Grundlagen der Systembiologie und der Modellierung epigenetischer Prozesse

Sonja J. Prohaska

Bioinformatics Group
Institute of Computer Science
University of Leipzig

October 18, 2010

A System

- microlevel of a system is *elements* + *relations*
- macrolevel of a system is the system as a whole
- it can show "behavior" involving input, throughput, and output of material, energy or information
- a system may contain sub-sytems and may itself constitute an element of a higher-order system
- its boundaries are subjective
- it is a expedient model adopted to the subject of study

Reductionism and Holism

Reductionism

- the behavior of a system is determined by its elements alone
- emergent phenomena can be explained completely by the elements and their relations
- a theory A can be reduce to a theory B if the laws of A can be derived from the lows of B

Holism

- the behavior of the elements is determined by the system as a whole
- emergent phenomena cannot be deduced from the properties of the elements alone
- the whole is more than the sum of its parts

Systems Biology

- tries to explain organisms as a whole
- focuses on structures and dynamics of cellular functions instead of static observations of single elements
- aims at quantivication of biological processes
- explicitly takes processes into account that interact with the process of interest

Approaches to Study Biological Systems

Top-down Approach

- measure thousands of reactants with high-throughput methods in parallel
- find structures in the data
- formulate a hypothesis that gives a mechanistic explaination
- validate the hypothesis by specific experiments

Bottom-up Approach

- describe the components and their interactions
- apply mathematical modelling to predict the behavior of the system
- compare the predictions with observations

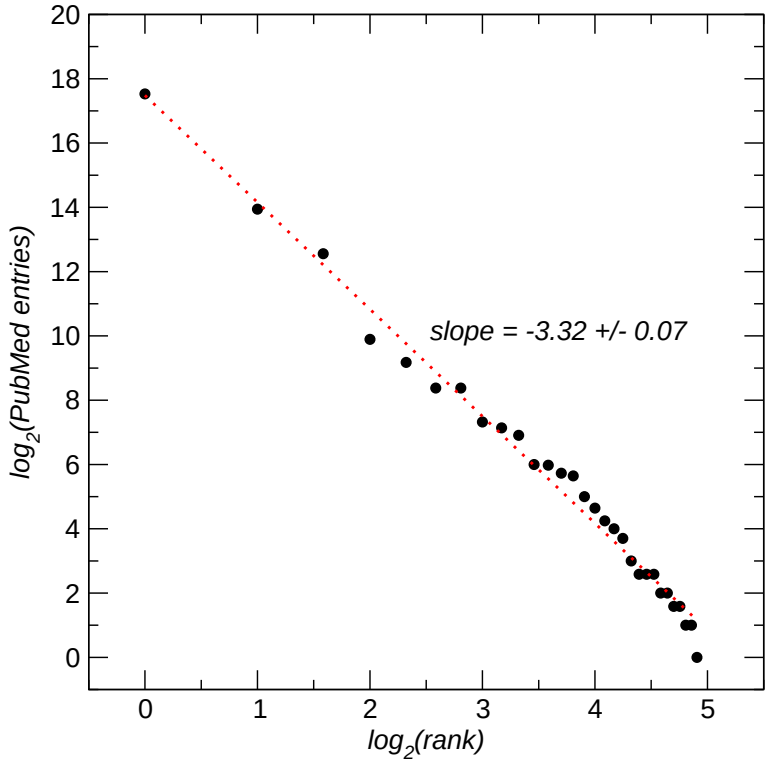
Large Data Collections

- suffix *-ome* refers to "all constituents considered collectively"
- should address a well-defined collection of biological objects/constituents
- should use a suite of (high-throughput) technologies allowing parallel measurements
- should catalog all constituents and their quantities in a sample
- should make the catalog available as a database

- the corresponding research field is labeled *-omics*
- *functional -omics*: ascribing biological function to the individual objects
- *comparative -omics*: cross-species comparison
- *computational -omics*: computational and statistical methods to analyze the large amounts of data

