V11 –
8. Function Annotation and Protein Synthesis

- Gene Ontology: annotate function to gene and gene products, e.g. to differentially expressed genes
- Similarity of GO Terms
- Translation of Proteins

Tue, Nov. 26, 2019

The Gene Ontology (GO)

Ontologies are **structured vocabularies**.

The Gene Ontology consists of

3 non-redundant areas:

- Biological process (BP)

- molecular function (MF)

- cellular component (localisation).

Shown here is a part of the BP

vocabulary.

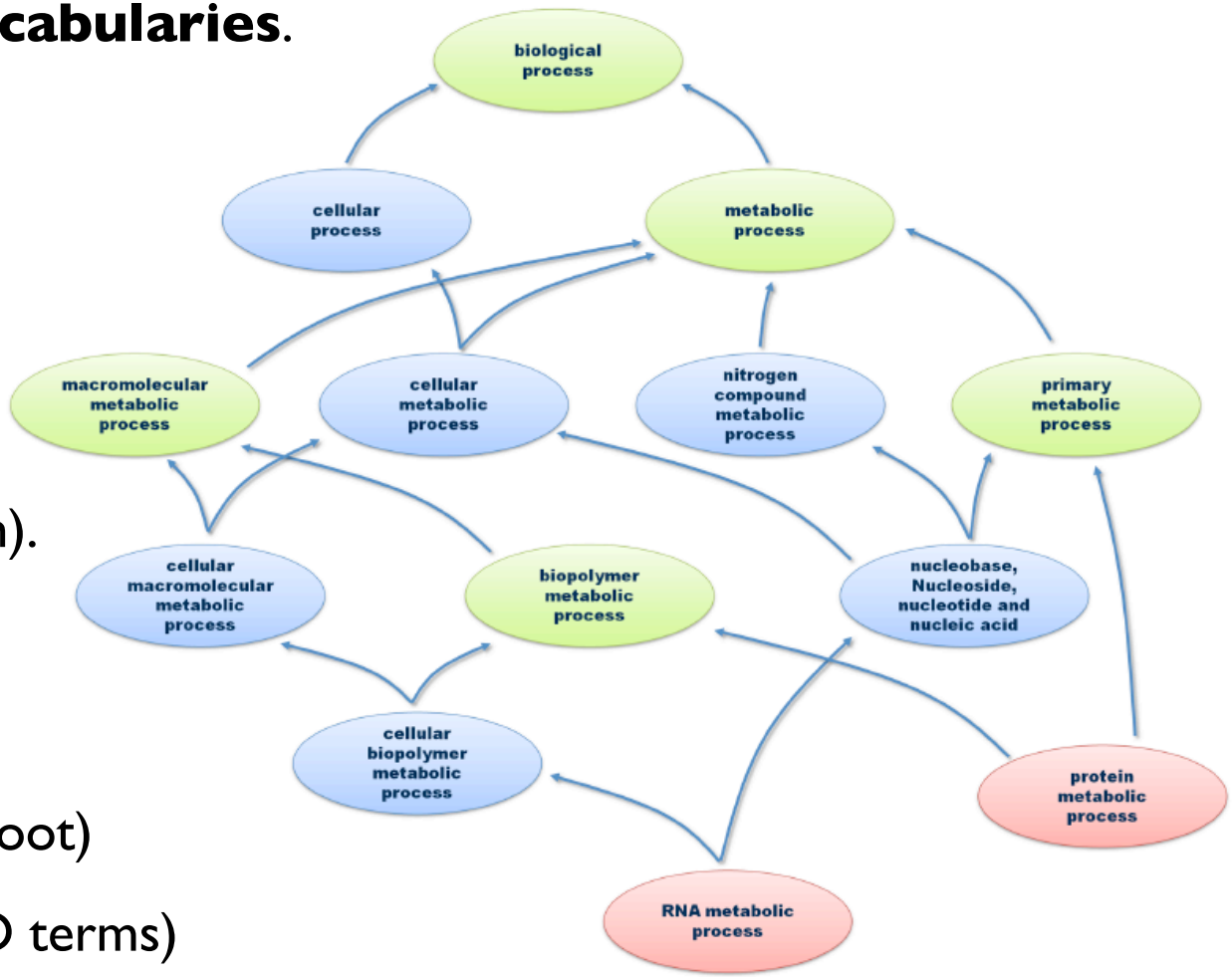
At the top: most general term (root)

Red: tree leafs (very specific GO terms)

Green: common ancestor

Blue: other nodes.

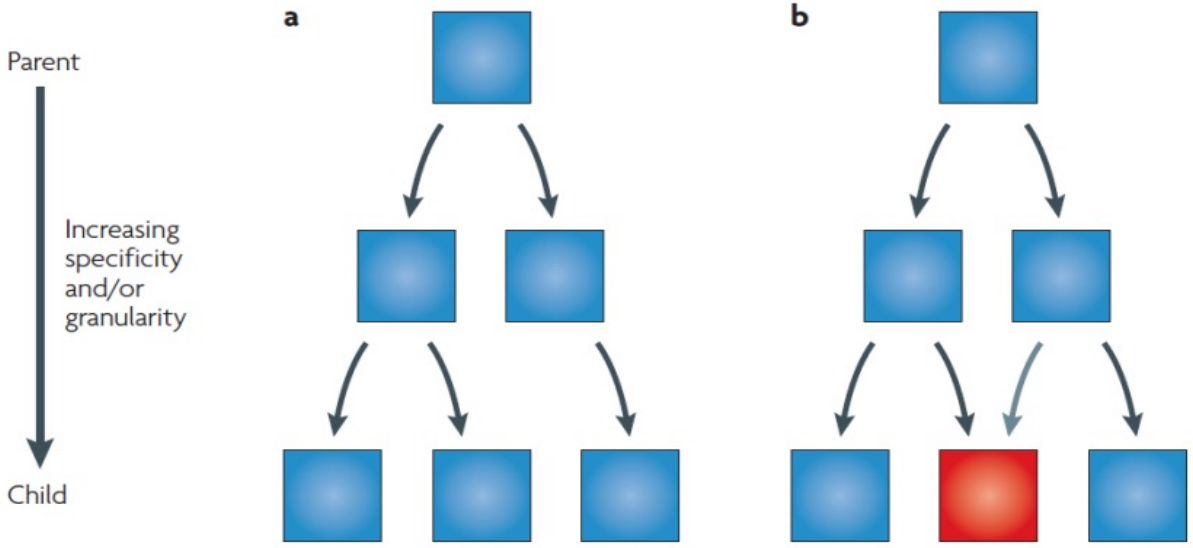
Arcs: relations between parent and child nodes



PhD Dissertation

Andreas Schlicker (UdS, 2010)

Simple tree vs. cyclic graphs



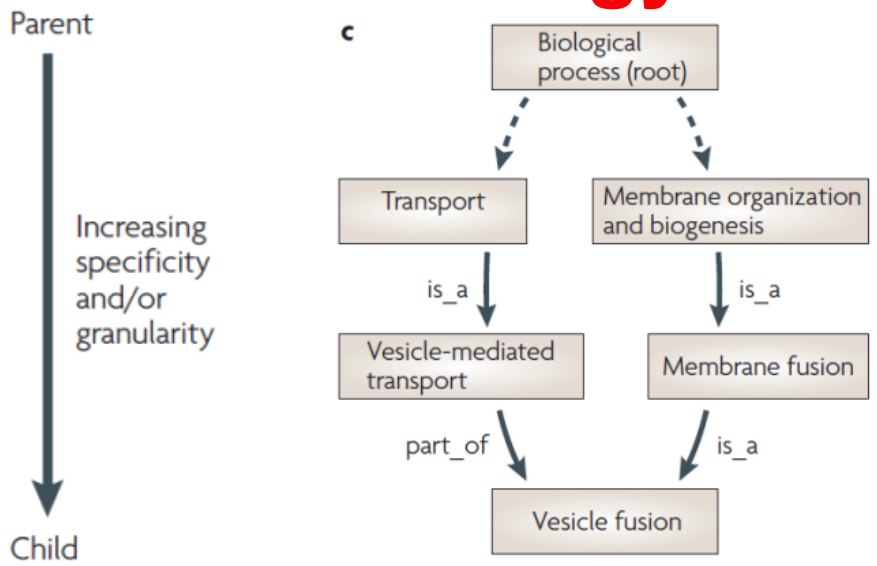
Boxes represent nodes; arrows represent edges.

a | An example of a simple **tree**, in which each child has only one parent and the edges are directed.
That is, there is a source (parent) and a destination (child) for each edge.

b | A **directed acyclic graph** (DAG), in which each child can have one or more parents.
The **red-colored node** has **multiple parents**. The additional edge is colored grey.

Rhee et al. (2008) Nature Rev. Genet. 9: *509*

Gene Ontology is a directed acyclic graph



An example of the node
vesicle fusion
in the BP ontology with multiple parentage.

(Arrows point into the wrong direction.)

Dashed edges : there are other nodes not shown between the nodes and the root node.

Root : node with no incoming edges, and at least one leaf.

Leaf node : a terminal node with no children (vesicle fusion).

Similar to a simple tree, a DAG has directed edges and does not have cycles.

Depth of a node : length of the longest path from the root to that node.

Height of a node: length of the longest path from that node to a leaf.

