

Sequence and Omics Data

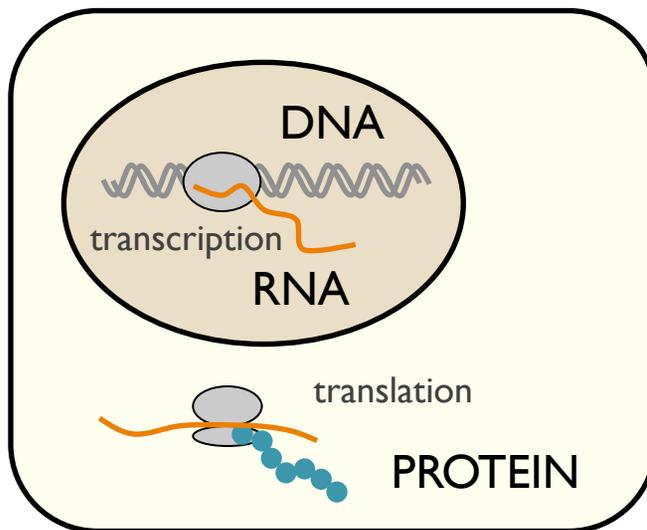
Cydney Nielsen



University of British Columbia
British Columbia Cancer Agency
Vancouver . British Columbia . Canada

Big Picture

eukaryotic cell



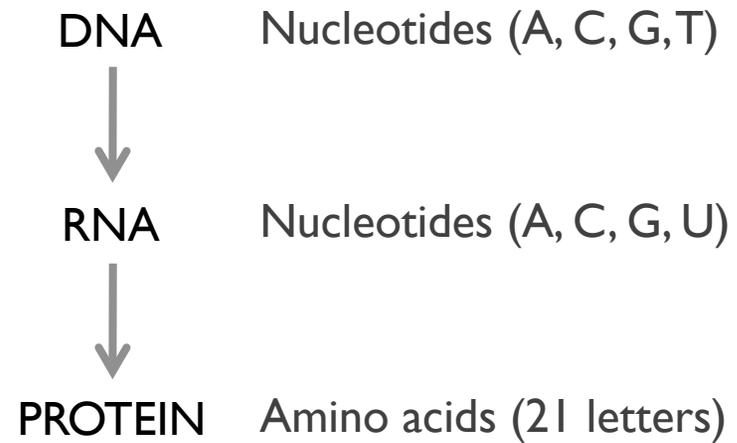
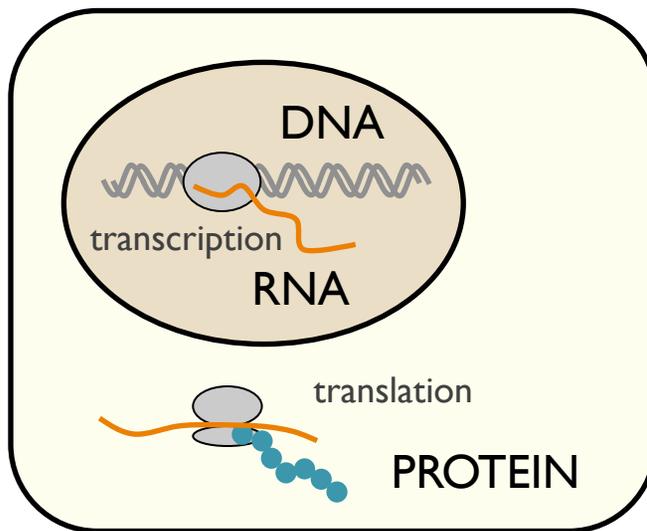
Genes (DNA sequences) are transcribed into RNA molecules (*expressed*)