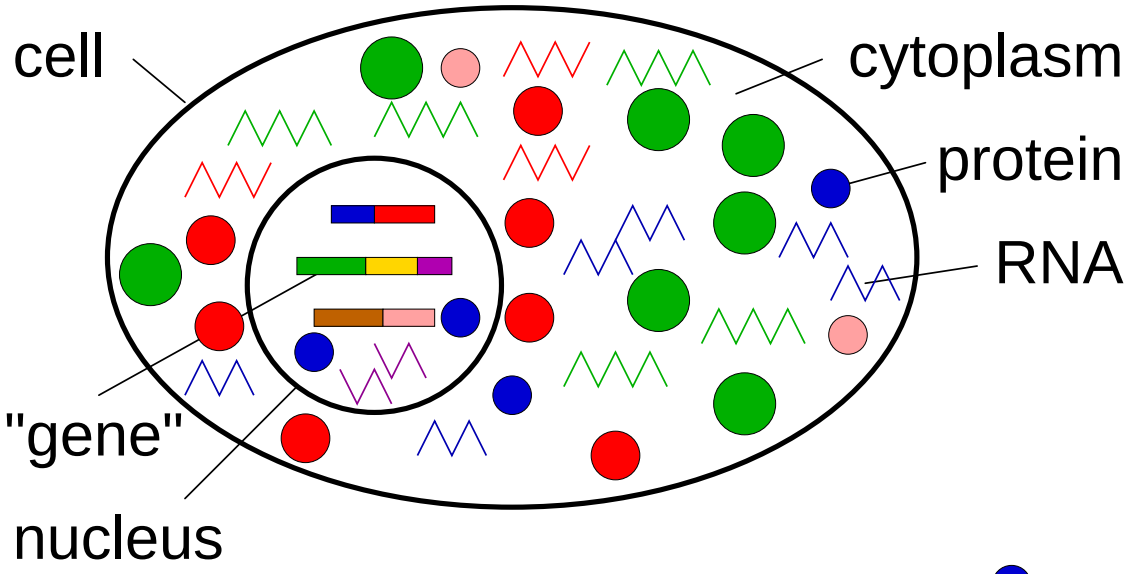
-omics

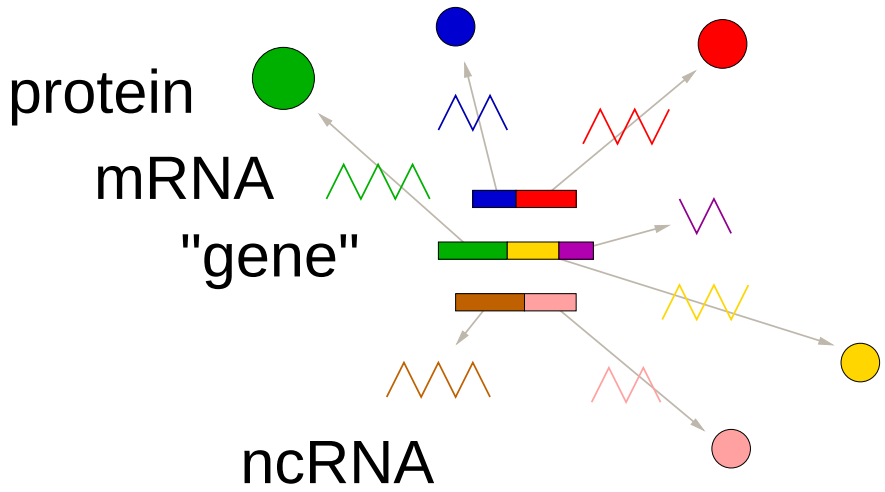
	PubMed entries		
	-ome[s]		*-omics*
	number	since	number
genome	189019	1943	55750
proteome	15756	1995	23343
transcriptome	6022	1997	778
metabolome	950	1998	1686
interactome	578	1999	43
epigenome	375	1987	189
secretome	333	2000	8
peptidome	160	2001	158
phenome	141	1989	102
glycome	120	2000	479
lipidome	64	2001	279
orfeome	63	2000	1
degradome	53	2003	22
cellome	32	2002	68
fluxome	25	1999	21
regulome	19	2004	2
variome	16	2006	–
toponome	13	2003	7
transportome	8	2004	–
modificome	6	2006	3
translatome	6	2001	2
localizome	6	2002	–
ribonome	4	2002	10
RNome	4	2005	54
morphome	3	1996	1
recombinome	3	2006	–
signalome	2	2001	–
expressome	2	2007	–
foldome	1	2009	1