Rhee et al. (2008) Nature Rev. Genet. 9: *509*

relationships in GO

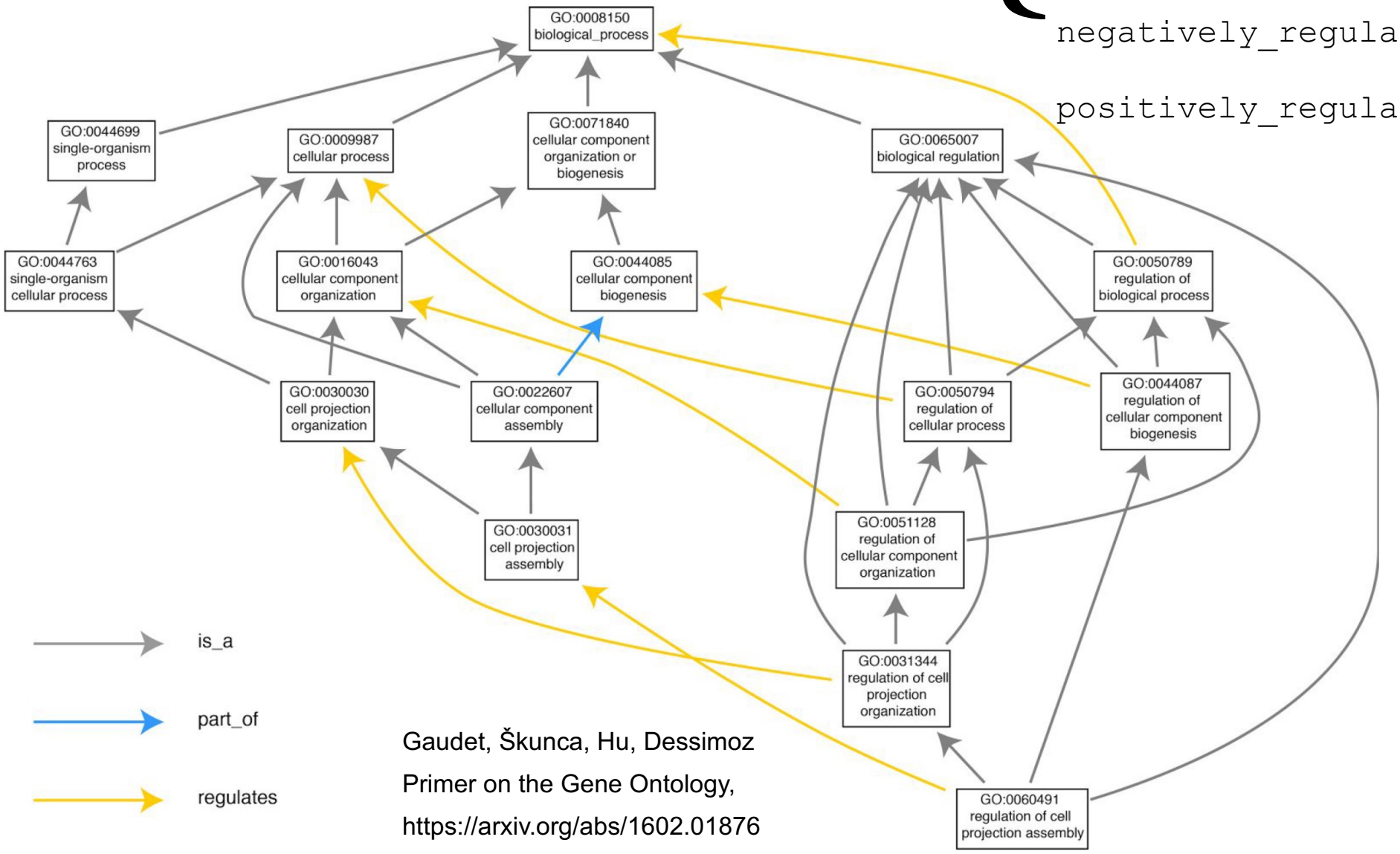
is_a

is a part_of

Gene X { regulates relationship

negatively_regulates

positively_regulates



Gaudet, Škunca, Hu, Dessimoz

Primer on the Gene Ontology,

https://arxiv.org/abs/1602.01876

Where do the Gene Ontology annotations come from?

Evidence code	Evidence code description	Source of evidence	Manually checked	Current number of annotations*
IDA	Inferred from direct assay	Experimental	Yes	71,050
IEP	Inferred from expression pattern	Experimental	Yes	4,598
IGI	Inferred from genetic interaction	Experimental	Yes	8,311
IMP	Inferred from mutant phenotype	Experimental	Yes	61,549
IPI	Inferred from physical interaction	Experimental	Yes	17,043
ISS	Inferred from sequence or structural similarity	Computational	Yes	196,643
RCA	Inferred from reviewed computational analysis	Computational	Yes	103,792
IGC	Inferred from genomic context	Computational	Yes	4
IEA	Inferred from electronic annotation	Computational	No	15,687,382
IC	Inferred by curator	Indirectly derived from experimental or computational evidence made by a curator	Yes	5,167
TAS	Traceable author statement	Indirectly derived from experimental or computational evidence made by the author of the published article	Yes	44,564
NAS	Non-traceable author statement	No 'source of evidence' statement given	Yes	25,656
ND	No biological data available	No information available	Yes	132,192
NR	Not recorded	Unknown	Yes	1,185

*October 2007 release

Rhee et al. Nature Reviews Genetics 9, 509-515 (2008)

IEA: Inferred from Electronic Annotation

The evidence code **IEA** is used for all inferences made without human supervision, regardless of the method used.

The IEA evidence code is by far the **most abundantly** used **evidence** code.

Guiding idea behind computational function annotation:

genes with similar sequences or structures are likely to be **evolutionarily related**.

Thus, assuming that they largely kept their ancestral function, they might still have **similar functional roles** today.

Gaudet, Škunca, Hu, Dessimoz
Primer on the Gene Ontology,
https://arxiv.org/abs/1602.01876.
Published in : Methods in Molecular Biology
Vol1446 (2017) – **open access**!

Significance of GO annotations

Very **general GO terms** such as "`cellular metabolic process`" are annotated to many genes in the genome.

Very **specific terms** belong to a few genes only.

→ One needs to compare how **significant** the occurrence of a GO term is in a given set of genes compared to a randomly selected set of genes of the same size.

This is often done with the **hypergeometric test**.

PhD Dissertation Andreas Schlicker (UdS, 2010)

Hypergeometric test

$$\text{p-value} = \sum_{i=k_\pi}^{min(n,K_\pi)} \frac{\binom{K_\pi}{i}\binom{N-K_\pi}{n-i}}{\binom{N}{n}}$$

The hypergeometric test is a statistical test.

It can be used to check e.g. whether a biological annotation π is **statistically significant enriched** in a given test set of genes compared to the full genome.

- N : number of genes in the genome
- n : number of genes in the test set
- K_π : number of genes in the genome with annotation π.
- k_π : number of genes in test set with annotation π.

The hypergeometric test provides the **likelihood** that k_π or more genes that were **randomly selected** from the genome also have annotation π.

Hypergeometric test

Select $i \geq k_\pi$ genes with annotation π from the genome.

There are K_π such genes.

The other $n - i$ genes in the test set do NOT have annotation π. There are $N - K_\pi$ such genes in the genome.

$$\text{p-value} = \sum_{i=k_\pi}^{min(n,K_\pi)} \frac{\binom{K_\pi}{i}\binom{N-K_\pi}{n-i}}{\binom{N}{n}}$$

The sum runs from k_π elements to the maximal possible number of elements.

This is either the number of genes with annotation π in the genome (K_π) or the number of genes in the test set *(n).*

number of possibilities for selecting *n* elements from a set of *N* elements.

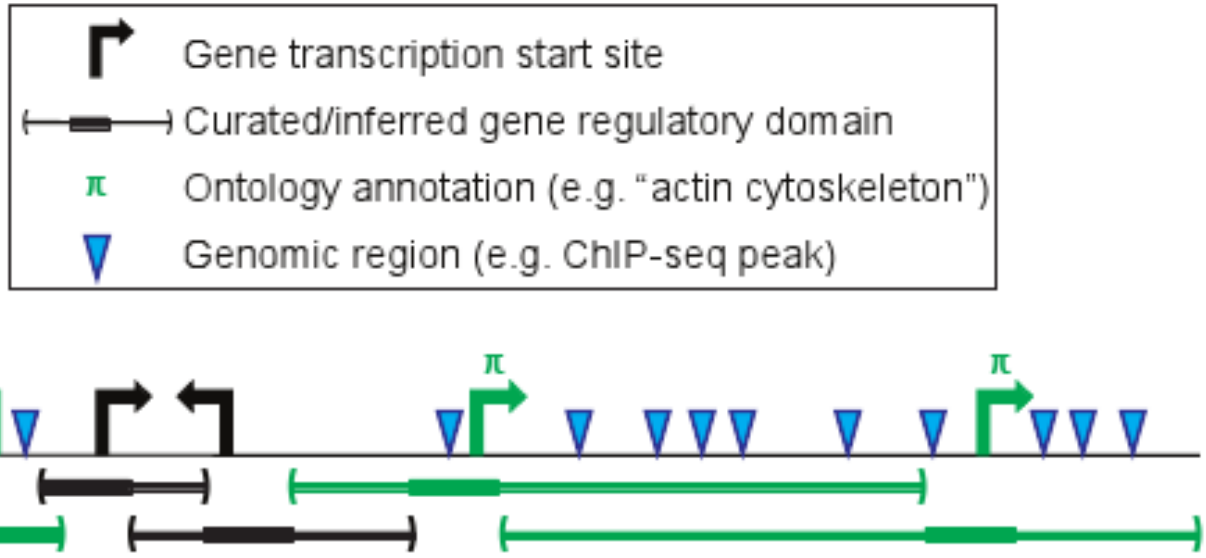
This correction is applied if the sequence of drawing the elements is not important.

http://great.stanford.edu/

http://www.schule-bw.de/

Example

p-value = $\sum_{i=k_-}^{min(n,K_\pi)} \frac{\binom{K_\pi}{i}\binom{N-K_\pi}{n-i}}{\binom{N}{n}}$



Is annotation π significantly enriched in the test set of 3 genes?

Hypergeometric test over genes
N = 6 total genes
K_π = 3 genes annotated with π
n = 3 genes with an associated genomic region
k_π = 3 genes annotated and with a genomic region
P-value = 0.05

Yes! p = 0.05 is (just) significant.

http://great.stanford.edu/

Comparing GO terms

The hierarchical structure of the GO allows to compare proteins annotated to different terms in the ontology, as long as the terms have relationships to each other.

Terms located close together in the ontology graph (i.e., with a few intermediate terms between them) tend to be **semantically more similar** than those further apart.

One could simply count the **number of edges** between 2 nodes as a measure of their similarity.

However, this is problematic because not all regions of the GO have the same **term resolution**.

