

# Biomolecular analytics in the context of experimental biology

Ruedi Aebersold

Methoden der Biomolekularen  
Analytik

November 27th, 2017

# Teaching goal

---

- You can explain how bioanalytical techniques support different approaches to experimental biology
- You understand that experimental techniques evolve along with the evolution of biological paradigms

# Biological knowledge comes from experiments

---

- The facts you learn come from experiments that were done, mostly over the last century

# Principles of experimental biology

---

- Background knowledge: Theory, paradigm, first principles
- **Formulate a hypothesis** (hypotheses can never be proven, just rejected at a certain probability)
- Design experiment to test hypothesis
- Reinforce, adjust or reject theory based on experiment results

## Major principles of molecular biology

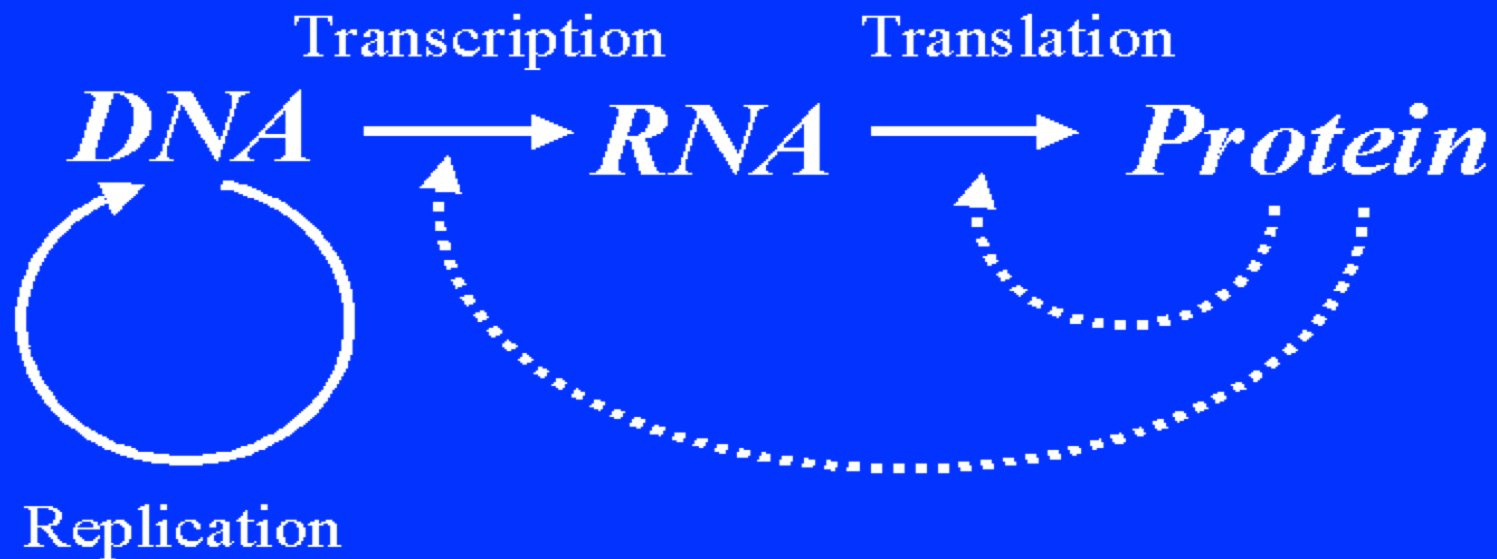
---

Mendel's Laws	1865	Mendelian inheritance
Avery	1944	DNA as transforming principle
Beadle & Tatum	1941	One gene – one protein – one function
Central Dogma (Crick)	1958	DNA -> RNA -> protein
Pauling	1949	Sickle cell anemia as a molecular disease: Mechanistically connecting genomic variability to phenotypic variability