RNA molecules transcribed from a gene often called *transcripts*

RNA transcripts are translated into *proteins*

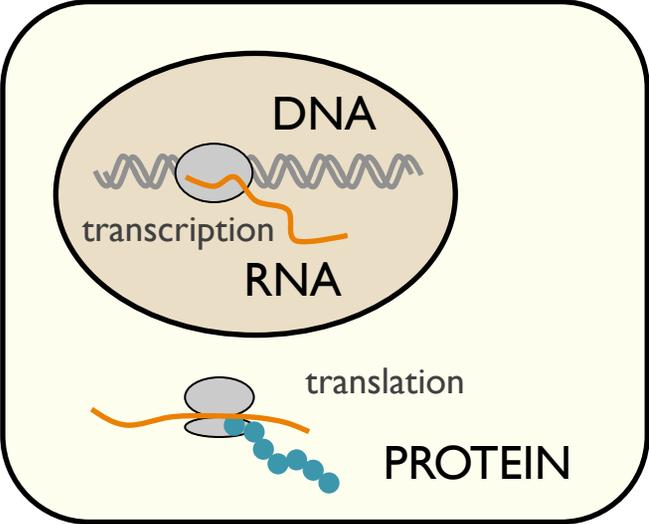
Big Picture

eukaryotic cell

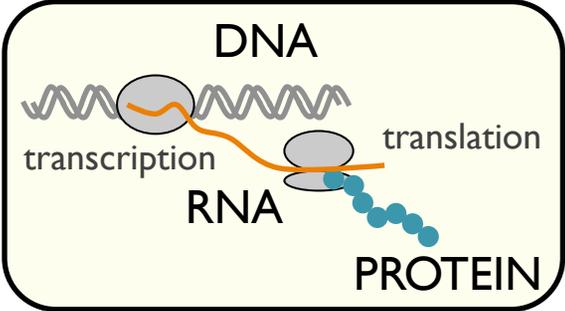


Big Picture

eukaryotic cell



prokaryotic cell

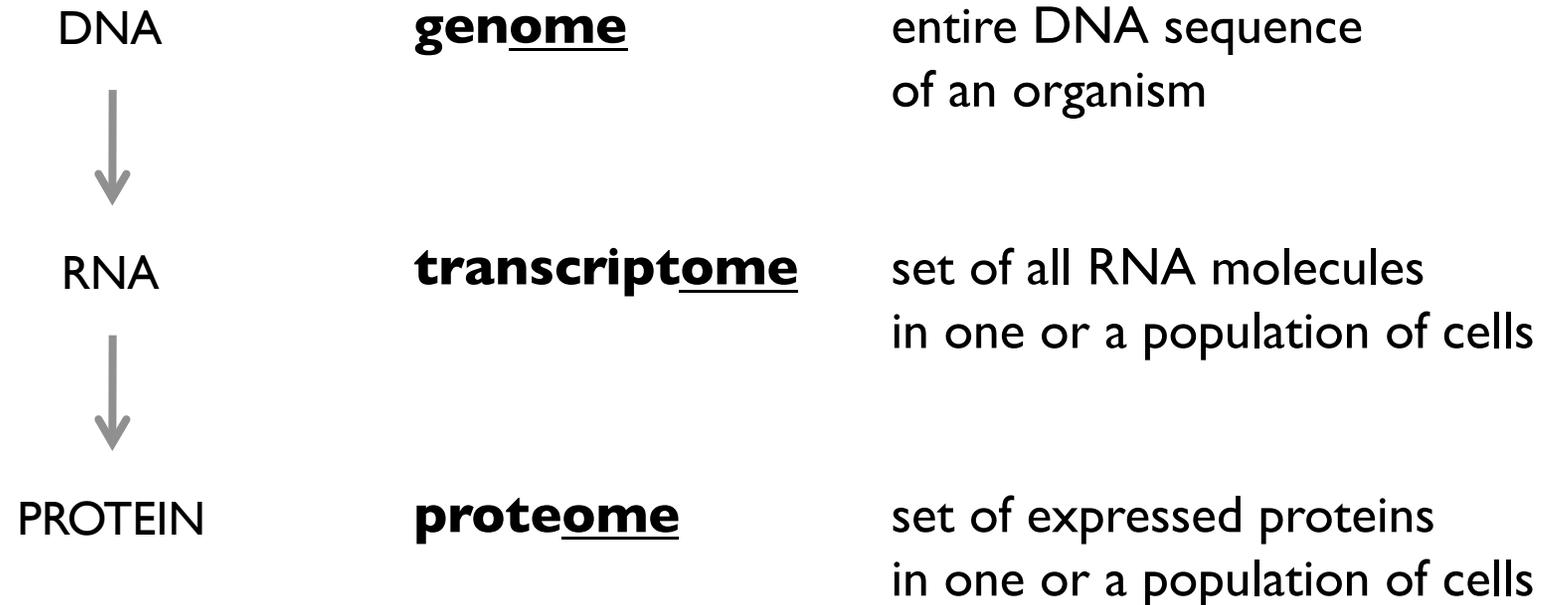


High-Throughput Technologies

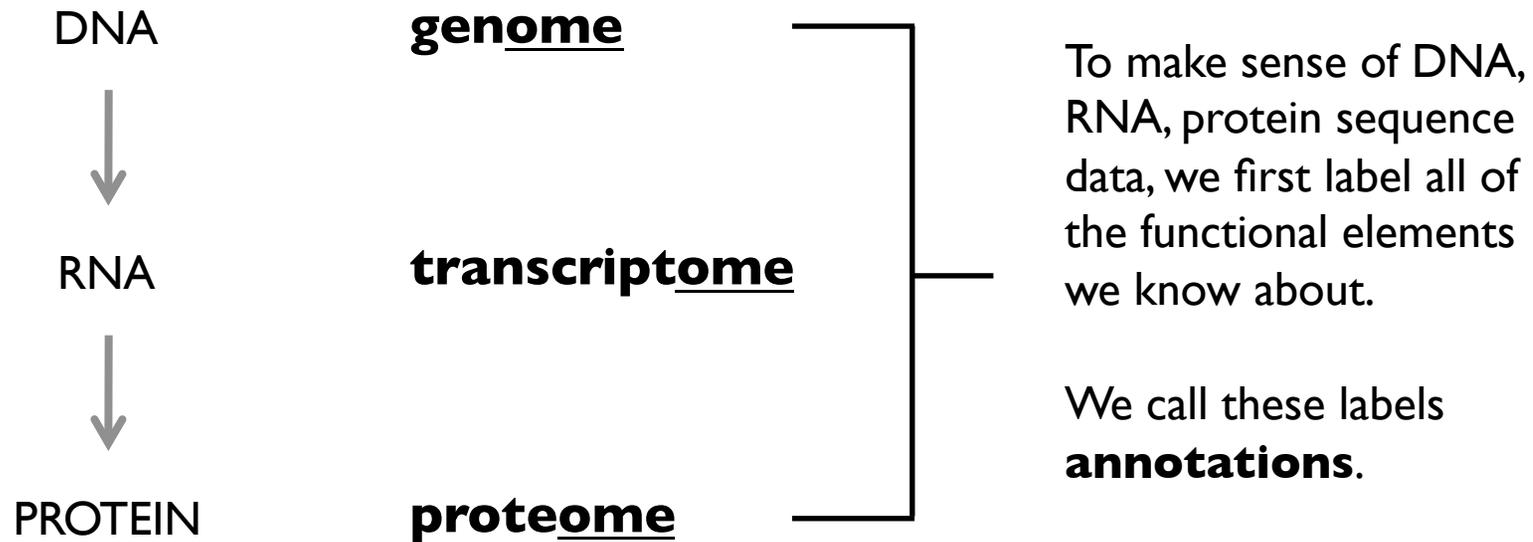


<http://www.bcgsc.ca/platform>

Omics Data – High-Throughput Measurements

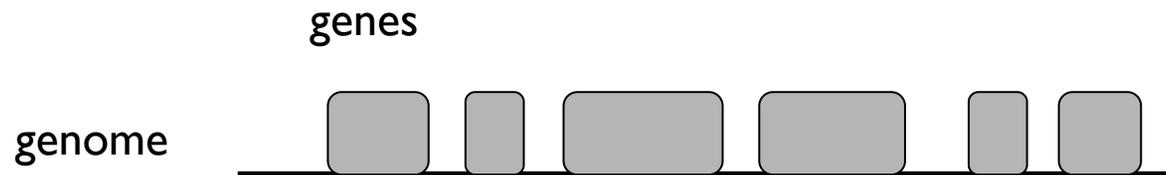
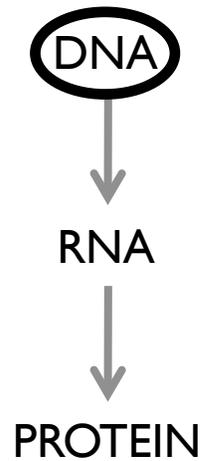


Annotations



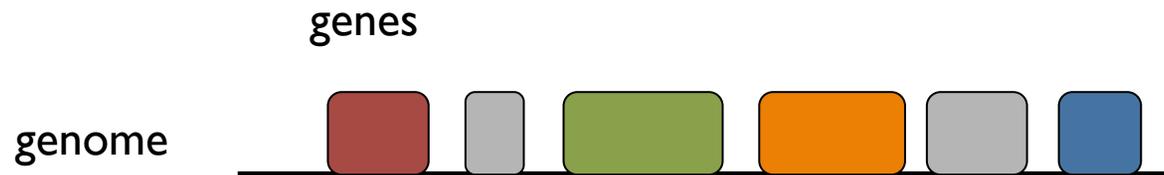
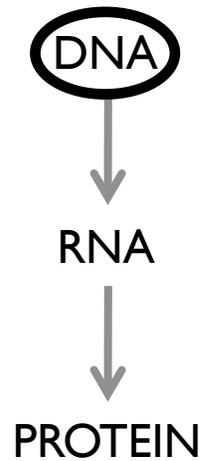
Paper 3: COMBat: Visualizing Co-Occurrence of Annotation Terms