Usage of *-ome* and *-omics* terms in the scientific literature. PubMed was queried on Fri Jan 29 2010 for "*ome or *omes" and "*omics" for each of the terms below. The distribution of *-ome* and *-omics* terms follows a power law. Only a handful of top-ranking terms are commonly used.

potential constituents





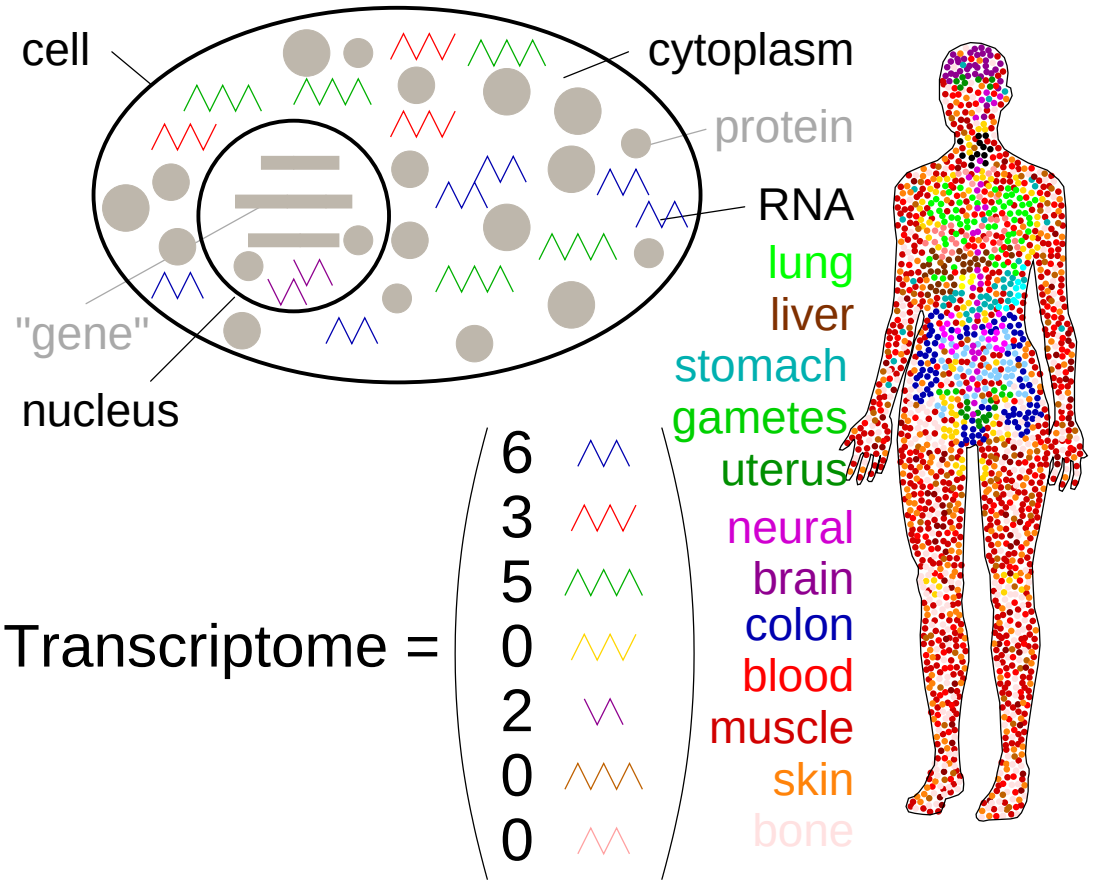
a constituents *-ome*

... is an n-dimensional vector, where n is the number of total constituents of the whole system in all possible (healthy) conditions.