Gaudet, Škunca, Hu, Dessimoz
Primer on the Gene Ontology,
https://arxiv.org/abs/1602.01876

Information content of GO terms

The **likelihood** of a node *t* can be defined in 2 ways:

How many genes have annotation *t* relative to the root node?

$$p_{anno}(t) = \frac{\cdot\, occur(t)}{occur(root)}$$

Number of GO terms in subtree below *t* relative to number of GO terms in tree

$$p_{graph}(t) = \frac{D(t)}{D(root)}$$

The likelihood takes values between 0 and 1 and increases monotonic from the leaf nodes to the root.

Define **information content** of a node from its likelihood:

$$IC(t) = -\log p(t)$$

A rare node has high information content.

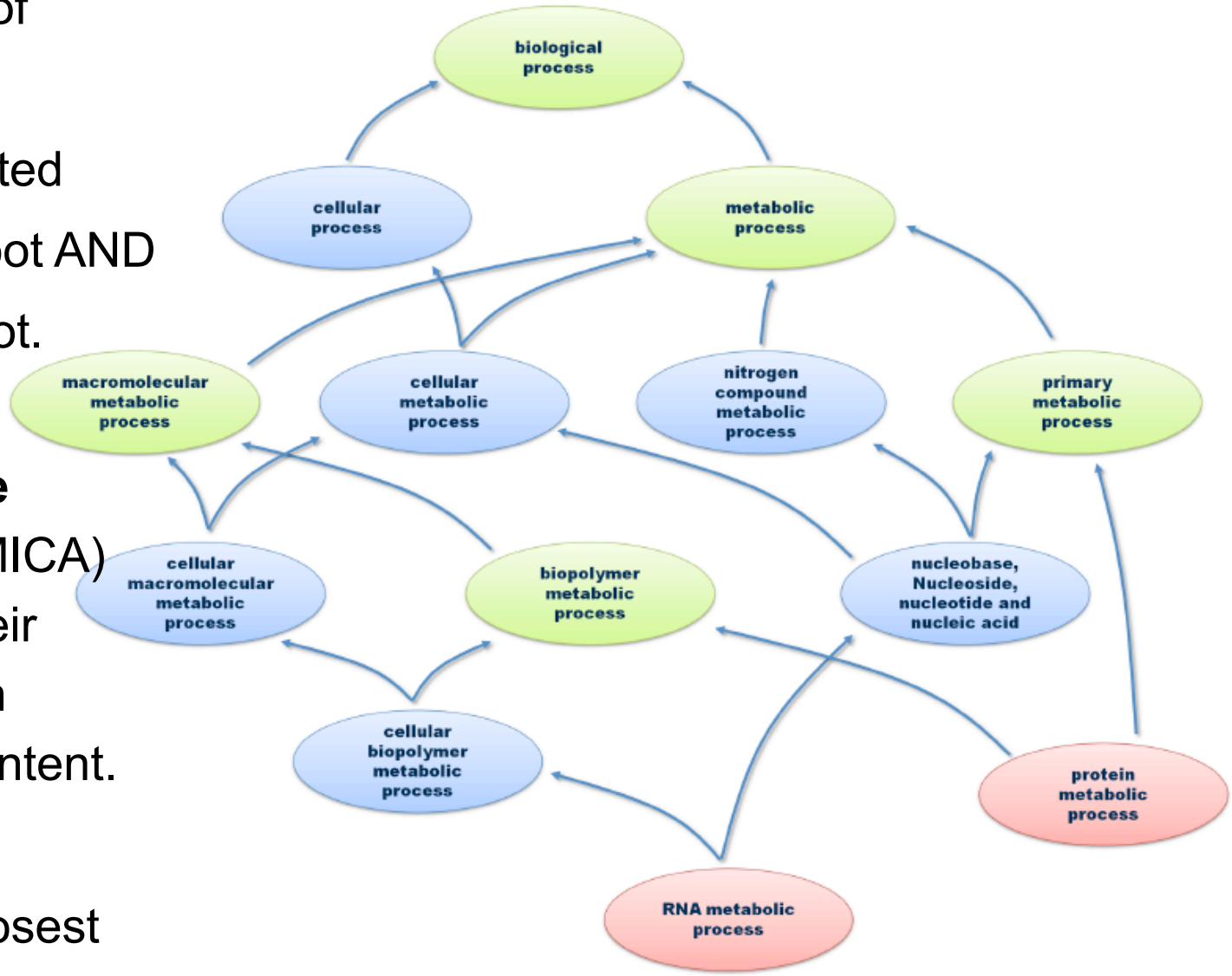
PhD Dissertation Andreas Schlicker (UdS, 2010)

Common ancestors of GO terms

Common ancestors of two nodes t_1 and t_2 : all nodes that are located on a path from t_1 to root AND on a path from t_2 to root.

The **most informative common ancestor** (MICA) of terms t_1 und t_2 is their common ancestor with highest information content.

Typically, this is the closest common ancestor.



PhD Dissertation

Andreas Schlicker (UdS, 2010)

Measure functional similarity of GO terms

Lin *et al.* defined the **similarity** of two GO terms t_1 und t_2

based on the information content of the most informative common ancestor (MICA)

$$sim_{Rel}(t_1,t_2) = \frac{2 \cdot IC(MICA)}{IC(t_1)+IC(t_2)}$$

If MICAs are close to the two GO terms, they receive a high similarity score.

Schlicker *et al*. defined the following variant:

$$sim_{Rel}(t_1,t_2) = \frac{2 \cdot IC(MICA)}{IC(t_1)+IC(t_2)} \cdot (1 - p(MICA))$$

where the term similarity is weighted with the counter-probability of the MICA.

By this, shallow annotations (low “depth” in the tree, slide #4) receive less relevance than MICAs further away from the root.

PhD Dissertation Andreas Schlicker (UdS, 2010)

Measure functional similarity of two genes

Two genes or two sets of genes *A* und *B* typically have more than 1 GO annotation each. → Consider similarity of all terms *i* and *j*:

$$s_{ij} = sim(GO_i^A, GO_j^B), \forall i \in 1,...,N, \forall j \in 1,...,M.$$

and select the maxima in all rows and columns:

$$rowScore(A,B) = \frac{1}{N}\sum_{i=1}^{N} \max_{1 \le j \le M} s_{ij}, \qquad GOscore_{avg}^{BMA}(A,B) = \frac{1}{2} \cdot (rowScore(A,B) + columnScore(A,B))$$

$$columnScore(A,B) = \frac{1}{M}\sum_{j=1}^{M} \max_{1 \le i \le N} s_{ij}. \quad GOscore_{max}^{BMA}(A,B) = max(rowScore(A,B), columnScore(A,B))$$

Compute *funsim*-Score from scores for BP tree and MF tree:

$$funsim(A,B) = \frac{1}{2} \cdot \left[\left(\frac{BPscore}{\max(BPscore)} \right)^2 + \left(\frac{MFscore}{\max(MFscore)} \right)^2 \right]$$

PhD Dissertation Andreas Schlicker (UdS, 2010)

GO is inherently incomplete

The Gene Ontology is a representation of the **current state of knowledge**; thus, it is very **dynamic**.

The ontology itself is constantly being improved to more accurately represent biology across all organisms.

The ontology is augmented as new discoveries are made.

The **creation of new annotations** occurs at a rapid pace, aiming to keep up with published work.

Despite these efforts, the information contained in the GO database is necessarily **incomplete**.

Thus, absence of evidence of function does not imply absence of function.

This is referred to as the **Open World Assumption**

Gaudet, Dessimoz,
Gene Ontology: Pitfalls, Biases, Remedies
https://arxiv.org/abs/1602.01876

Summary

- The GO is the **gold-standard** for **computational annotation of gene function**.

- It is continuously updated and refined.

- **Hypergeometric test** is most often used to compute **enrichment** of GO terms in gene sets

- **Semantic similarity** concepts allow measuring the **functional similarity** of genes. Selecting an optimal definition for semantic similarity of 2 GO terms and for the mixing rule depends on what works best in practice.

- Functional gene annotation based on GO is affected by a number of **biases**.

Rates of mRNA transcription and protein translation

ARTICLE

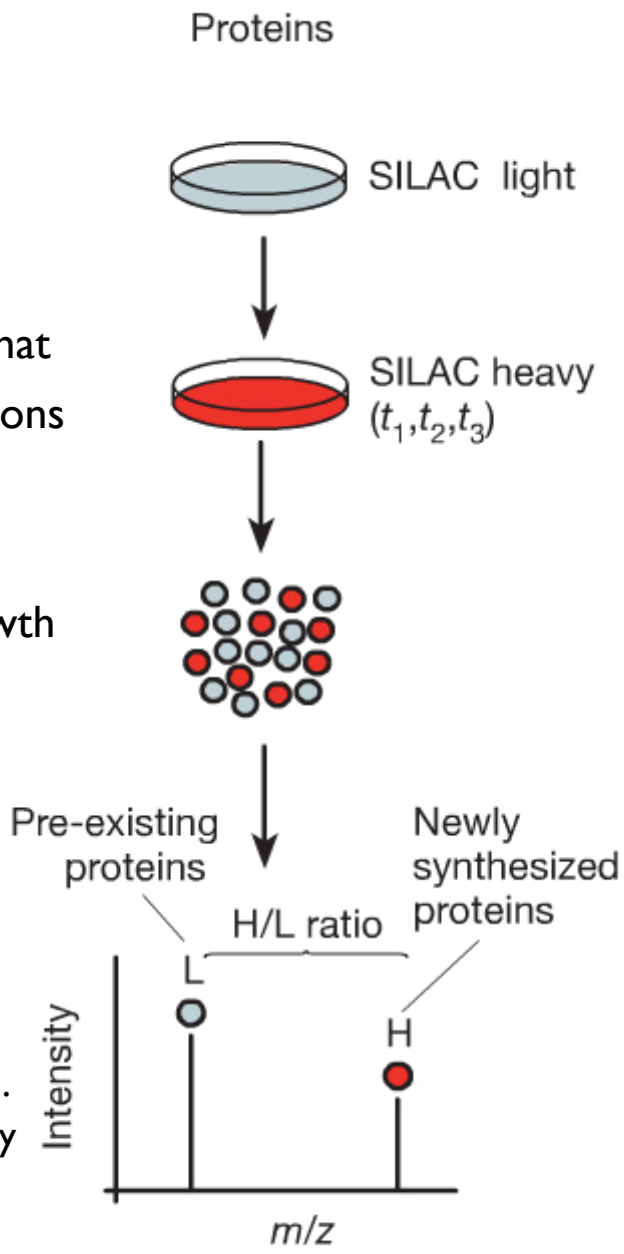
doi:10.1038/nature10098

Global quantification of mammalian gene expression control

Björn Schwanhäusser[1], Dorothea Busse[1], Na Li[1], Gunnar Dittmar[1], Johannes Schuchhardt[2], Jana Wolf[1], Wei Chen[1] & Matthias Selbach[1]

SILAC: „stable isotope labelling by amino acids in cell culture“ means that cells are cultivated in a medium containing **heavy** stable-isotope versions of **essential amino acids**.

