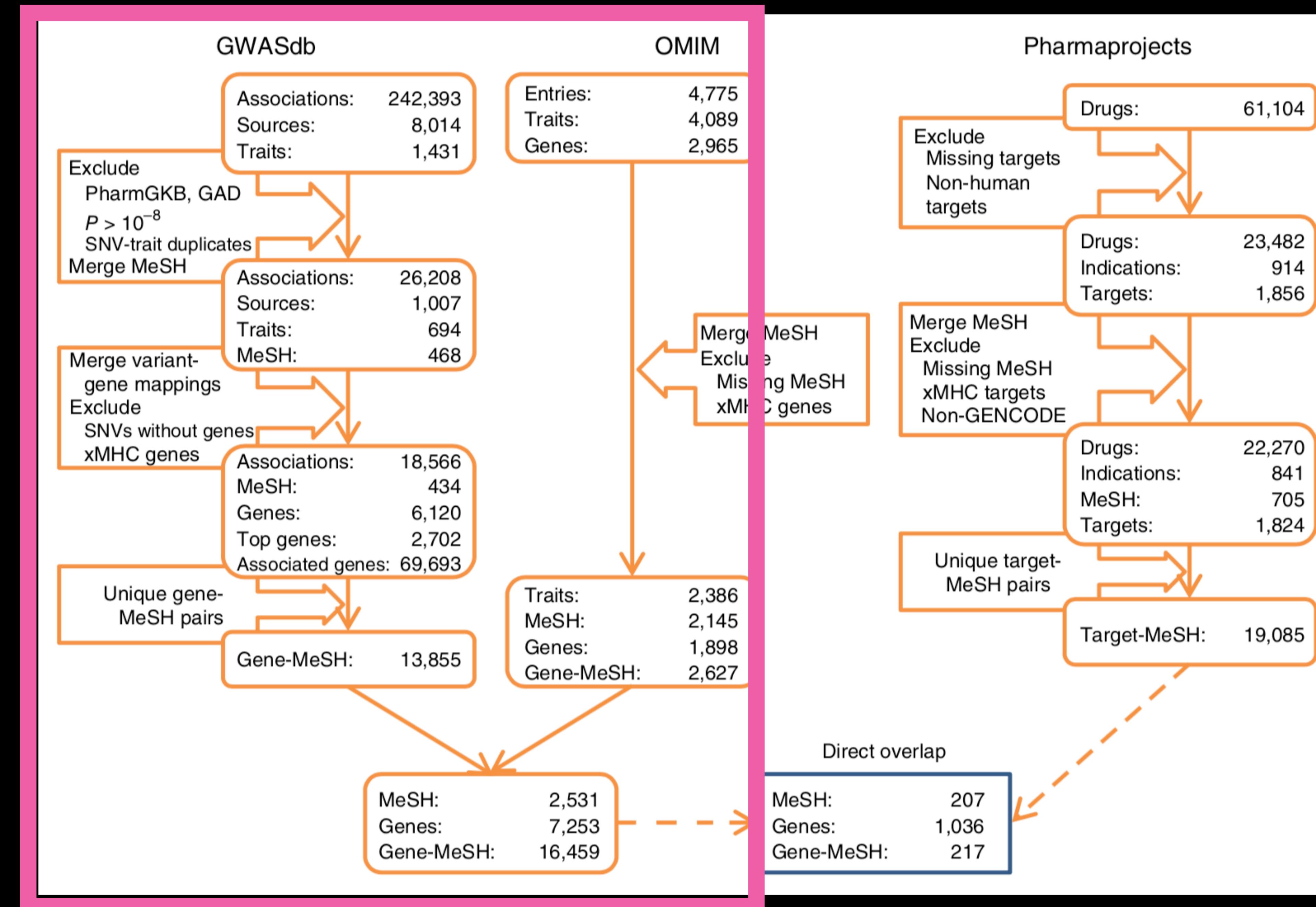
## Step 2: Authors curate words used to describe phenotypes.

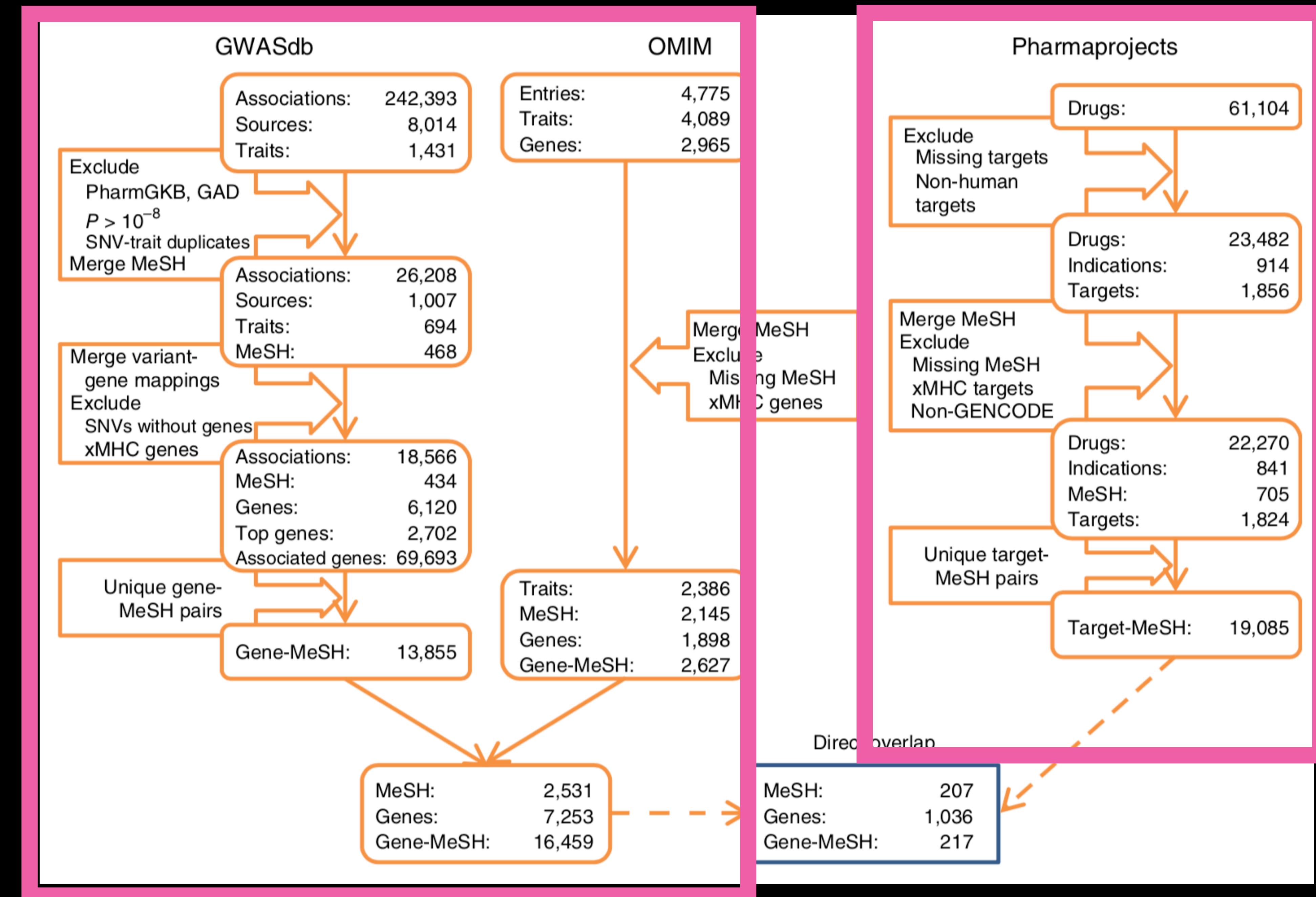
To allow comparisons among all data sources, we manually mapped all traits to the most specific Medical Subject Heading (MeSH) terms applicable.

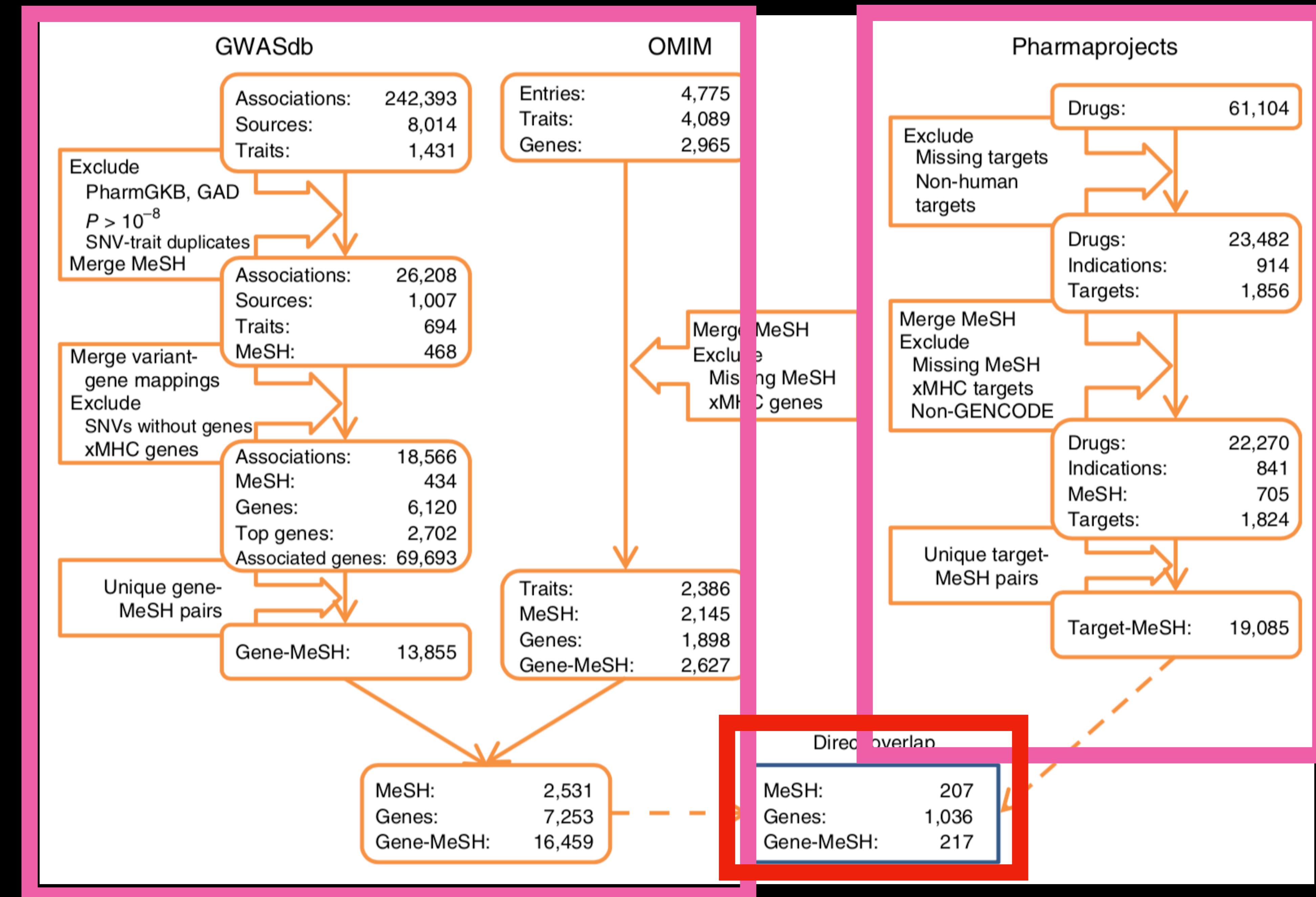
# Step 3: Authors test if polymorphisms affect genes.









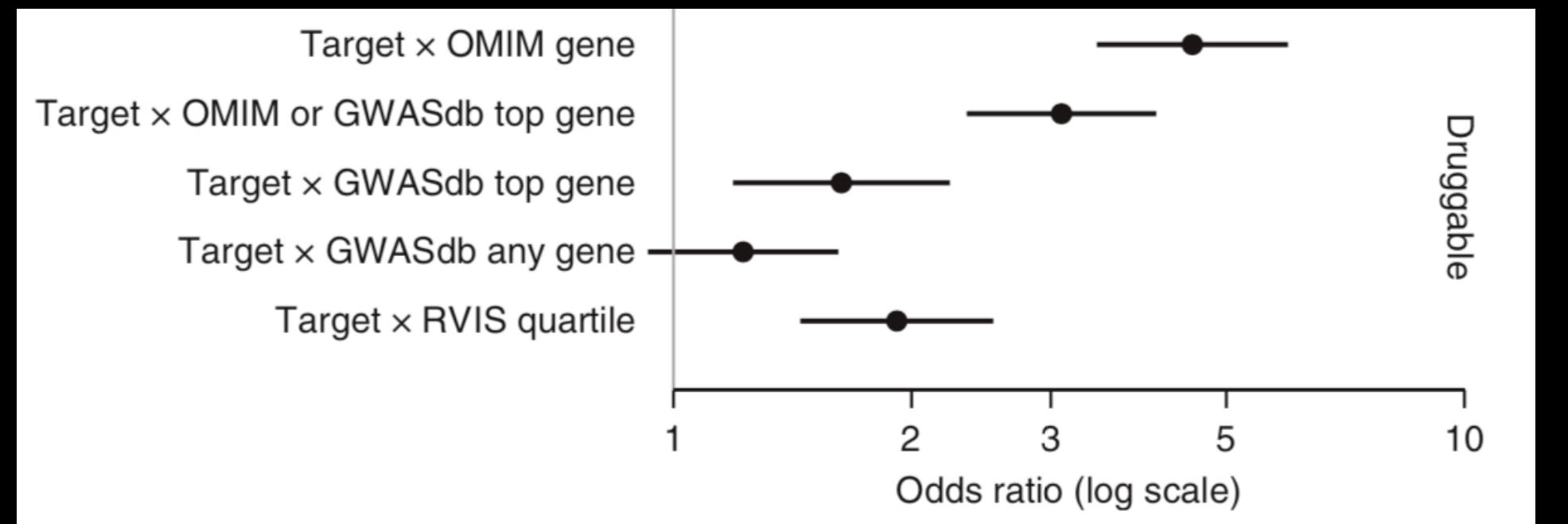


# Research questions

distorted literature

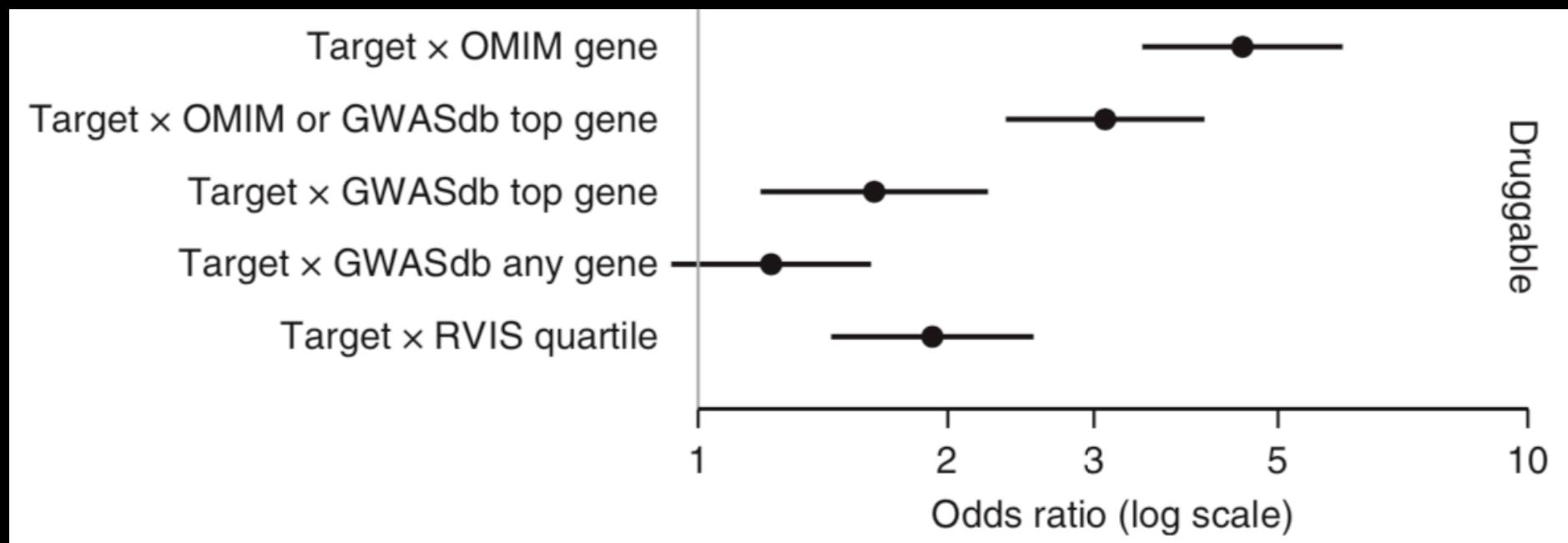
Example: Can the success of drug trials be predicted, if ignoring small-scale studies?

# Support for drug targets.



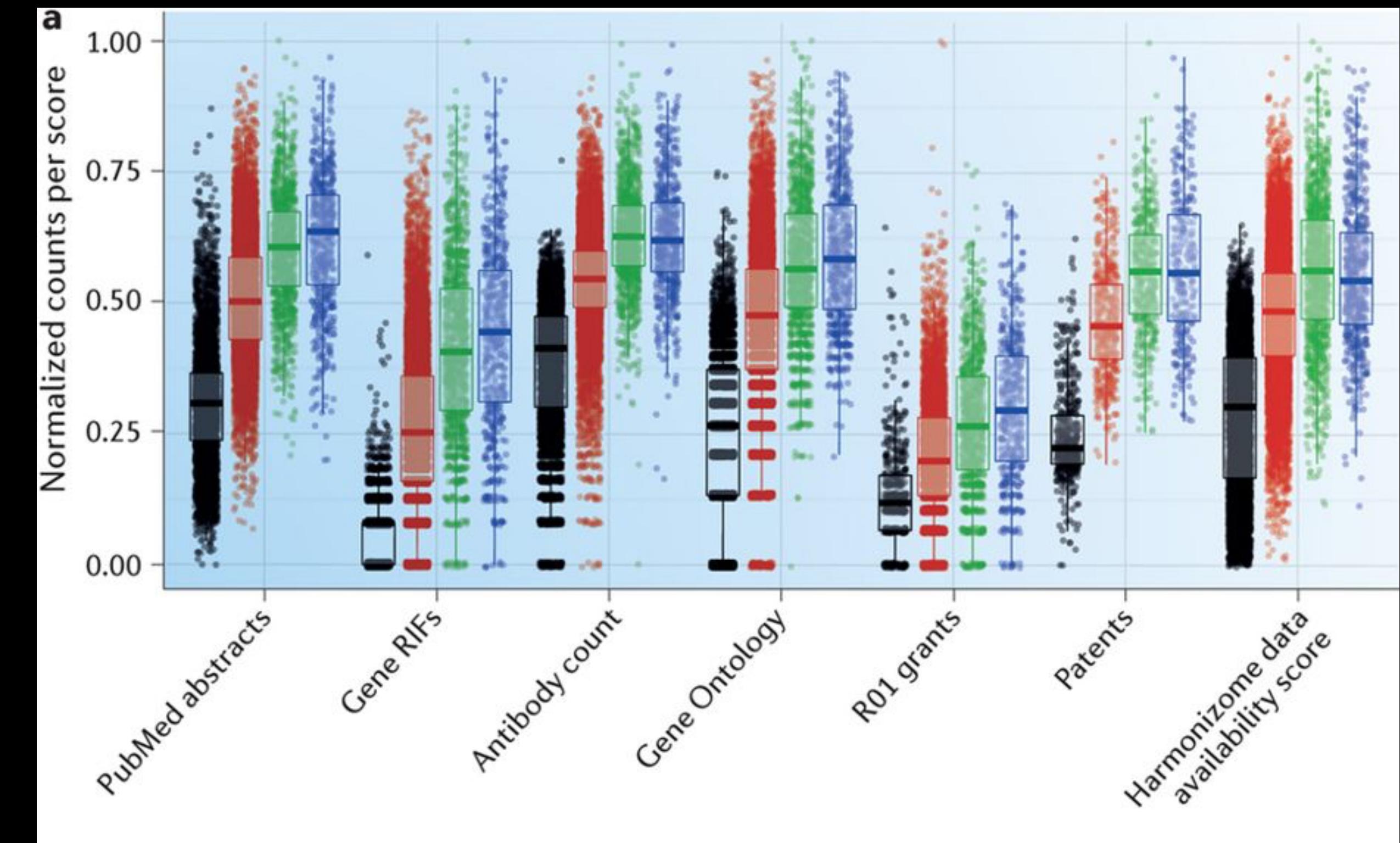
Nelson et al. 2014

# Support for drug targets.



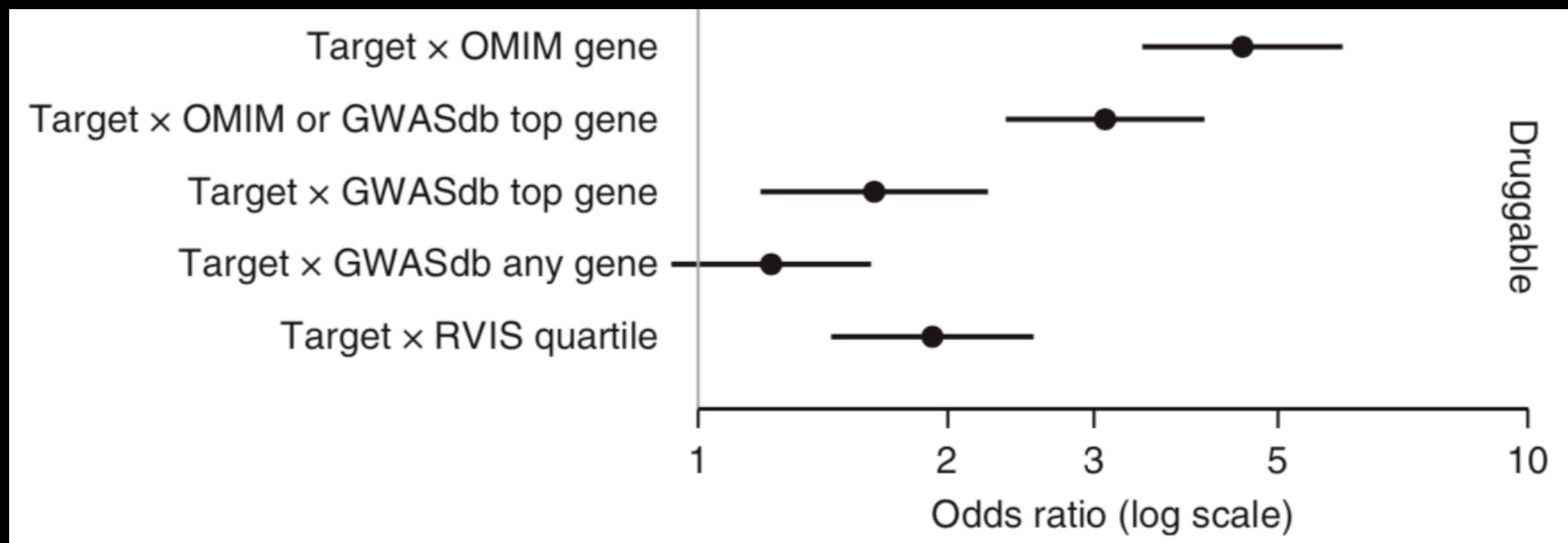
Nelson et al. 2014

## Interactions with drugs vs. grants, patents, literature.



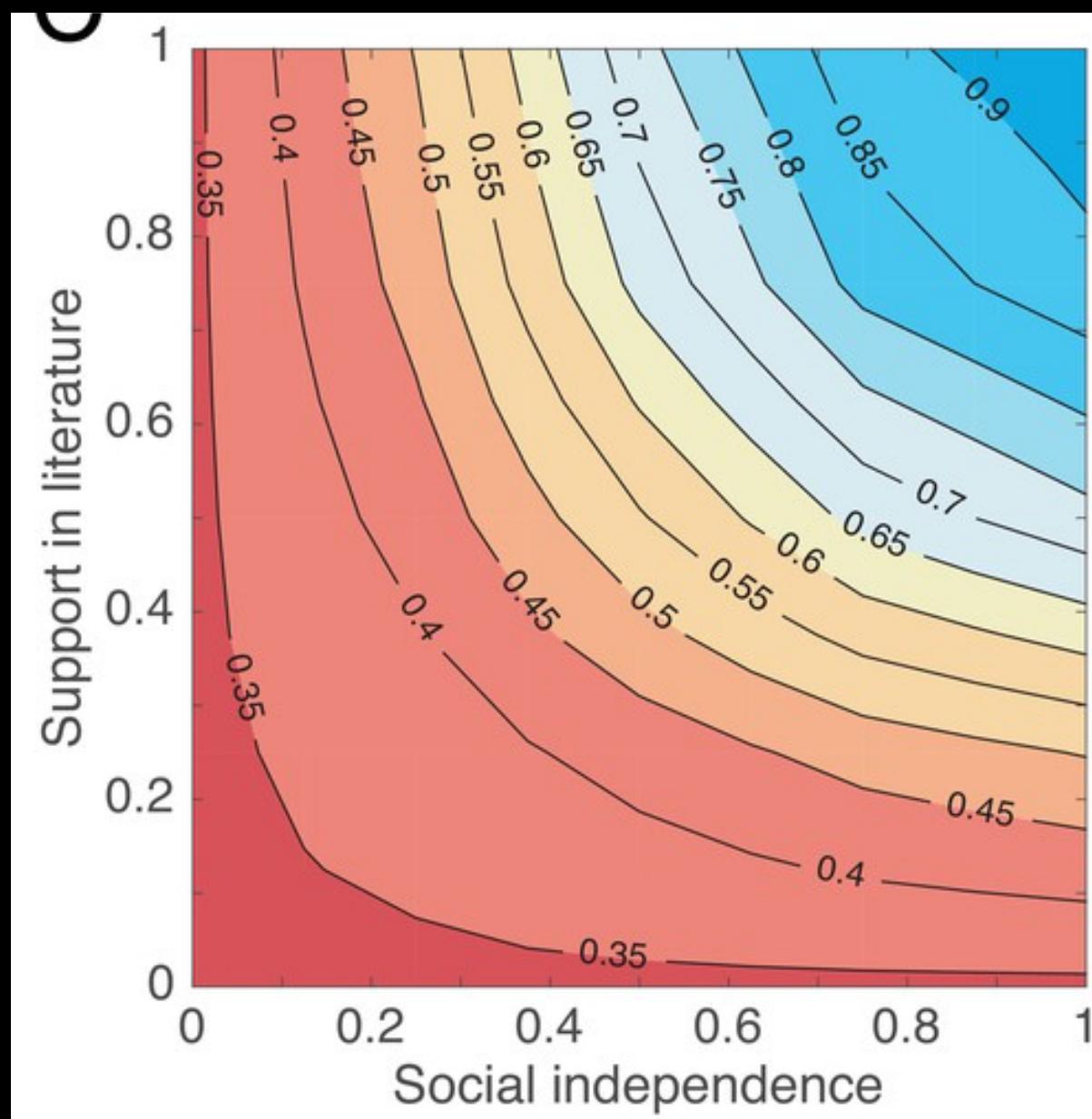
Oprea et al. 2018

# Support for drug targets.



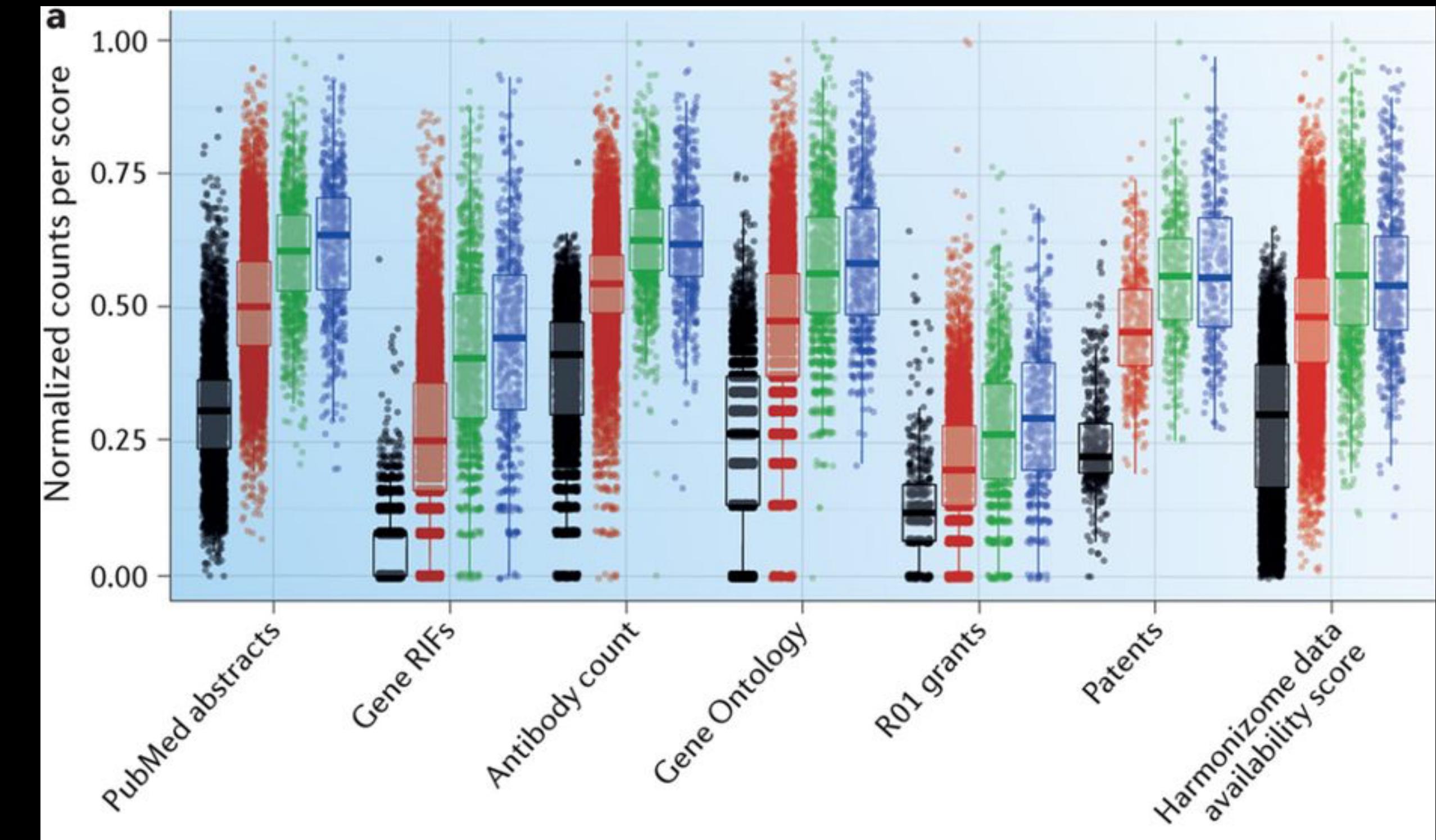
Nelson et al. 2014

Drug-> gene expression reproducibility and social interactions between scientists.



Danchev et al. 2019

Interactions with drugs vs. grants, patents, literature.