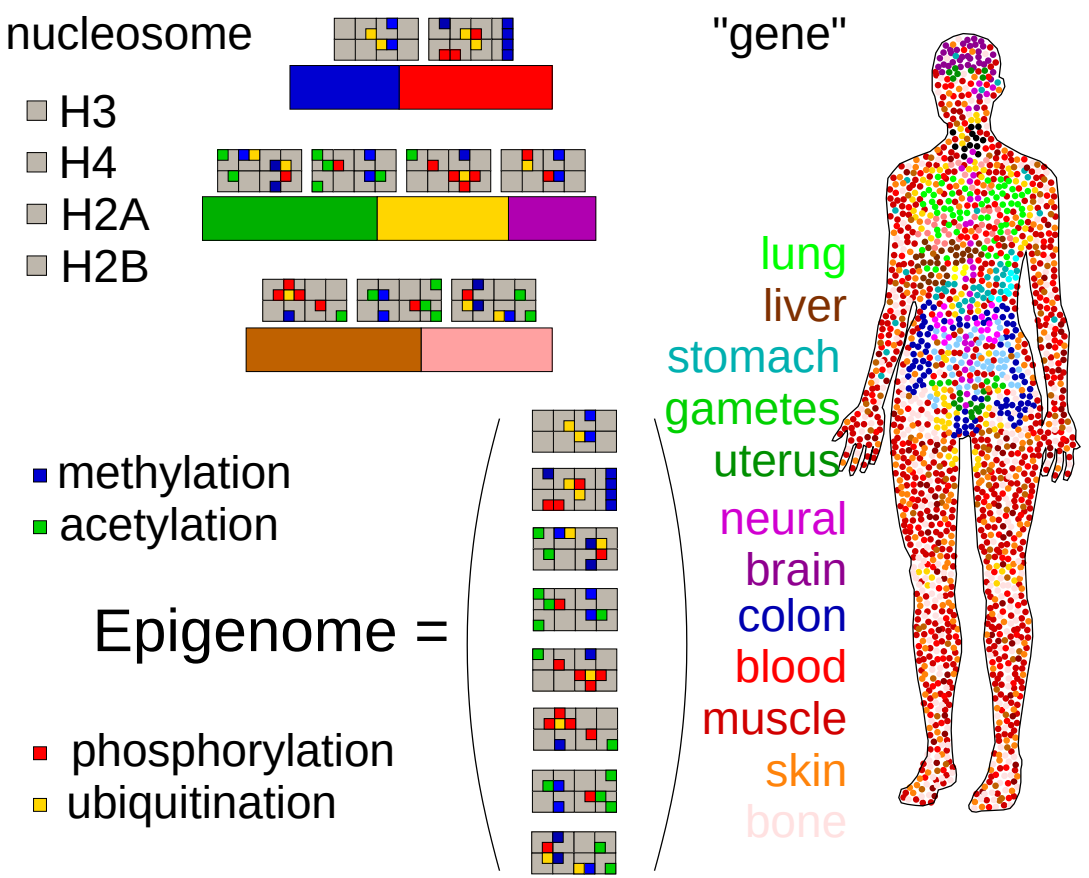
When is the catalog complete?

The state of the *-ome* is the absolute or relative amount of each constituent at a specific time in a specific sample.