One of the central notions in biology

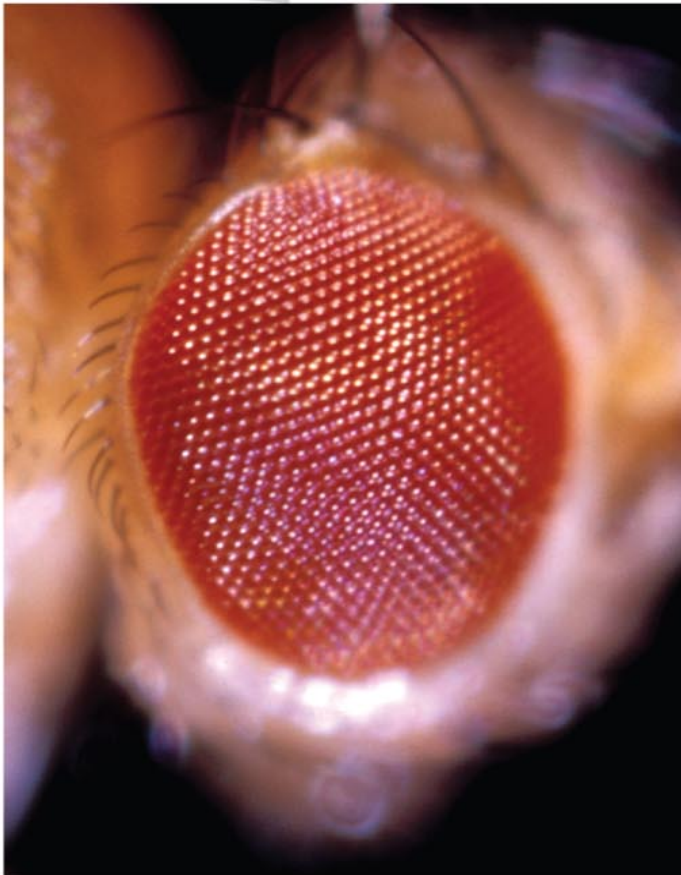
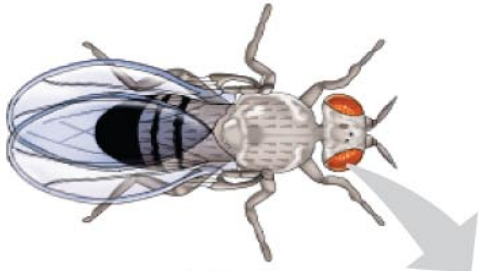
---

## The Central Dogma of Modern Molecular Biology

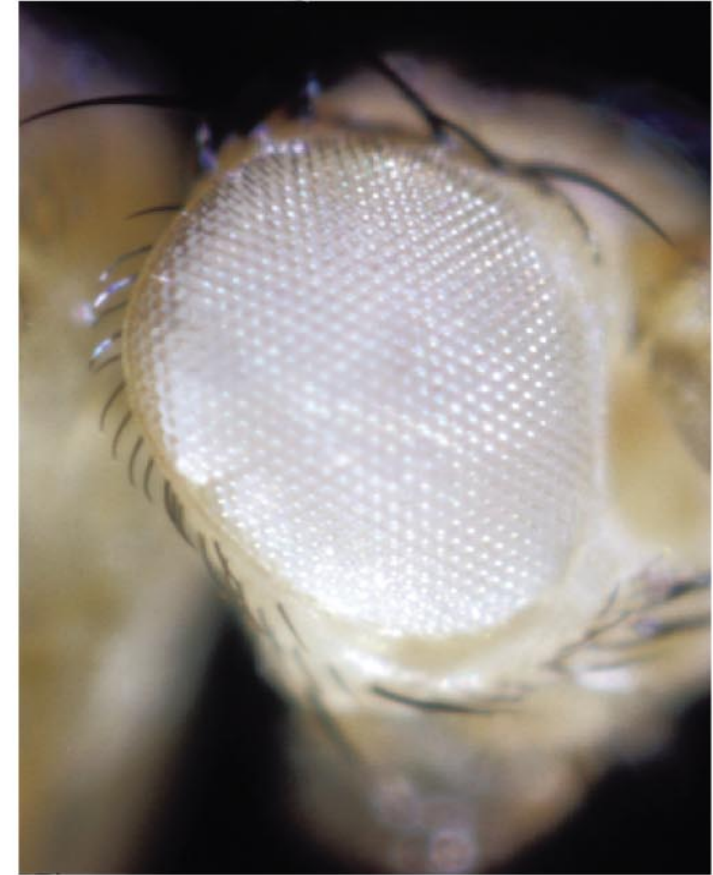
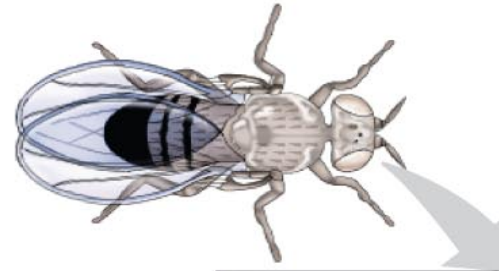