Annotations



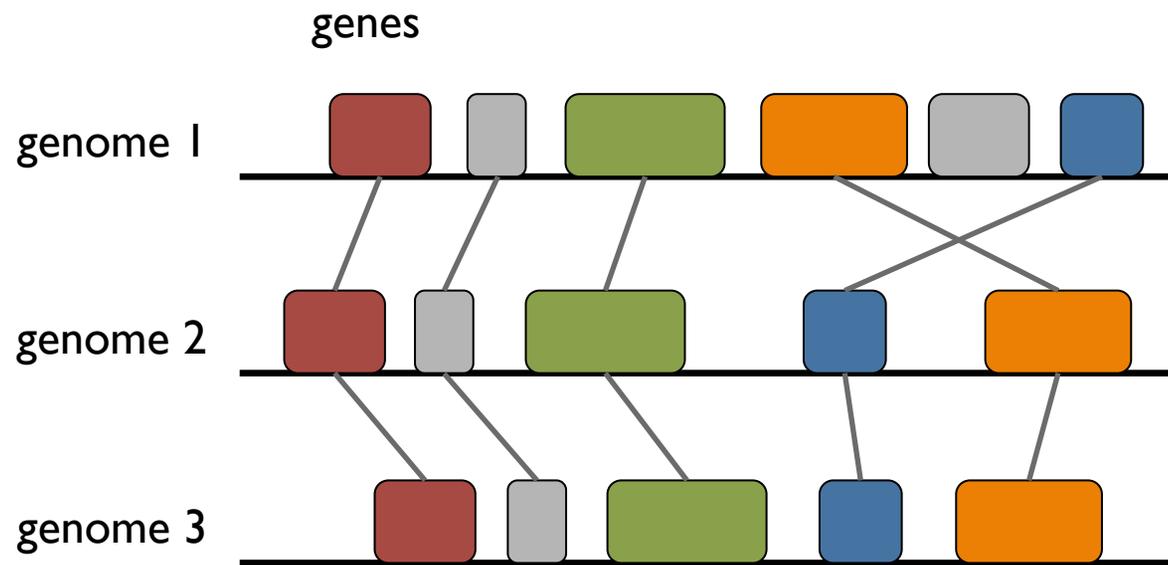
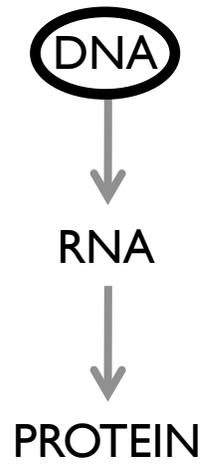
For well studied genomes (e.g. *E. coli*, mouse, human) can *annotate* the genome sequence with known genes

Annotations

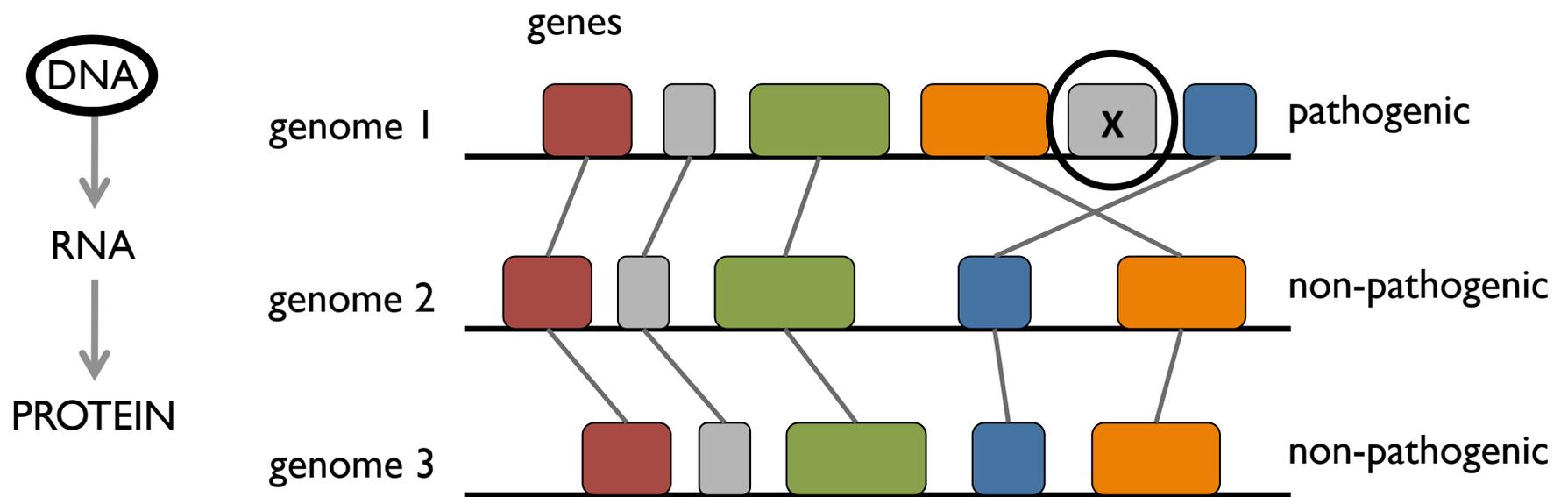


For well studied genomes (e.g. *E. coli*, mouse, human) can *annotate* the genome sequence with known genes and with known functions

Annotations



Annotations



Presence/absence patterns of genes in different genomes can inspire hypotheses about the underlying biology:

Does gene X drive the pathogenicity of bacteria with genome 1 ?

Annotations

Rich online sources of annotations for DNA, RNA, and proteins

Examples



NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

DNA & RNA genes, genomes & variation	Gene expression RNA, protein & metabolite expression	Proteins sequences, families & motifs
Structures Molecular & cellular structures	Systems reactions, interactions & pathways	Chemical biology chemogenomics & metabolomics
Ontologies taxonomies & controlled vocabularies	Literature Scientific publications & patents	Other software cross-domain tools & resources