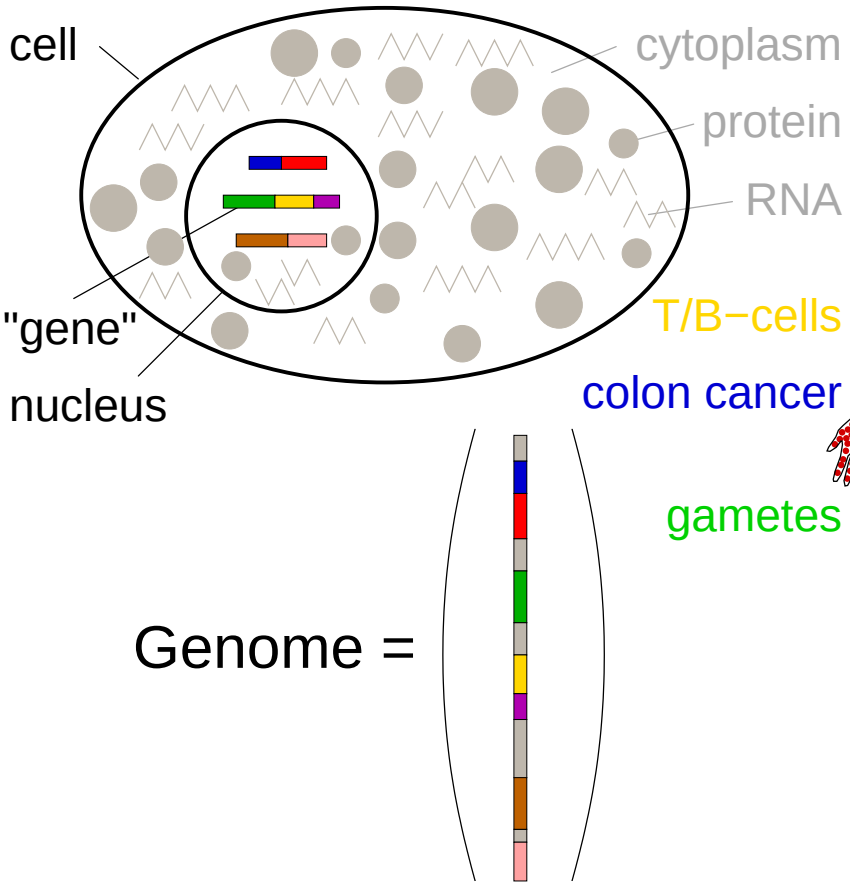
Transcriptomics

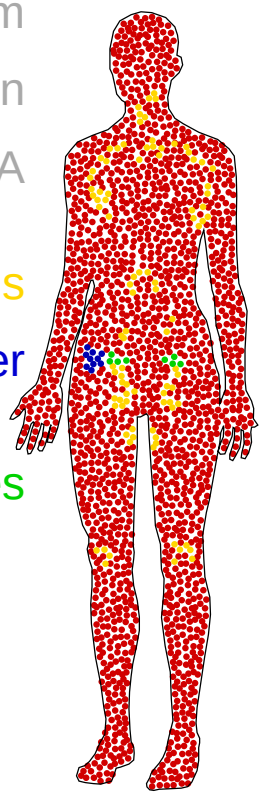


Epigenomics



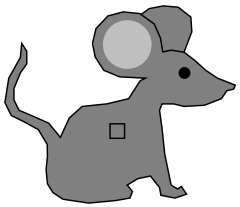
Genomics



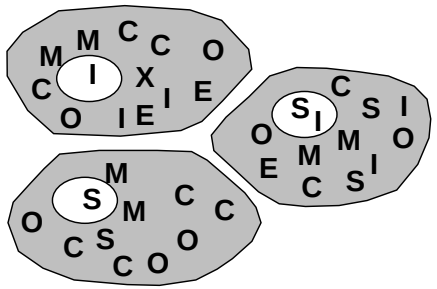


exception from the rule: “genome” combines “gene” with the “chromosome” (i.e. “color body”, from the Greek “chromo”, “color”, and “soma”,”body”) – all genes considered collectively?