# Genotype– Phenotype

---

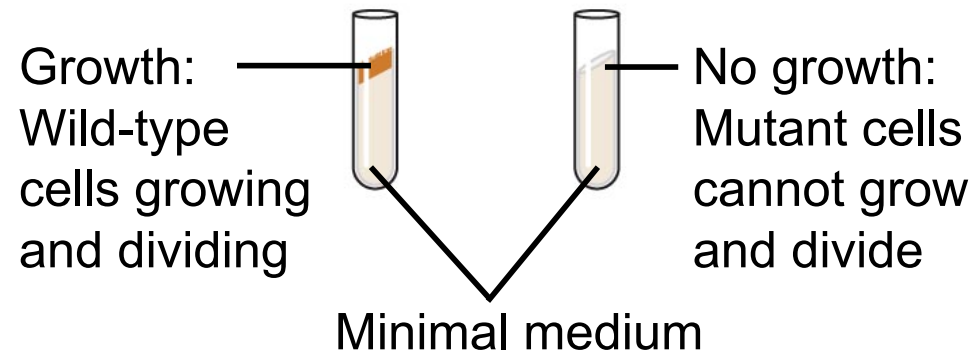


Wildtype (w<sup>+</sup>)



Mutant (w)

## EXPERIMENT



Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

- Phenotype: Growth or cell death
- Mutation: Change in DNA
- Medium: Solution sustaining cell growth



## RESULTS

### Classes of *Neurospora crassa*

















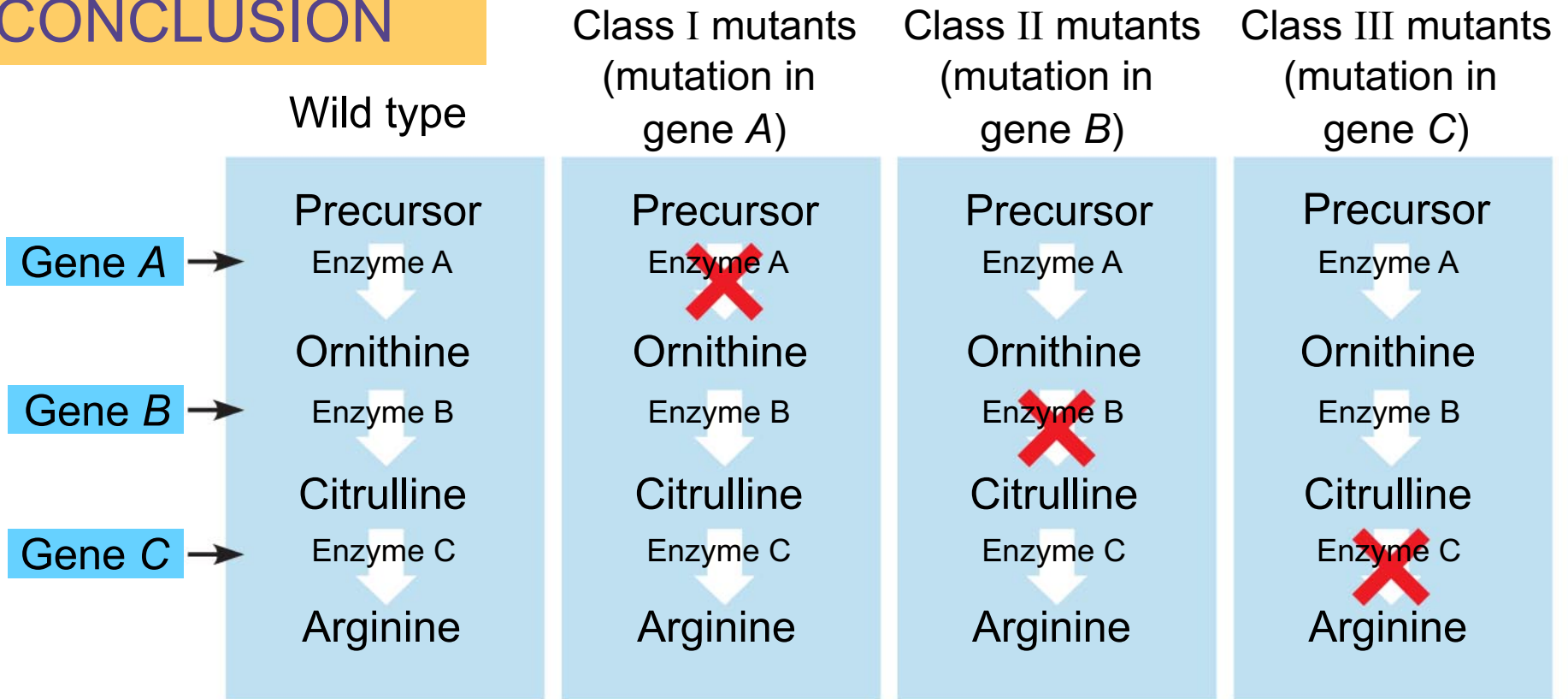
Condition	Wild type	Class I mutants	Class II mutants	Class III mutants
				