Annotations

Rich online sources of annotations for DNA, RNA, and proteins

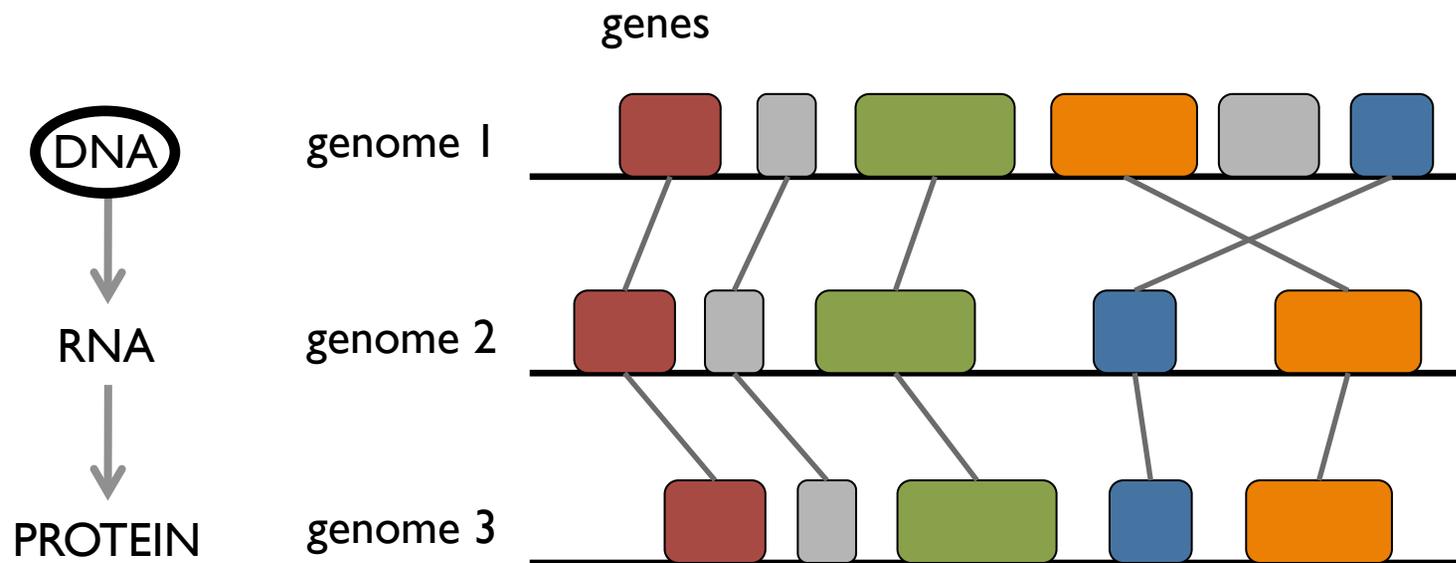
Challenges

How to integrate diverse resources?

How to facilitate query and exploration of diverse annotations?

Paper 3: COMBat: Visualizing Co-Occurrence of Annotation Terms

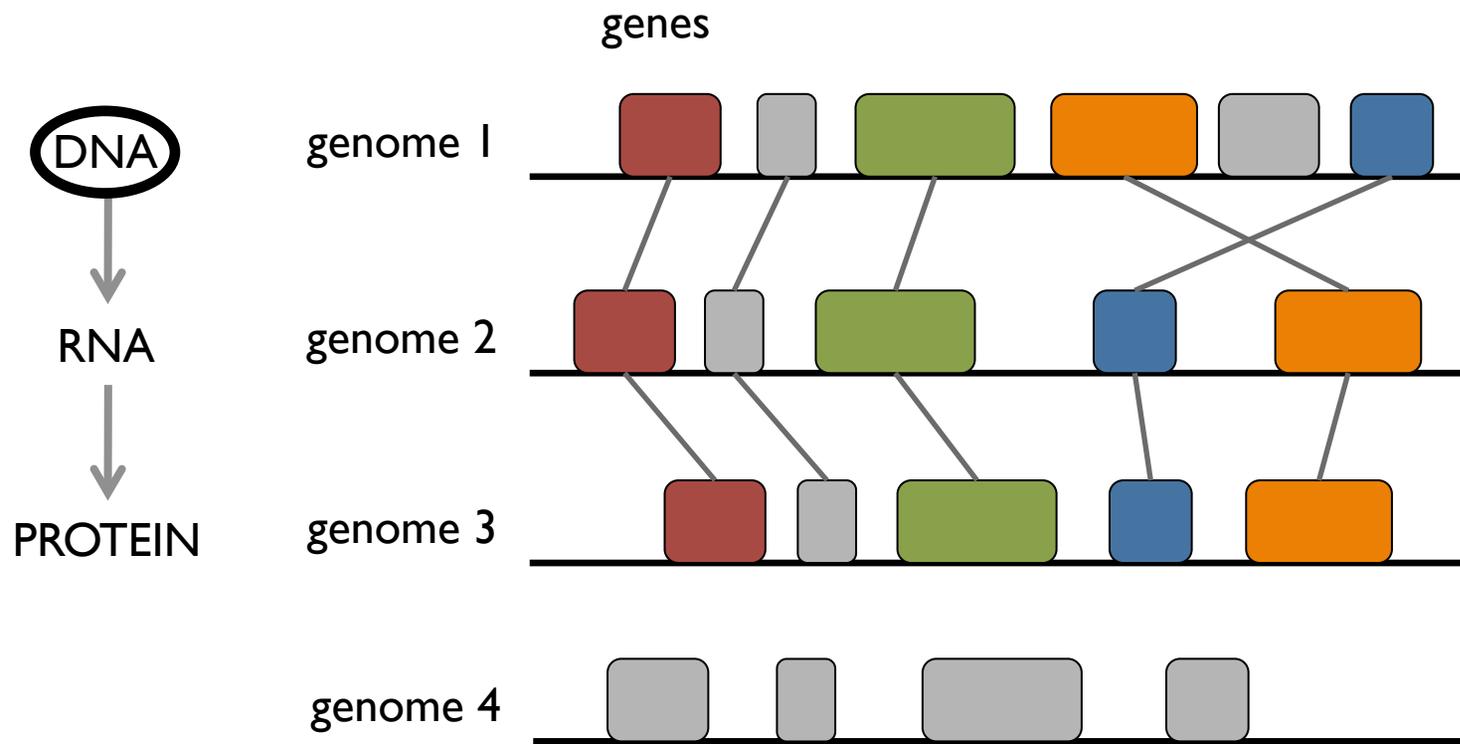
Multiple Sequence Alignments



Paper 2:

Large-Scale Multiple Sequence Alignment Visualization through Gradient Vector Analysis