When non-labelled (i.e. light) cells are transferred to heavy SILAC growth medium, newly synthesized proteins incorporate the heavy label while pre-existing proteins remain in the light form.



Quantification of protein turnover and levels. Mouse fibroblasts were pulse-labelled with heavy amino acids (SILAC). Protein turnover is quantified by mass spectrometry.

Schwanhäuser et al.
Nature 473, 337 (2011)

Rates of mRNA transcription and protein translation

ARTICLE

doi:10.1038/nature10098

Global quantification of mammalian gene expression control

Björn Schwanhäusser[1], Dorothea Busse[1], Na Li[1], Gunnar Dittmar[1], Johannes Schuchhardt[2], Jana Wolf[1], Wei Chen[1] & Matthias Selbach[1]

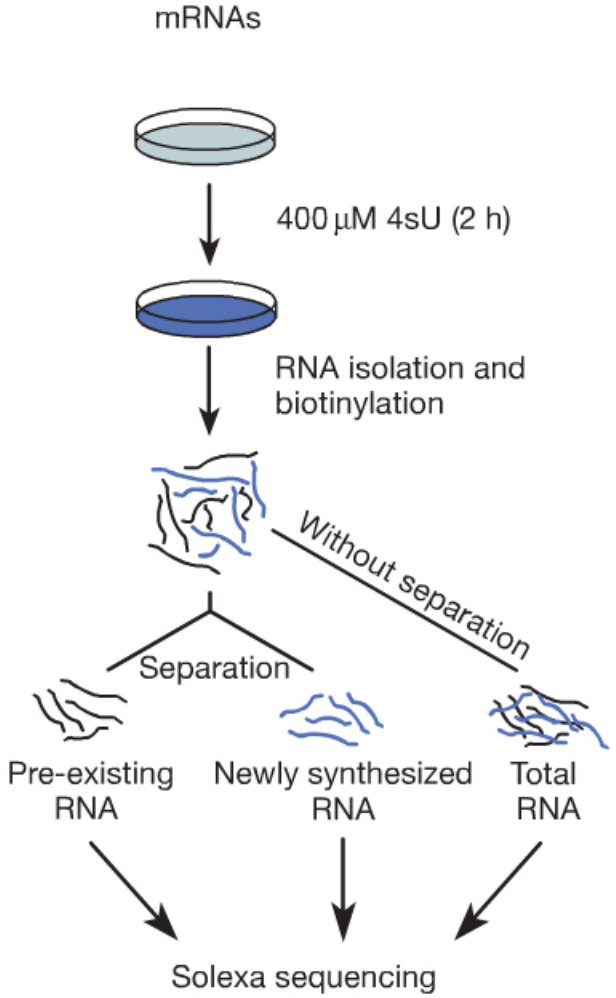
Quantification of mRNA turnover and levels. Mouse fibroblasts were pulse-labelled with the nucleoside **4-thiouridine** (4sU). mRNA turnover is quantified by next-generation sequencing.

The 4sU-labeled RNA fraction is thiol-specifically biotinylated generating a disulfide bond between biotin and the newly transcribed RNA.

'Total cellular RNA' can then be quantitatively separated into labeled ('newly transcribed') and unlabeled ('pre-existing') RNA with high purity using streptavidin-coated magnetic beads.

Finally, labeled RNA is recovered from the beads by simply adding a reducing agent (e.g. dithiothreitol) cleaving the disulfide bond and releasing the newly transcribed RNA from the beads.

Rädle, J Vis Exp. 2013; (78): 50195.



Rates of mRNA transcription and protein translation

84,676 peptide sequences were identified by MS and assigned to 6,445 unique proteins.

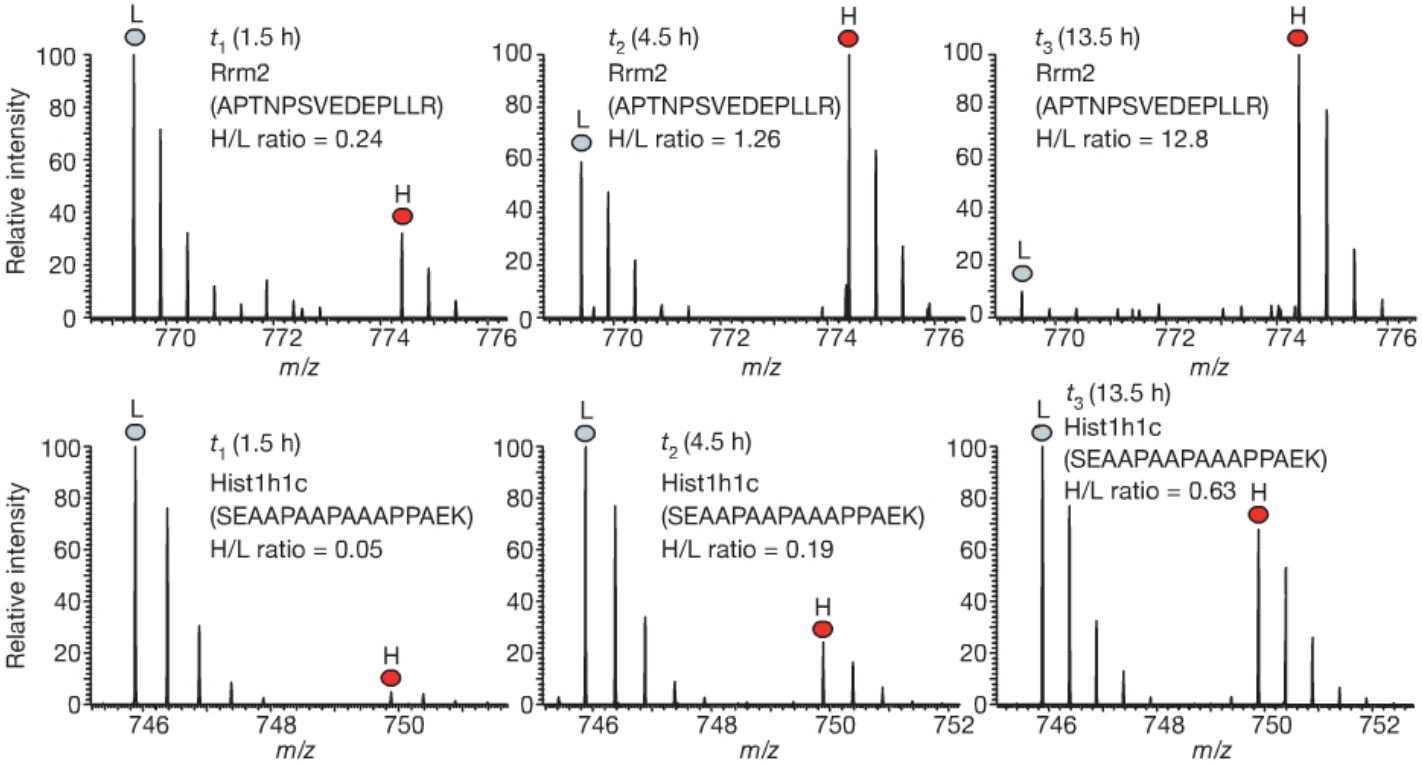
5,279 of these proteins were quantified by at least 3 heavy to light (H/L) peptide ratios belonging to these proteins.

Mass spectra of peptides for two proteins (x-axis: mass over charge ratio).

Over time, the heavy to light (H/L) ratios increase.

You should understand these spectra!

Top: **high-turnover protein**



Schwanhäuser et al. Nature 473, 337 (2011)

Bottom: **low-turnover protein**, slow synthesis, long half-life

Protein half-lifes and decay rates

Consider ratio r of protein with heavy amino acids (P_H) and light amino acids (P_L):

$$r = \frac{P_H}{P_L}$$

Assume that proteins labelled with light amino acids decay exponentially with degradation rate constant k_{dp} :

$$P_L = P_0 e^{-k_{dp} t} .$$

Express (P_H) as difference between total number of a specific protein P_{total} and P_L:

$$P_H(t) = P_{total}(t) - P_L(t)$$

Assume that P_{total} doubles during duration of one cell cycle (which lasts t_∞):

$$P_H(t) = P_{total}(t) - P_L(t) = P_0 2^{t/t_{cc}} - P_L(t) ,$$

$$r = \frac{P_H}{P_L} = \frac{P_0}{P_L} 2^{\frac{t}{t_{cc}}} - 1$$

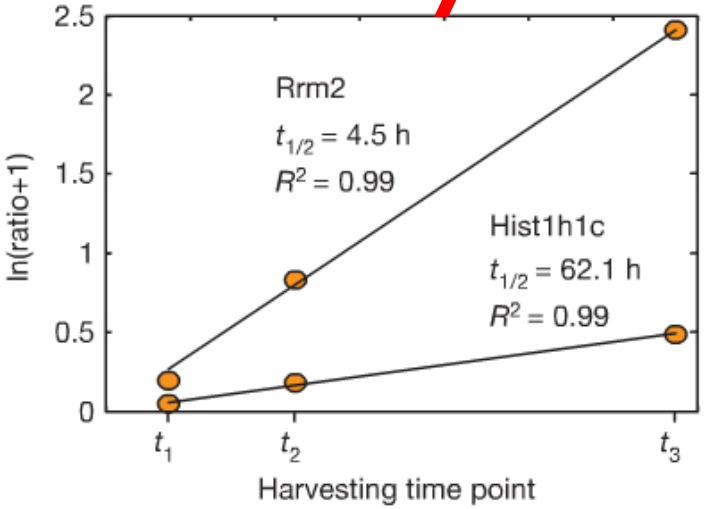
$$\frac{P_H}{P_L} + 1 = \frac{P_0}{P_L} 2^{\frac{t}{t_{cc}}}$$