				
				
				

Fig. 17-2c

## CONCLUSION



# Summary

---

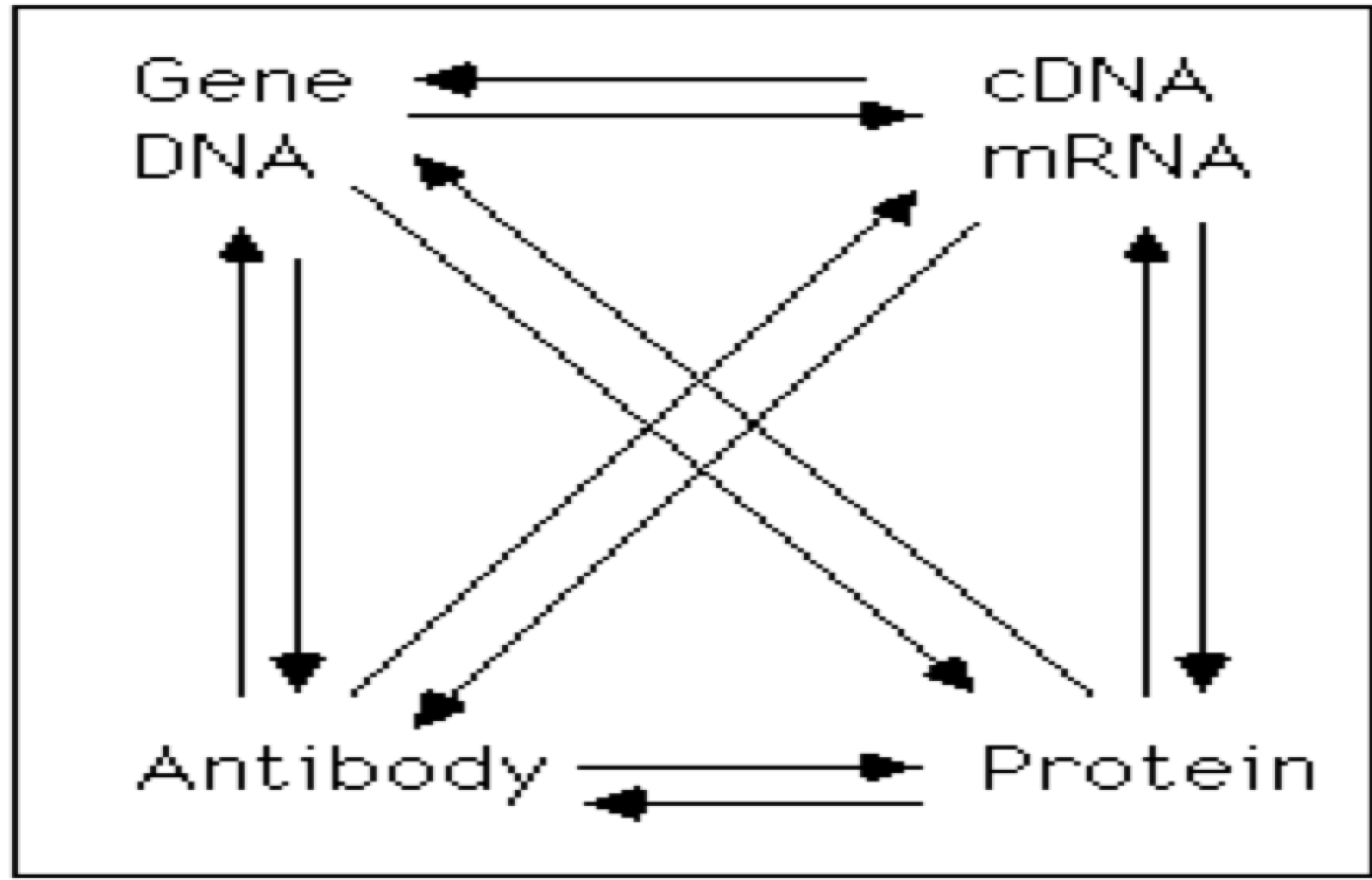
- The Beadle-Tatum experiment shows that a lesion at a genomic locus affects enzymes that in turn determine a phenotype
- This led to the “*one gene–one enzyme hypothesis*” which states that each gene leads to the generation of a protein that carries out a function and determines a phenotype
- The *one gene–one enzyme hypothesis* was later expanded to the *one gene–one protein hypothesis* and established the link between genotype and phenotype