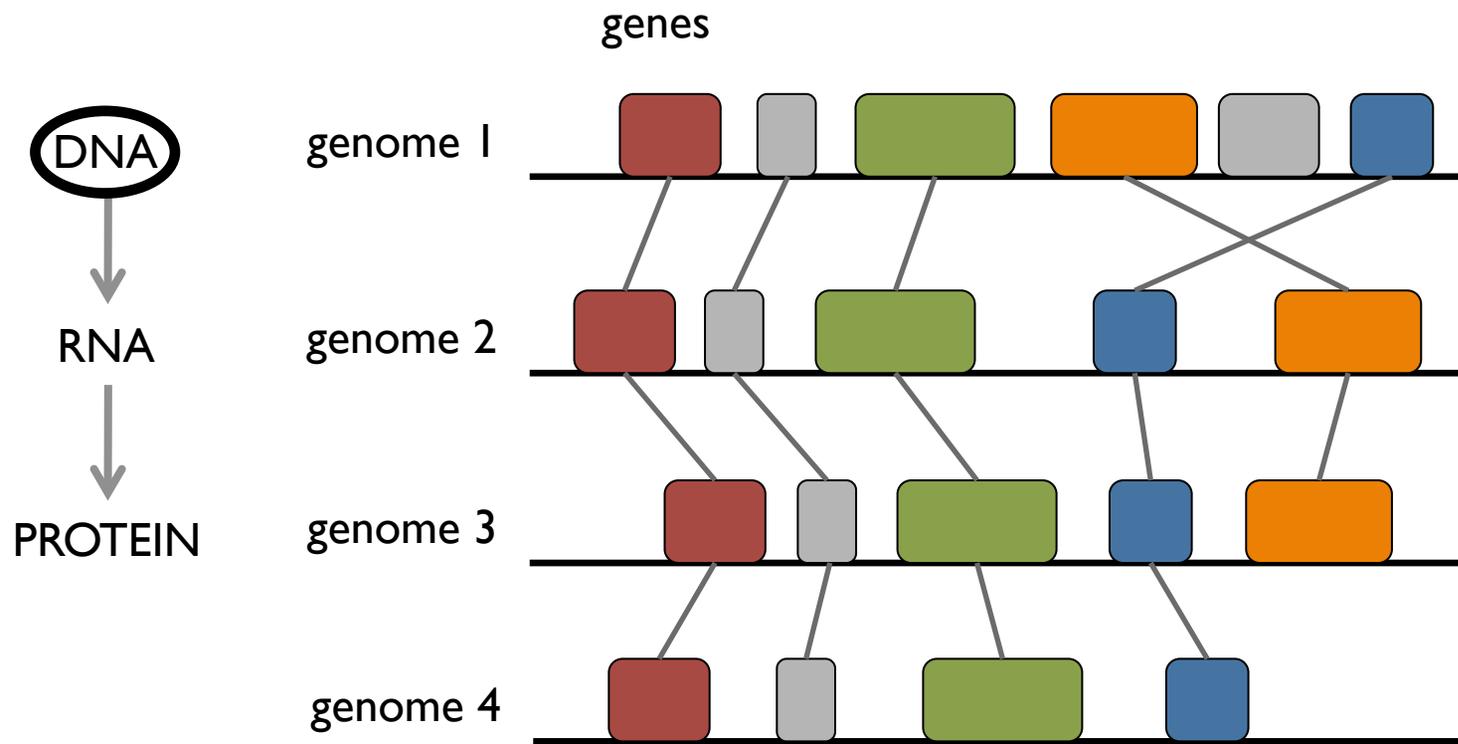
Multiple Sequence Alignments



Newly sequenced genome

Computational gene predictions but no known functions

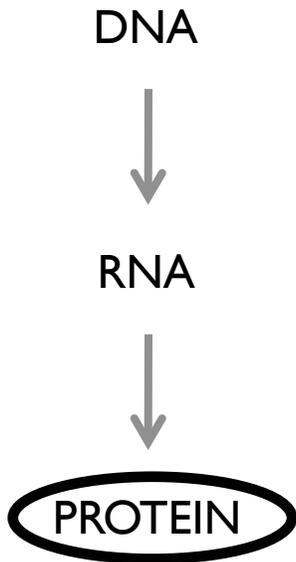
Multiple Sequence Alignments



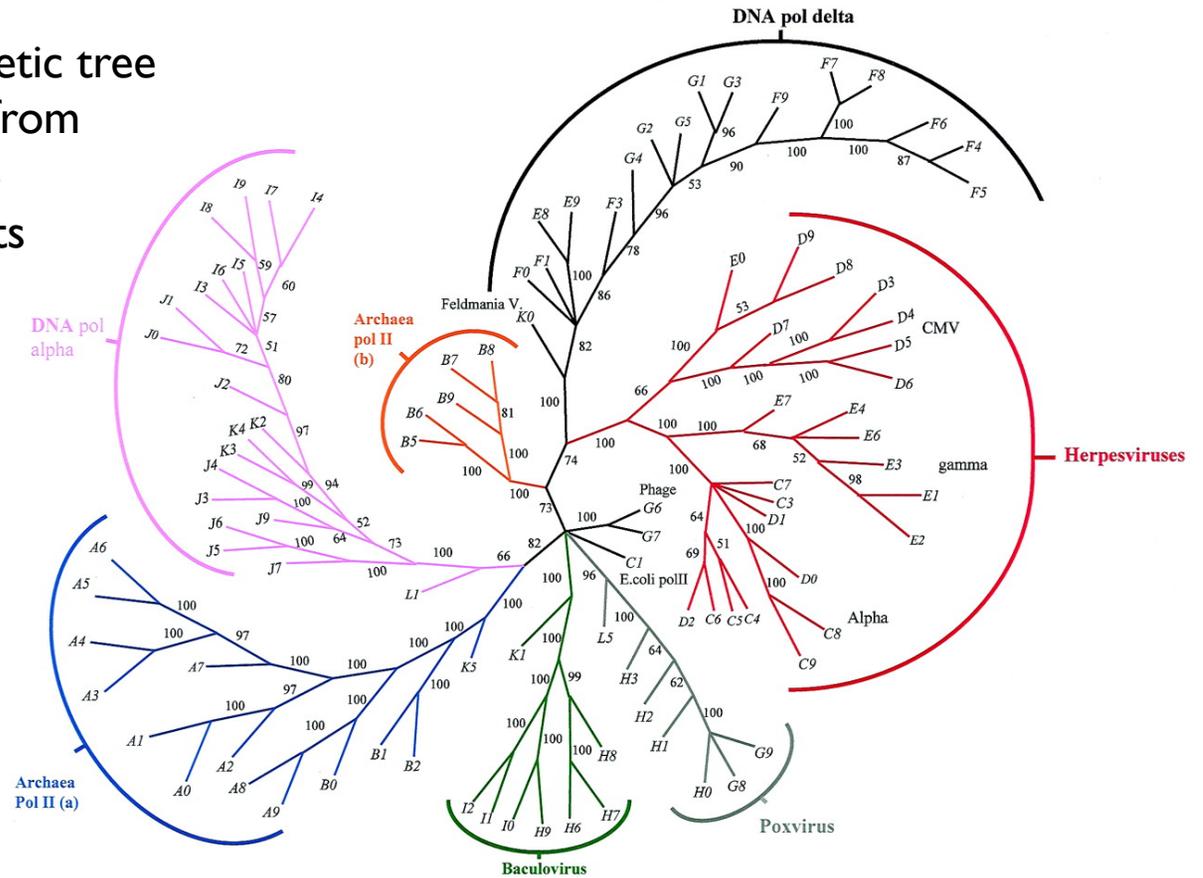
Alignments allow you to use sequence **conservation** to infer **function**

Multiple Sequence Alignments

Sequence similarity suggests evolutionarily relationships



Phylogenetic tree
inferred from
sequence
alignments



Multiple Sequence Alignments

Challenges

How to represent many thousand aligned sequences?

How to indicate potential errors introduced by the alignment algorithms?