Oprea et al. 2018

# Research questions

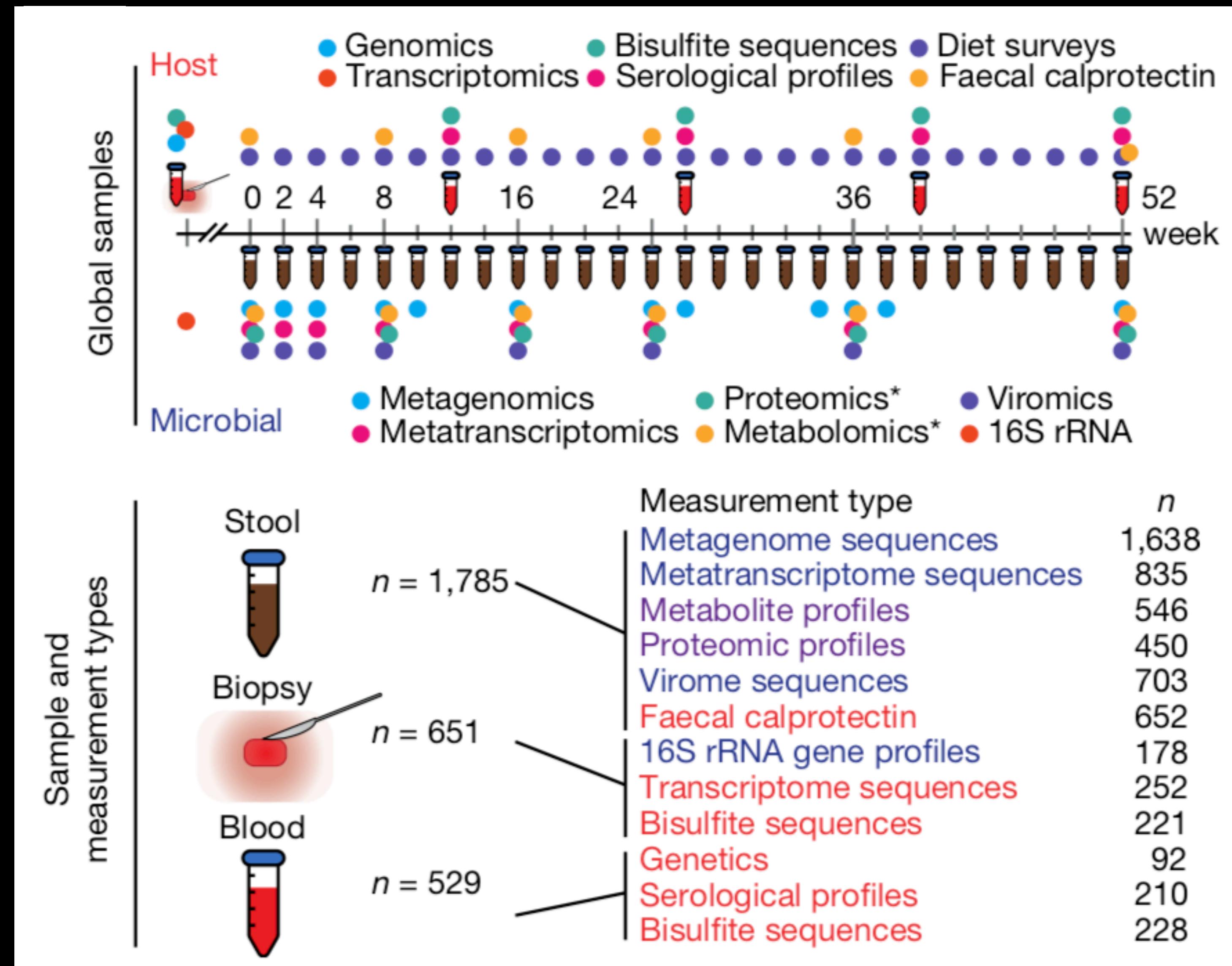
orientation

# Research questions

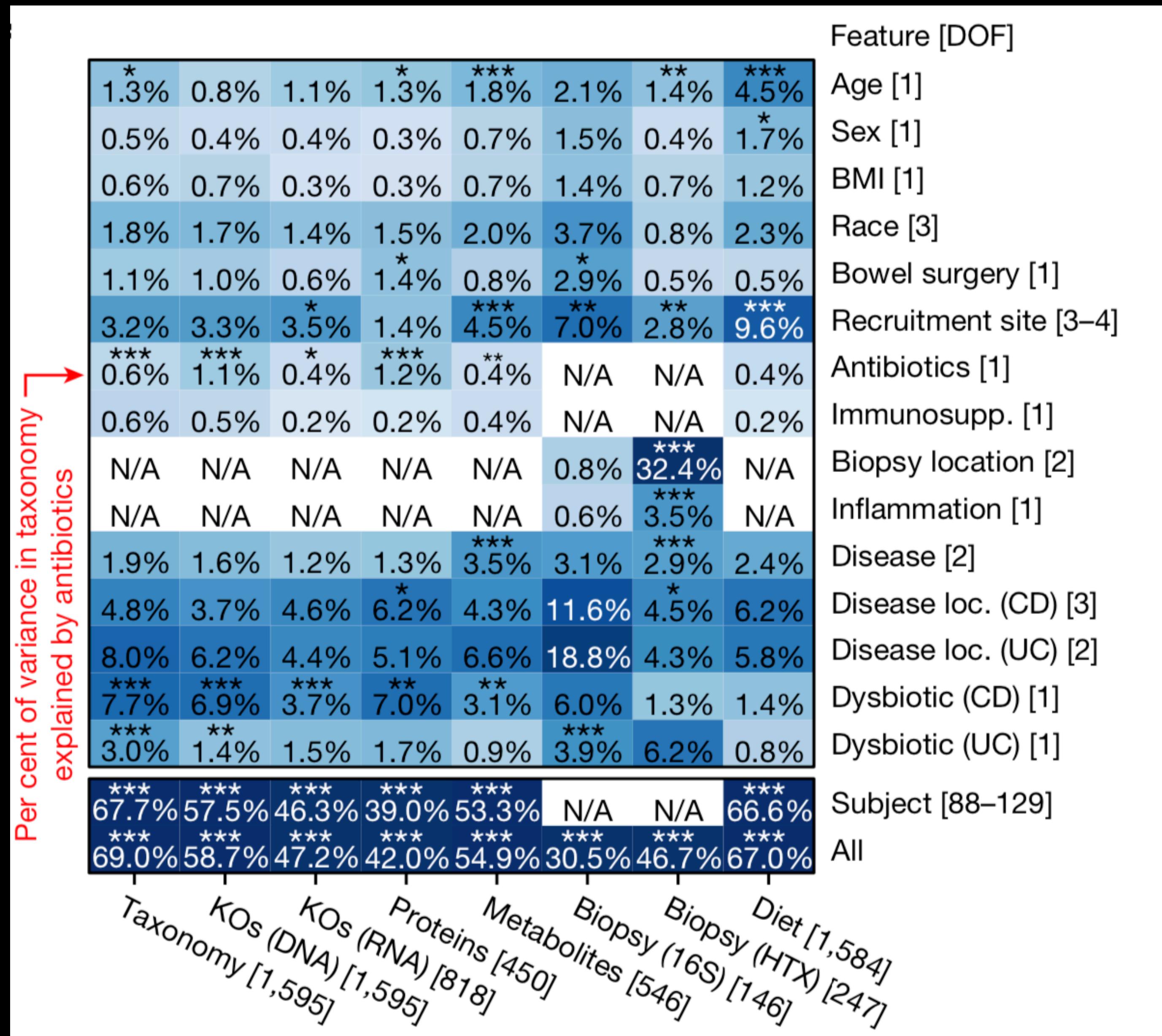
orientation

Example: What happens  
during inflammatory bowel  
disease?

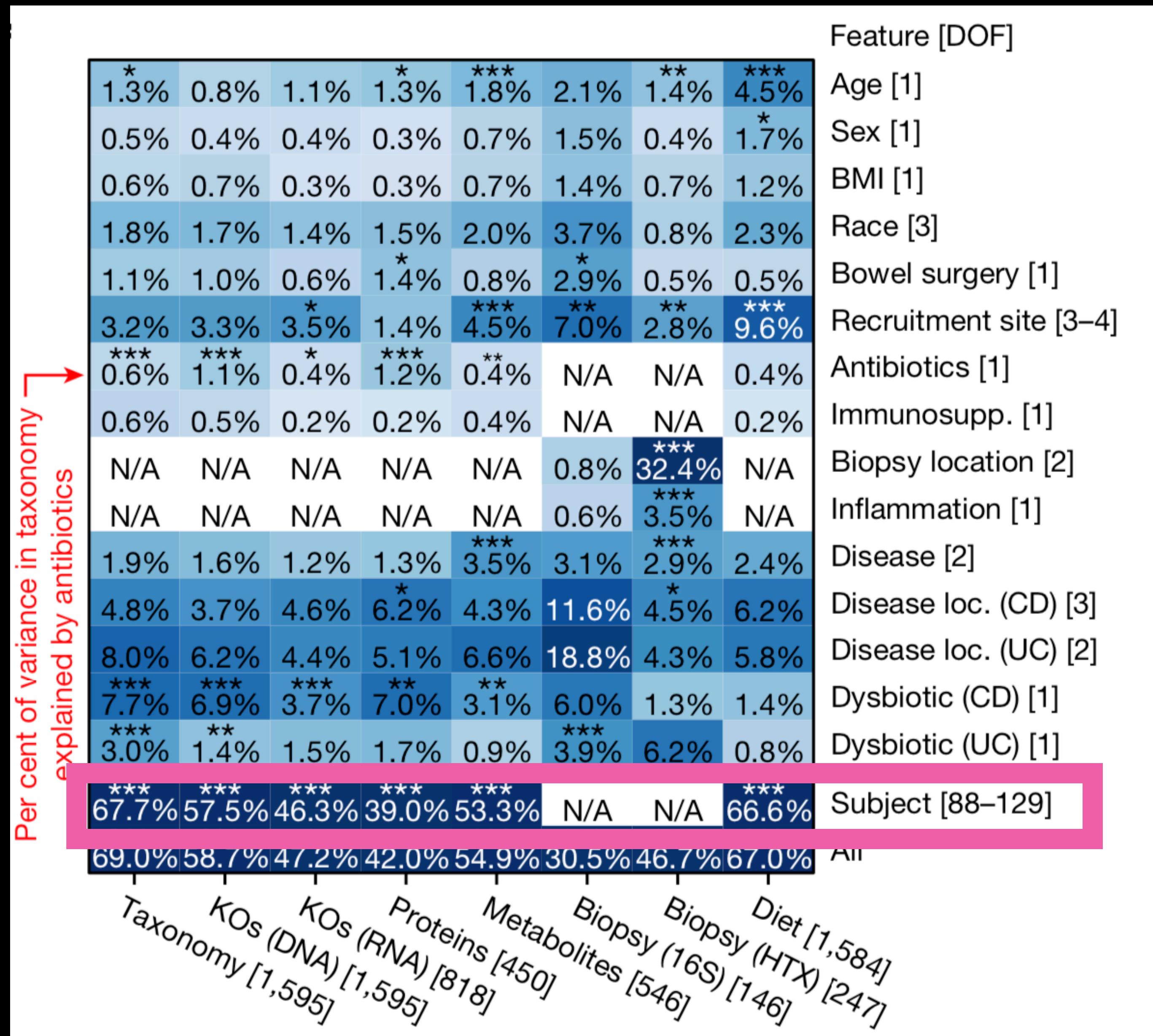
# What happens during inflammatory bowel disease?



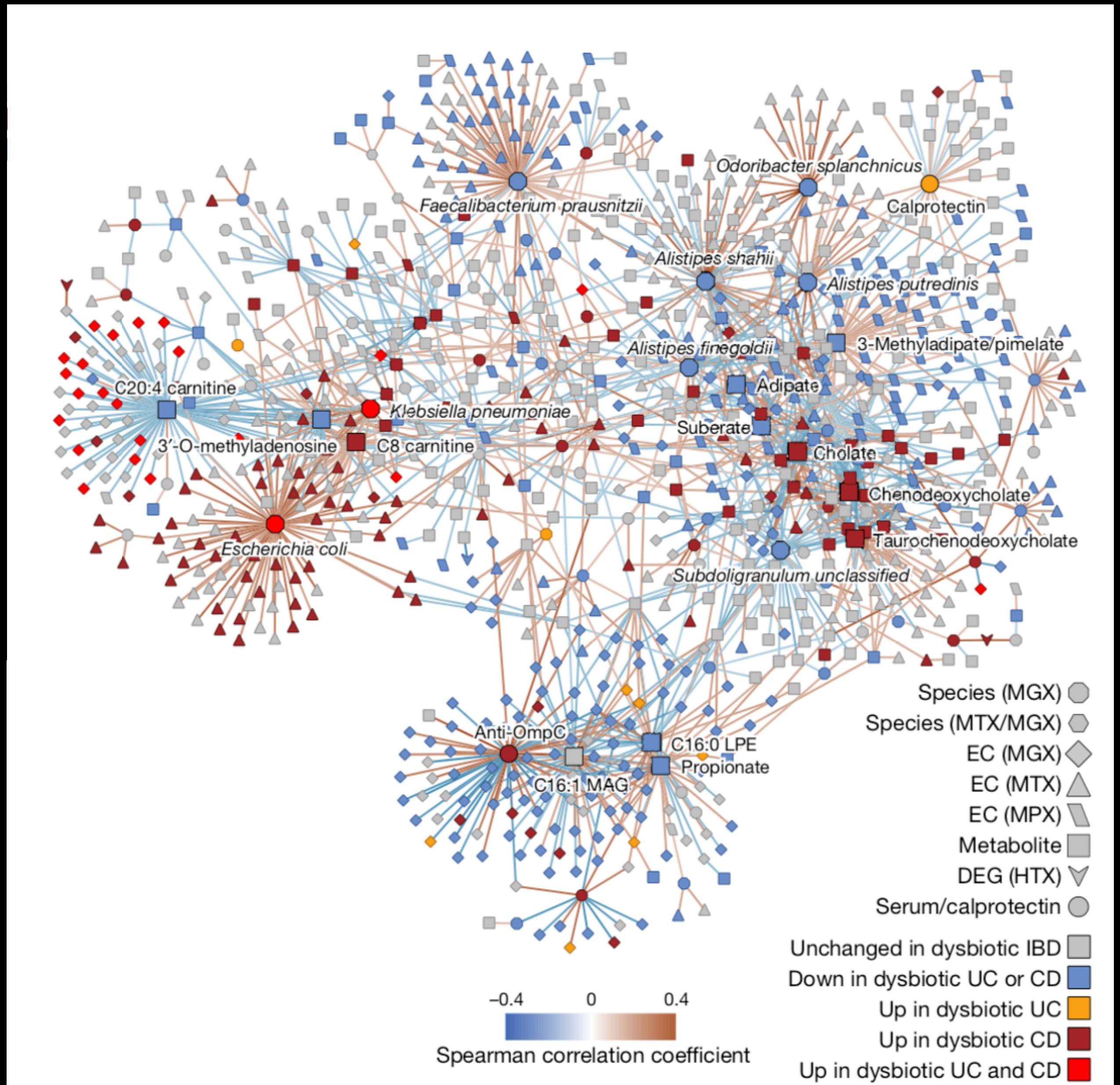
# Observation: most differences are subject-specific.



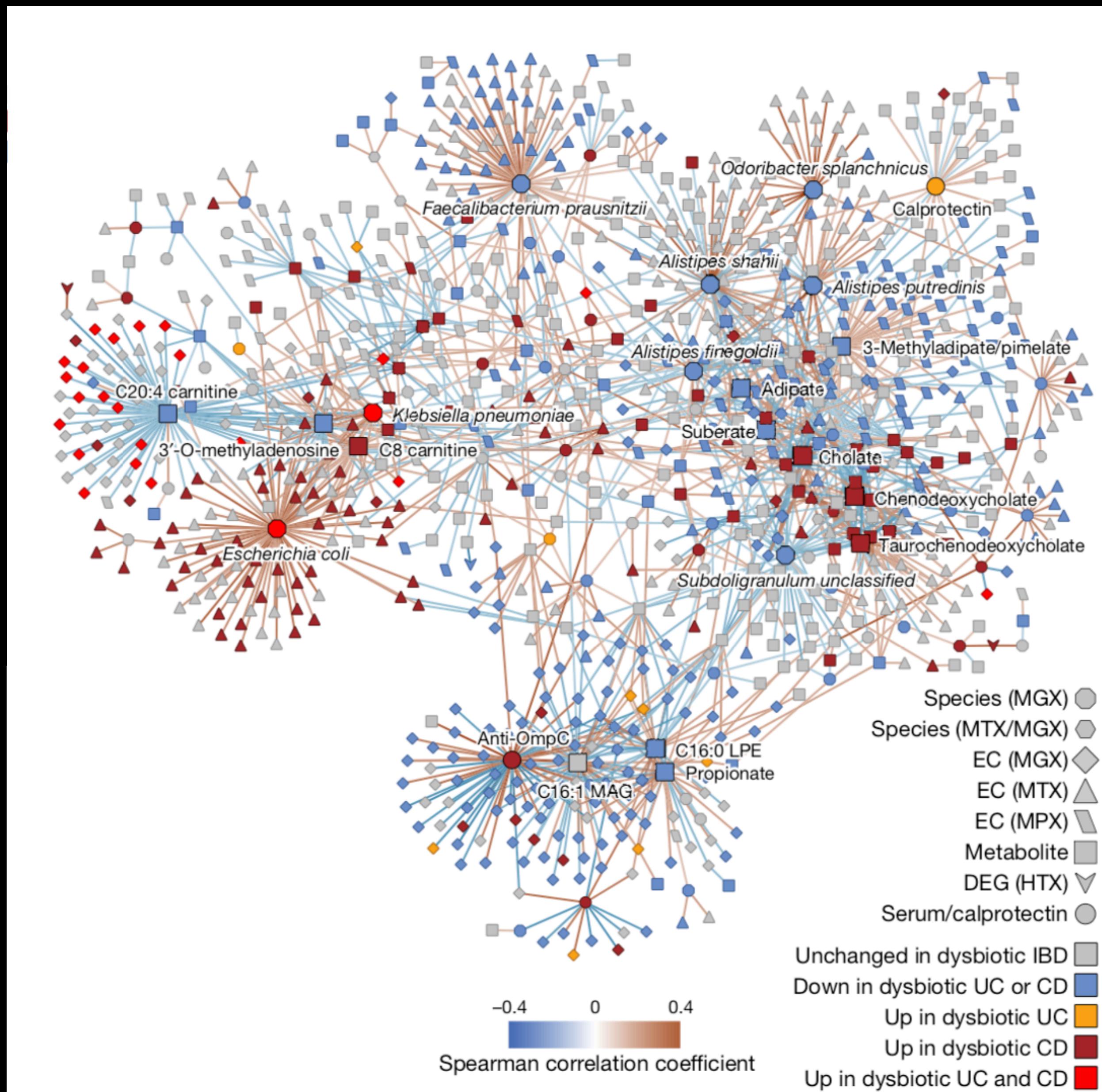
# Observation: most differences are subject-specific.



# Observation: Some readouts correlate (still).

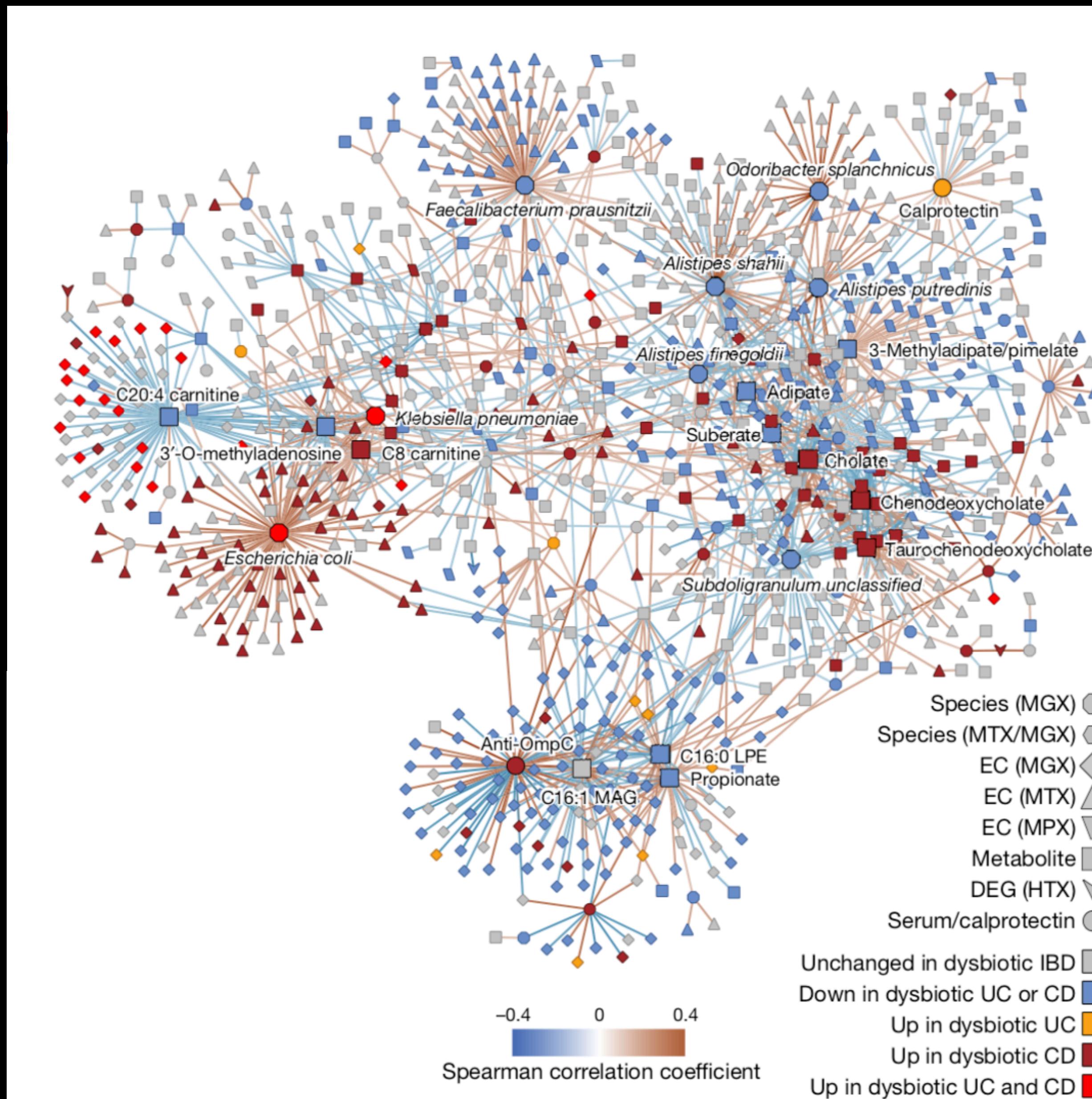


# Observation: Some readouts correlate (still).



We stress that it has not yet been determined whether these multi-omic features of the microbiome can predict disease events before their occurrence and that the disease-relevant time scales of distinct molecular events have not been identified (for example, static host genetics, relatively slow epigenetics or microbial growth, rapid host and microbial transcriptional changes).

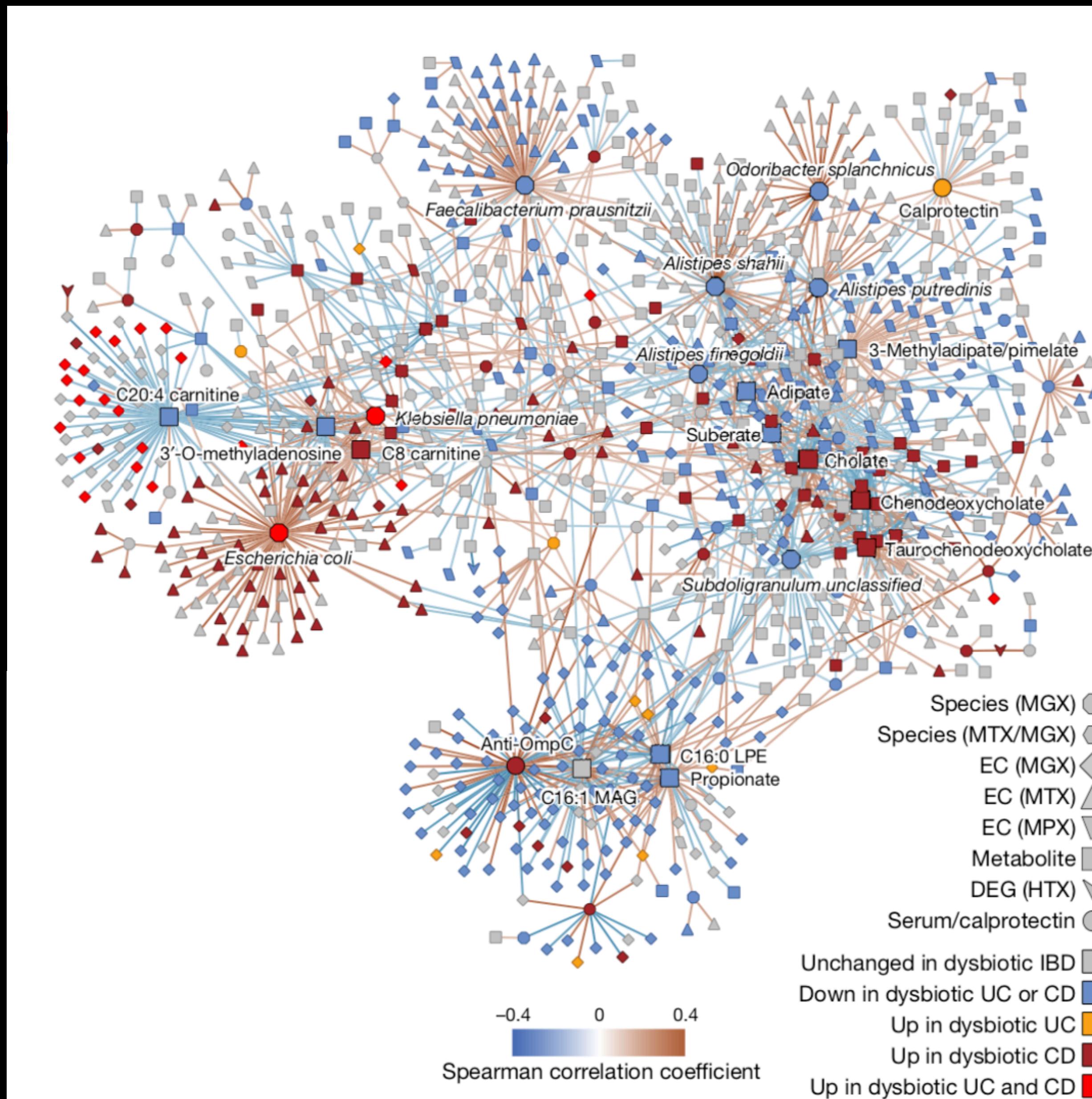
# Observation: Some readouts correlate (still).



We stress that it has not yet been determined whether these multi-omic features of the microbiome can predict disease events before their occurrence and that the disease-relevant time scales of distinct molecular events have not been identified (for example, static host genetics, relatively slow epigenetics or microbial growth, rapid host and microbial transcriptional changes).

It will be most important to take these molecular results back to the clinic,

# Observation: Some readouts correlate (still).



We stress that it has not yet been determined whether these multi-omic features of the microbiome can predict disease events before their occurrence and that the disease-relevant time scales of distinct molecular events have not been identified (for example, static host genetics, relatively slow epigenetics or microbial growth, rapid host and microbial transcriptional changes).

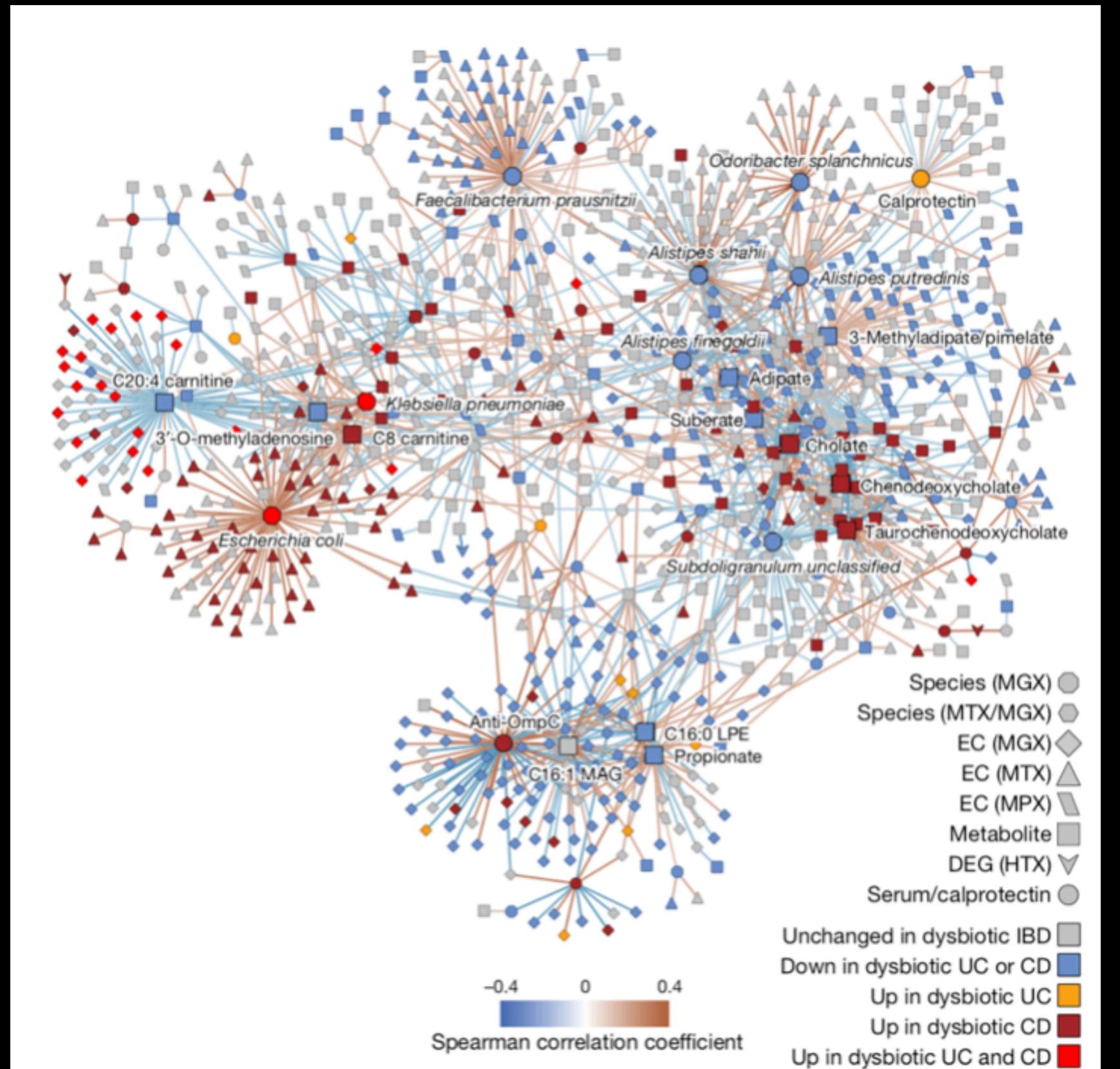
It will be most important to take these molecular results back to the clinic,

**Data is publicly available.**

Example: What happens  
during inflammatory bowel  
disease?