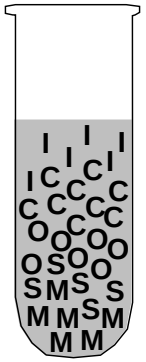
Interpretability of Transcriptiome Data



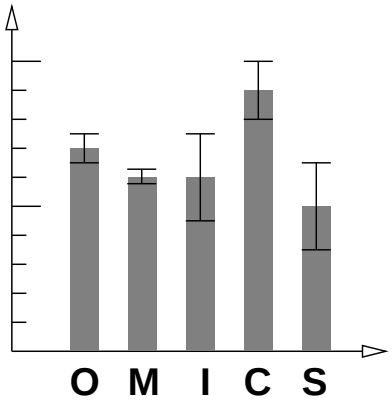
Spatio−Temporal Sample



Individual Variation



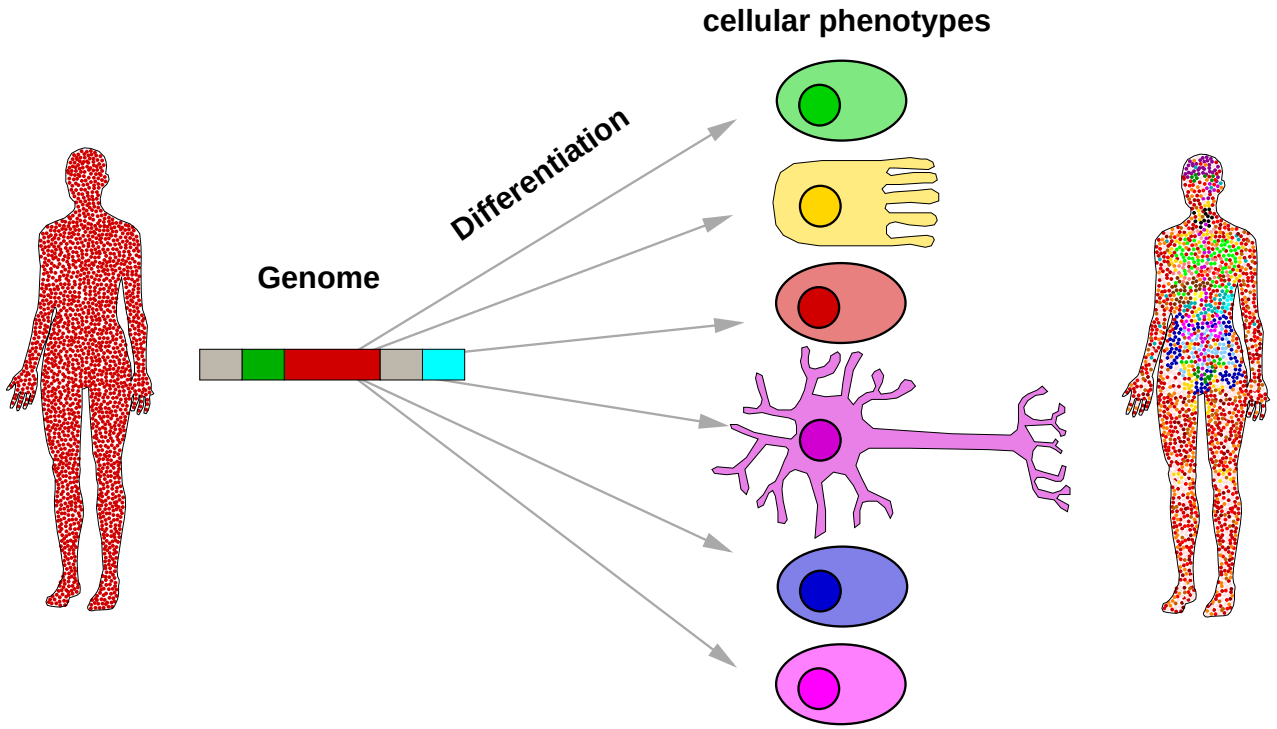
Completeness



Precision

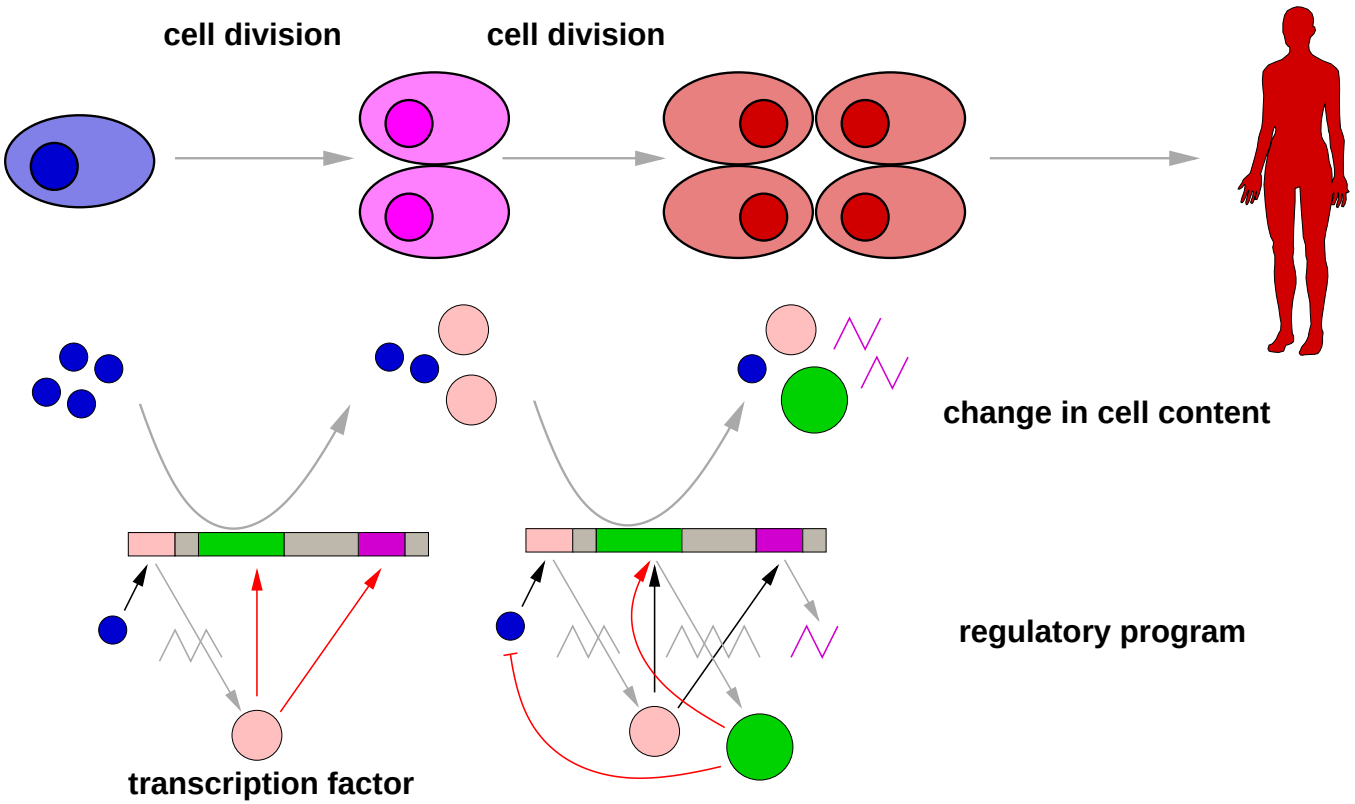
- depends on the sample that was taken e.g. from a whole mouse
- individual cells can contribute variation (see “C” and “S”)
- the completeness of the list is strongly dependent on the measurement technique (in the figure only “one-stroke characters” are analyzed). Transcripts that are not recognized (see “E” and “X”) will systematically be missed.
- rare transcripts may not be negectible (“E” and “X” necessary to form “MEXICO” and “MEIOSIS”)