# Consequences for biopolymer analysis

---

- We need a method to purify a specific protein activity
- We need to (partially) sequence the corresponding protein
- We need to generate a gene probe
- We need to use the probe to isolate the corresponding gene from a gene library
- We need to sequence the respective gene
- Subsequently, we can experiment with the respective gene
- Focus on one or a few genes, throughput is not important but precision is

# The Micro-chemical Cycle



# Reductionist research approach

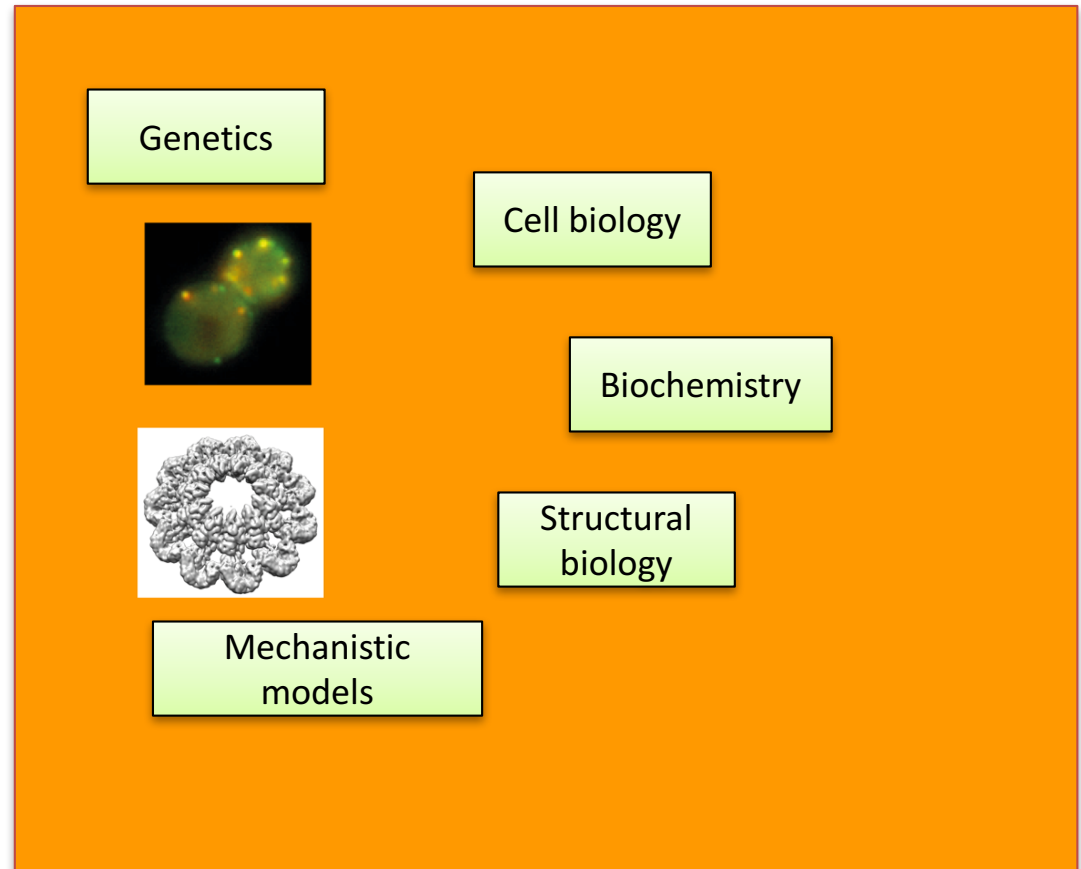
---

## Reductionist world: Focus on specific molecules

### Main **questions**:

Can we discover the gene/protein that codes/carries out a specific function

Ability to measure all molecules and to find the one associated with a specific function