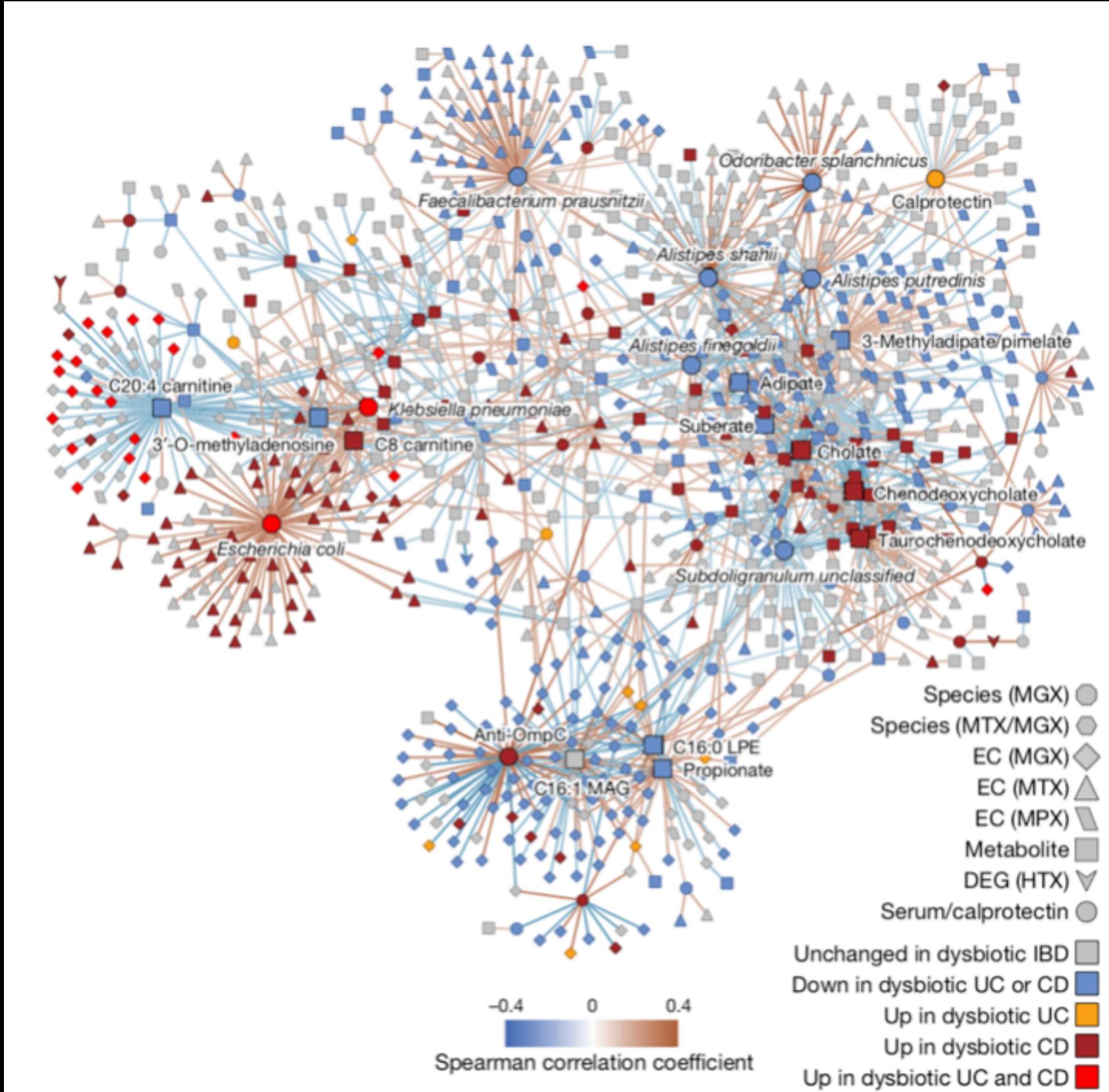
orientation

# Hypotheses for host-microbiome-metabolome interactions



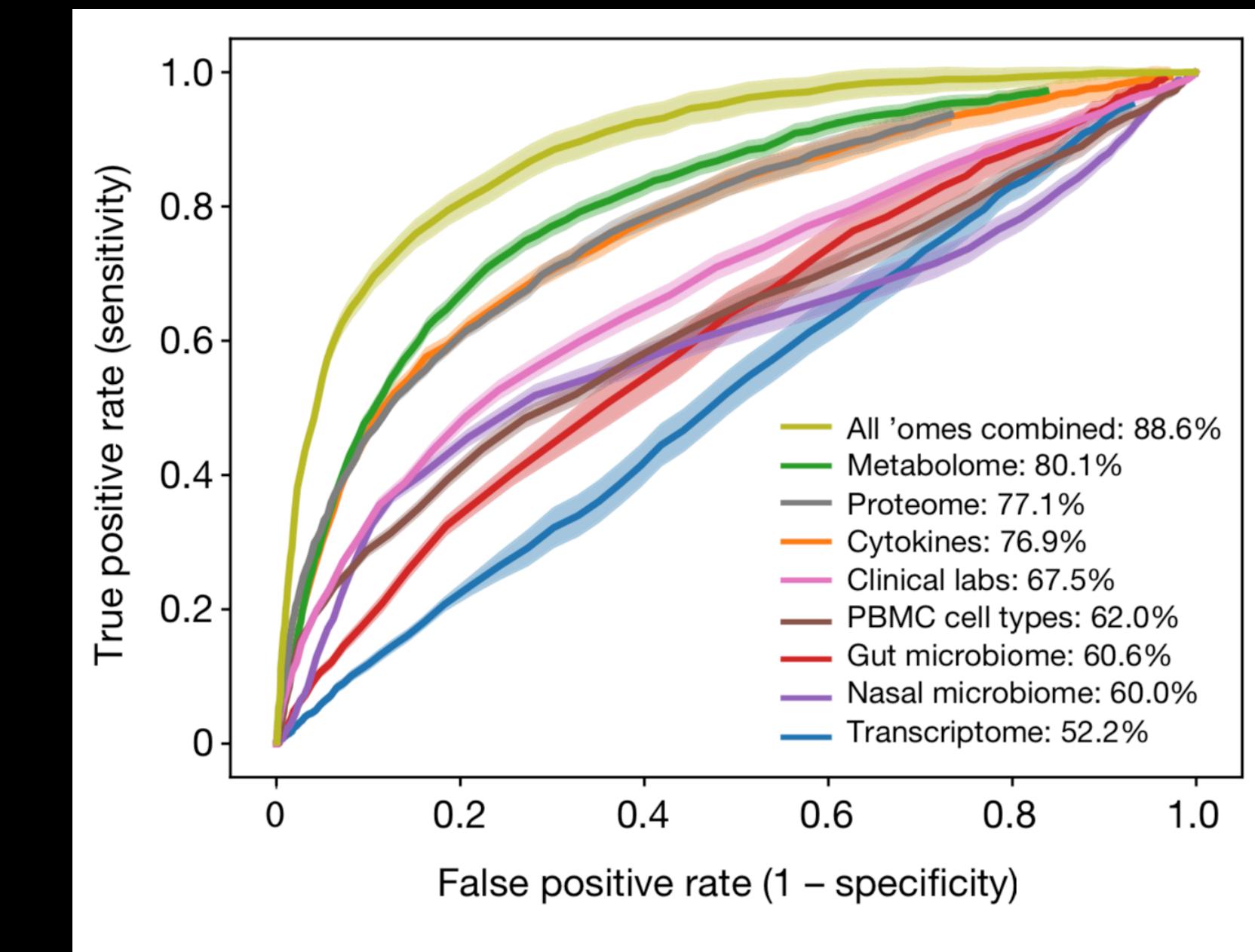
Lloyd-Price et al. 2019

# Hypotheses for host-microbiome-metabolome interactions



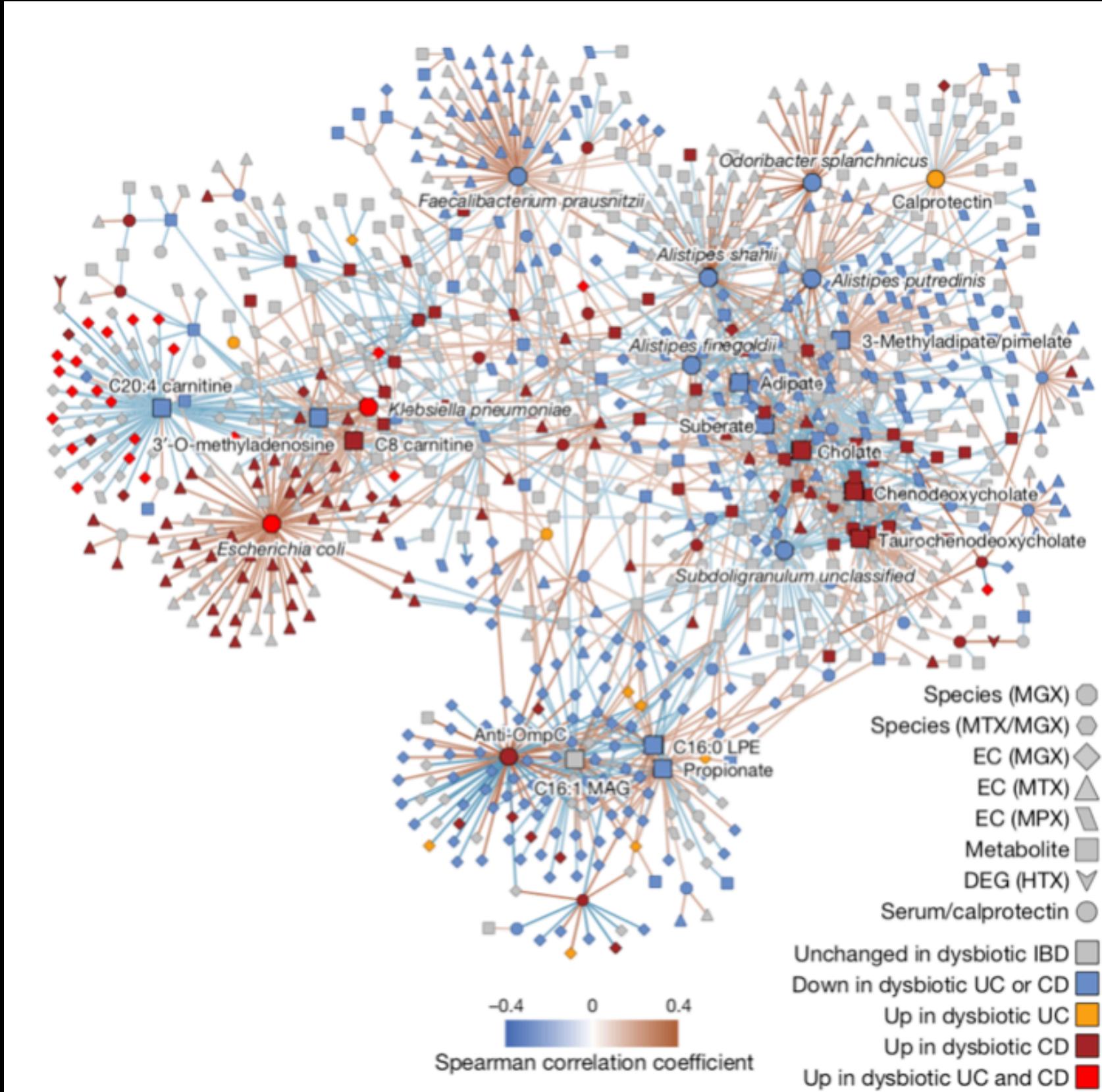
Lloyd-Price et al. 2019

# Diabetes (and respiratory infections)



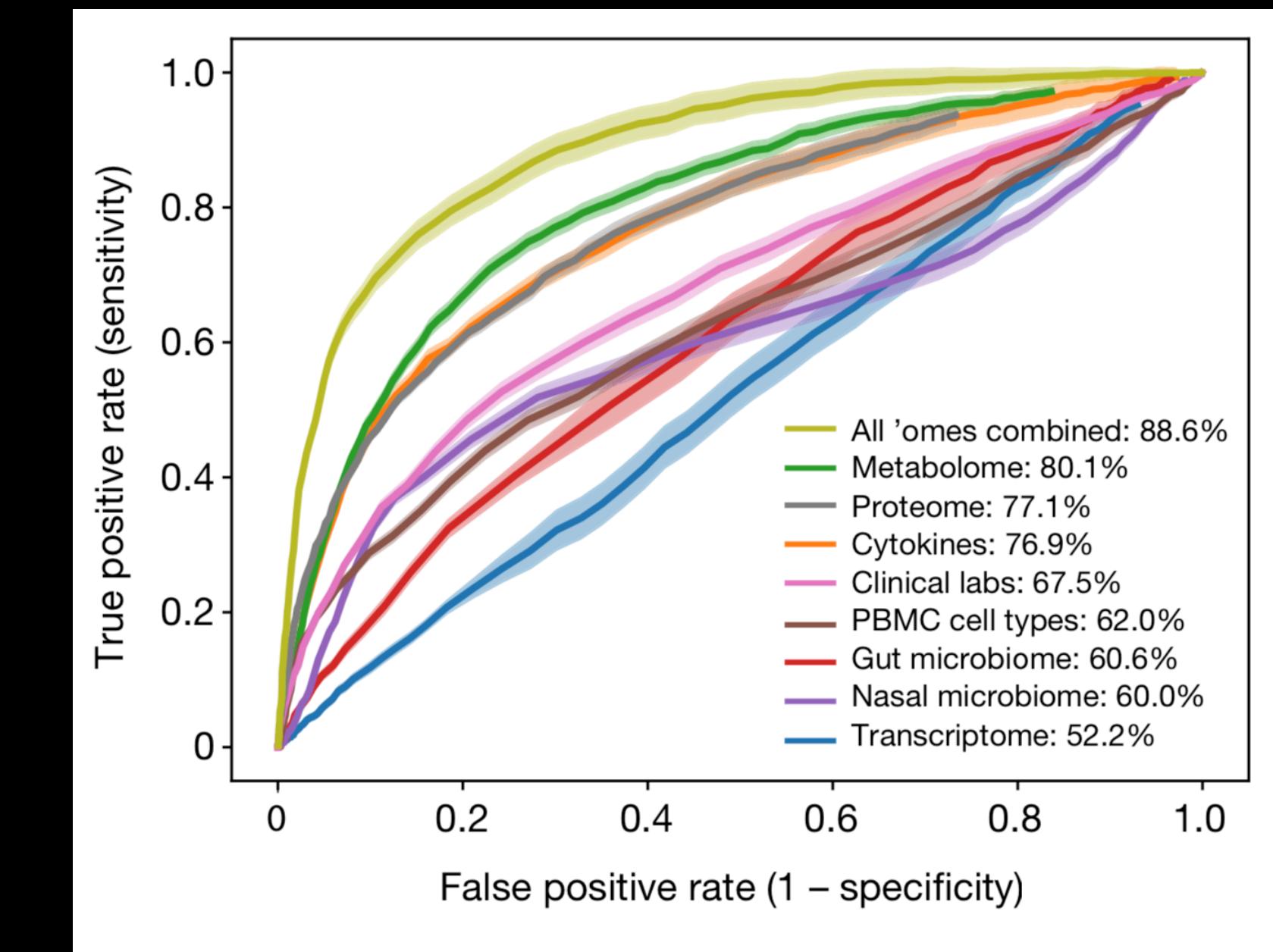
Zhou et al. 2019

# Hypotheses for host-microbiome-metabolome interactions



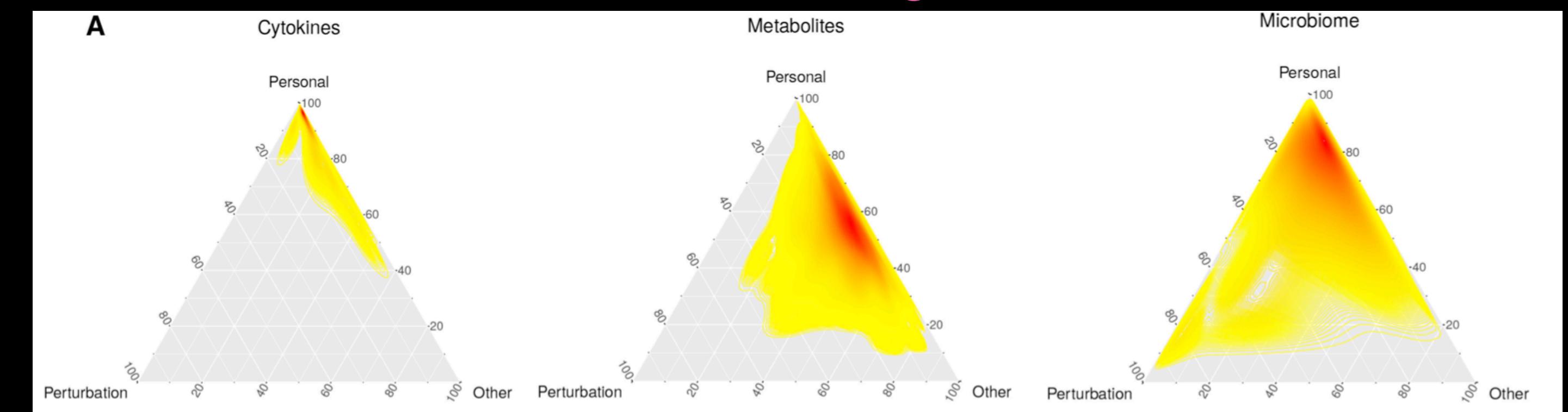
Lloyd-Price et al. 2019

# Diabetes (and respiratory infections)



Zhou et al. 2019

# Dieting



Piening et al. 2018

# **Research questions**

scales

distorted literature

orientation

**-> Combine qualitatively  
different data!**

## Research questions

scales

distorted literature

orientation

-> Combine qualitatively  
different data!

2 minutes:

Discuss with neighbor:

What types of data  
can be combined?

Can you think of more  
than those discussed?



**Questions /  
Would like to  
know more  
about ...**

# Research questions

scales  
distorted literature  
orientation

-> Combine qualitatively  
different data!

# Technology

organizing  
analytical approaches  
understanding limits

-> How to (start) working  
with multi-omics!

Technology

organizing

no-one can be an expert  
in everything

Technology

organizing

Technology

no-one can be an expert  
in everything

organizing

try to store data  
in meaningful ways

# Technology

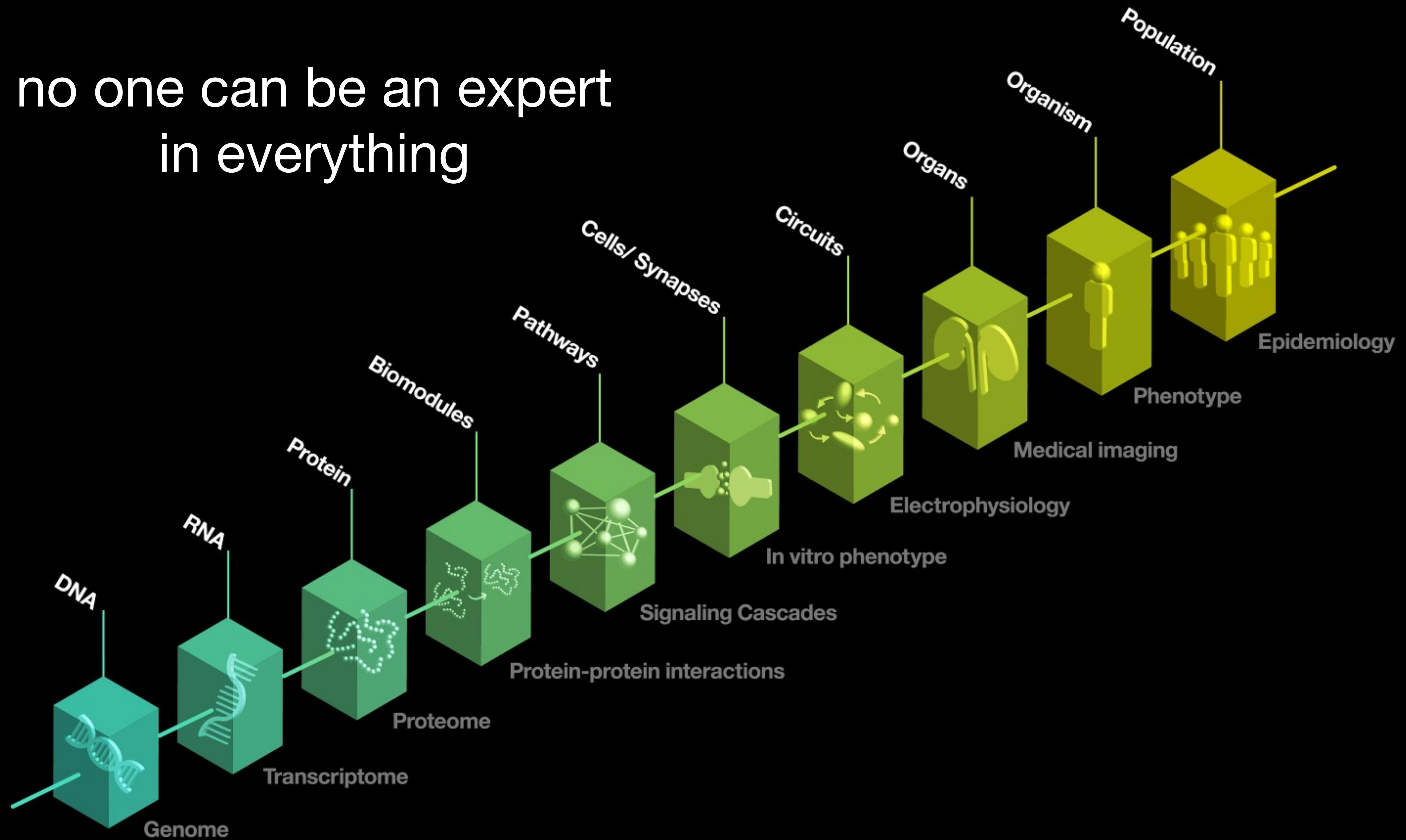
no-one can be an expert  
in everything

organizing

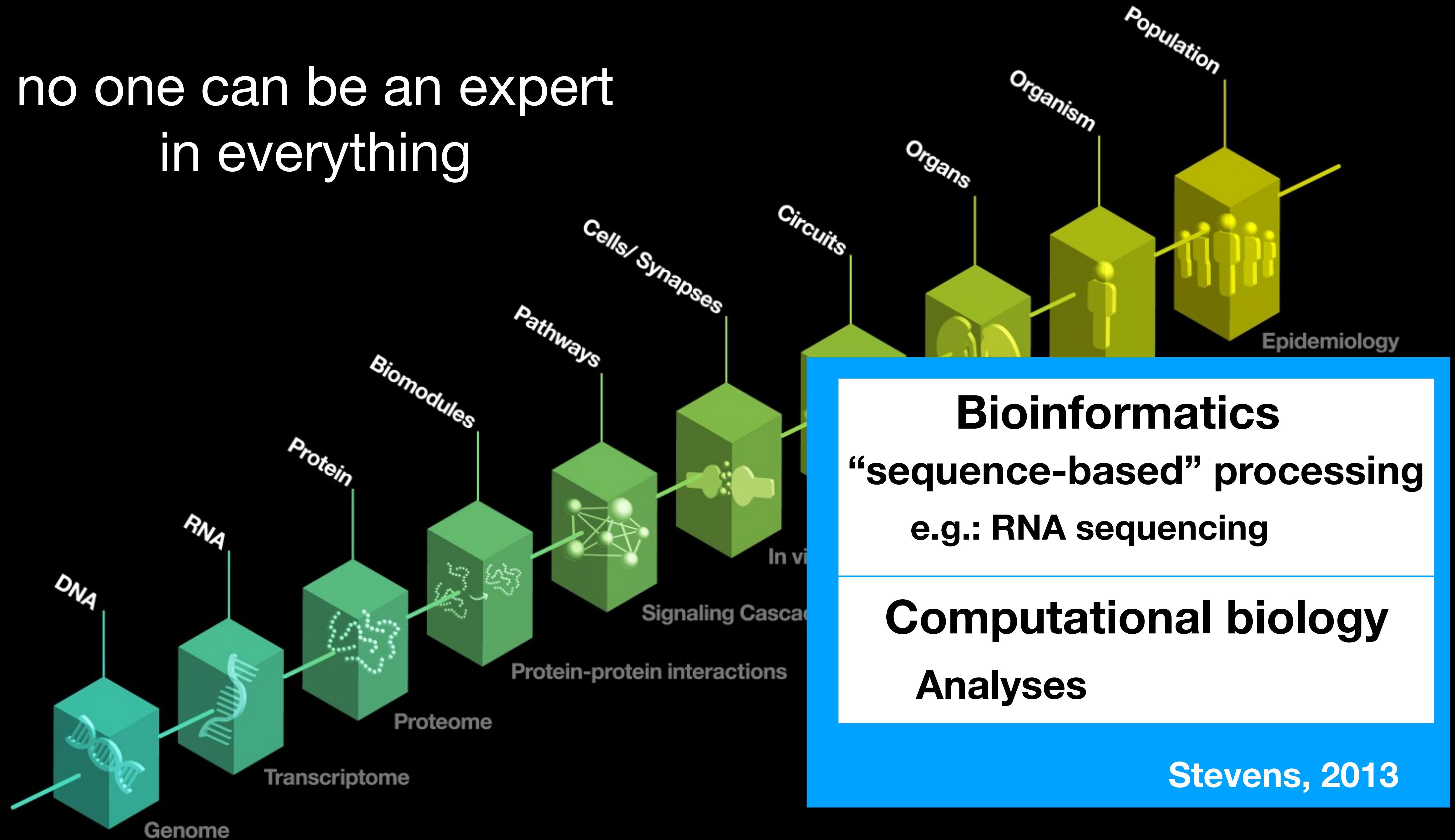
try to store data  
in meaningful ways

write wrappers to anticipate  
changes in data

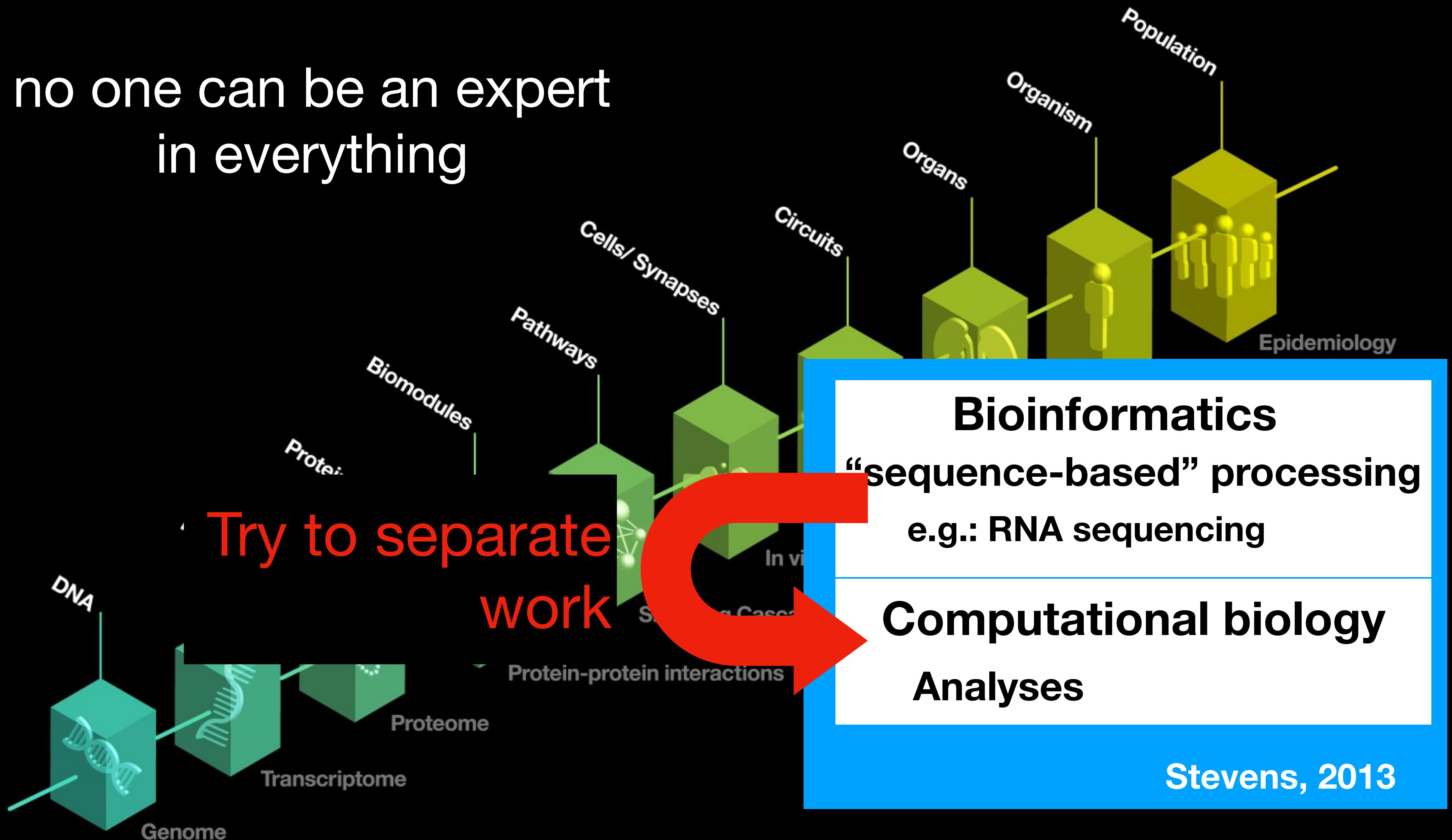
no one can be an expert  
in everything



no one can be an expert  
in everything



no one can be an expert  
in everything

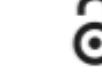


# Technology

no-one can be an expert  
in everything

organizing

try to store data  
in meaningful ways

 OPEN ACCESS

 Check for updates

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

**(Same fundamental principles whether a spreadsheet or database table.)**

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

**(Same fundamental principles whether a spreadsheet or database table.)**

Be consistent.

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

(Same fundamental principles whether a spreadsheet or database table.)

Be consistent.  
Just put one thing in a cell.

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

(Same fundamental principles whether a spreadsheet or database table.)

Be consistent.  
Just put one thing in a cell.  
Create a data dictionary.

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

(Same fundamental principles whether a spreadsheet or database table.)

Be consistent.  
Just put one thing in a cell.  
Create a data dictionary.