- several orders of magnitude separate the high and low abundant transcripts and complicate the quantification process
- the measurement technique a source of imprecision, making it difficult to distinguished true variation from uncertainty

Differentiation



... is the transition between functionally different states. Even though "all" cells have the same genome, the regulatory program induces different cellular phenotypes and cell functions.

Regulatory Program



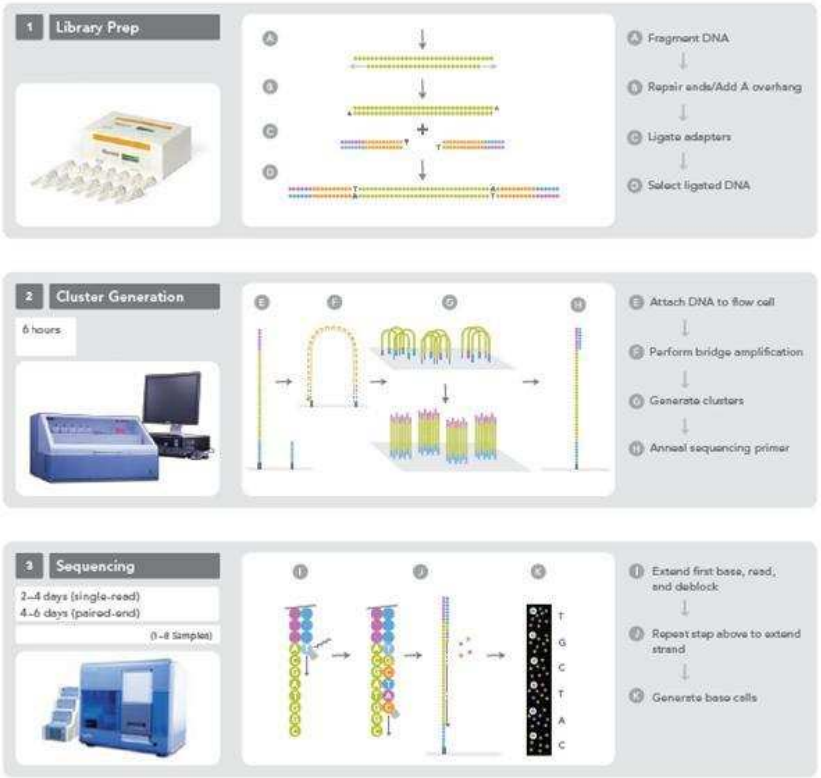
Changes in the amount of constituants correlate with the transition from one state $t-1$ to another state t. The regulatory program is understood as a markov process, meanig that the state t depends only on the state $t-1$.