take *ln* on both sides

$$\ln(ratio + 1) = \ln \frac{P_0}{P_L} 2^{\frac{t}{t_{cc}}} = \ln e^{k_{dp} t} + \ln 2^{\frac{t}{t_{cc}}} = k_{dp} t + \ln 2^{\frac{t}{t_{cc}}}$$

$$ln(ratio + 1) = k_{dp} t + \frac{t}{t_{cc}} ln2 = t \times \left(k_{dp} + \frac{ln2}{t_{cc}} \right) \quad ln(ratio + 1)\, t = t^2 \times \left(k_{dp} + \frac{ln2}{t_{cc}} \right)$$

$$ln(ratio + 1)\, t = t^2 \times \left(k_{dp} + \frac{ln2}{t_{cc}} \right)$$



Consider m intermediate time points:

$$k_{dp} = \frac{\sum_{i=1}^{m} \log_e (r_{t_i} + 1) t_i}{\sum_{i=1}^{m} t_i^2} - \frac{\log_e 2}{t_{cc}} ,$$

From k_{dp} we get the desired half-life:

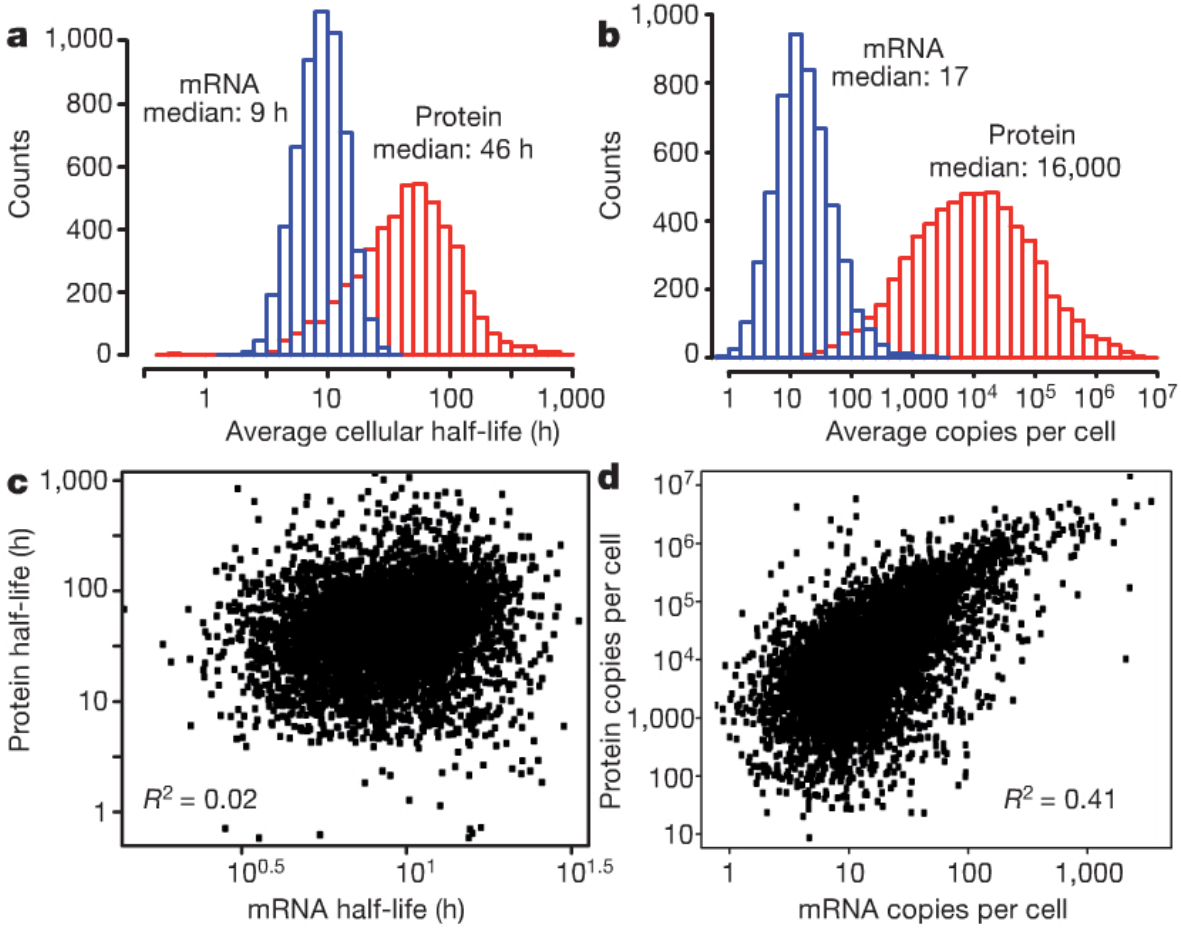
$$T_{1/2} = \frac{\log_e 2}{k_{dp}} .$$ because this gives

$$P_L = P_0 e^{-k_{dp} t} = P_0 e^{-k_{dp} \frac{log_e 2}{k_{dp}}} = P_0 e^{log_e \frac{1}{2}} = \frac{1}{2} P_0$$

The same is done to compute mRNA half-lives (not shown).

Schwanhäuser et al. Nature 473, 337 (2011)

mRNA and protein levels and half-lives



a, b, Histograms of mRNA (blue) and protein (red) half-lives (a) and levels (b).

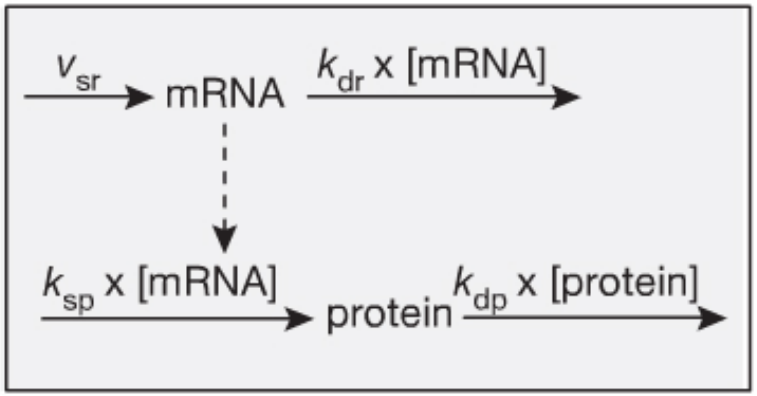
Proteins were on average 5 times more stable (46h vs. 9h) and 900 times more abundant than mRNAs.

(right) mRNA and protein levels showed reasonable correlation (R^2 = 0.41)
(left) However, there was practically no correlation of protein and mRNA half-lives.

Schwanhäuser et al. Nature 473, 337 (2011)

Mathematical model of transcription and translation

A widely used minimal description of the dynamics of transcription and translation includes the synthesis and degradation of mRNA and protein, respectively



$$\frac{dR}{dt} = v_{sr} - k_{dr}R$$

$$\frac{dP}{dt} = k_{sp}R - k_{dp}P$$

The mRNA (R) is synthesized with a constant rate v_{sr} and degraded proportional to their numbers with rate constant k_{dr}.

The protein level (P) depends on the number of mRNAs, which are translated with rate constant k_{sp}.

Protein degradation is characterized by the rate constant k_{dp}.

The synthesis rates of mRNA and protein are calculated from their measured half lives and levels.

Schwanhäuser et al. Nature 473, 337 (2011)

Computed transcription and translation rates

Top

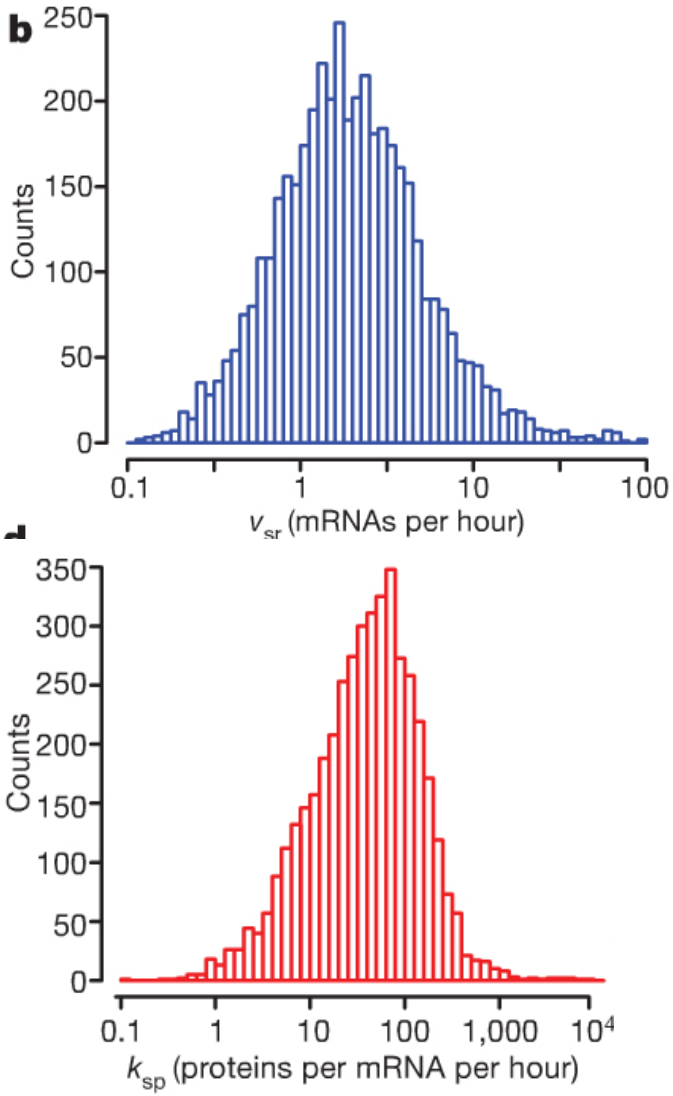
Average cellular **transcription rates** predicted by the model span two orders of magnitude.