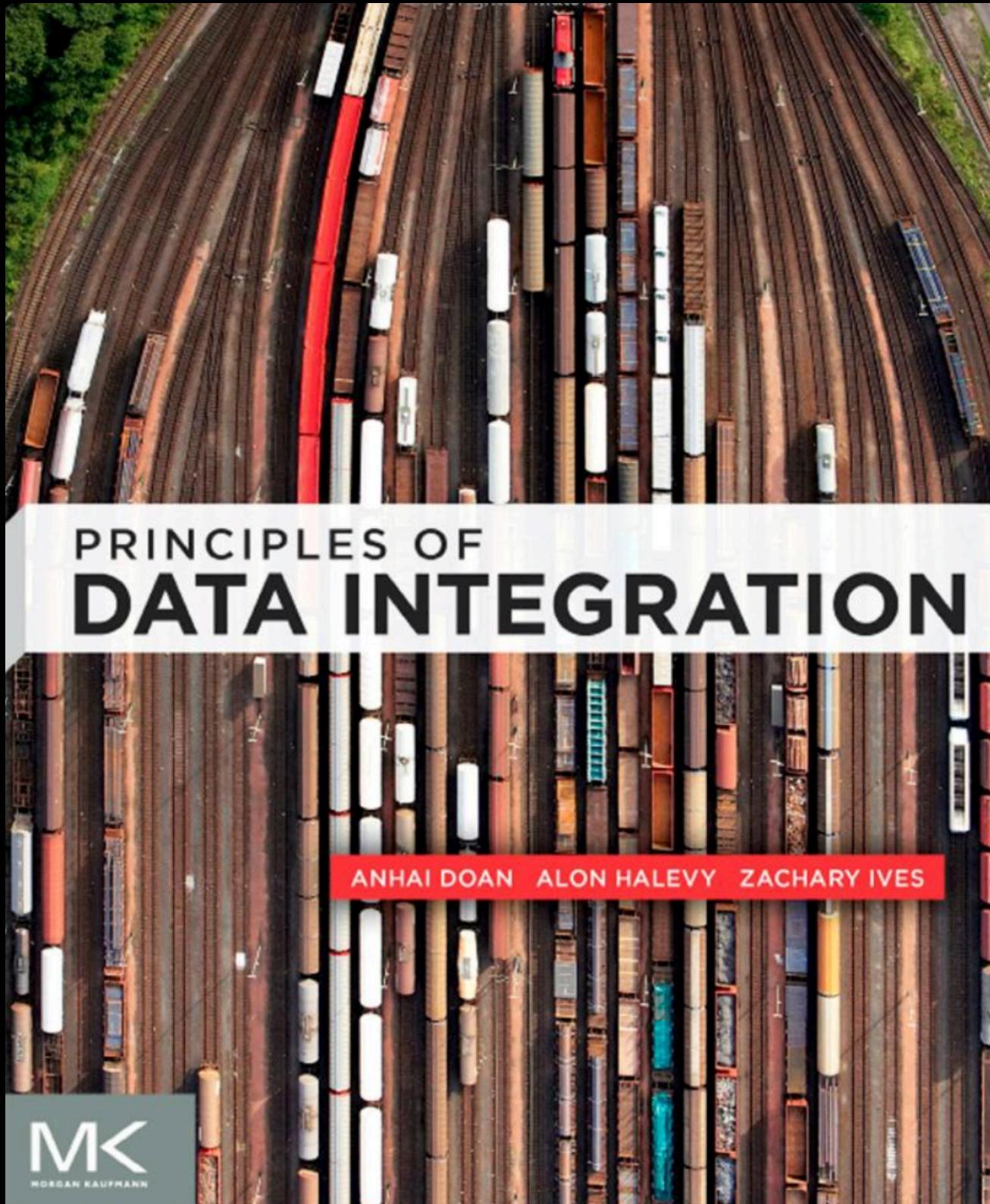
note: there might be domain-specific conventions  
(e.g.: tidy data, models, storage formats...)

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

(Same fundamental principles whether a spreadsheet or database table.)



Be consistent.  
Just put one thing in a cell.  
Create a data dictionary.

note: there might be domain-specific conventions  
(e.g.: tidy data, models, storage formats...)

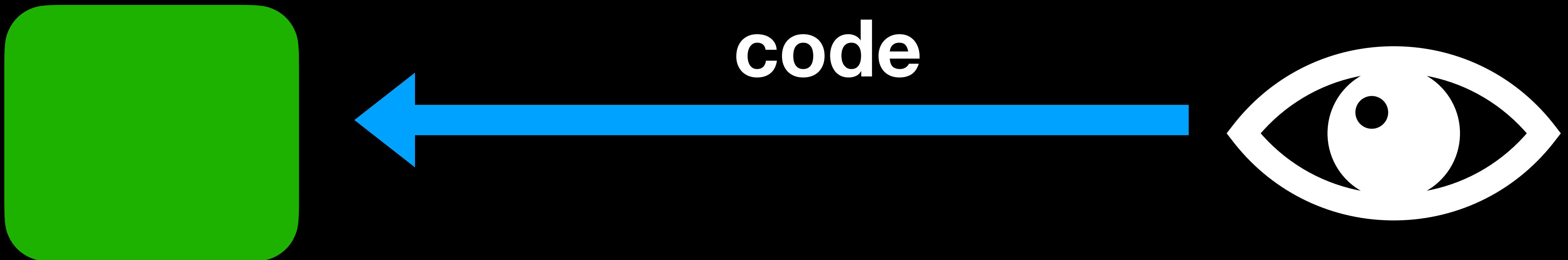
# Technology

no-one can be an expert  
in everything

organizing

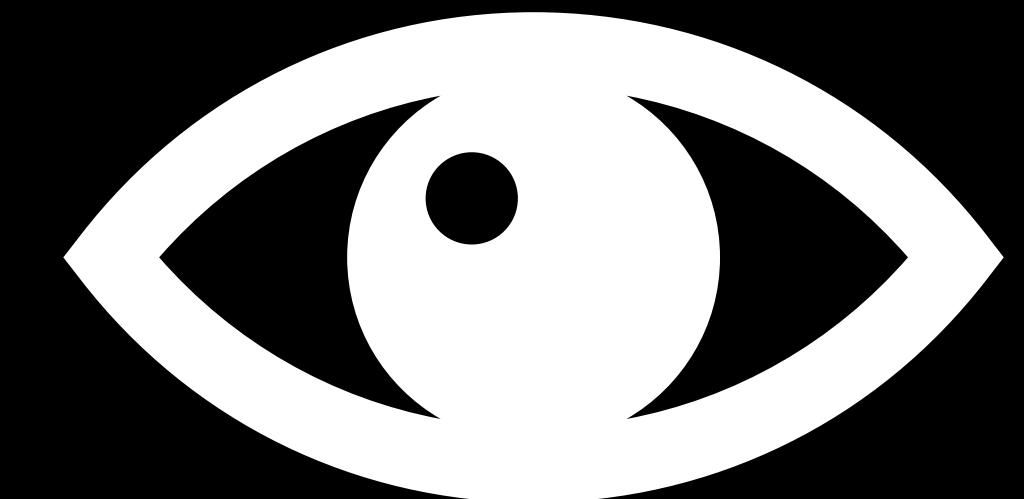
try to store data  
in meaningful ways

**write wrappers to anticipate  
changes in data**

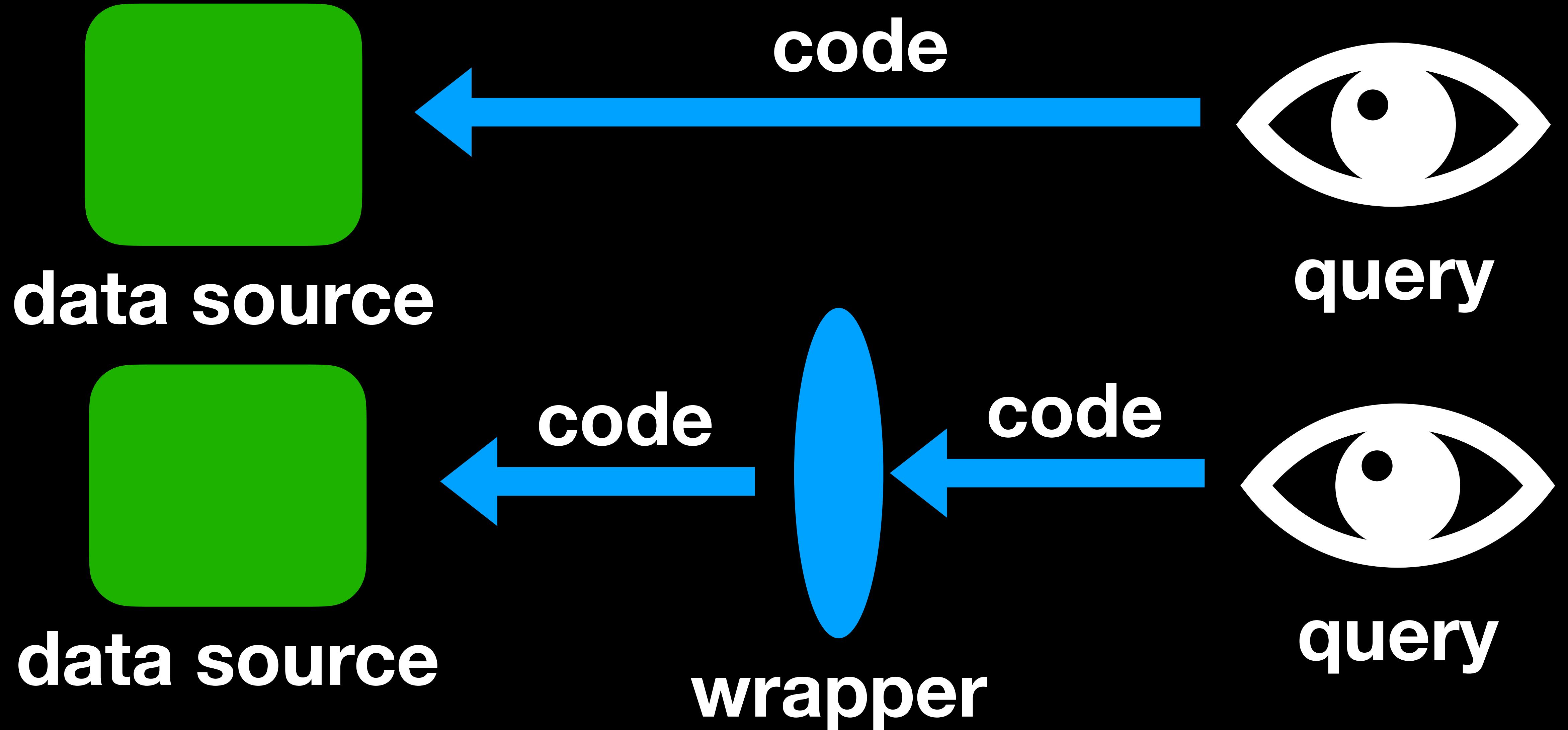


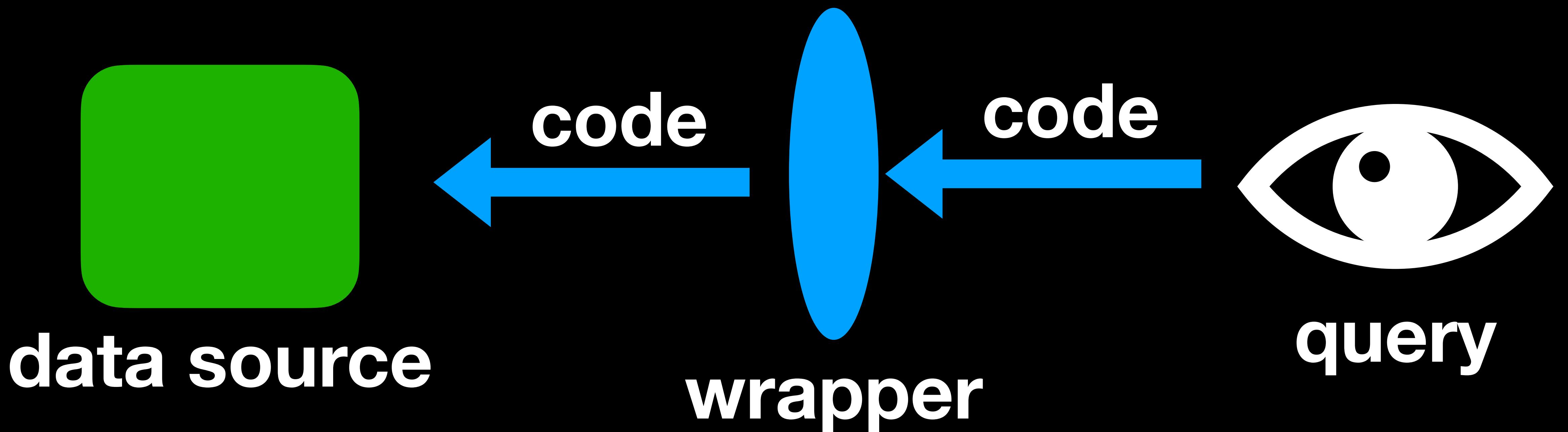
**data source**

**code**

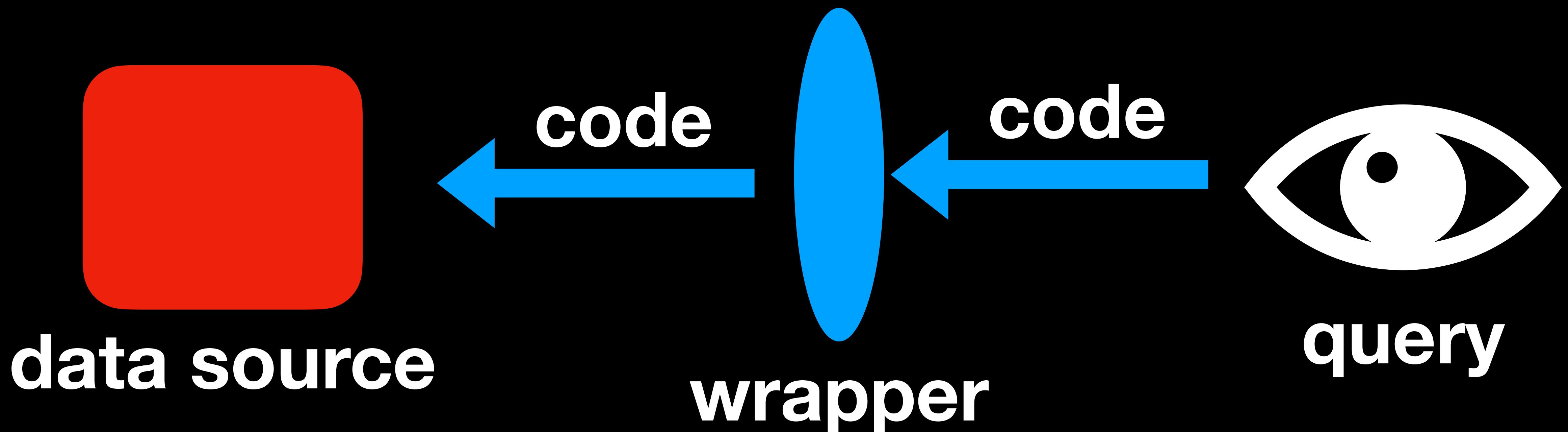


**query**



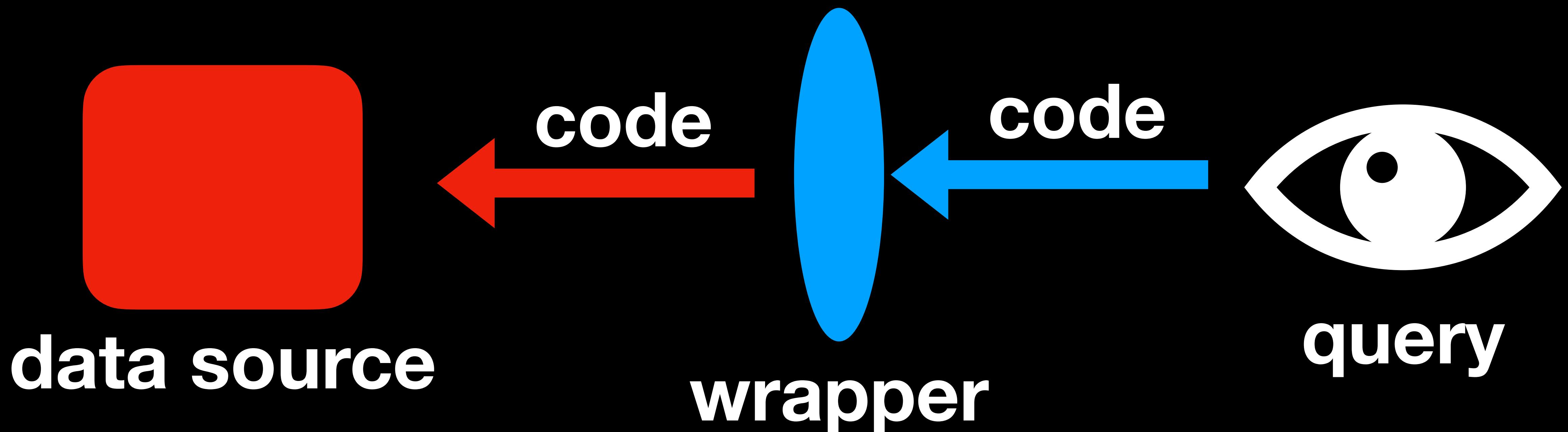


**other / new  
differently  
formatted**



**other / new  
differently  
formatted**

**optimize  
speed if  
frequently used**



no-one can be an expert  
in everything

try to store data  
in meaningful ways

write wrappers to anticipate  
changes in data

no-one can be an expert  
in everything

try to store data  
in meaningful ways

write wrappers to anticipate  
changes in data

1 minutes:

Discuss with neighbor:

Which tricks could  
save you or  
others time?