DNA

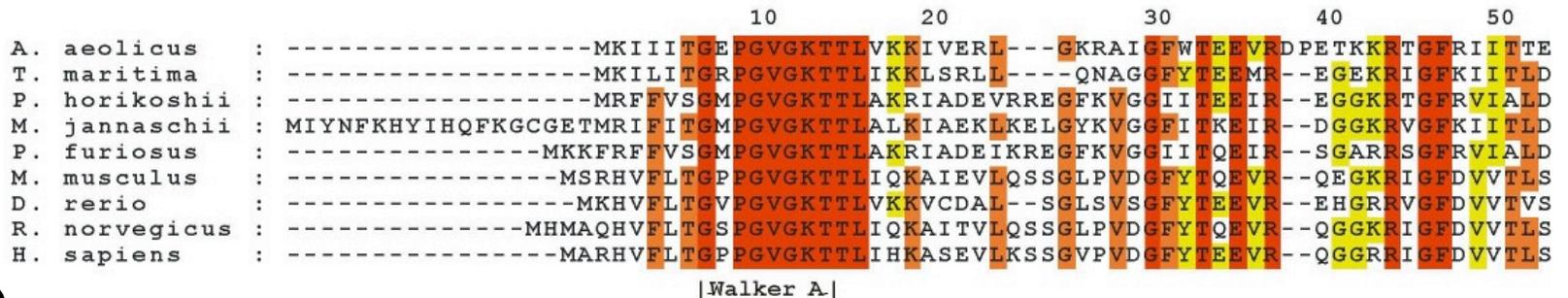


RNA



PROTEIN

Walker loop
phosphate-binding motif

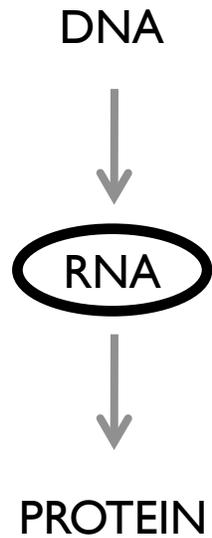


Roßbach et al. BMC Structural Biology 2005

Paper 2:

Large-Scale Multiple Sequence Alignment Visualization through Gradient Vector Analysis

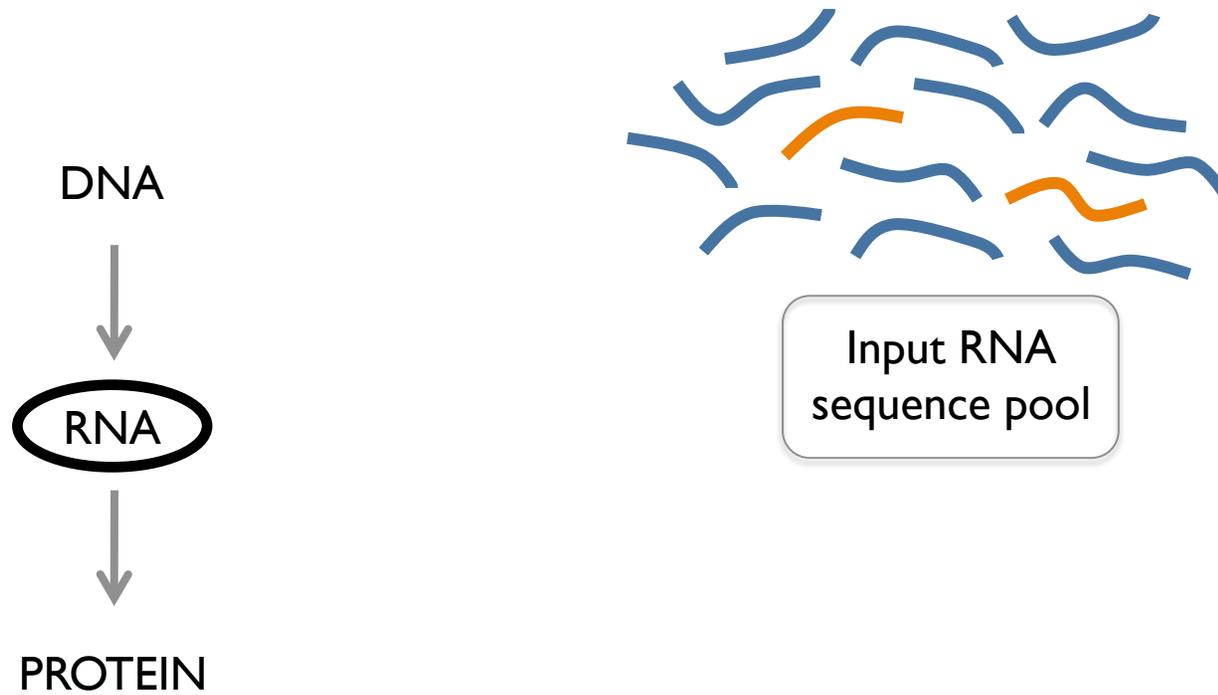
In vitro selection



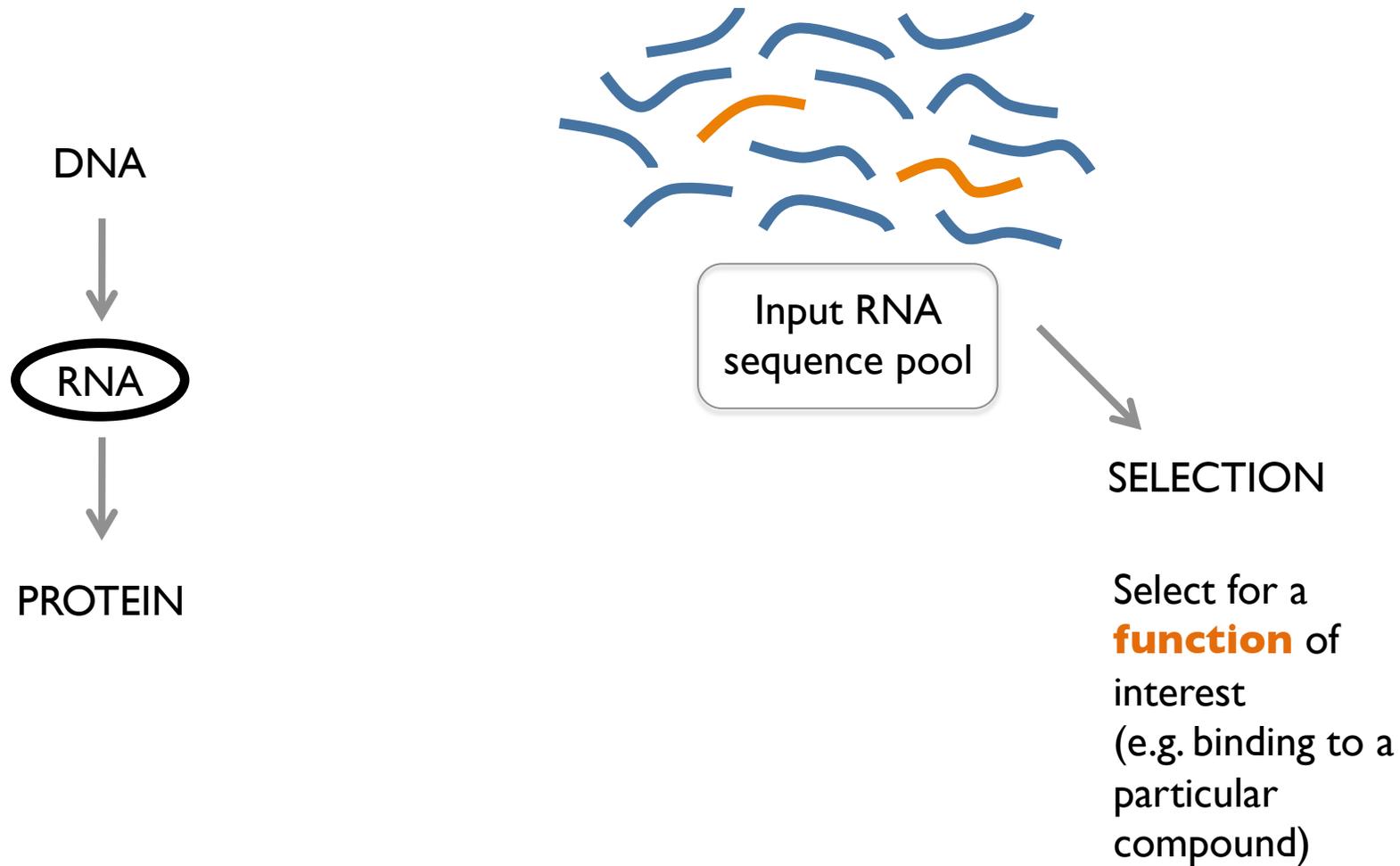
Paper 1:

invis: Exploring High-Dimensional RNA Sequences from In Vitro Selection

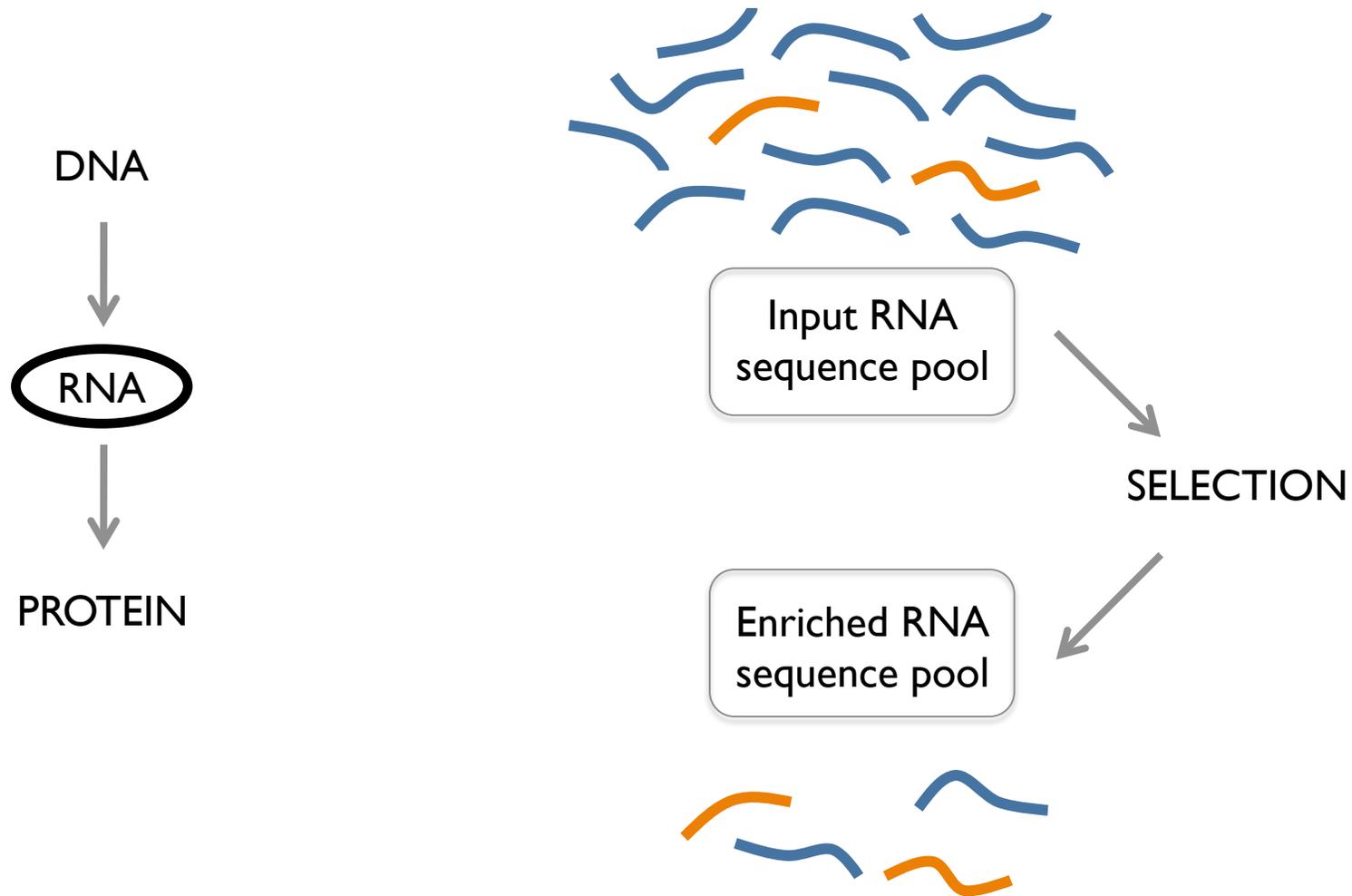
In vitro selection



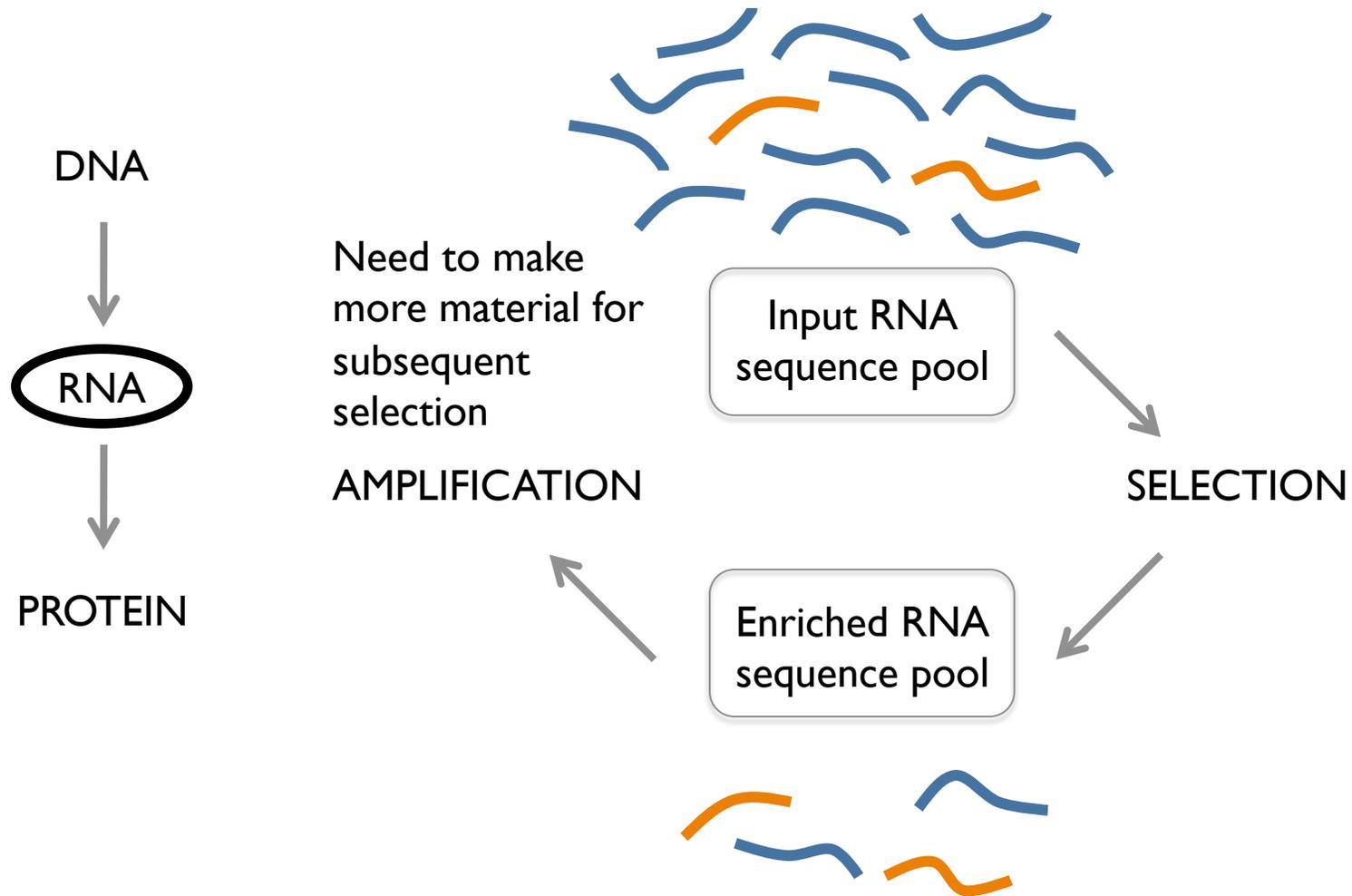
In vitro selection



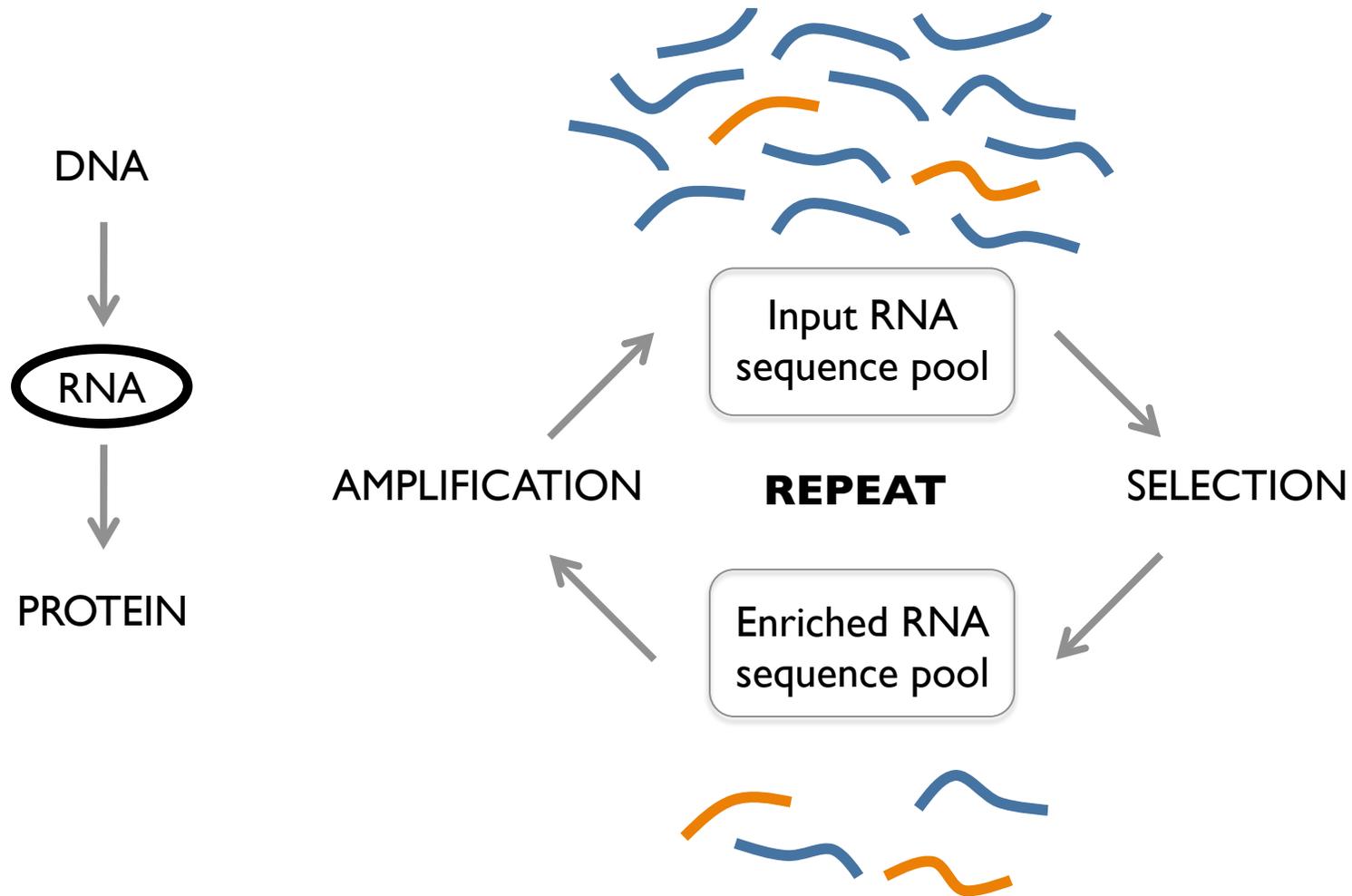
In vitro selection



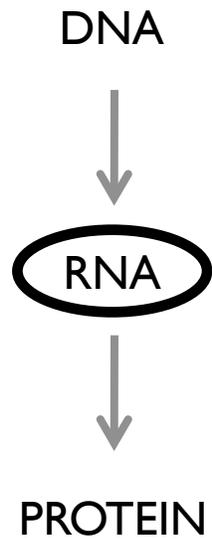
In vitro selection



In vitro selection



In vitro selection



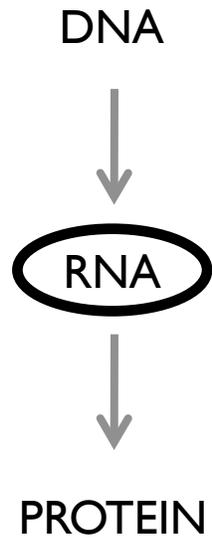
Random sequence pool

- Allows for discovery of previously unknown functional RNAs

Randomly mutated versions of a known functional sequence

- Allows you to explore the diversity of sequences that maintain the target functionality

In vitro selection



Challenges

How to visualize and explore high dimensional sequence space?
How to reveal its overall structure while also showing details at the individual sequence level?

Paper 1:

invis: Exploring High-Dimensional RNA Sequences from In Vitro Selection

Sequence and Omics Data Papers

Paper 1:

invis: Exploring High-Dimensional RNA Sequences from In Vitro Selection

Paper 2:

Large-Scale Multiple Sequence Alignment Visualization through Gradient Vector Analysis

Paper 3:

COMBat: Visualizing Co-Occurrence of Annotation Terms