The median is about 2 mRNA molecules per hour (**very slow**!).

An extreme example is the protein Mdm2 of which more than 500
mRNAs per hour are transcribed.

Bottom

The median **translation rate** constant
is about 40 proteins per mRNA
per hour



Calculated translation rate constants are not uniform

Schwanhäuser et al. Nature 473, 337 (2011)

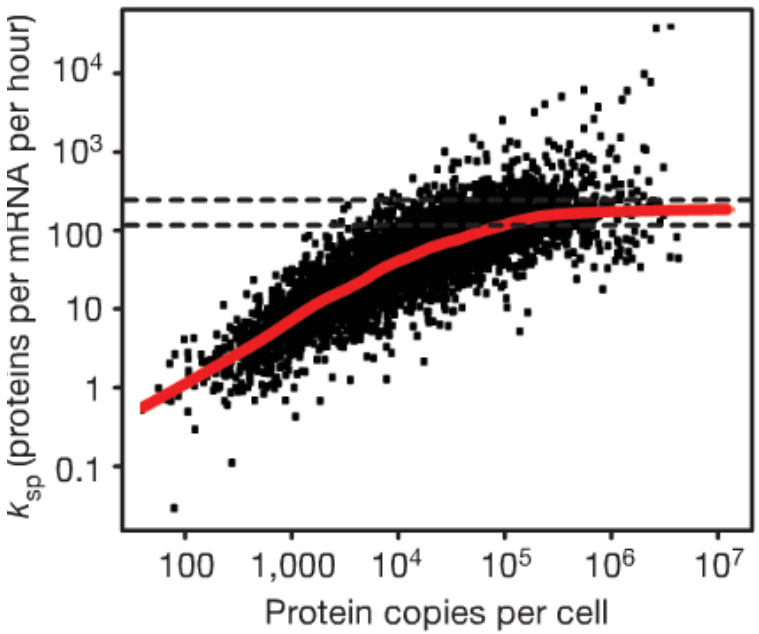
Maximal translation constant

Abundant proteins are translated about 100 times more efficiently than those of low abundance

Translation rate constants of abundant proteins saturate between approximately 120 and 240 proteins per mRNA per hour.

The maximal translation rate constant in mammals is not known.

The estimated maximal translation rate constant in sea urchin embryos is 140 copies per mRNA per hour, which is surprisingly close to the prediction of this model.



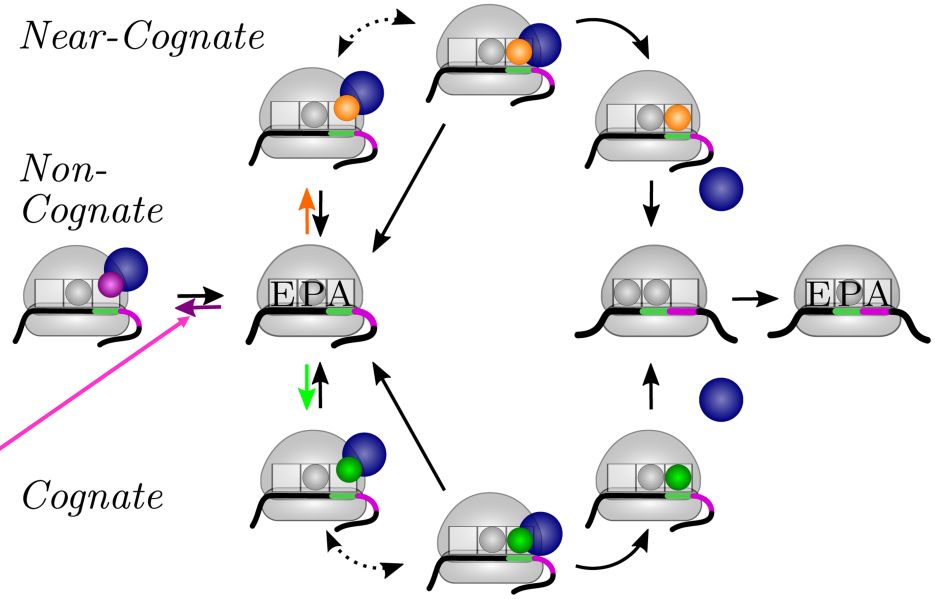
Schwanhäuser et al. Nature 473, 337 (2011)

Ribosomal mRNA translation

Elongation cycle of a ribosome (gray dome) translating an mRNA.

Aminoacyl-tRNA (small gray, **green**, **purple**, or **orange** sphere) is delivered to the ribosome in a ternary complex with the elongation factor EF-Tu (larger **blue** sphere) and GTP (not shown).

In addition to the initial binding site, the ribosome has 3 tRNA binding sites, the A, P, and E sites.

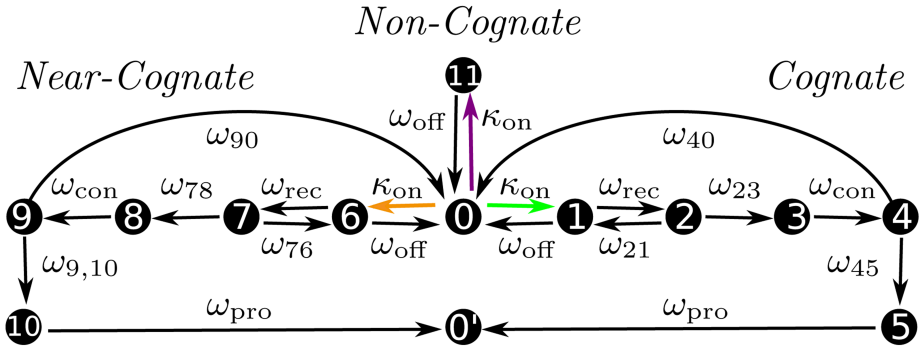


The elongation cycle of translation starts when the A site of the ribosome has arrived at a new codon (**green**) of the mRNA. The ribosome then binds a ternary complex with a tRNA that may be cognate (sequence matches), near-cognate, or non-cognate to this codon.

As a consequence, the elongation cycle exhibits 3 different branches corresponding to 3 different reaction pathways:
(left) A non-cognate ternary complex is again released from the initial binding site of the ribosome;
(top) A near-cognate ternary complex is usually rejected but is very rarely used to elongate the peptide chain; and
(bottom) A cognate ternary complex may also be rejected but is typically used for elongation of the peptide chain.

Rudorf et al. (2014)
PLoS Comput Biol 10: e1003909.

mRNA translation modelled as Markov process



(Left) All transition rates of this Markov model could be measured for *E.coli* in vitro. ω_{rec} : recognition rate, ω_{con} : conformational rate

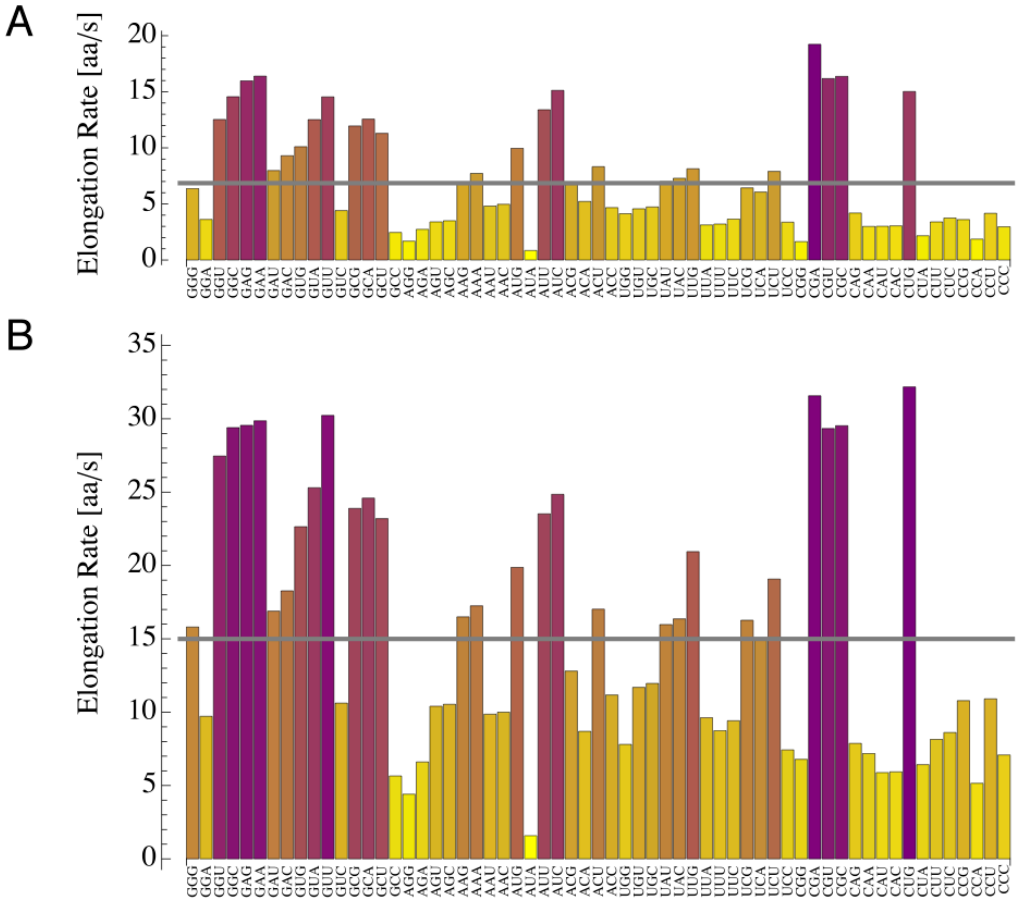
What are the in vivo rates?