# Technology

analytical approaches

# Technology

different integration  
strategies

analytical approaches

# Technology

different integration  
strategies

mapping between datasets

analytical approaches

# Technology

different integration  
strategies

mapping between datasets

list of tools

analytical approaches

# Technology

different integration  
strategies

analytical approaches



**Figure 1.** Overview of multi-omics clustering approaches.

Pavlidis et al., 2001

Rappoport and Shamir, 2018

Stuart and Satija, 2019

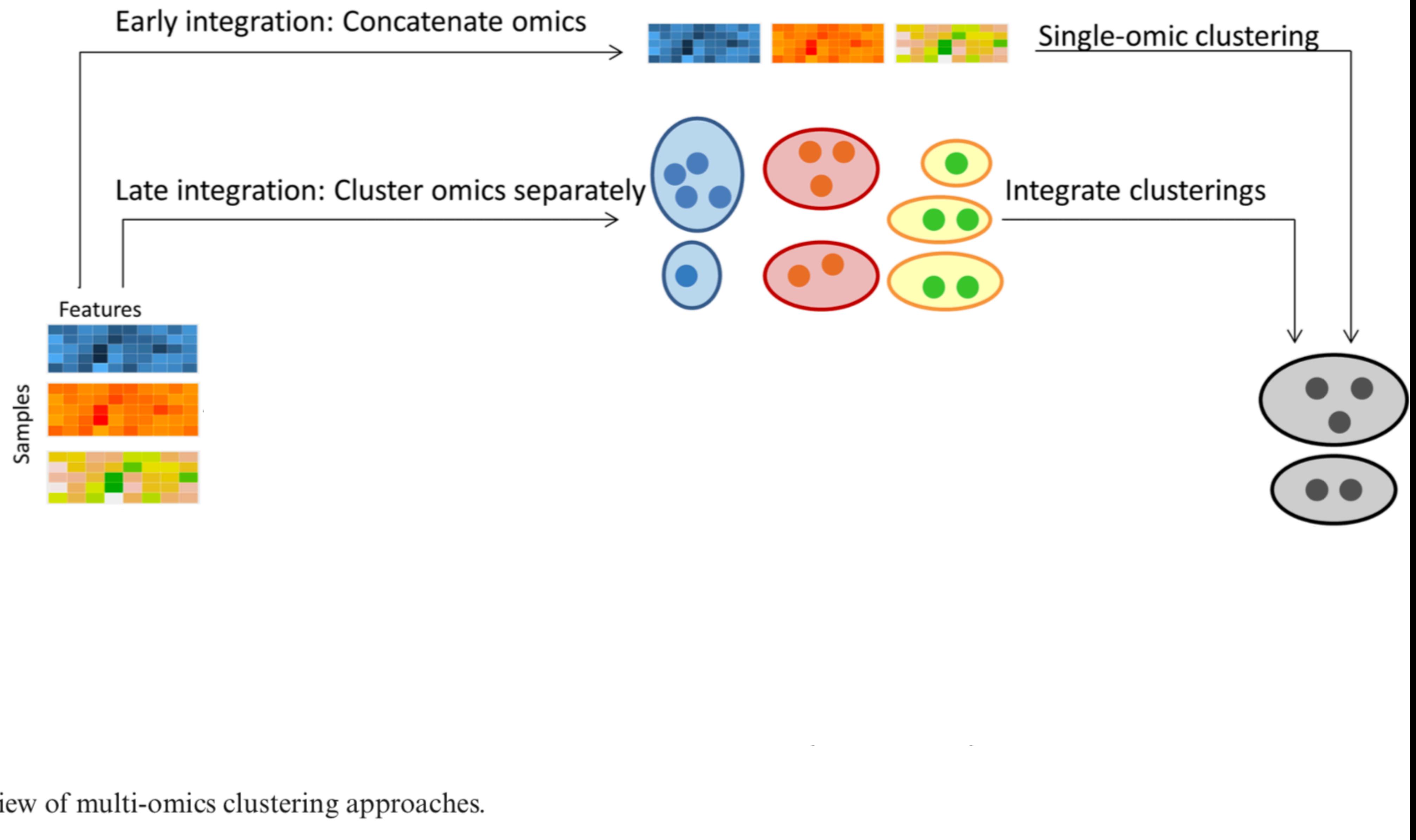


**Figure 1.** Overview of multi-omics clustering approaches.

Pavlidis et al., 2001

Rappoport and Shamir, 2018

Stuart and Satija, 2019

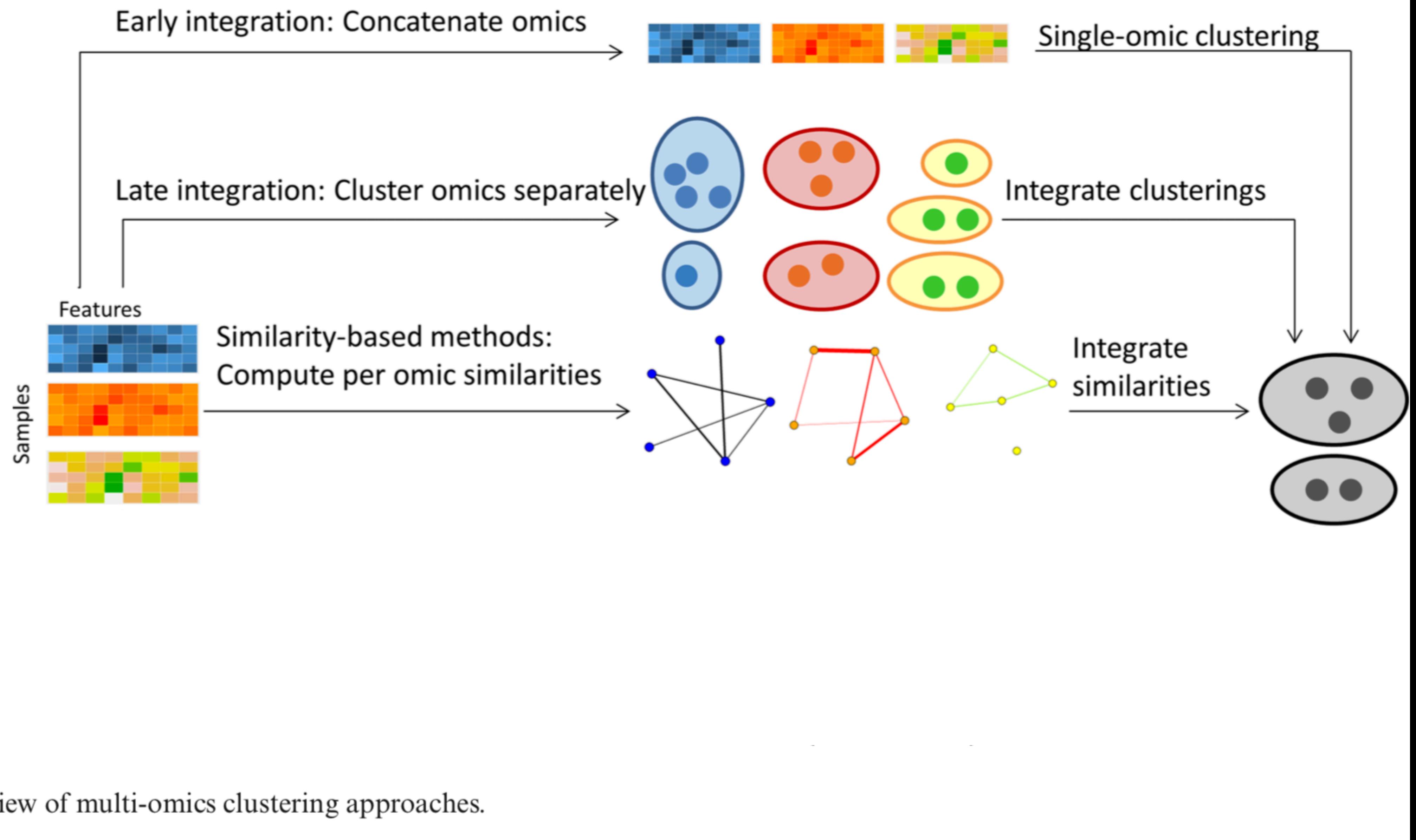


**Figure 1.** Overview of multi-omics clustering approaches.

Pavlidis et al., 2001

Rappoport and Shamir, 2018

Stuart and Satija, 2019

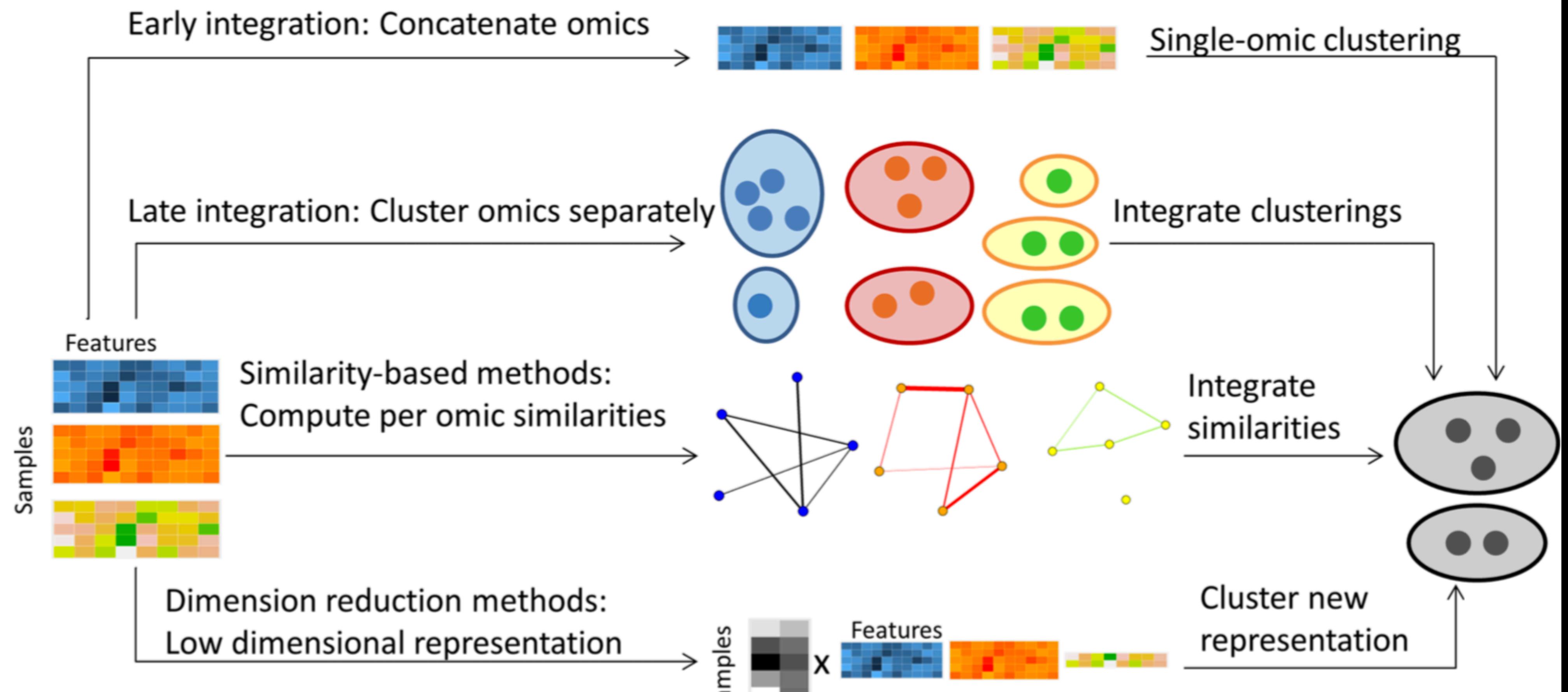


**Figure 1.** Overview of multi-omics clustering approaches.

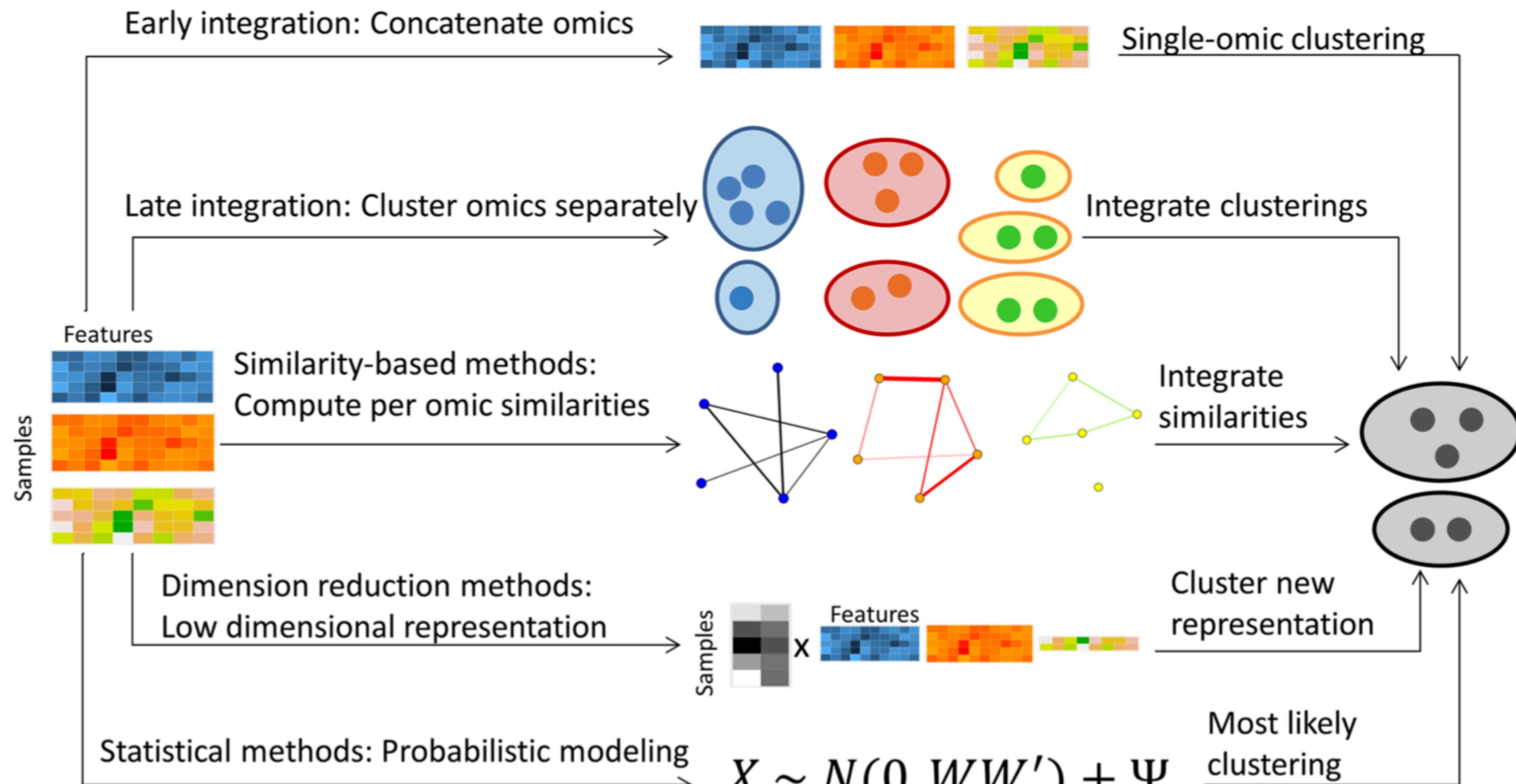
Pavlidis et al., 2001

Rappoport and Shamir, 2018

Stuart and Satija, 2019



**Figure 1.** Overview of multi-omics clustering approaches.



**Figure 1.** Overview of multi-omics clustering approaches.

Pavlidis et al., 2001

Rappoport and Shamir, 2018

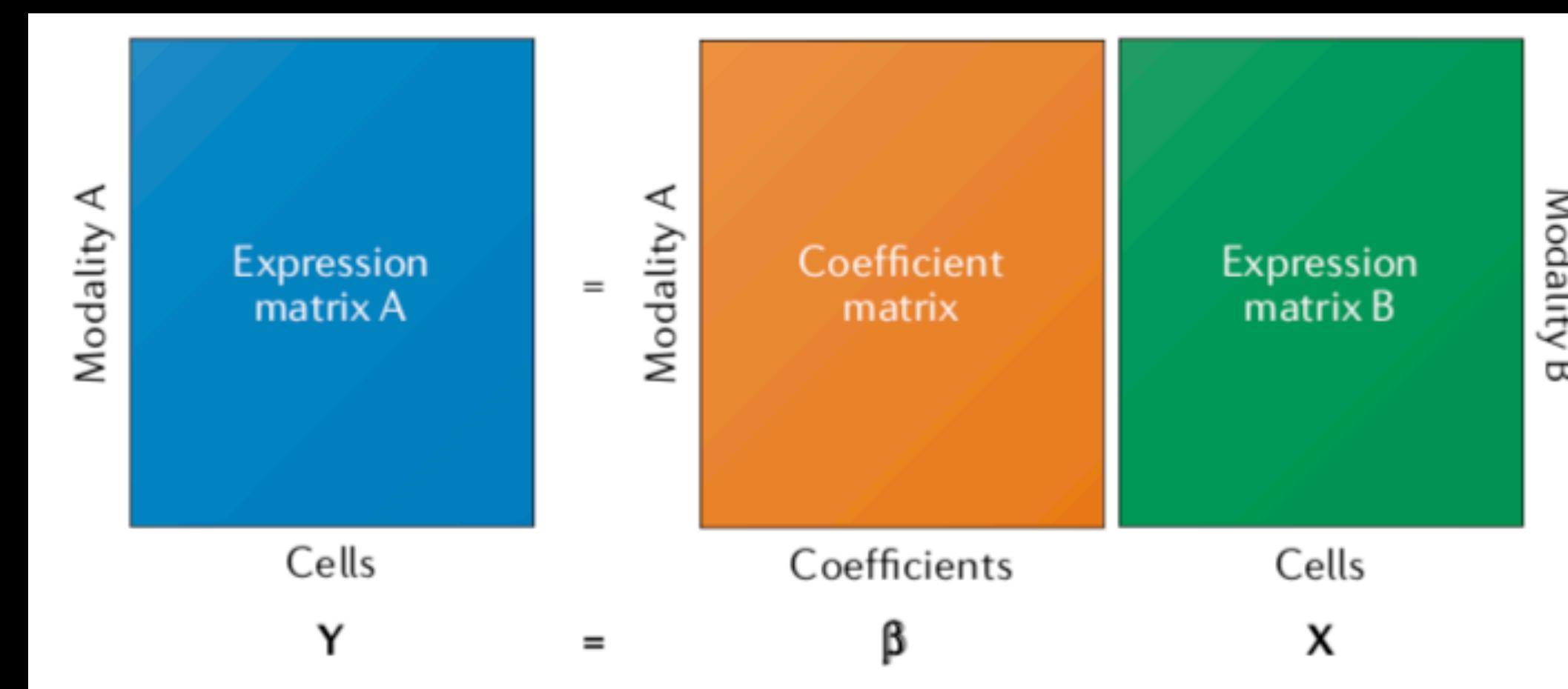
Stuart and Satija, 2019

# Technology

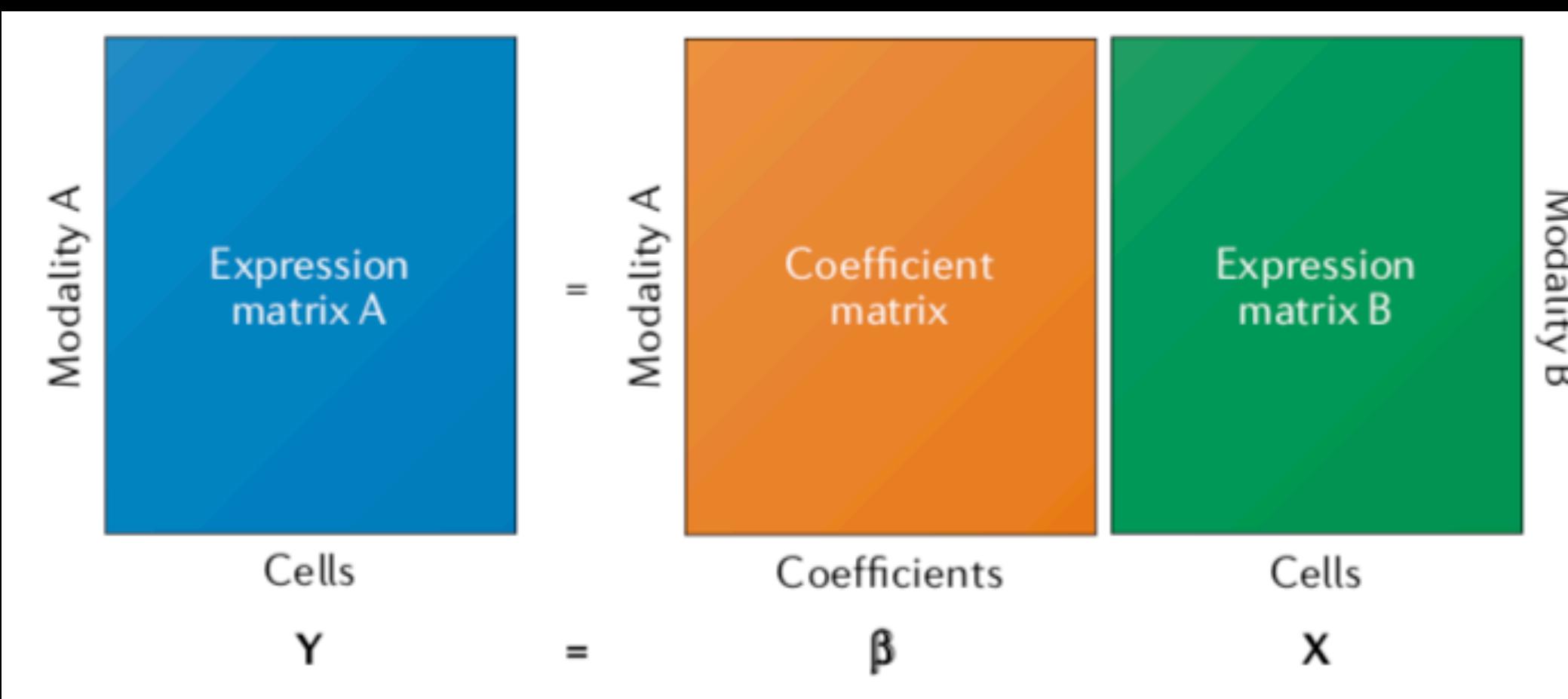
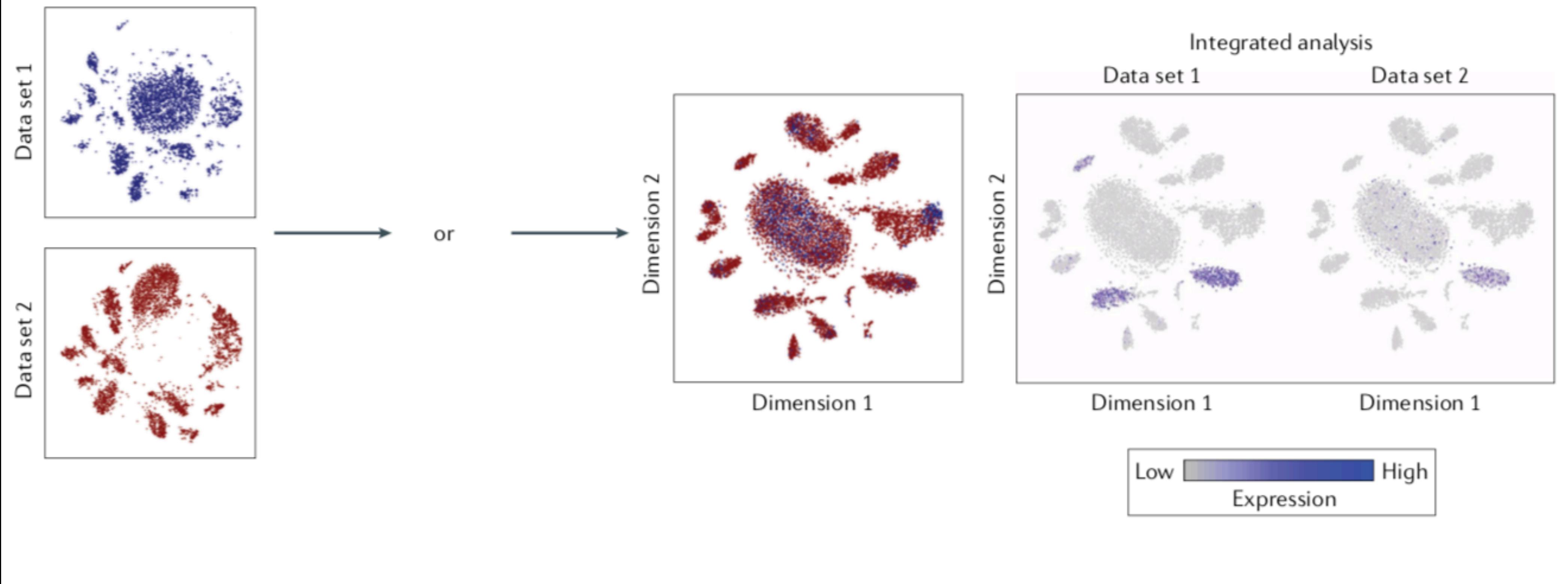
different integration  
strategies

mapping between datasets

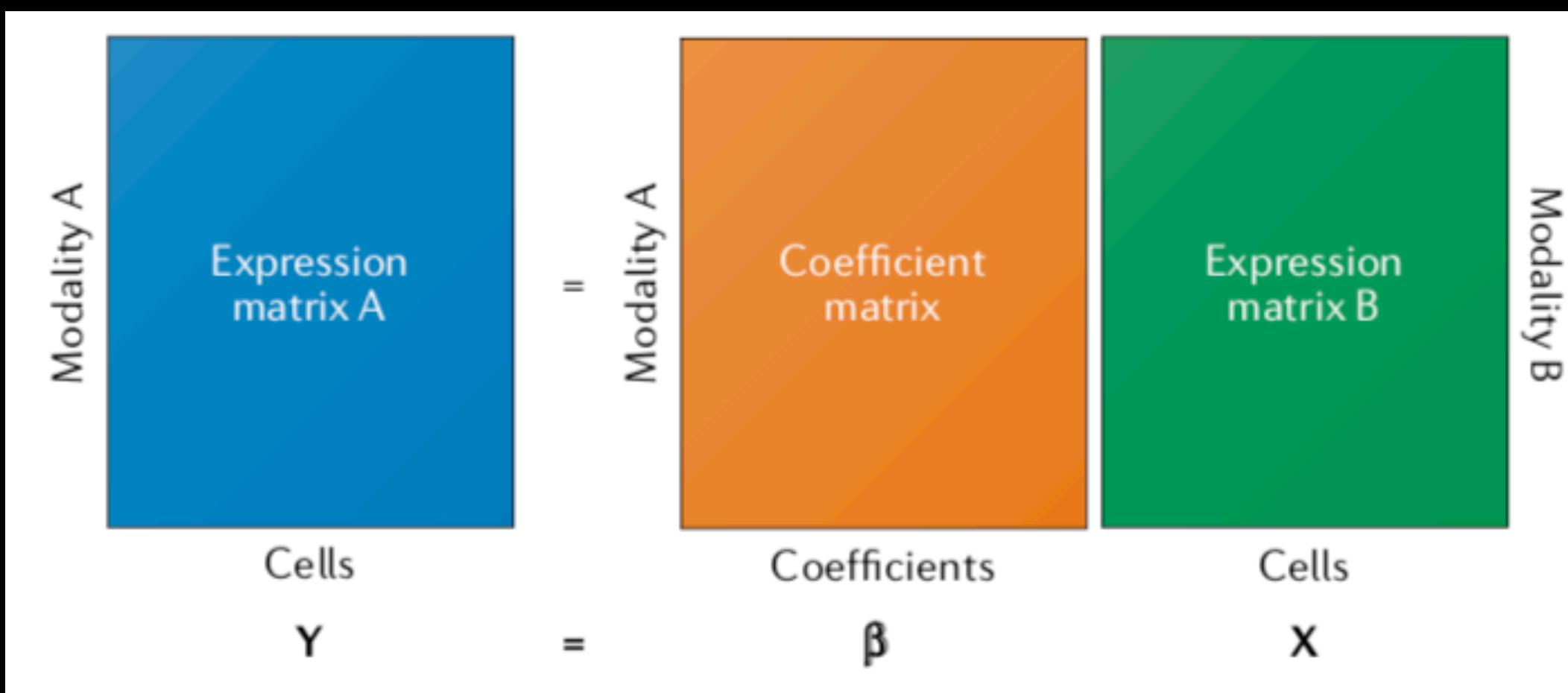
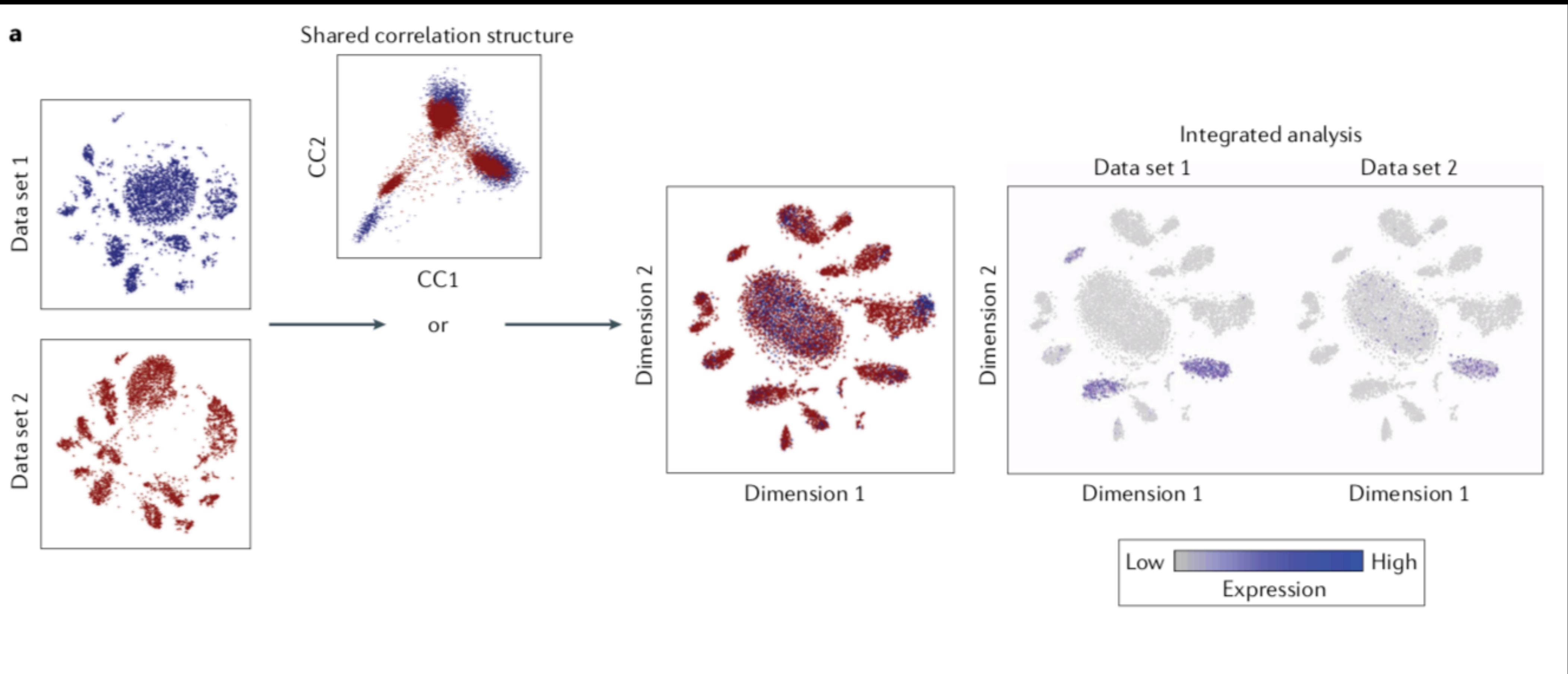
analytical approaches



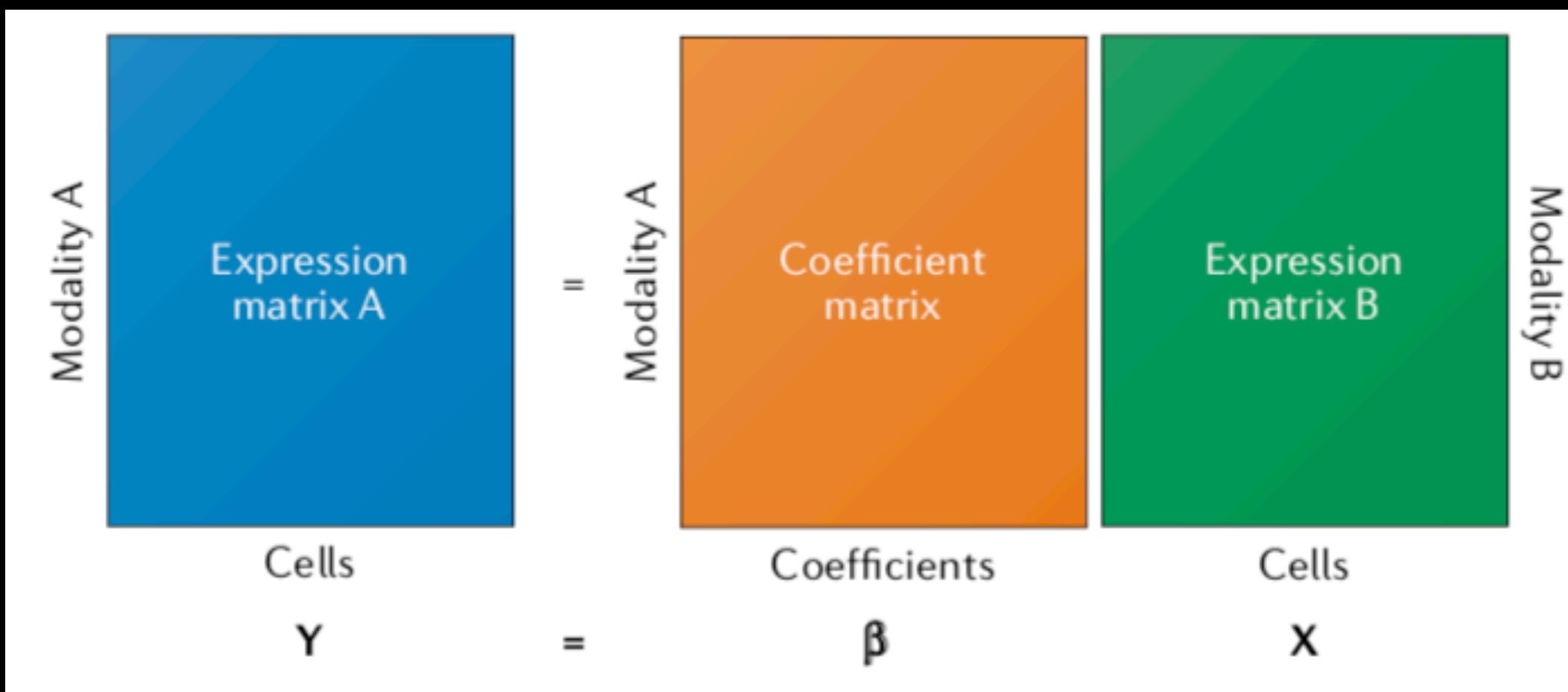
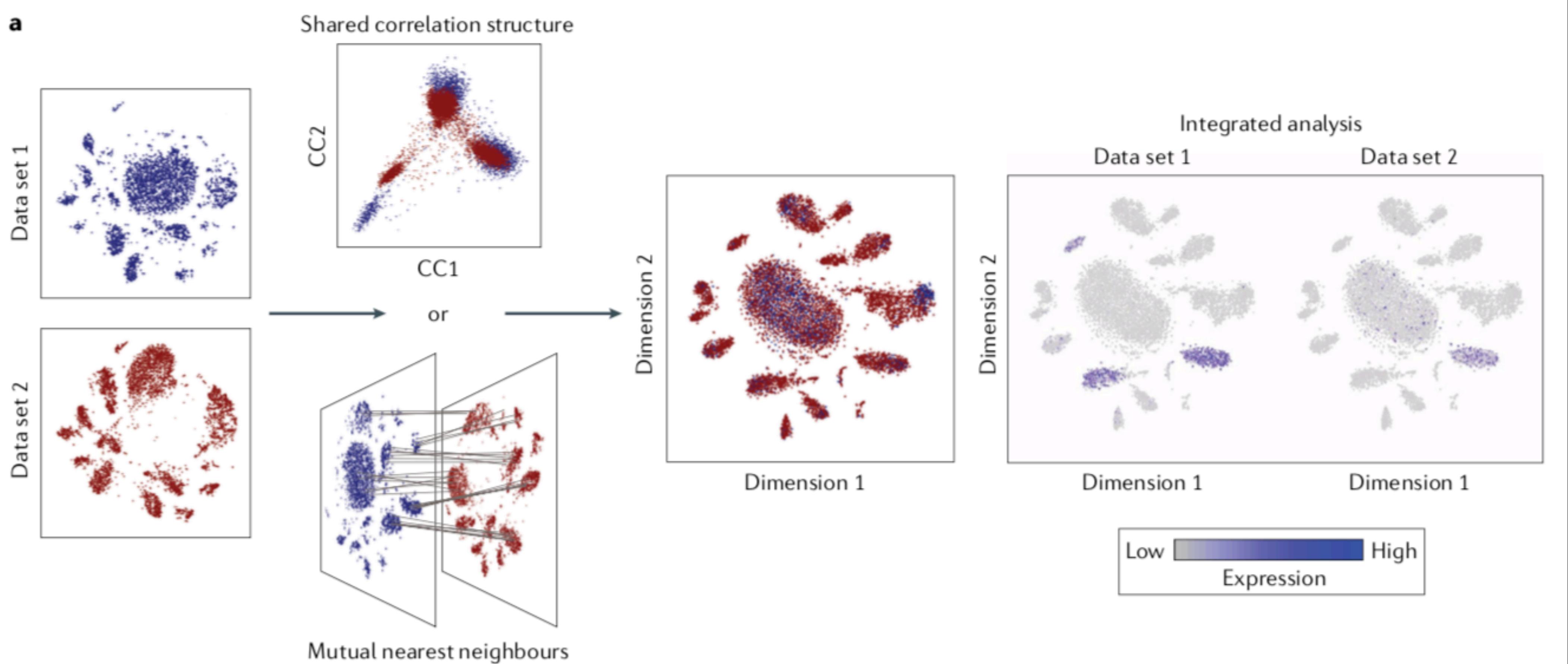
**Assumption: values in one matrix ('-ome') can be derived from values of another matrix ('-ome') through transformations.**

**a**

**Assumption: values in one matrix ('-ome') can be derived from values of another matrix ('-ome') through transformations.**



**Assumption: values in one matrix ('-ome') can be derived from values of another matrix ('-ome') through transformations.**



**Assumption: values in one matrix ('-ome') can be derived from values of another matrix ('-ome') through transformations.**

# Technology

different integration  
strategies

mapping between datasets

analytical approaches

list of tools

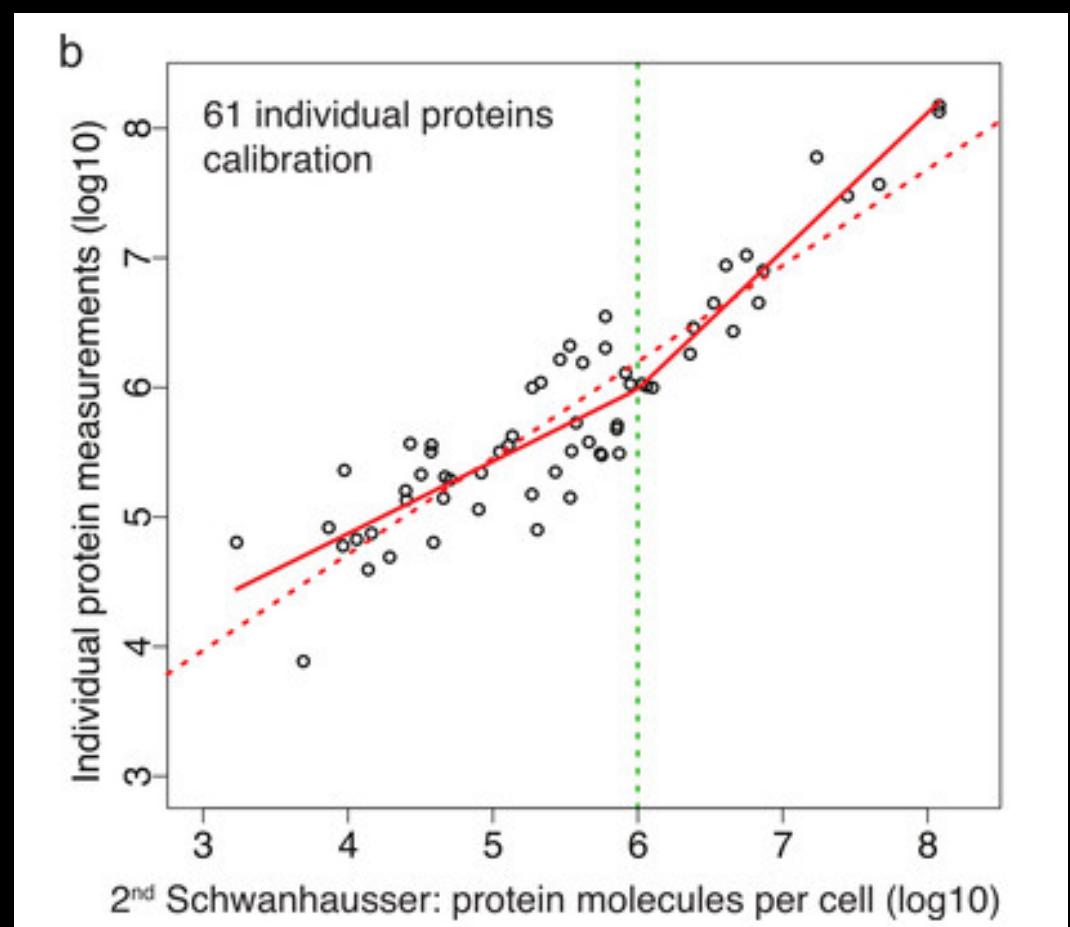
<b>Early integration</b>			<b>Early integration</b>				
LRcluster*	Data originate from low rank matrix, omic data distributions modeled based on it	(16)	R	LRcluster*	Data originate from low rank matrix, omic data distributions modeled based on it	(16)	R
Structured sparsity	Linear transformation projects data into a cluster membership orthogonal matrix	(17)	Matlab	Structured sparsity	Linear transformation projects data into a cluster membership orthogonal matrix	(17)	Matlab
<b>Alternate optimization</b>			<b>Alternate optimization</b>				
MV k-means, MV EM	Alternating $k$ -means and EM. Each iteration is done w.r.t. a different view	(18)	NA	MV k-means, MV EM	Alternating $k$ -means and EM. Each iteration is done w.r.t. a different view	(18)	NA
<b>Late integration</b>			<b>Late integration</b>				
COCA	Per omic clustering solutions integrated with hierarchical clustering	(19)	NA	COCA	Per omic clustering solutions integrated with hierarchical clustering	(19)	NA
Late fusion using latent models	Per omic clustering solutions integrated with PLSA	(20)	NA	Late fusion using latent models	Per omic clustering solutions integrated with PLSA	(20)	NA
PINS*	Integration of co-clustering patterns in different omics. The clusterings are based on perturbations to the data	(21)	R	PINS*	Integration of co-clustering patterns in different omics. The clusterings are based on perturbations to the data	(21)	R
<b>Similarity-based methods</b>			<b>Similarity-based methods</b>				
Spectral clustering generalizations	Generalizations of similarity based spectral clustering to multi-omics data	(22–25)	Matlab	Spectral clustering generalizations	Generalizations of similarity based spectral clustering to multi-omics data	(22–25)	Matlab
Spectral clustering with random walks	Generalizations of spectral clustering by random walks across similarity graphs	(26,27)	NA	Spectral clustering with random walks	Generalizations of spectral clustering by random walks across similarity graphs	(26,27)	NA
SNF*	Integration of similarity networks by message passing	(28,29)	R, Matlab	SNF*	Integration of similarity networks by message passing	(28,29)	R, Matlab
rMKL-LPP*	DR using multiple kernel learning; similarities maintained in lower dimension	(30)	**	rMKL-LPP*	DR using multiple kernel learning; similarities maintained in lower dimension	(30)	**
<b>Dimension reduction</b>			<b>Dimension reduction</b>				
General DR framework	General framework for integration with DR	(31)	NA	General DR framework	General framework for integration with DR	(31)	NA
JIVE	Variation in data partitioned into joint and omic-specific	(32)	Matlab,R (33)	JIVE	Variation in data partitioned into joint and omic-specific	(32)	Matlab,R (33)
CCA*	DR to axes of max correlation between datasets. Generalizations: Bayesian, kernels, >2 omics, sparse solutions, deep learning, count data	(34–43), CCA for count data	R, two omics (44), R, multiple omics	CCA*	DR to axes of max correlation between datasets. Generalizations: Bayesian, kernels, >2 omics, sparse solutions, deep learning, count data	(34–43), CCA for count data	R, two omics (44), R, multiple omics
PLS	DR to axes of max covariance between datasets. Generalizations: kernels, >2 omics, sparse solutions, partition into omic-specific and joint variation	(45–52)	R, two omics, Matlab, multiple omics	PLS	DR to axes of max covariance between datasets. Generalizations: kernels, >2 omics, sparse solutions, partition into omic-specific and joint variation	(45–52)	R, two omics, Matlab, multiple omics
MCIA	DR to axes of max covariance between multi-omic datasets	(53)	R	MCIA	DR to axes of max covariance between multi-omic datasets	(53)	R
NMF generalizations*	DR using generalizations of NMF to multi-omic data	(54–57), EquiNMF, (58,59)	MultiNMF (Matlab)	NMF generalizations*	DR using generalizations of NMF to multi-omic data	(54–57), EquiNMF, (58,59)	MultiNMF (Matlab)
Matrix tri-factorization	DR. Each omic describes the relationship between two entities	(60)	NA	Matrix tri-factorization	DR. Each omic describes the relationship between two entities	(60)	NA
Convex methods	DR with convex objective functions, allowing unique optimum and efficient computation	(16,61,62)	Matlab	Convex methods	DR with convex objective functions, allowing unique optimum and efficient computation	(16,61,62)	Matlab
Low-rank tensor MV clustering	Factorization based on low-rank tensors	(63)	Matlab	Low-rank tensor MV clustering	Factorization based on low-rank tensors	(63)	Matlab
<b>Statistical methods</b>			<b>Statistical methods</b>				
iCluster/Plus/Bayes*	Data originate from low dimensional representation, which determines the distribution of the observed data	(15,64,65)	R	iCluster/Plus/Bayes*	Data originate from low dimensional representation, which determines the distribution of the observed data	(15,64,65)	R
PARADIGM	Probabilistic model of cellular pathways using factor graphs	(66)	REST API	PARADIGM	Probabilistic model of cellular pathways using factor graphs	(66)	REST API
Disagreement between clusters	Methods based mainly on hierarchical Dirichlet processes; clustering in different omics need not agree	(67–71)	BCC (R)	Disagreement between clusters	Methods based mainly on hierarchical Dirichlet processes; clustering in different omics need not agree	(67–71)	BCC (R)
Survival-based	Probabilistic model; patient survival data used in the clustering process	(72,73)	SBC (R)	Survival-based	Probabilistic model; patient survival data used in the clustering process	(72,73)	SBC (R)
<b>Deep learning</b>			<b>Deep learning</b>				
Deep learning methods	Neural networks used for integration. A variant of CCA, early integration and middle integration	(37,74,75)	DeepCCA (Python)	Deep learning methods	Neural networks used for integration. A variant of CCA, early integration and middle integration	(37,74,75)	DeepCCA (Python)