Discover molecule that has a specific function, e.g. growth factor, oncogen

# Consequences for biopolymer analysis

---

- What if we knew all genes, i.e. the genome??
  - We would know all functions
  - We would not need to fish for genes anymore
  - We would recognize defective genes
  - We could compare genes
- This thinking is the root of the genome project

# A draft of the human genome (2000)

---



...announced that the international Human Genome Project and Celera Genomics Corporation **have both completed an initial sequencing of the human genome -- the genetic blueprint for human beings.** He congratulated the scientists working in both the public and private sectors on this landmark achievement, which promises to lead to a new era of molecular medicine, an era that will bring new ways to prevent, diagnose, treat and cure disease.

Now, scientists will be able to use the working draft of the human genome to:

- \* **Alert patients that they are at risk for certain diseases.**
- \* **Reliably predict the course of disease.**
- **Precisely diagnose disease and ensure the most effective treatment is used.**
- **Developing new treatments at the molecular level.**

Press release US.  
Govt.

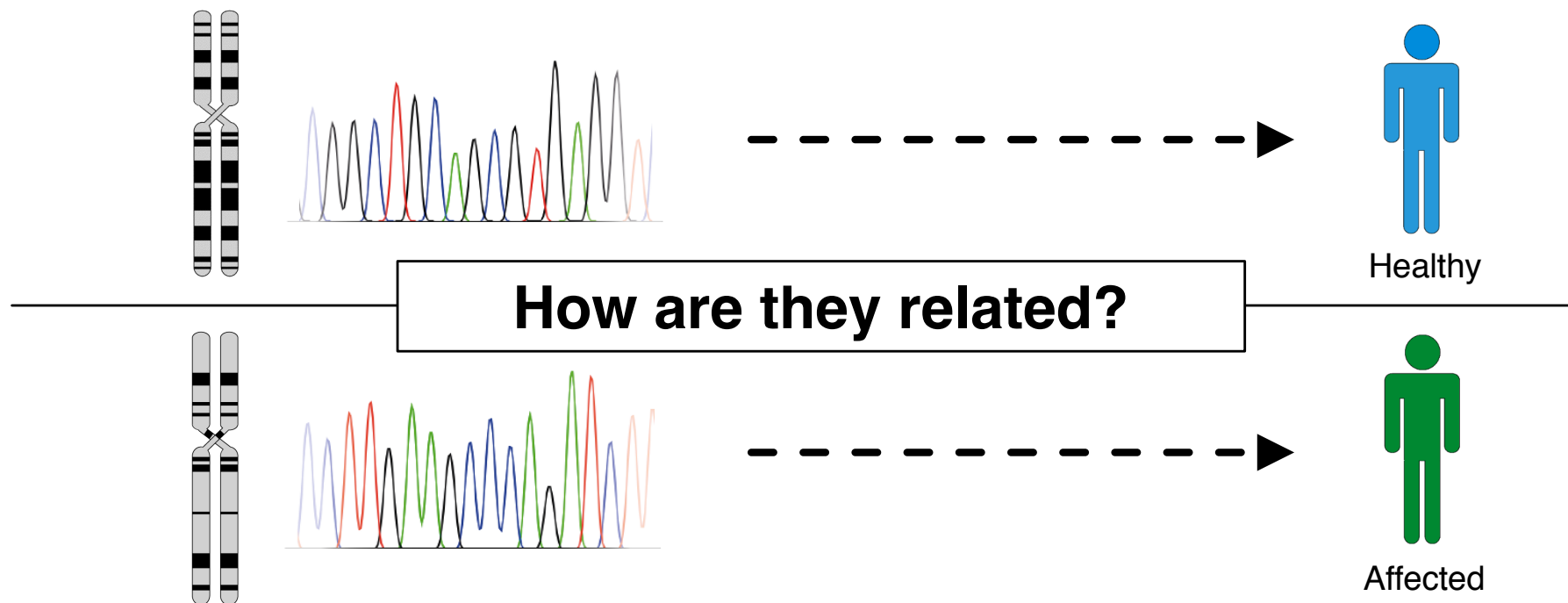


# Consequences for biopolymer analysis

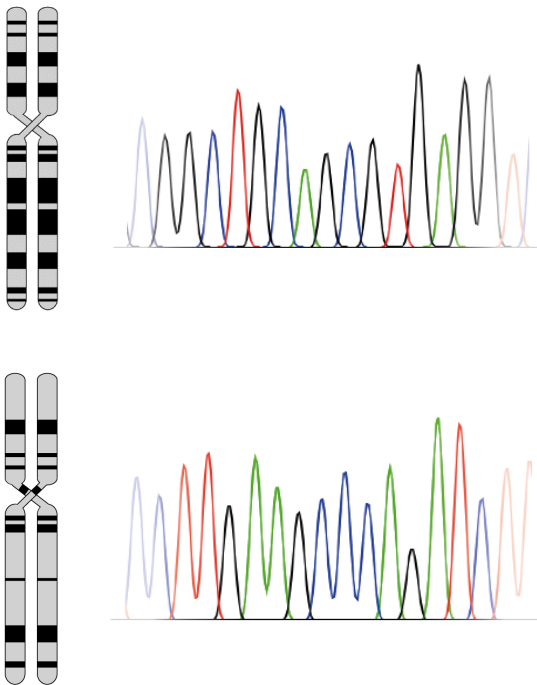
---

- To know all genes we need to sequence them once but very precisely (once and for all)
- The technology needs to be able to identify billions of nucleotides without error
- High throughput and high precision
- Sanger sequencing. Human genome was several year effort
- Shotgun mass spectrometry for proteins

## About phenotypes and their molecular origin



## About phenotypes and their molecular origin



- what is the effect of any inherited or somatic mutation on the phenotype?
- how do two or more independent mutations combine?
- how do the same inherited mutation affect different individuals with different genotypes?

**Questions we should be able to answer but can't**



Healthy



Affected

# Social Networks in biology and the real world

---

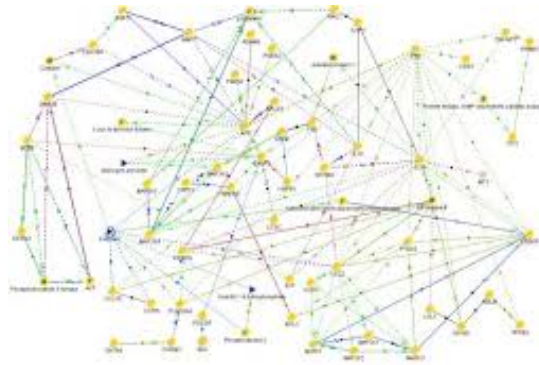


- Who interacts with whom?
- Who is the most important?
- What happens to the network when one node leaves?
- The crowd can do things that the individual cannot do
- Information resides in the network
- The individual and its context