From the model, one can deduce **codon-specific translation rates**:

(A) In-vitro values for high-fidelity buffer at 37°C.

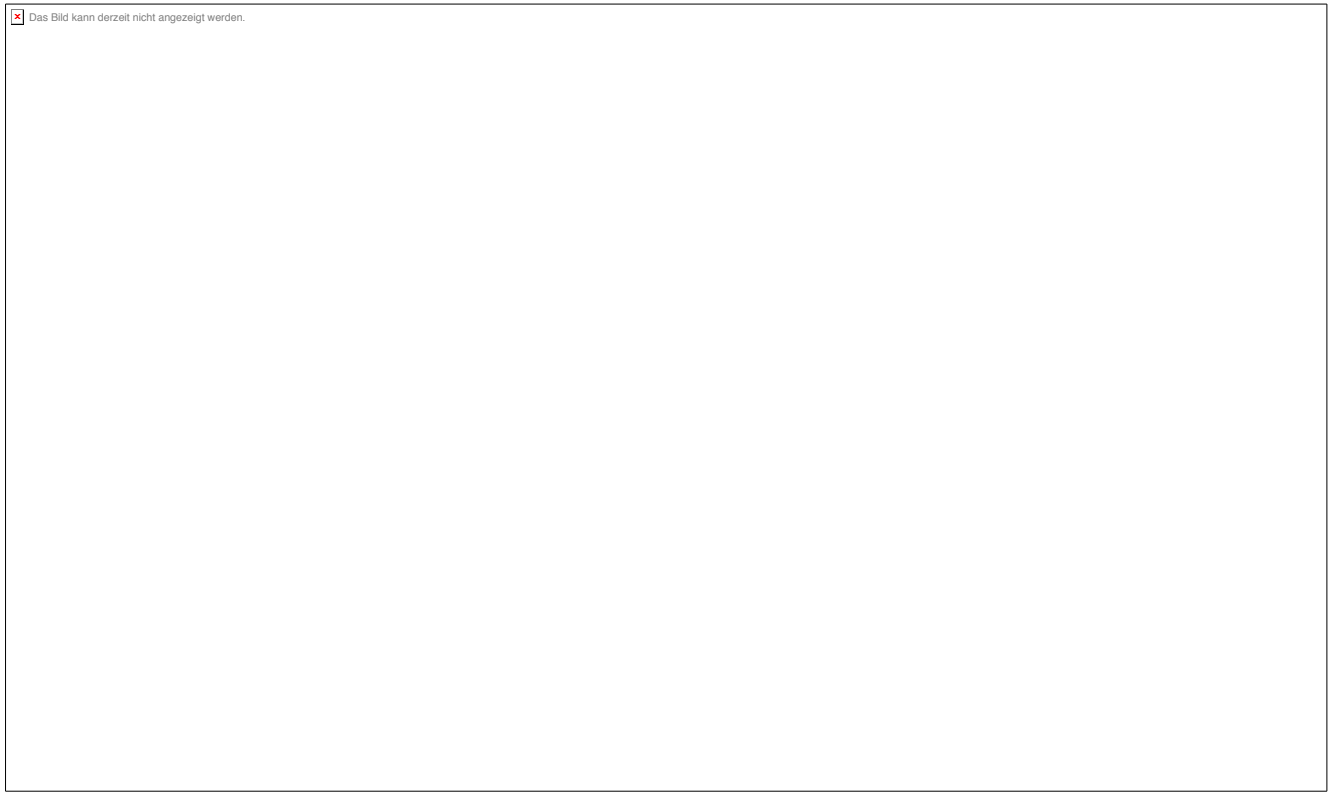
To derive in vivo rates from this, one adjusts the diffusion constant and uses exp. measured tRNA concentrations. This gave

(B) in-vivo values for *E. coli* at growth conditions of 0.7 dbl/h.



Rudorf et al. (2014) PLoS Comput Biol 10: e1003909.

Optimization of elongation rates



Sequences with alternative, synonymous codons are proposed from the original sequence and selected to maximize the protein expression score.

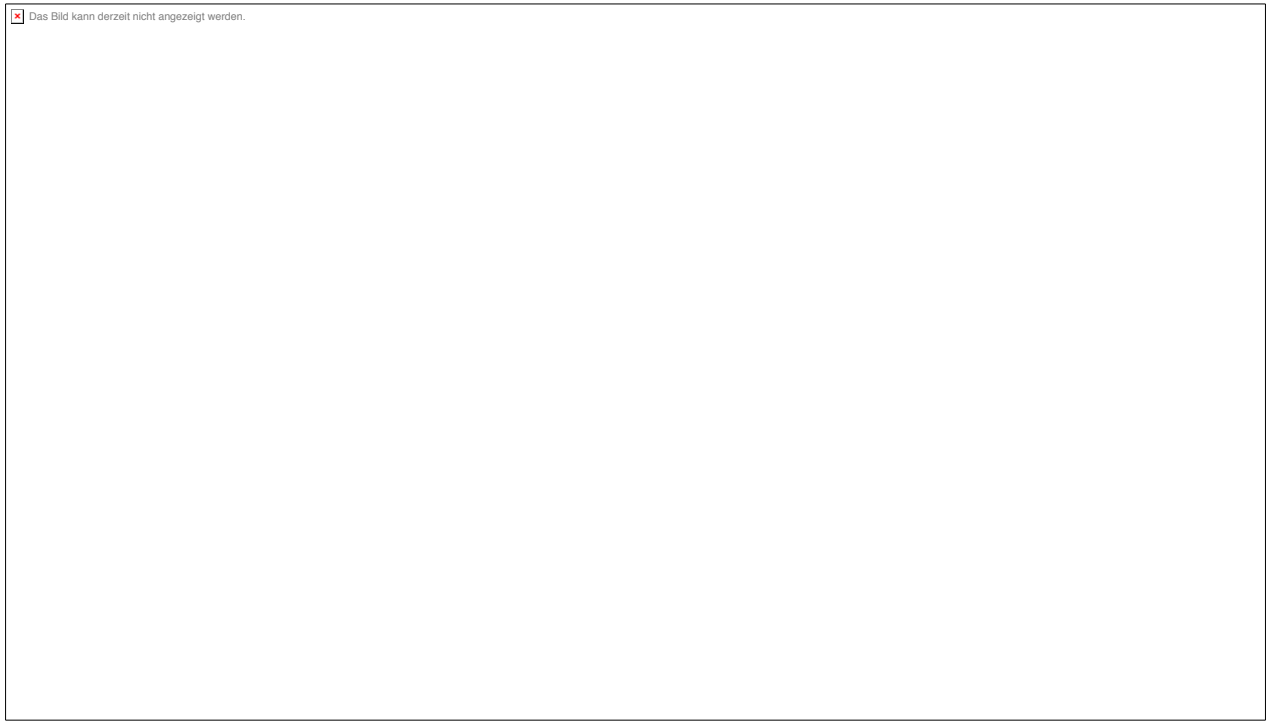
COSEM current: translation rate per mRNA transcript

mRNA secondary structure: mRNA folding energy in the first 30 codons of the 5′-end,

GC3 content: fraction of guanine and cytosine in the third nucleotide positions of all codons

Trösemeier et al. Sci. Rep. 9, 7511 (2019)

Derive codon-specific elongation rates



Protein expression of synthetic **ovalbumin** *(*main constituent of egg white*)* in *S. Typhimurium* *(*after artificial gene transfer*)*.

Measured protein abundance (Western blot) relative to wildtype compared to protein expression score relative to wildtype for *ova* variants.

Geneart (GeneOptimizer) from ThermoFisher is another tool to optimize codon usage.
This tool did not lead to increased Ova levels.

COSEM gave 3-4 fold increase.

Trösemeier et al. Sci. Rep. 9, 7511 (2019)

Codon-specific elongation rates in human

Table S9. Codon-specific elongation rates for HEK293.

Codon	Codon-specific elongation rate [s^{-1}]	Codon	Codon-specific elongation rate [s^{-1}]
AAA	21.21984	GAA	9.11984
AAC	14.12447	GAC	21.96470
AAG	11.23207	GAG	5.48026
AAU	14.12483	GAU	21.96470
ACA	10.48064	GCA	5.89915
ACC	14.78216	GCC	17.75464
ACG	7.33020	GCG	6.04496
ACU	14.78260	GCU	17.75464
AGA	1.59880	GGA	8.12605
AGC	15.08409	GGC	24.84194
AGG	7.37762	GGG	11.61666
AGU	15.08453	GGU	24.84194
AUA	2.51558	GUA	13.09163
AUC	11.82971	GUC	28.00956
AUG	34.86228	GUG	41.11098
AUU	11.81255	GUU	28.01011
CAA	12.59770	UAA	171.67736
CAC	18.93839	UAC	5.27462
CAG	27.28044	UAG	171.67736
CAU	18.93835	UAU	5.27632
CCA	8.84760	UCA	21.12999
CCC	1.00958	UCC	7.59376
CCG	5.92273	UCG	3.59940
CCU	1.00958	UCU	7.59376
CGA	17.13684	UGA	171.67736
CGC	7.35781	UGC	8.83135
CGG	10.10466	UGG	13.32441
CGU	7.35781	UGU	8.83135
CUA	3.35796	UUA	15.73559
CUC	0.71702	UUC	14.33256
CUG	16.39336	UUG	4.20173
CUU	0.71704	UUU	14.33295

UAA, UAG and UGA are stop codons.

The elongation rates for other codons are of similar magnitude (between 1 and 30 per second) as in *E.coli*.