Rappoport and Shamir, 2018

Great list of tools:

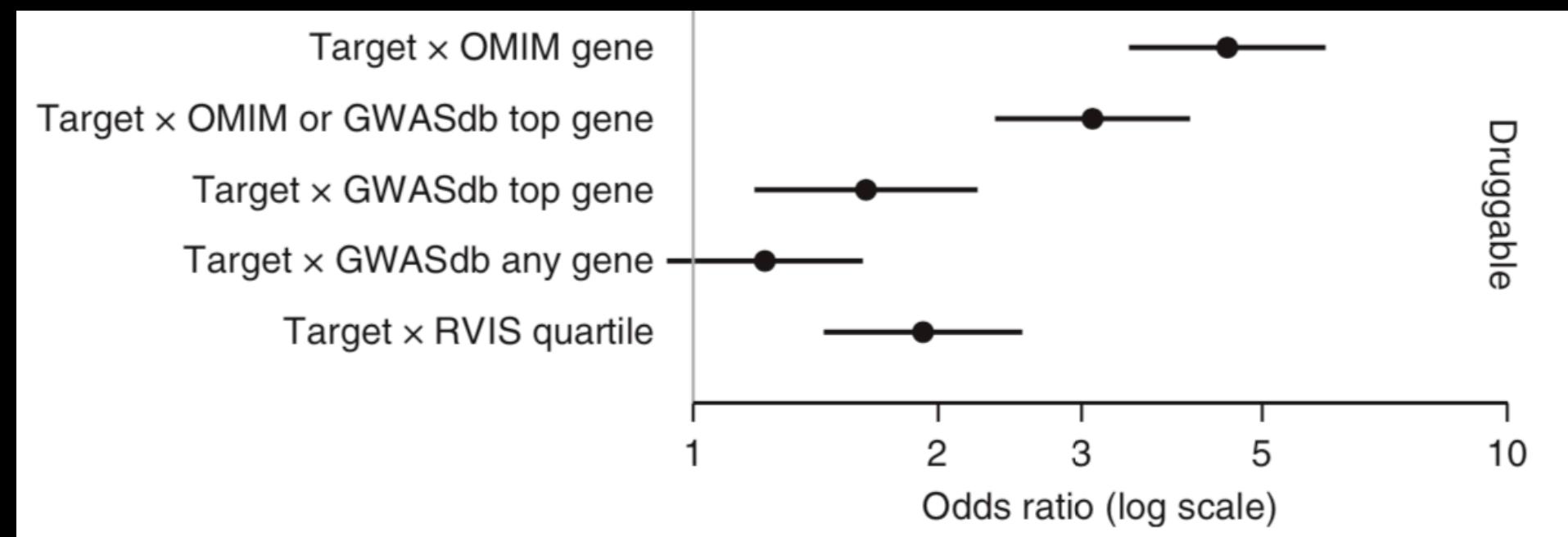
[github.com/mikelove/awesome-multi-omics](https://github.com/mikelove/awesome-multi-omics)

## Relation between transcriptome and proteome.



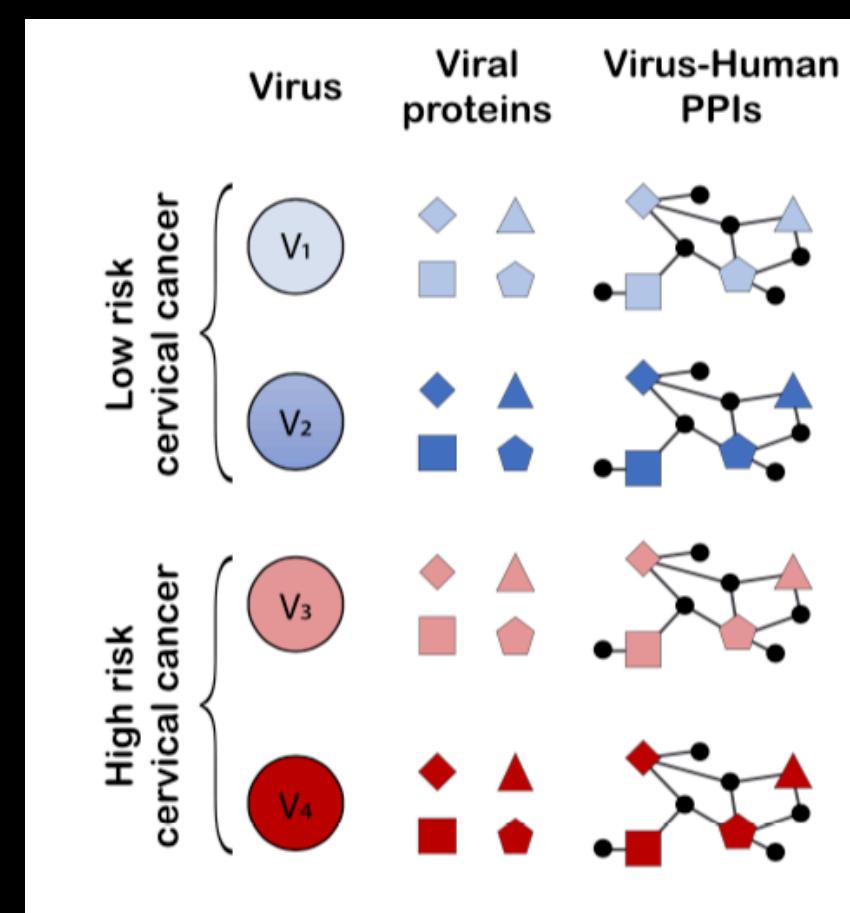
Vogel et al. 2010; Li et al. 2014f

## Support for drug targets.



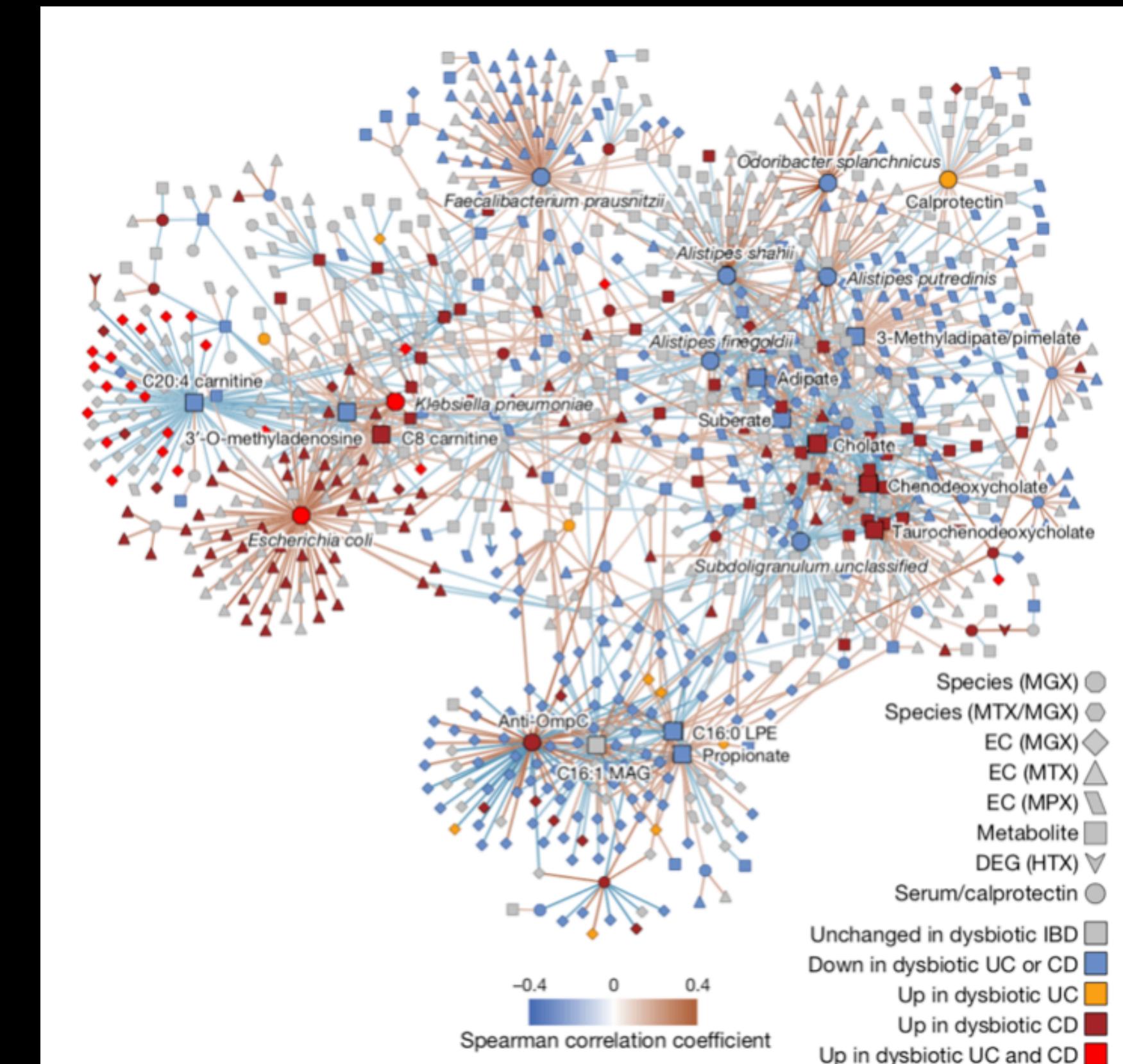
Nelson et al. 2014

## Disease-causing strains.



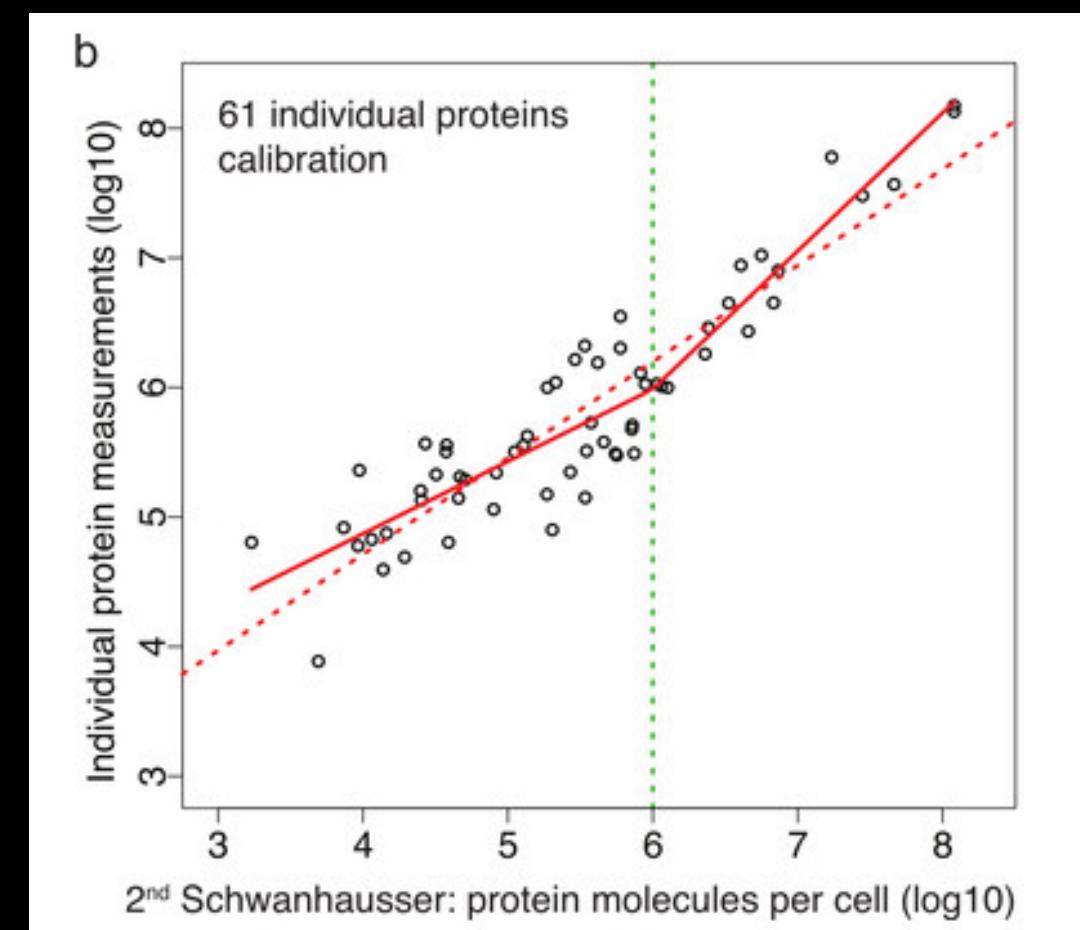
Lasso et al. 2019

## Hypotheses for host-microbiome-metabolome interactions



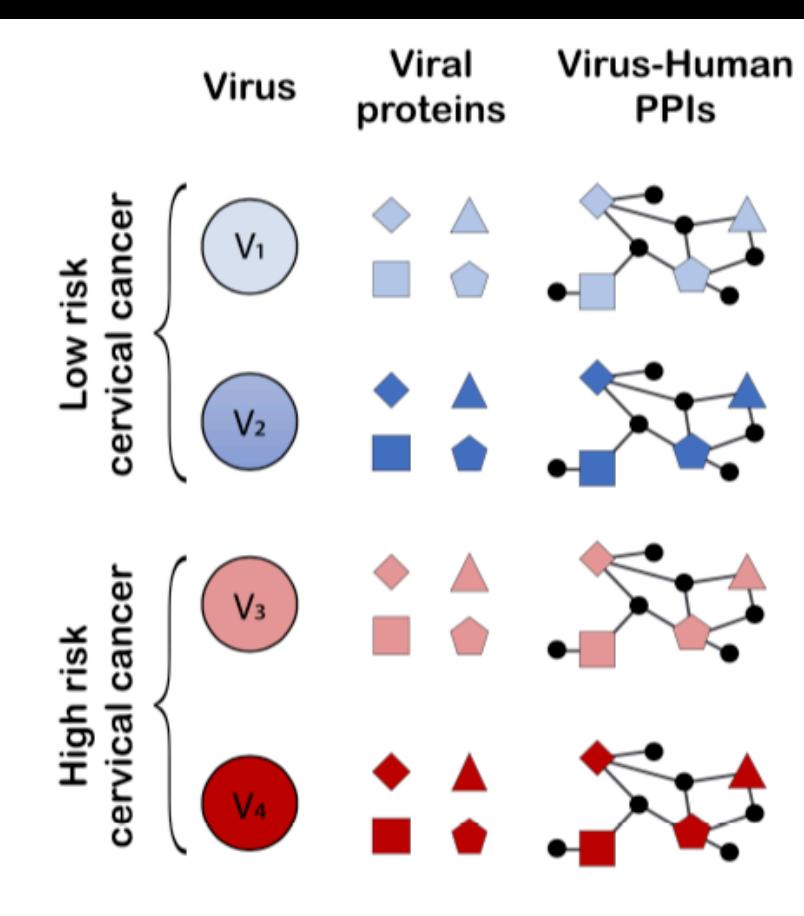
Lloyd-Price et al. 2019

## Relation between transcriptome and proteome.



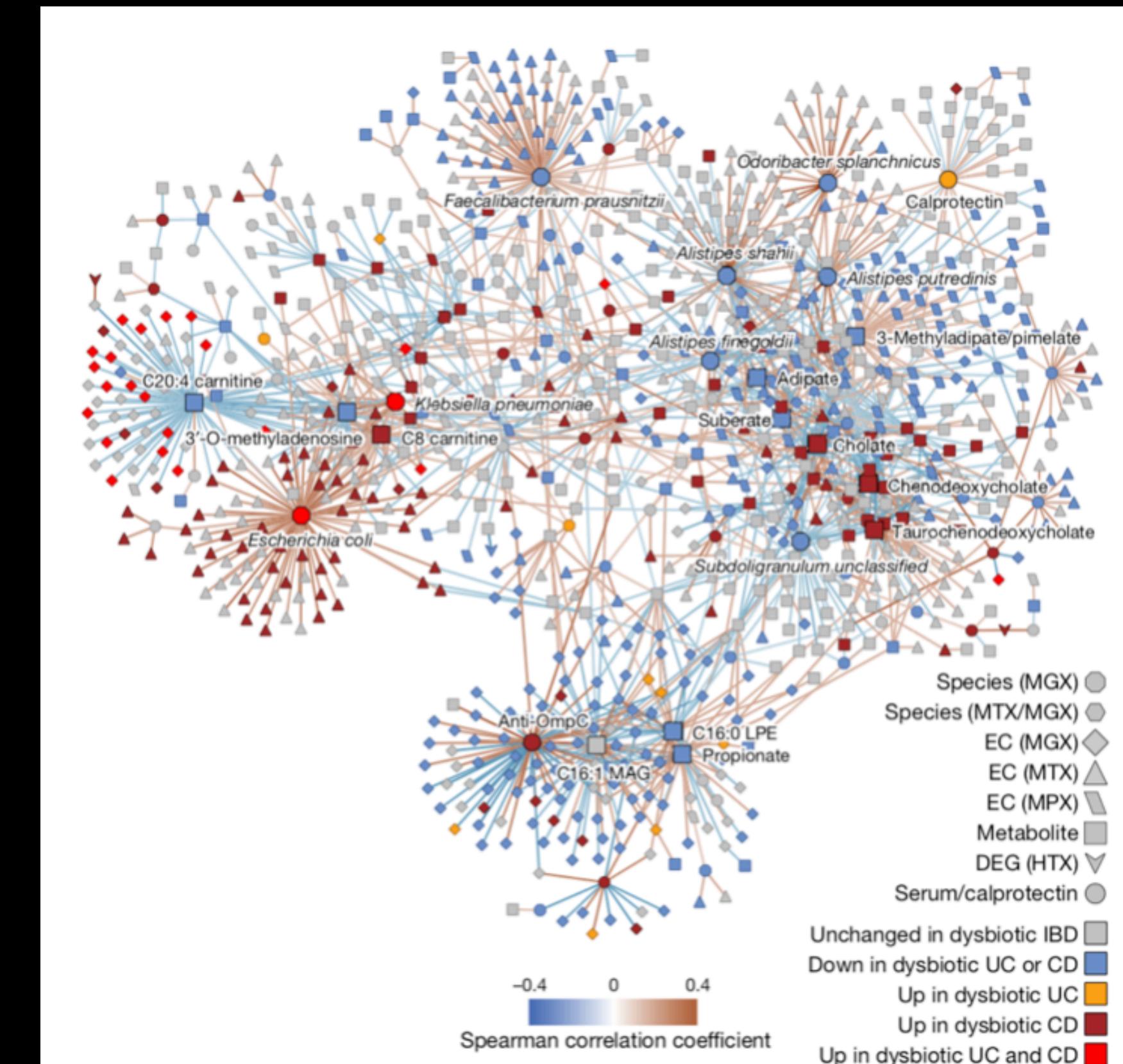
Vogel et al. 2010; Li et al. 2014f

## Disease-causing strains.



Lasso et al. 2019

## Hypotheses for host-microbiome-metabolome interactions



Lloyd-Price et al. 2019

Nelson et al. 2014

-> “standard” tools don’t fit many scientific questions.