Limitations of *-omics*

- technical limitations
  - caused by the measurement technology
  - uncertain connection between quantity of the constituent and the signal produced by the measurement technology (e.g. cross-hybridization on microarrays, low-abundant transcripts)
- limitations in the experimental design
  - assumptions about the occurence of the constituents and the universality of certain features (for practical reasons) (e.g. GeneChips, poly-T primers)
- conceptual limitations
  - design of an experiment is built on a "bad" concept
  - signal from the measurement produce complex aggregates of data
  - "what is it that we are measuring?"
- limitations in the analysis
  - partial loss of information from the measurment
  - analytical tasks may be algoritmically difficult (e.g. *in-situ* hybridization patterns)
  - quick-and-dirty analysis to pick low-hanging fruits

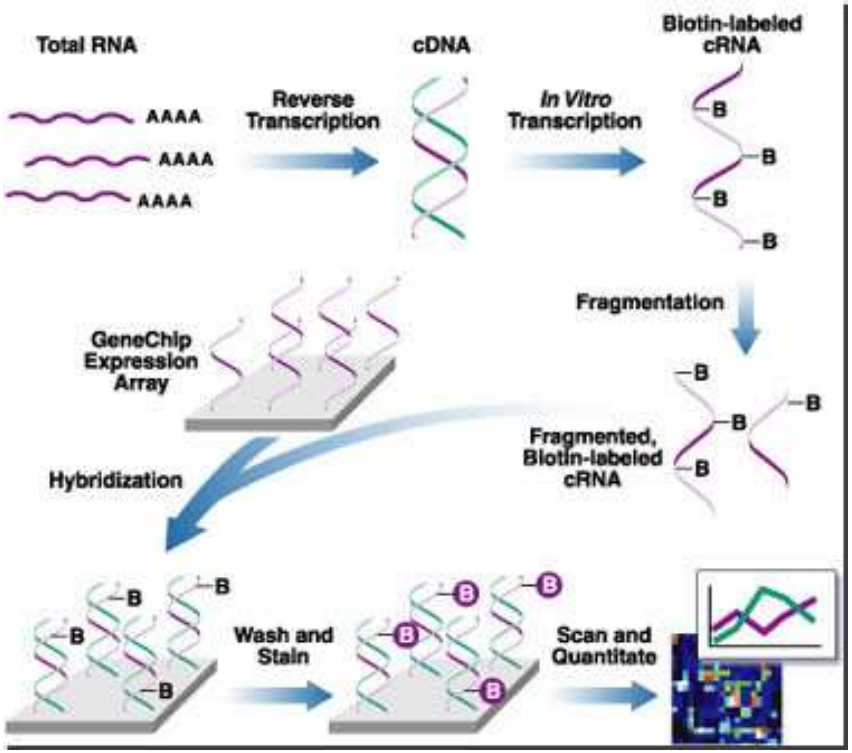
Two Transcriptome Experiments

HeLa cells
chop RNA, amplify with adapters



RNAseq using Solexa

whole mouse embryo
amplify with poly-T primers



GeneChip

Two Transcriptome Experiments

- How complete is the catalog obtained by the two methods?
- Do you expect to measure mRNA levels for histone mRNAs?
- Can you infer the amount of gene product obtained from a gene?
- Can you distinguish between alternative transcripts?
- Can you make statements about the transcription start site?
- If you are especially interested in zinc finger proteins, which problems do you expect?
- Are the expression levels comparable in the two experiments?
- What happens if the sum of all genes is devided up onto two GeneChips?

References

 [Prohaska, 20xx] Prohaska, SJ and Stadler PF. *The Use and Abuse of -Omes*.