## Molecules in context



Genotypic variability



**Molecular networks**



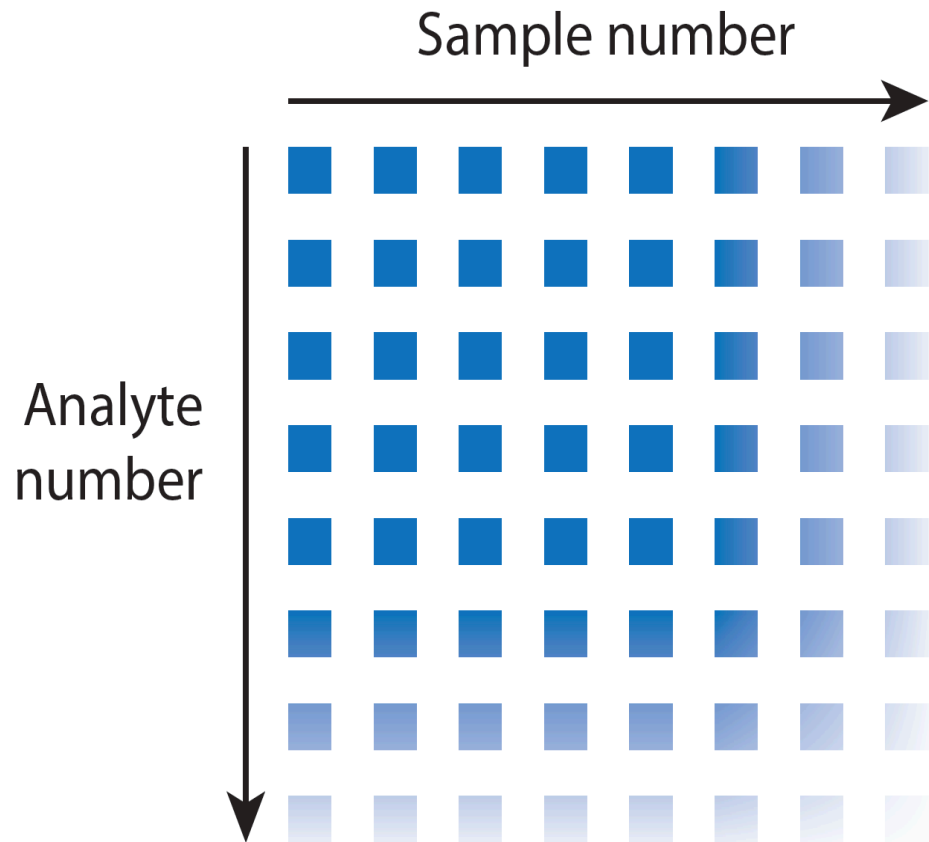
Phenotype



# Data matrix: The new currency of data driven approaches

---

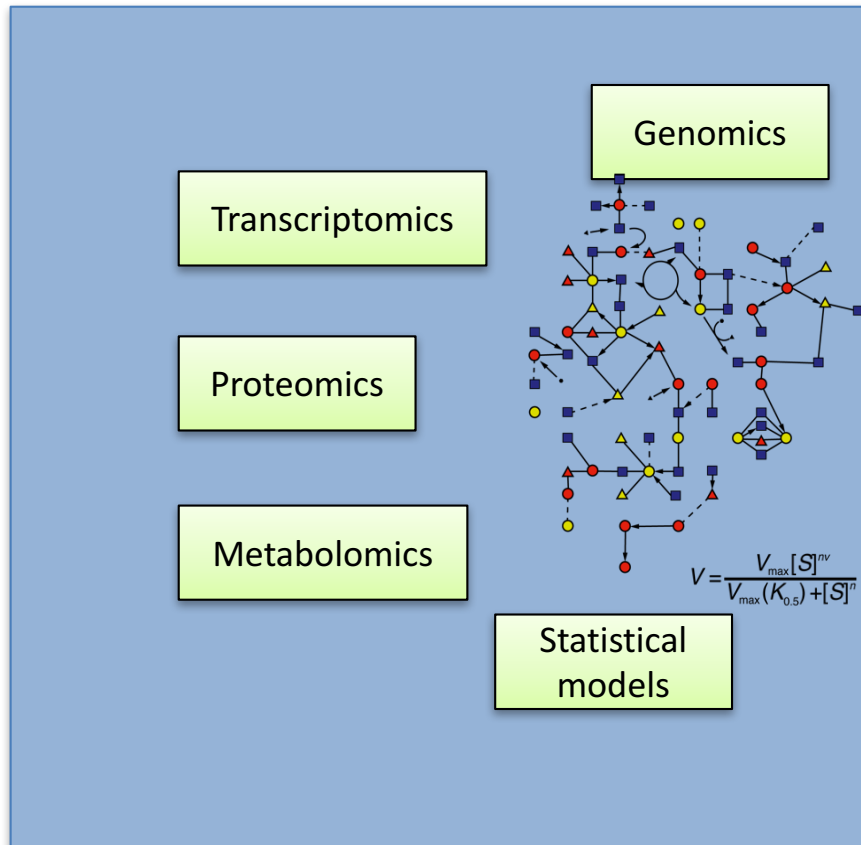
## Data matrix



- The data matrices open new ways for biological experimentation
- Support the use of machine learning and statistical tools to gain new insights

# Data Driven Research Approach

## BigData world: Focus on large datasets



### Main **questions**:

Can we discover the role of a gene/protein in the context of all other proteins (member of a society)

Main **technique** need to be fast and highly reproducible

Learn the behavior of a system from observing how it reacts to perturbations

# Consequences for biopolymer analysis

---

- To generate a data matrix biopolymers have to be precisely measured and quantified in a large number of conditions
- Few missing values
- High throughput and high accuracy
- Second third generation sequencing techniques for DNA
- Massively parallel targeting mass spectrometry techniques for proteins