# Technology

understanding limits

# Technology

practical

understanding limits

# Technology

practical

statistical

understanding limits

# Technology

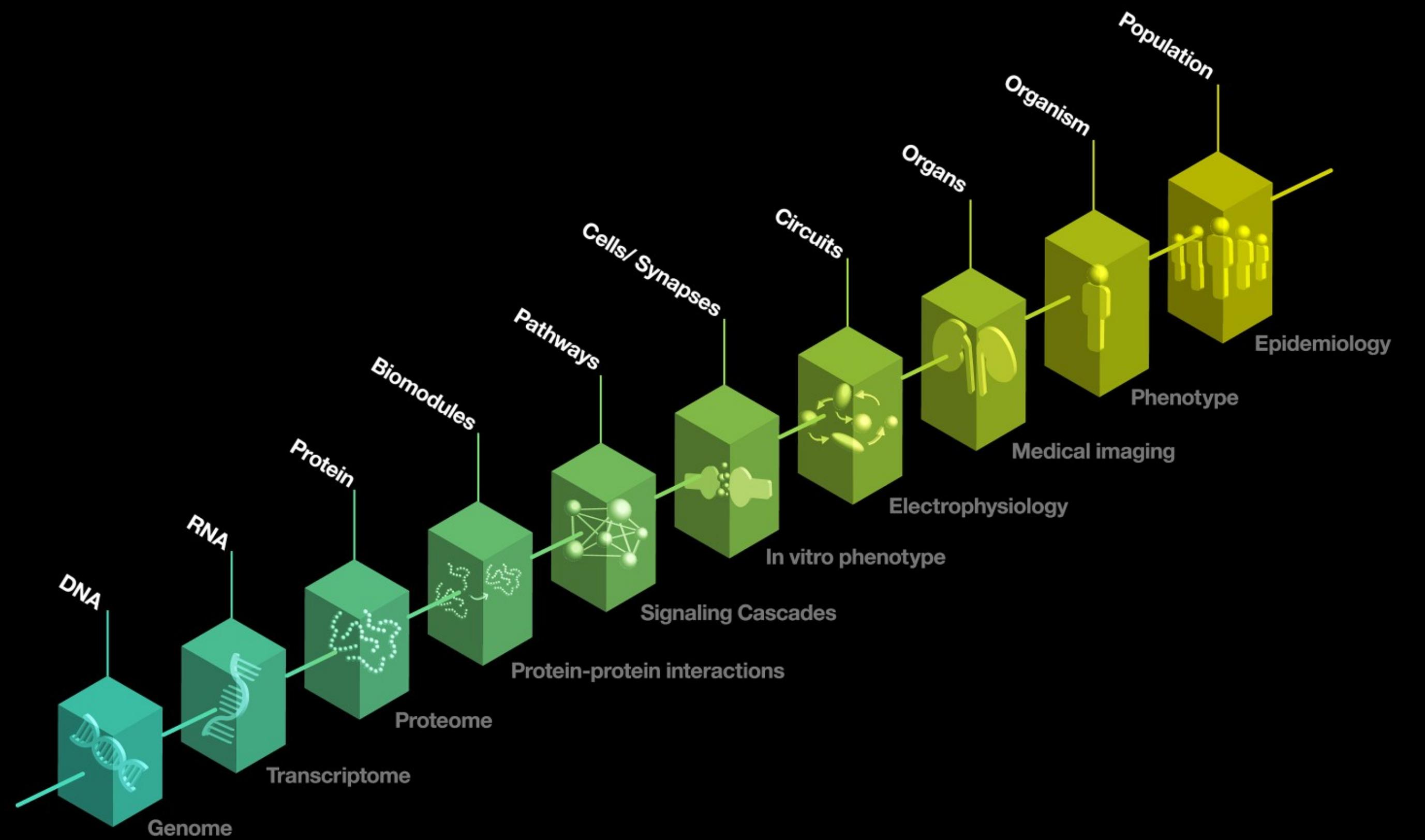
practical

statistical

understanding limits

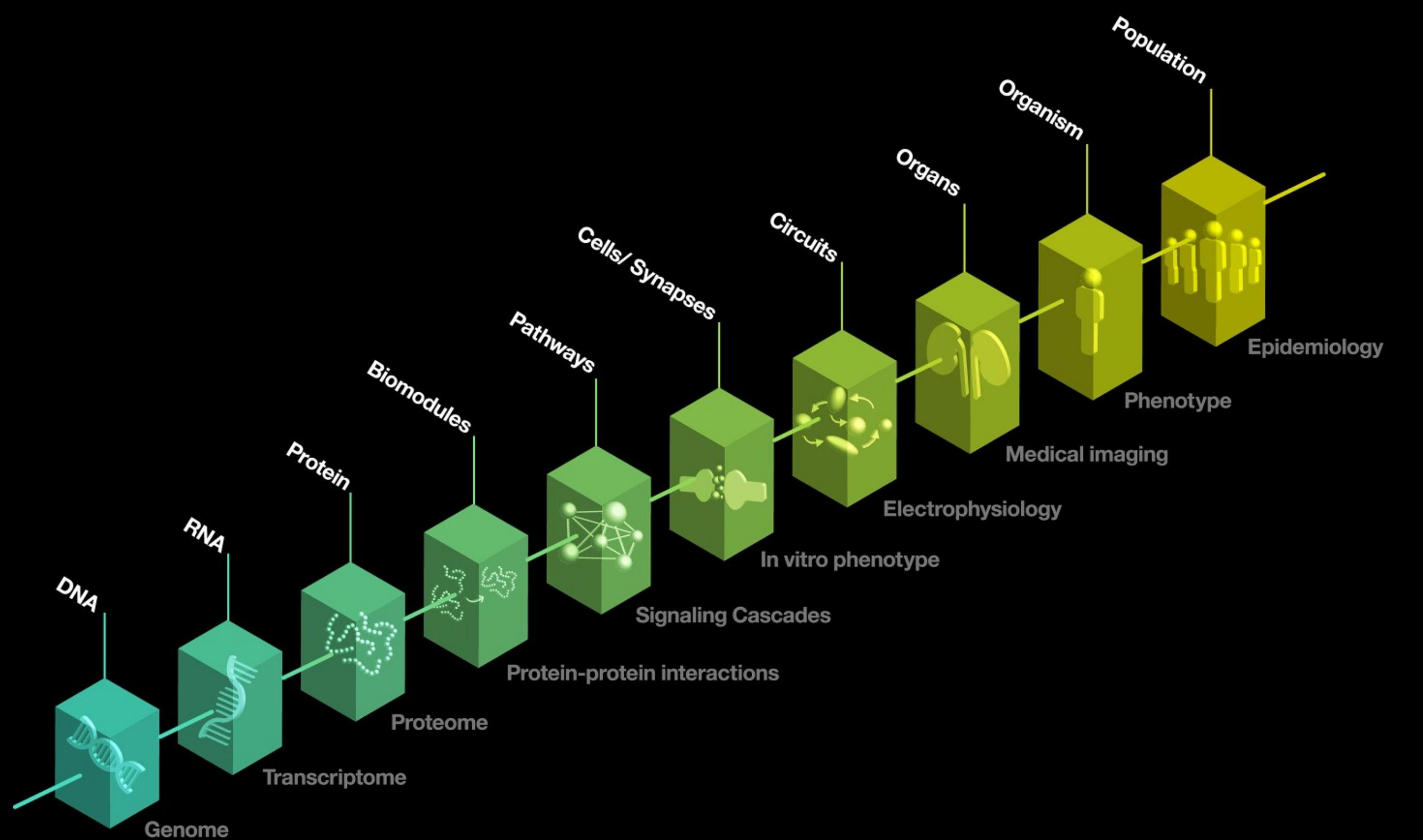
choosing methods

practical



practical

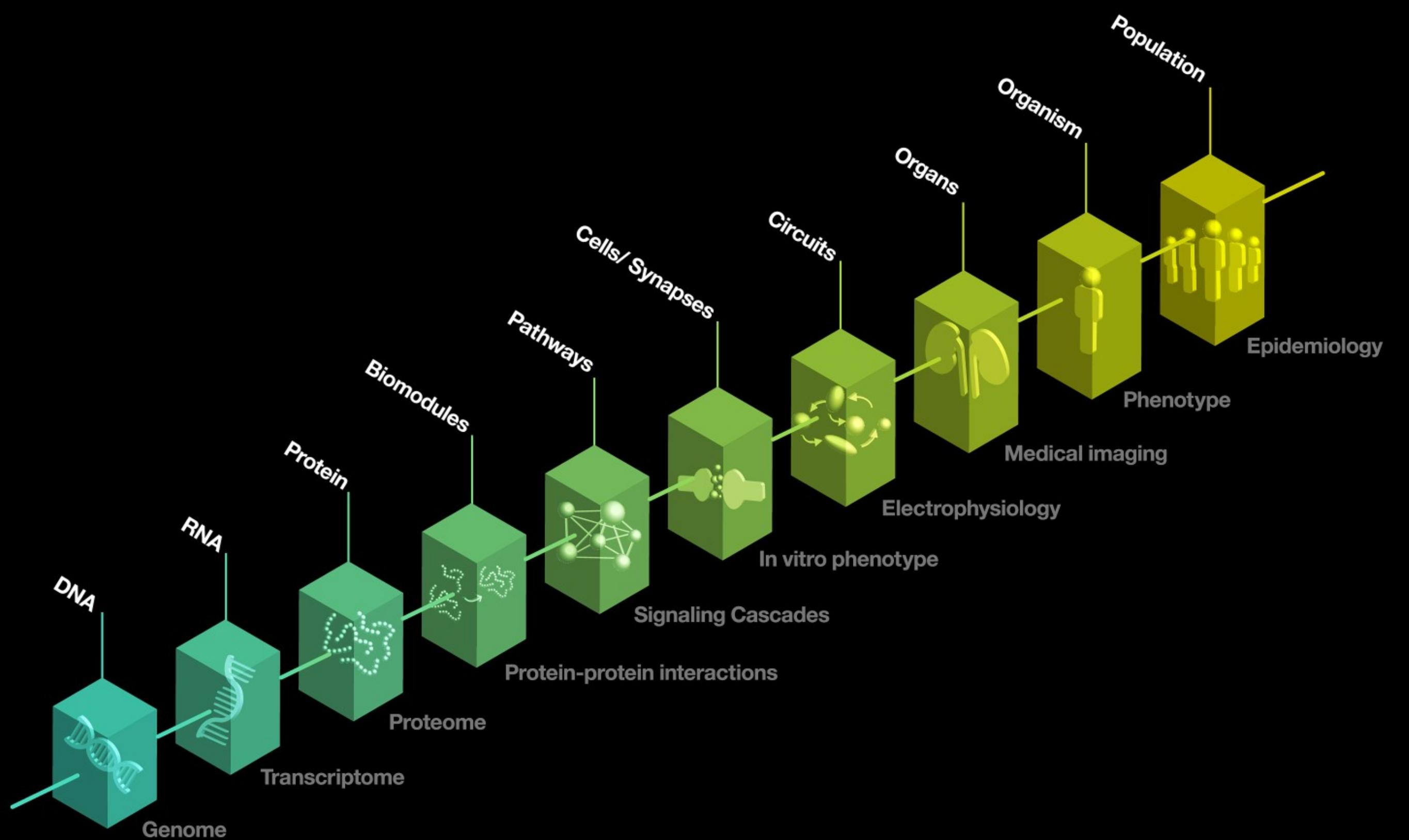
# expensive



practical

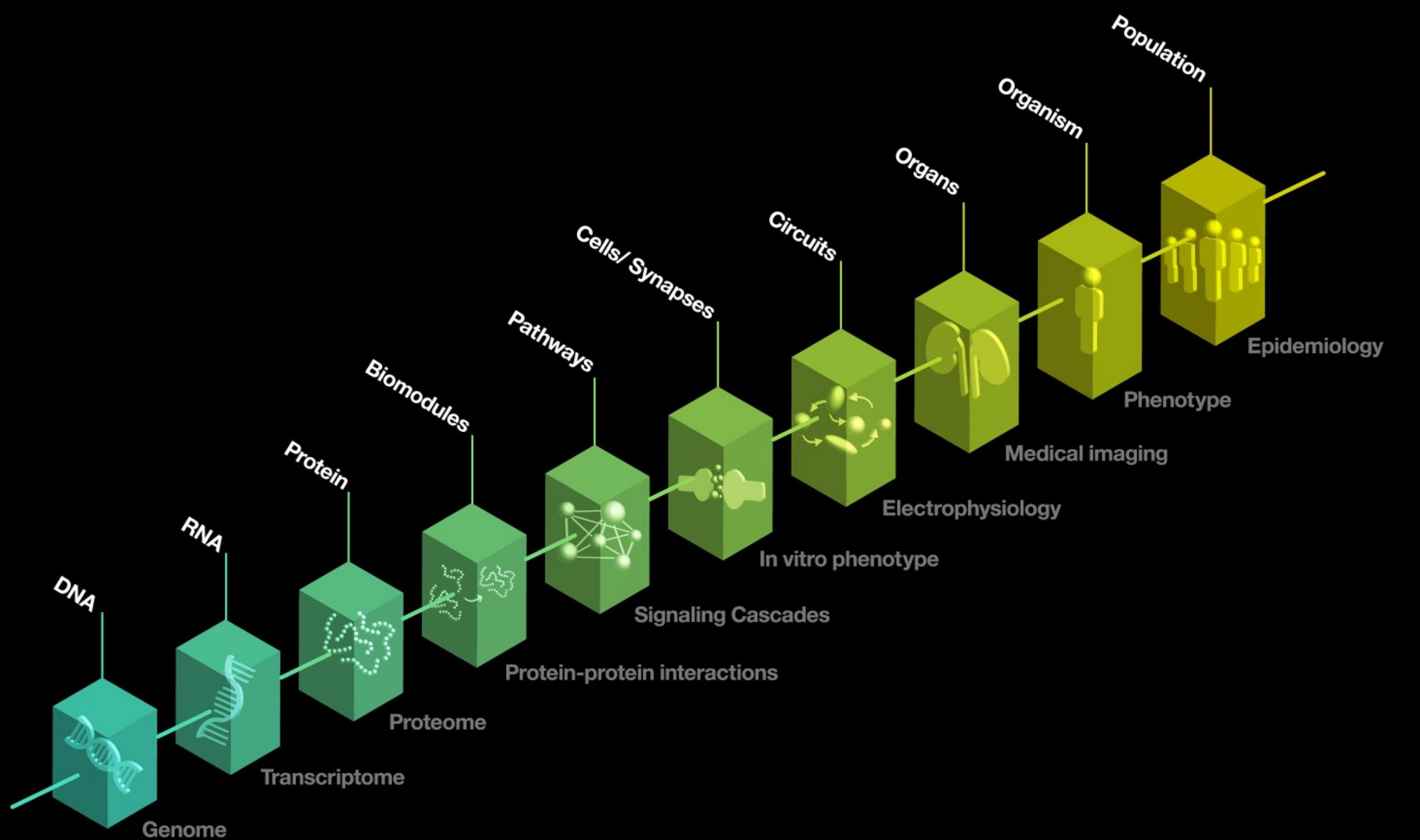
expensive

non-comparable sources



practical

practical

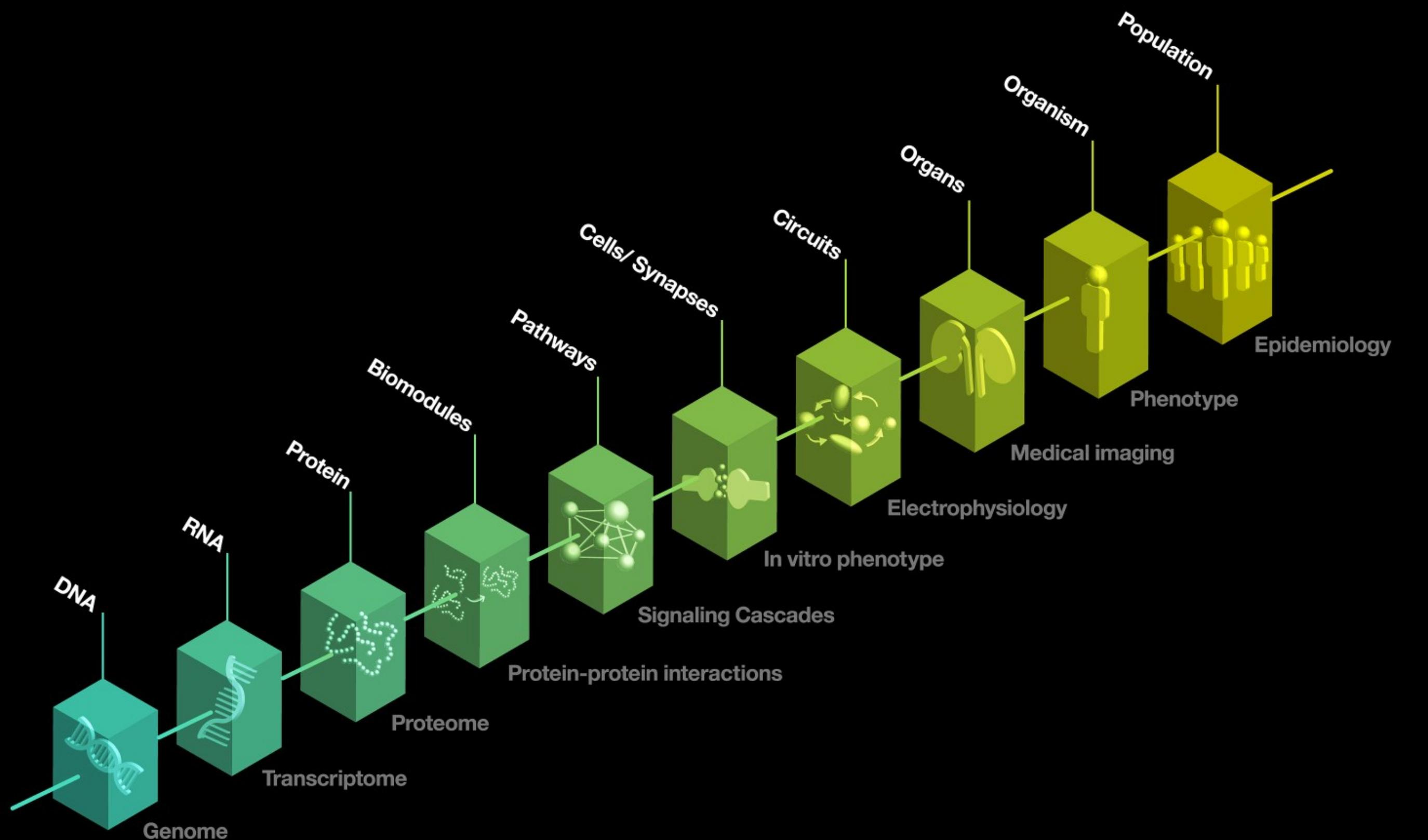


expensive

non-comparable sources

often not more useful

practical



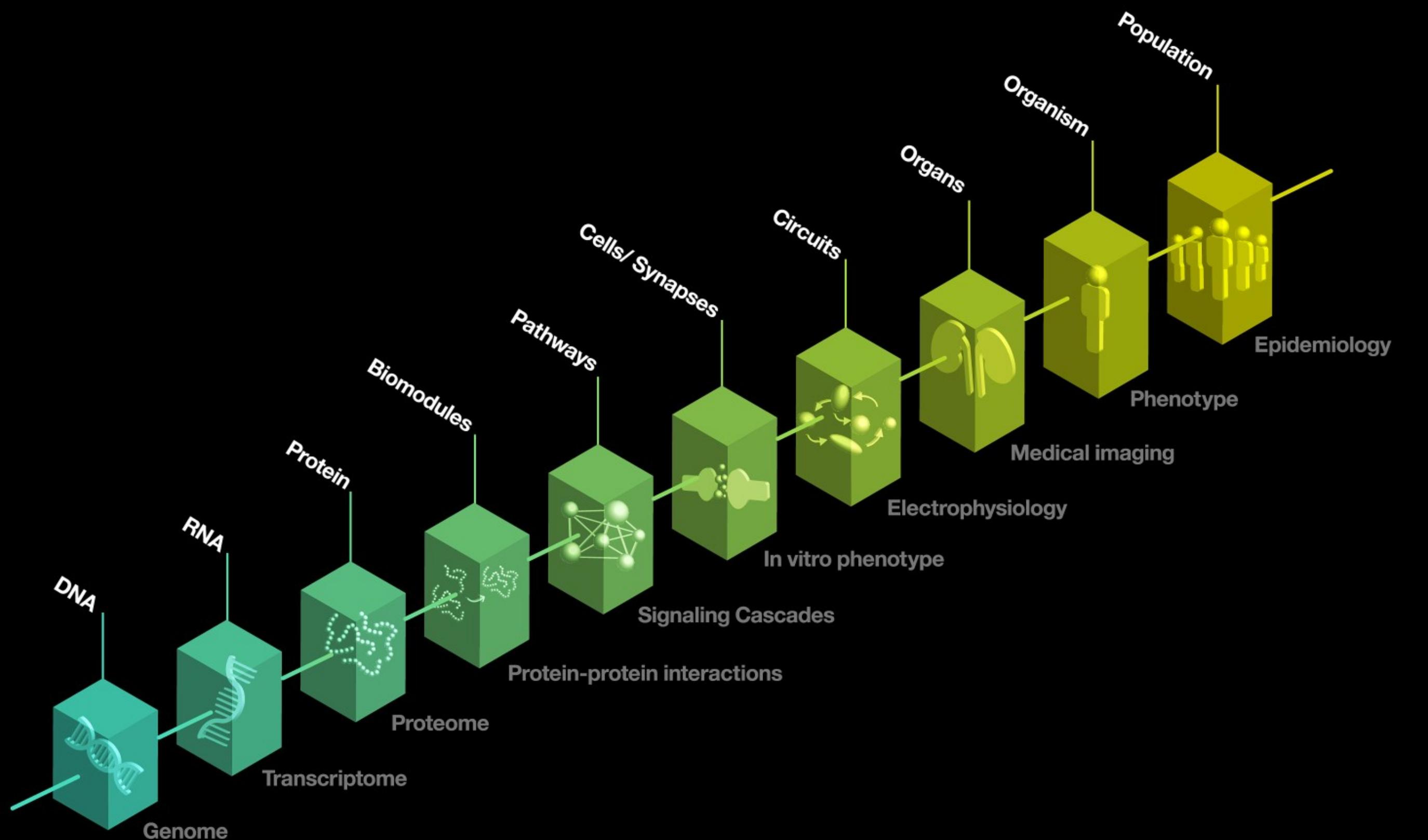
expensive

non-comparable sources

often not more useful

can require manual  
curation

practical



expensive

non-comparable sources

often not more useful

can require manual  
curation

needs many expertises

statistical

few biological  
replicates

The “**curse of -omics**” particularly  
haunts **multi-omics**.

many genes



statistical

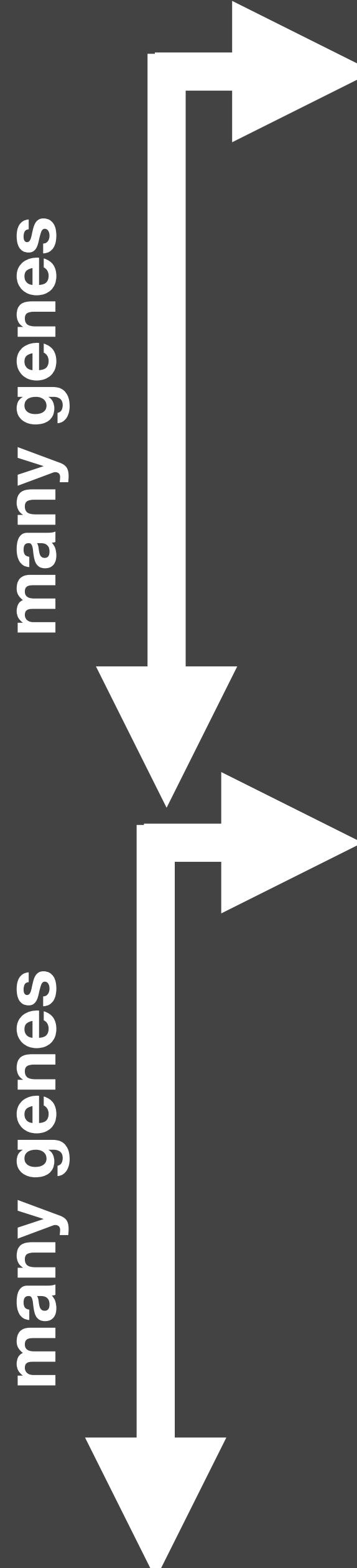
few biological  
replicates

The “**curse of -omics**” particularly  
haunts **multi-omics**.

many genes

many genes

statistical



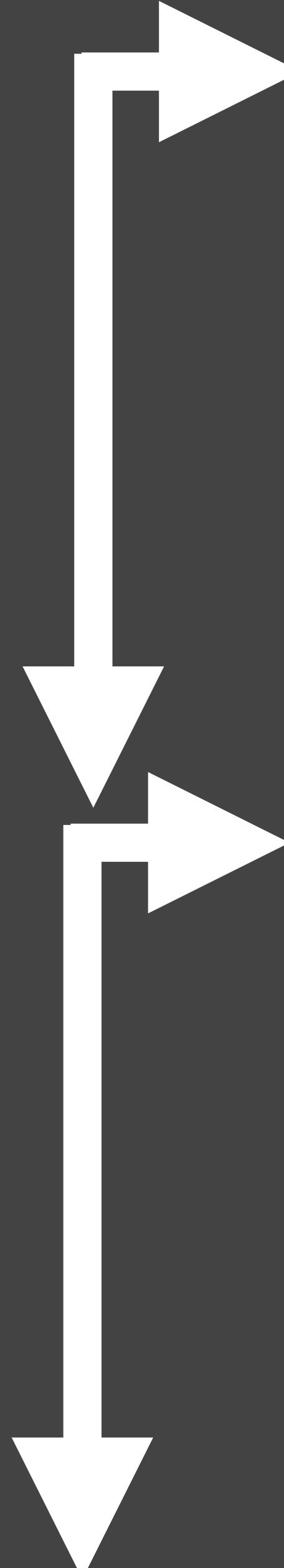
**few biological  
replicates**

The “**curse of -omics**” particularly  
haunts **multi-omics**.

**many genes**

**many genes**

**statistical**



**common work-arounds:**

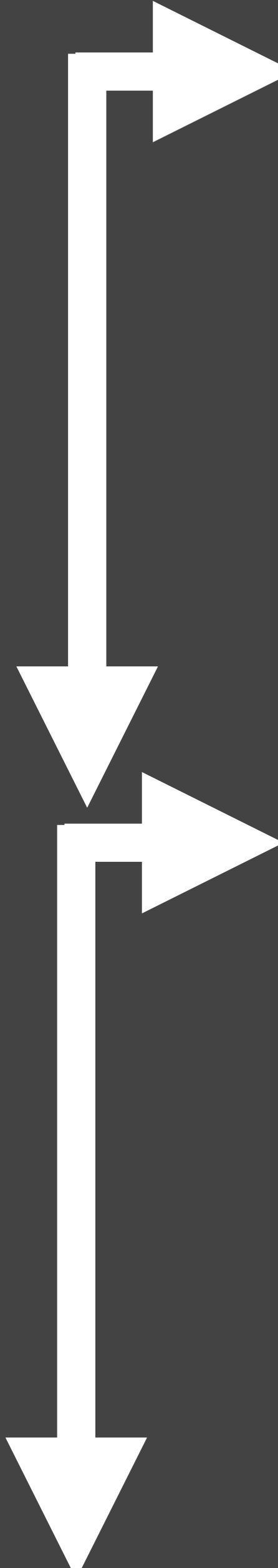
**few biological  
replicates**

The “**curse of -omics**” particularly  
haunts **multi-omics**.

**many genes**

**many genes**

**statistical**



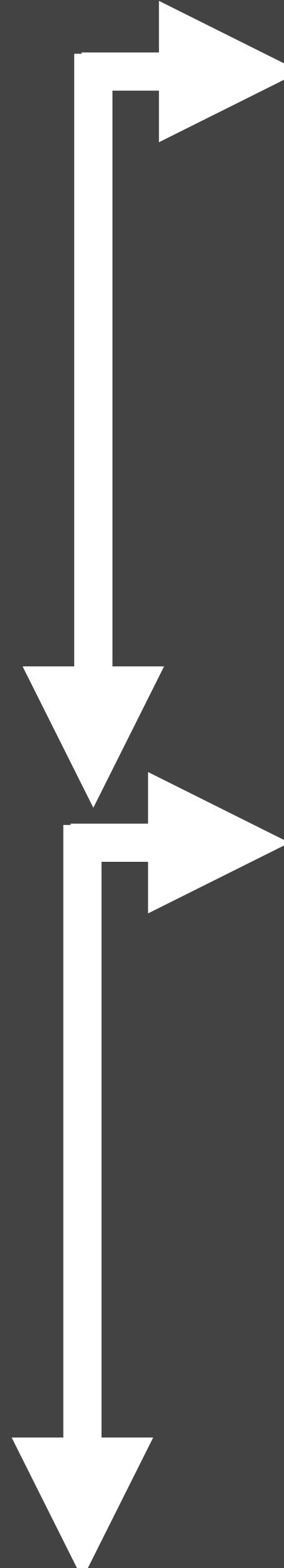
**common work-arounds:**

**regularization**

**few biological  
replicates**

The “**curse of -omics**” particularly  
haunts **multi-omics**.

**many genes**



**many genes**

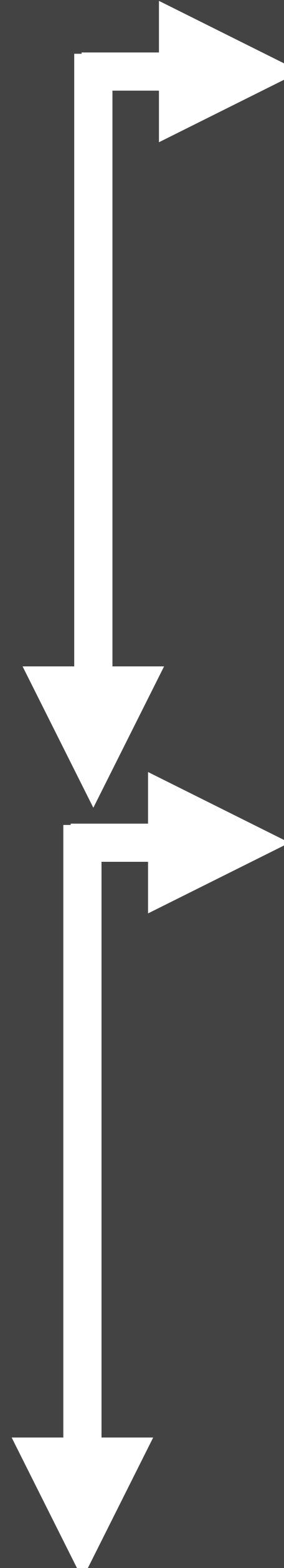
**statistical**

**common work-arounds:**  
**regularization**  
**dimensionality reduction**

**few biological  
replicates**

The “**curse of -omics**” particularly  
haunts **multi-omics**.

**many genes**



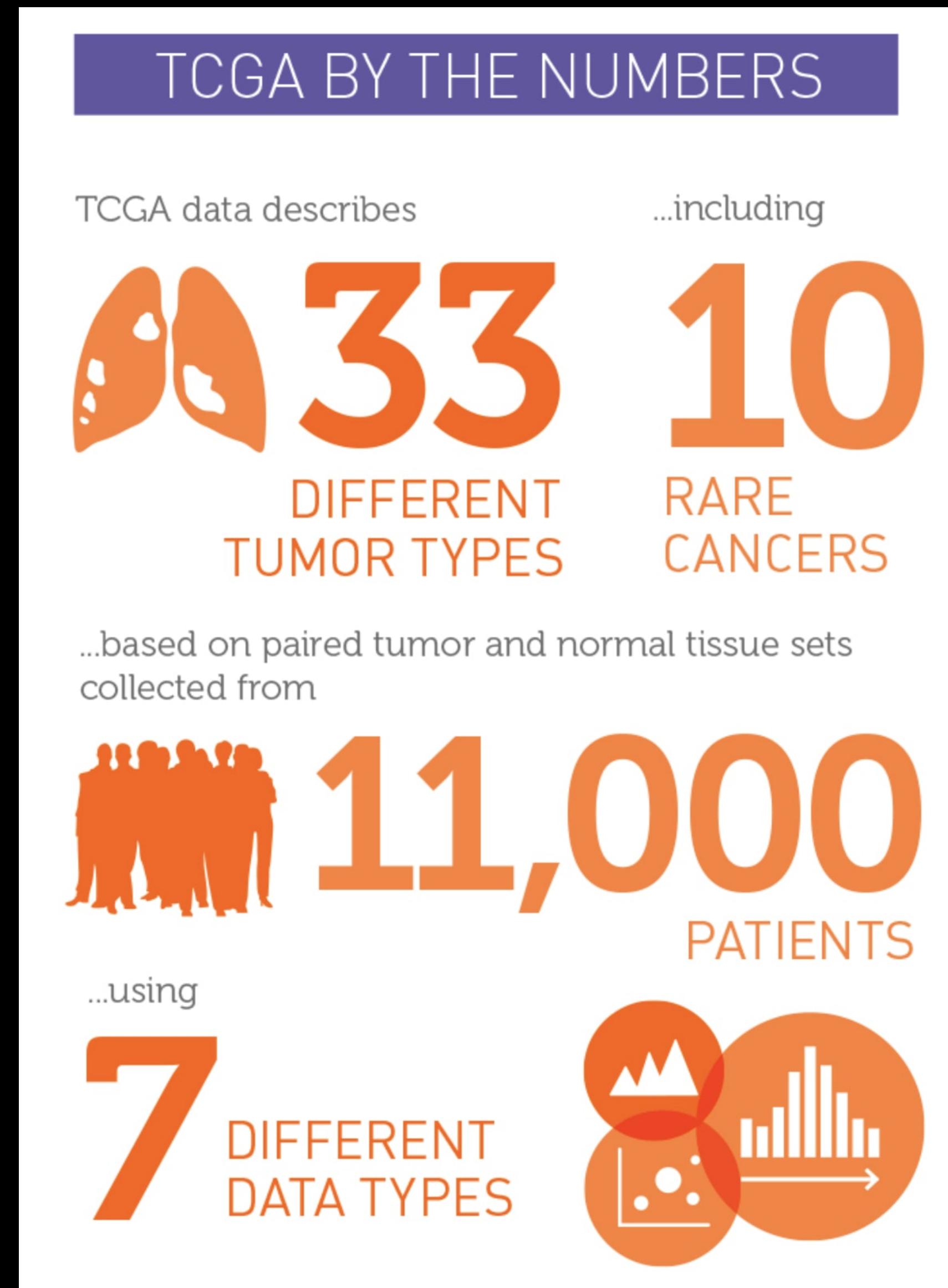
**common work-arounds:**

**regularization**  
**dimensionality reduction**  
**more observations**

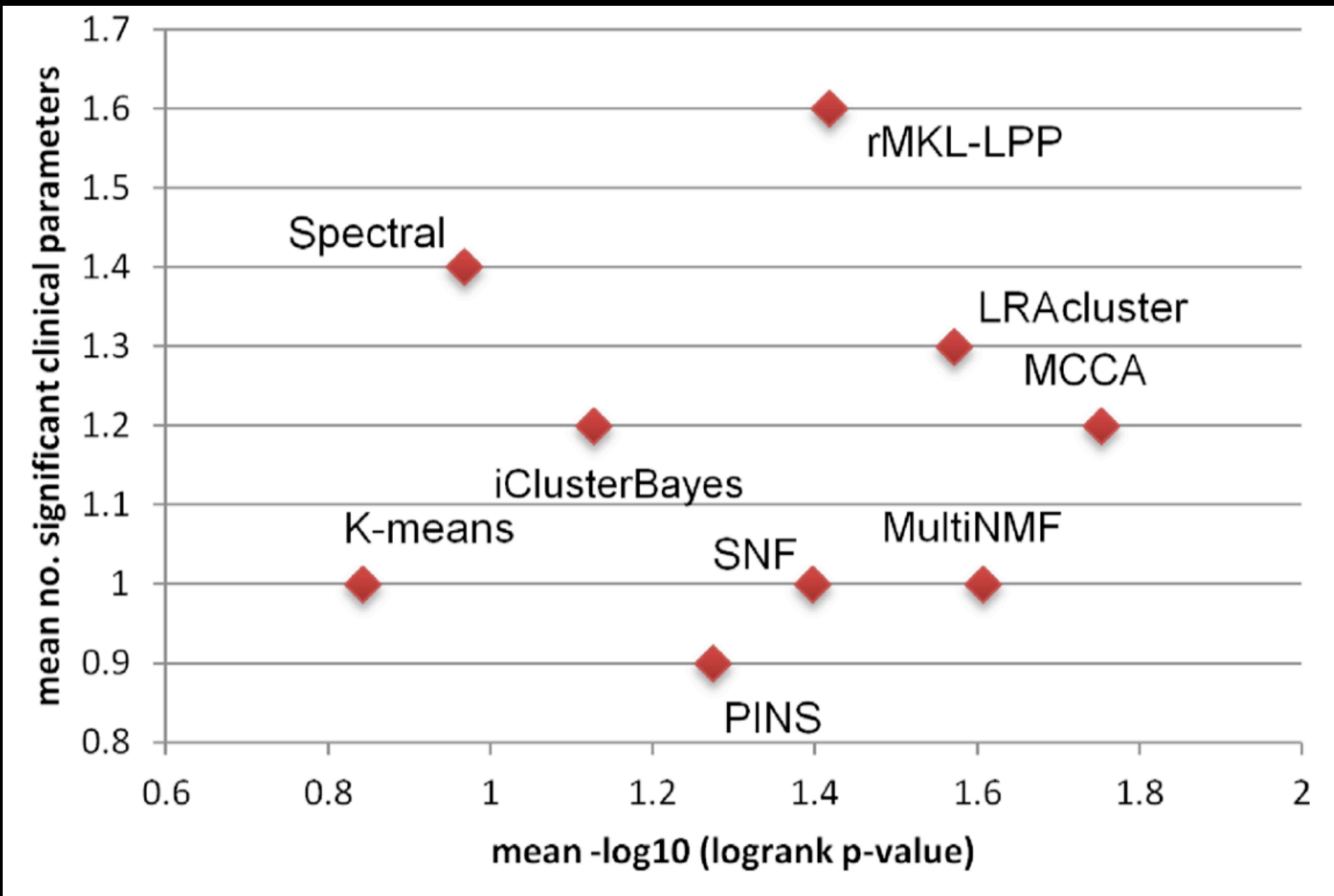
**Data of the Cancer Genome Atlas allows  
to benchmark different multi-omics approaches.**

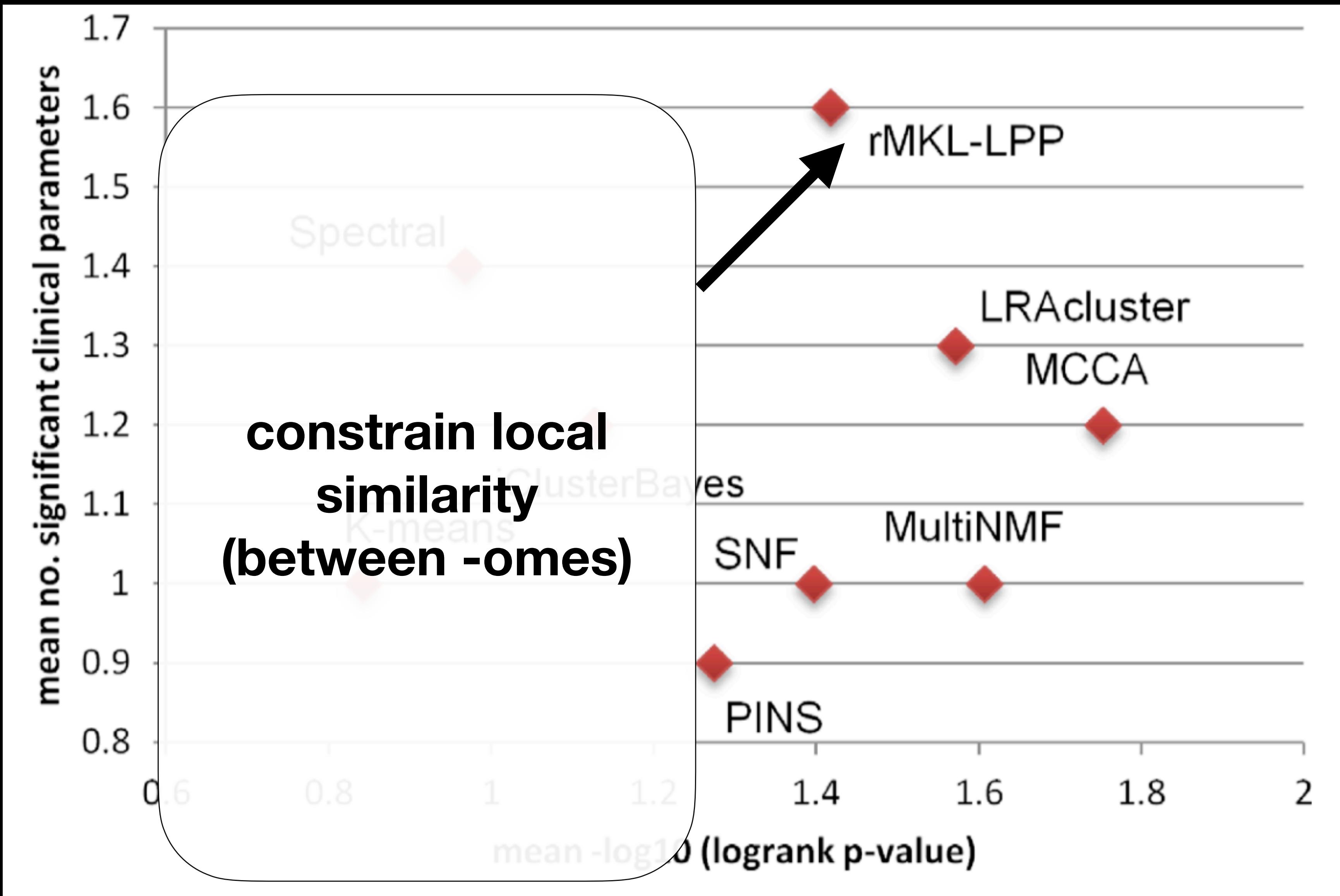
choosing methods

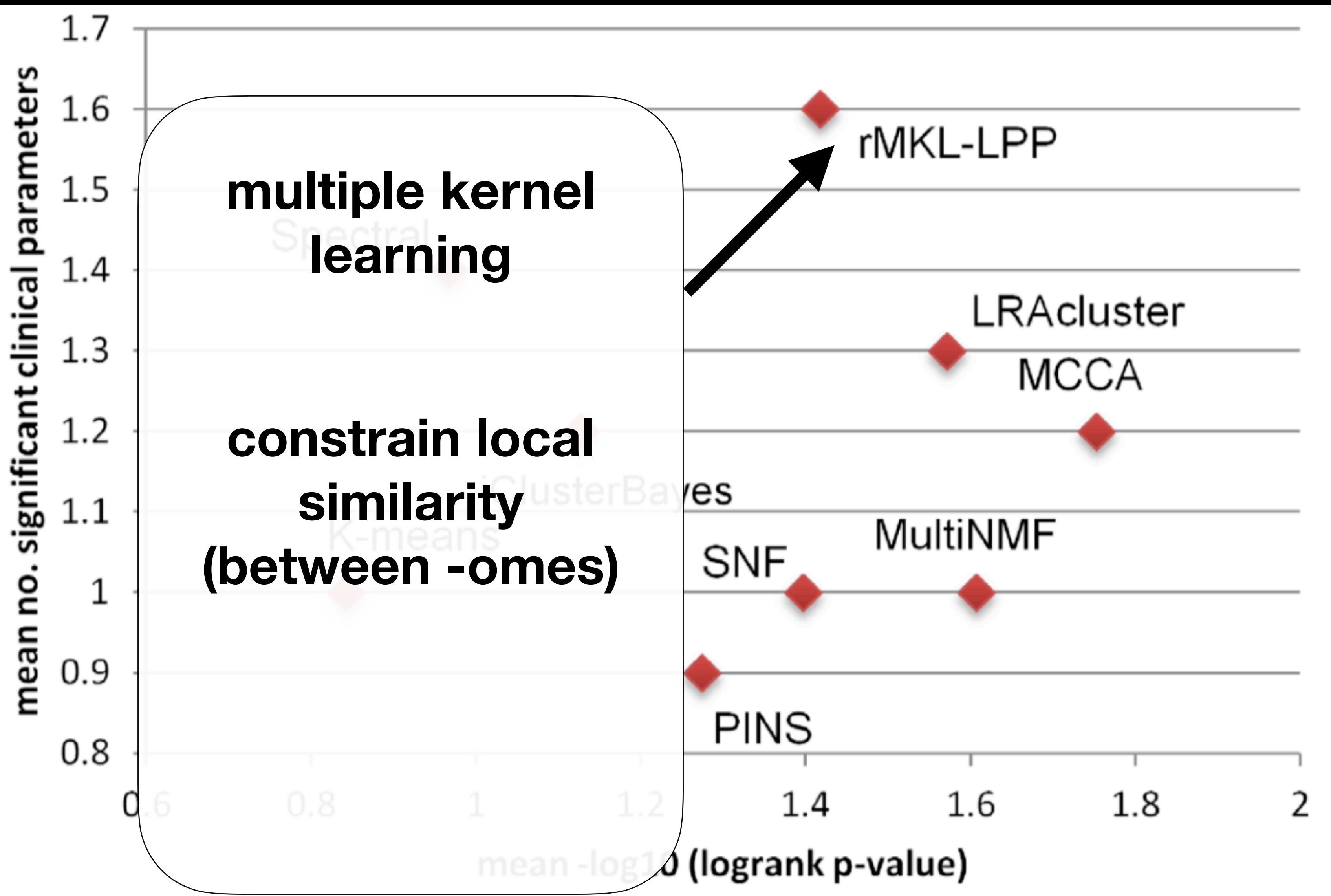
# Data of the Cancer Genome Atlas allows to benchmark different multi-omics approaches.

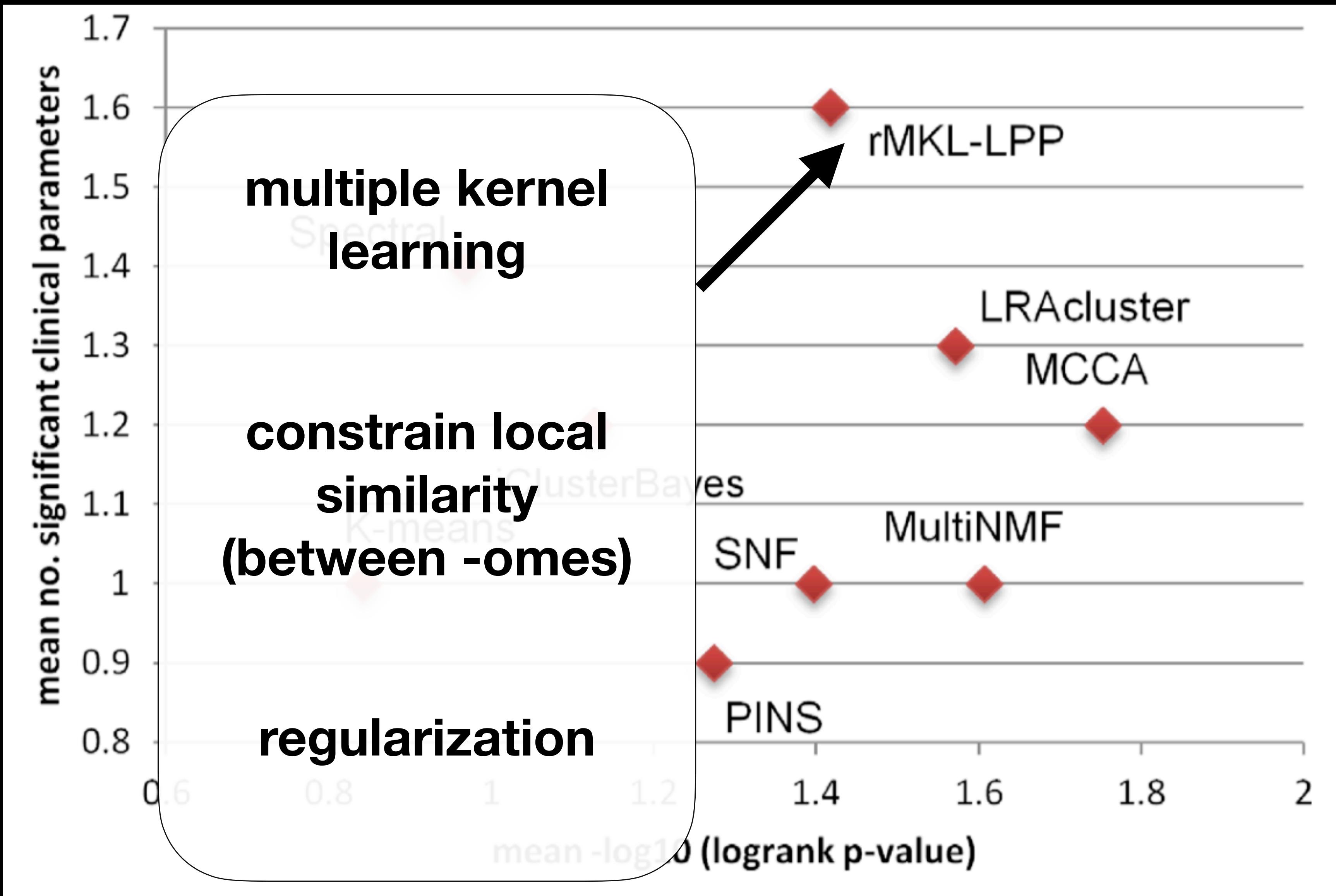


choosing methods



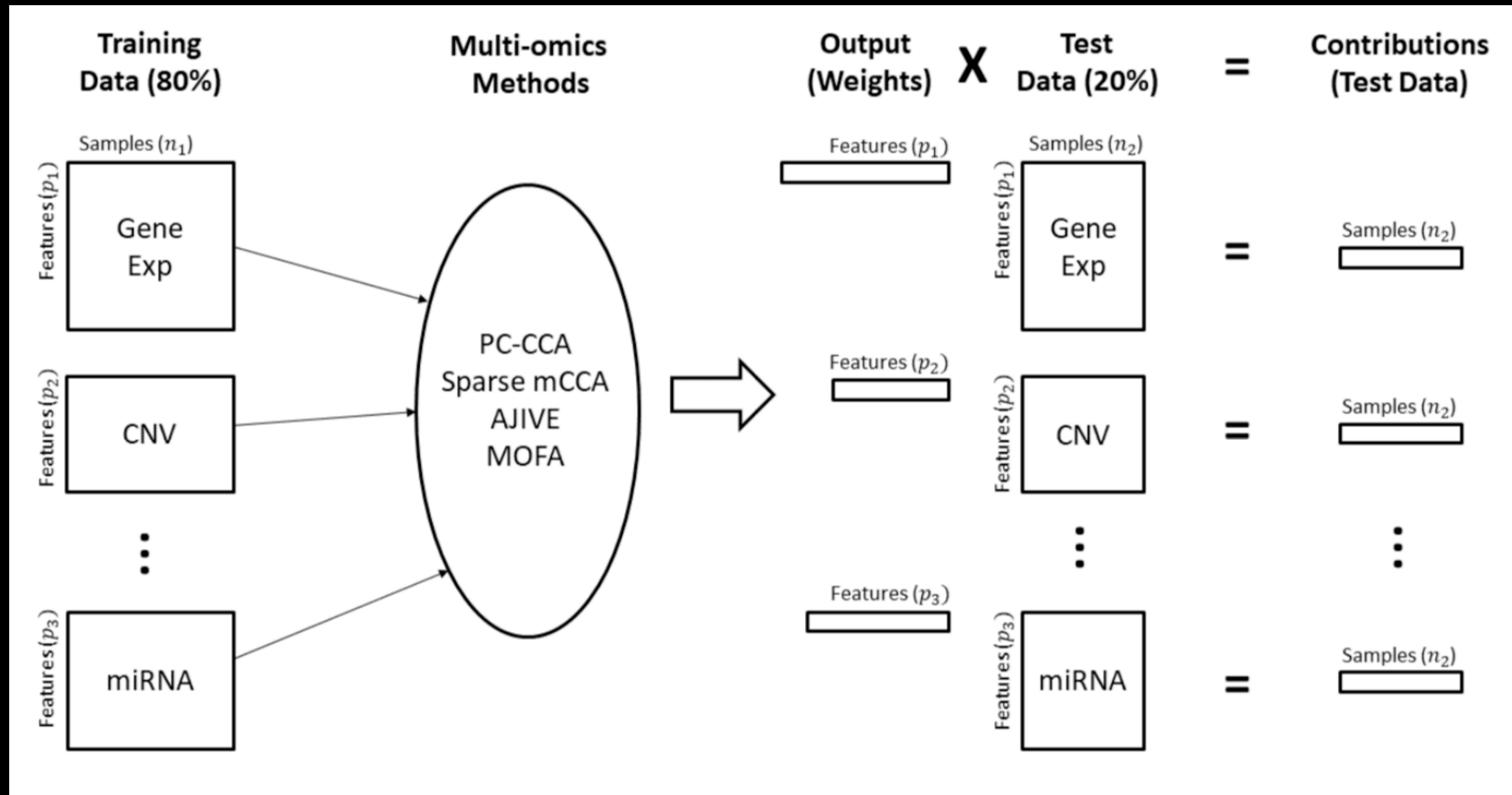




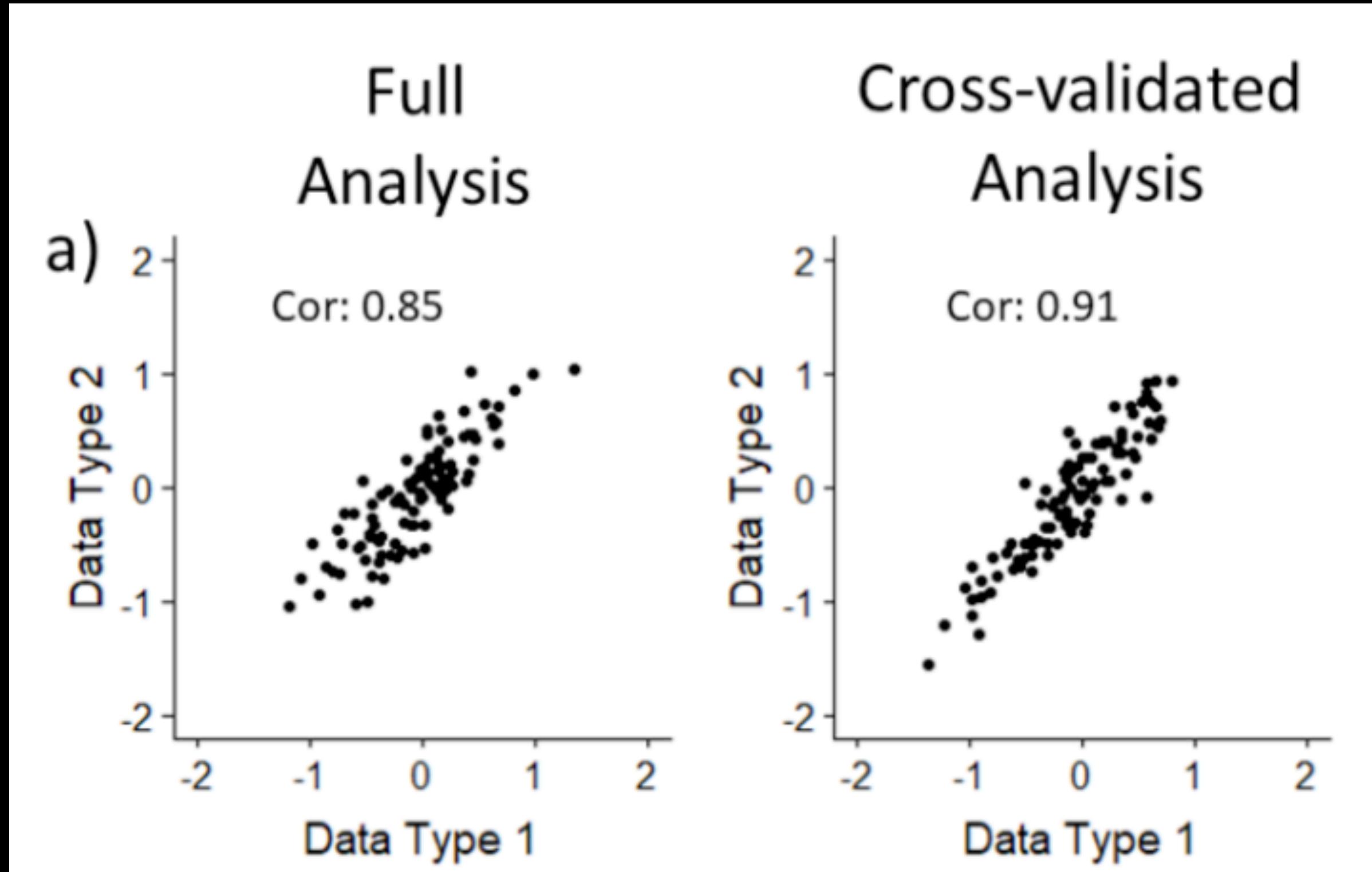


**If uncertain about the performance of multi-omics: Do cross-validation.**

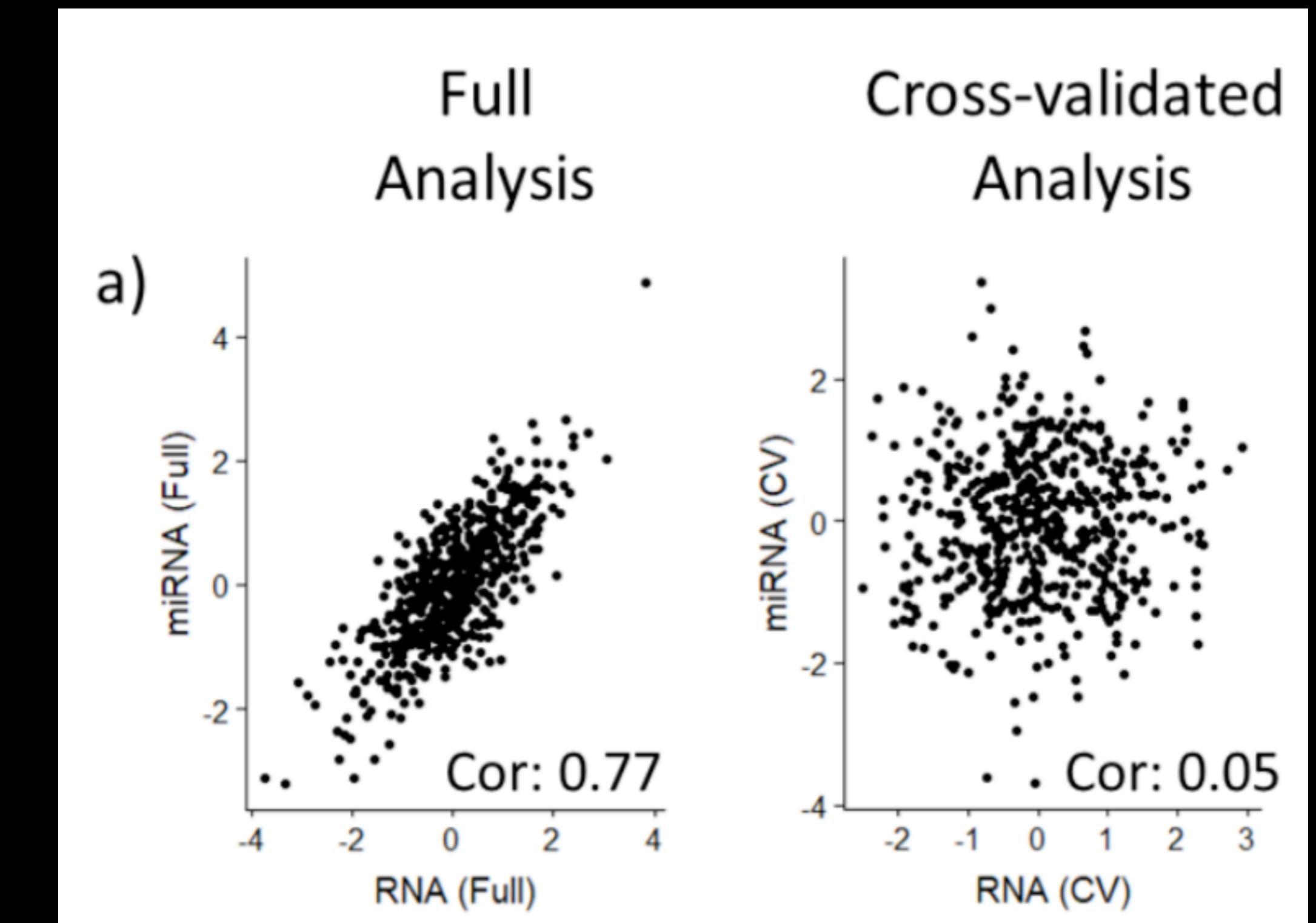
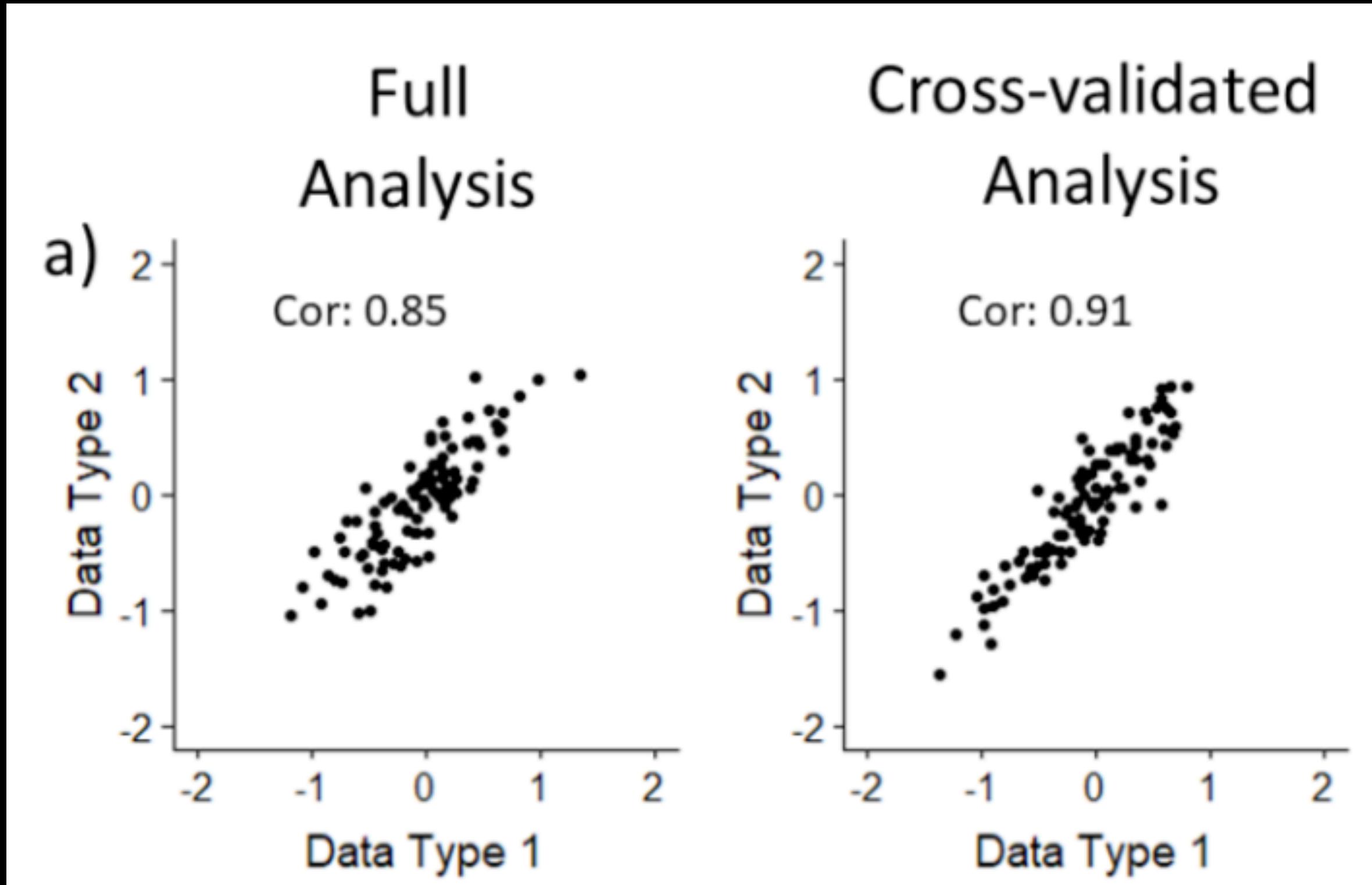
# If uncertain about the performance of multi-omics: Do cross-validation.



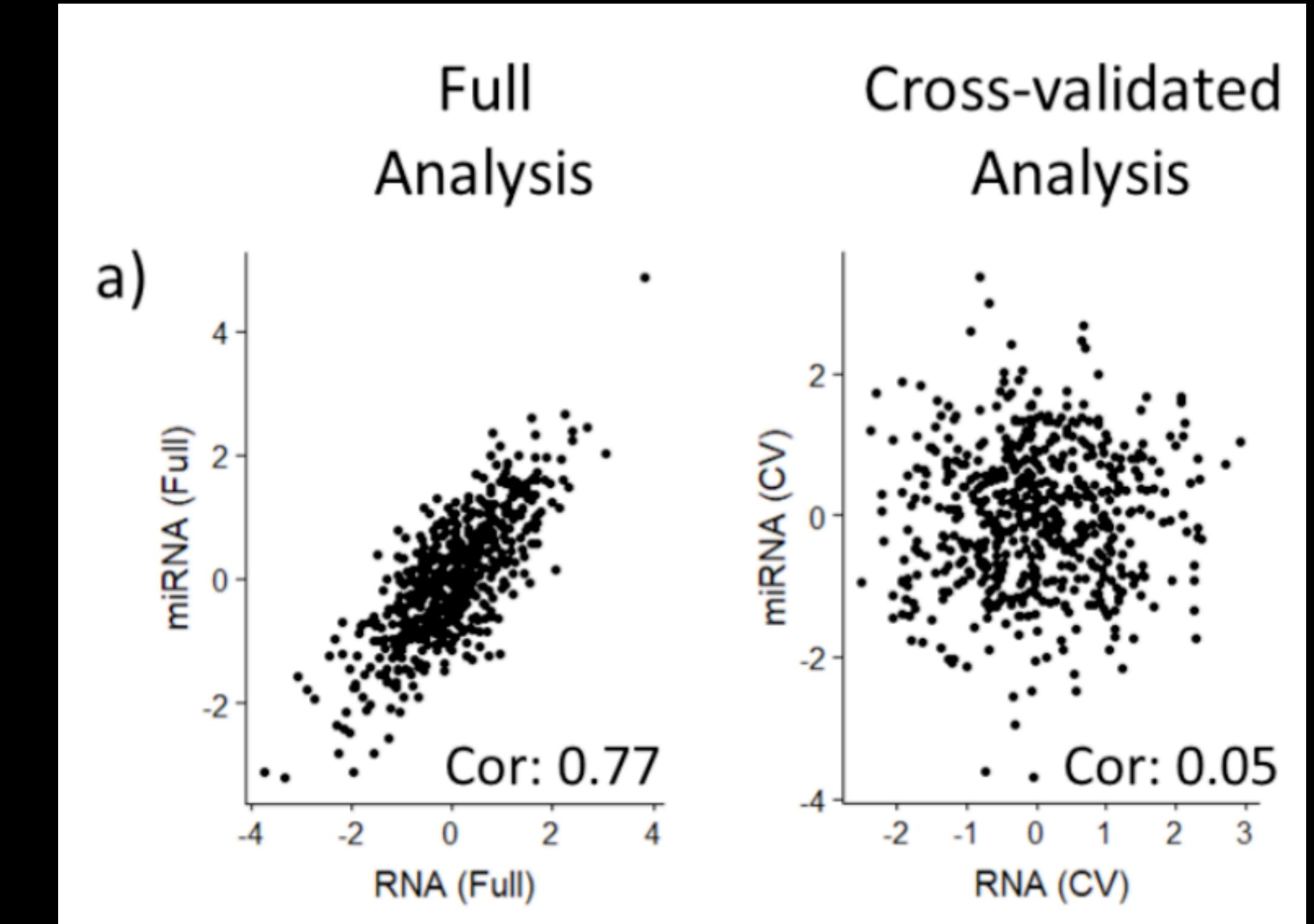
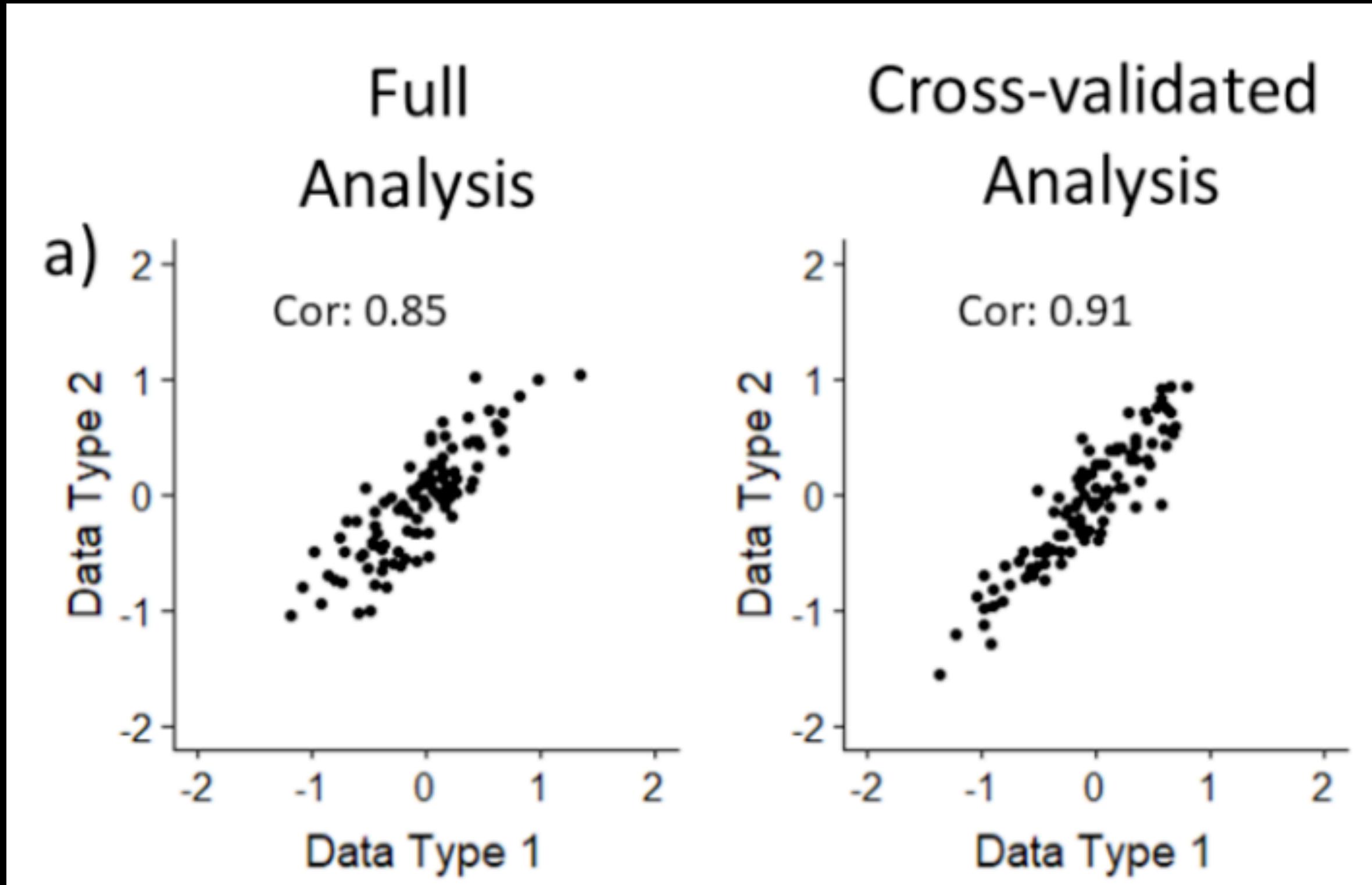
If uncertain about the performance of multi-omics: Do cross-validation.



If uncertain about the performance of multi-omics: Do cross-validation.



If uncertain about the performance of multi-omics: Do cross-validation.



Contemporary computational multi-omics require at least 50 samples!

# Technology

practical

statistical

choosing methods

understanding limits

**2 minutes:**  
**Discuss with neighbor.**

**Which technological  
considerations could also  
apply if measuring only  
a single “-ome”?**

# Two learning goals:

# Two learning goals:

Combining different data modalities (-‘omes’) can answer some interesting scientific questions.

# Two learning goals:

Combining different data modalities (-‘omes’) can answer some interesting scientific questions.

Approach multi-modal data carefully while looking for interesting scientific questions!



**Questions /  
Would like to  
know more  
about ...**