Trösemeier et al. Sci. Rep. 9, 7511 (2019)

Transcriptome / Proteome

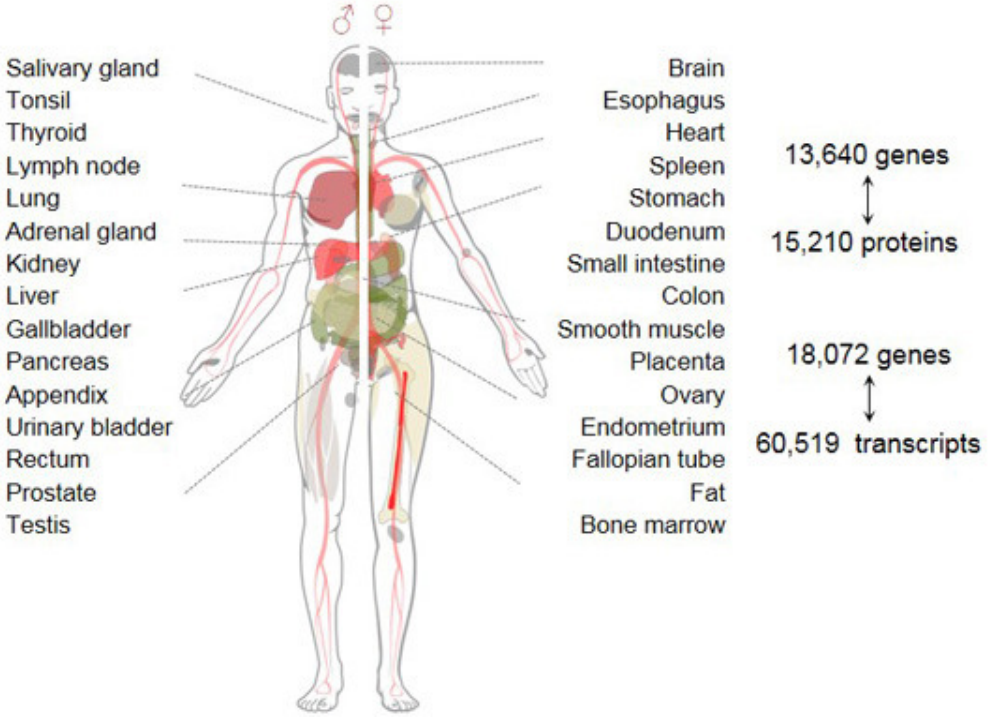
Article

TRANSPARENT PROCESS OPEN ACCESS

molecular systems biology

A deep proteome and transcriptome abundance atlas of 29 healthy human tissues

Dongxue Wang[1,†], Basak Eraslan[2,3,†], Thomas Wieland[4], Björn Hallström[5], Thomas Hopf[4], Daniel Paul Zolg[1], Jana Zecha[1], Anna Asplund[6], Li-hua Li[1], Chen Meng[1], Martin Frejno[1], Tobias Schmidt[1], Karsten Schnatbaum[7], Mathias Wilhelm[1], Frederik Ponten[6], Mathias Uhlen[5], Julien Gagneur[2,*], Hannes Hahne[4,**] & Bernhard Kuster[1,8,***]

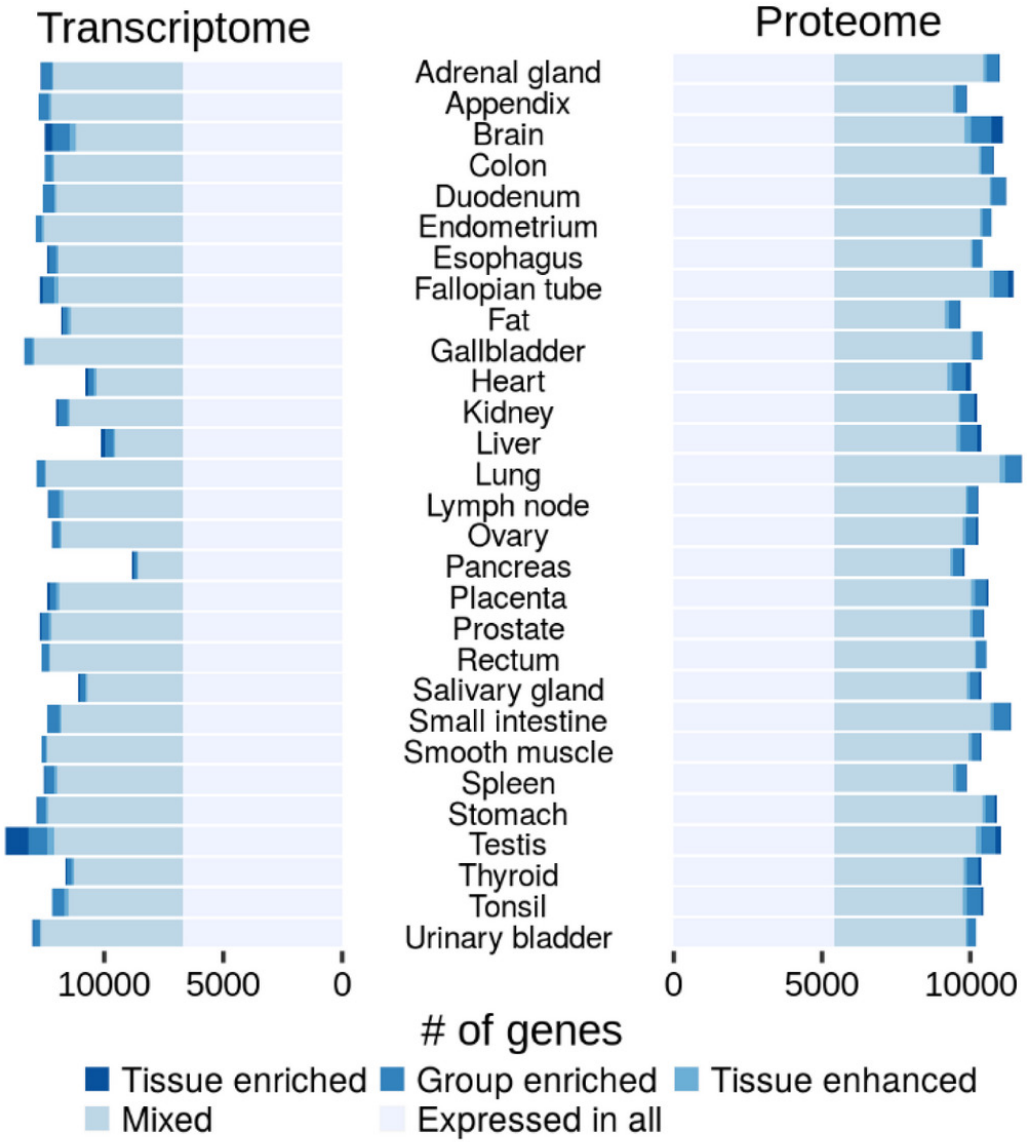


FINDINGS:

- hundreds of proteins, particularly in testis, could not be detected
- even for highly expressed mRNAs, few proteins show tissue specific expression,
- there exist strong differences between mRNA and protein quantities within and across tissues
- Protein expression is often more stable across tissues than that of transcripts.

Wang et al. Mol. Syst. Biol. 15, e8503 (2019)

Transcriptome / Proteome

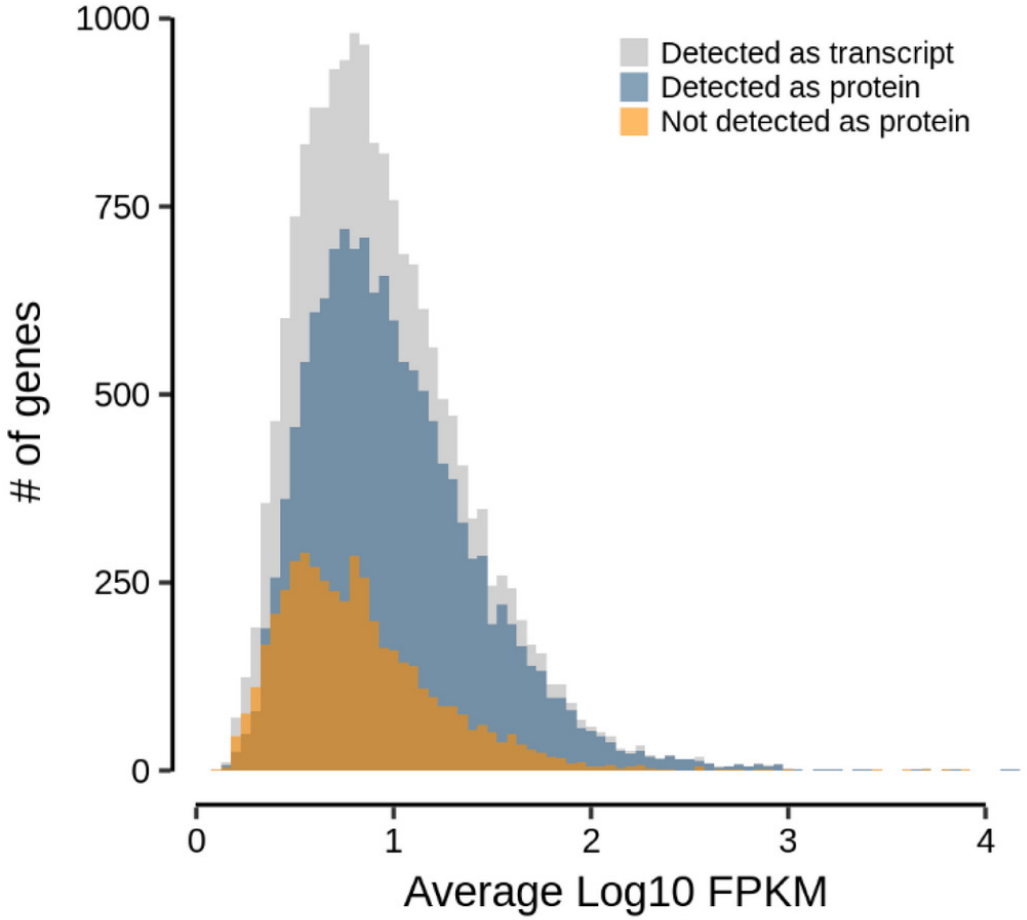


A large fraction of all represented genes was expressed in all tissues: 37% (6,725) at the transcript level and 39% (5,400) at the protein level.

However, 43% of all transcripts and 53% of all proteins showed elevated expression in one or more tissues.

Wang et al. Mol. Syst. Biol. 15, e8503 (2019)

Transcriptome / Proteome



Grey:
Abundance distribution of all transcripts detected in all tissues

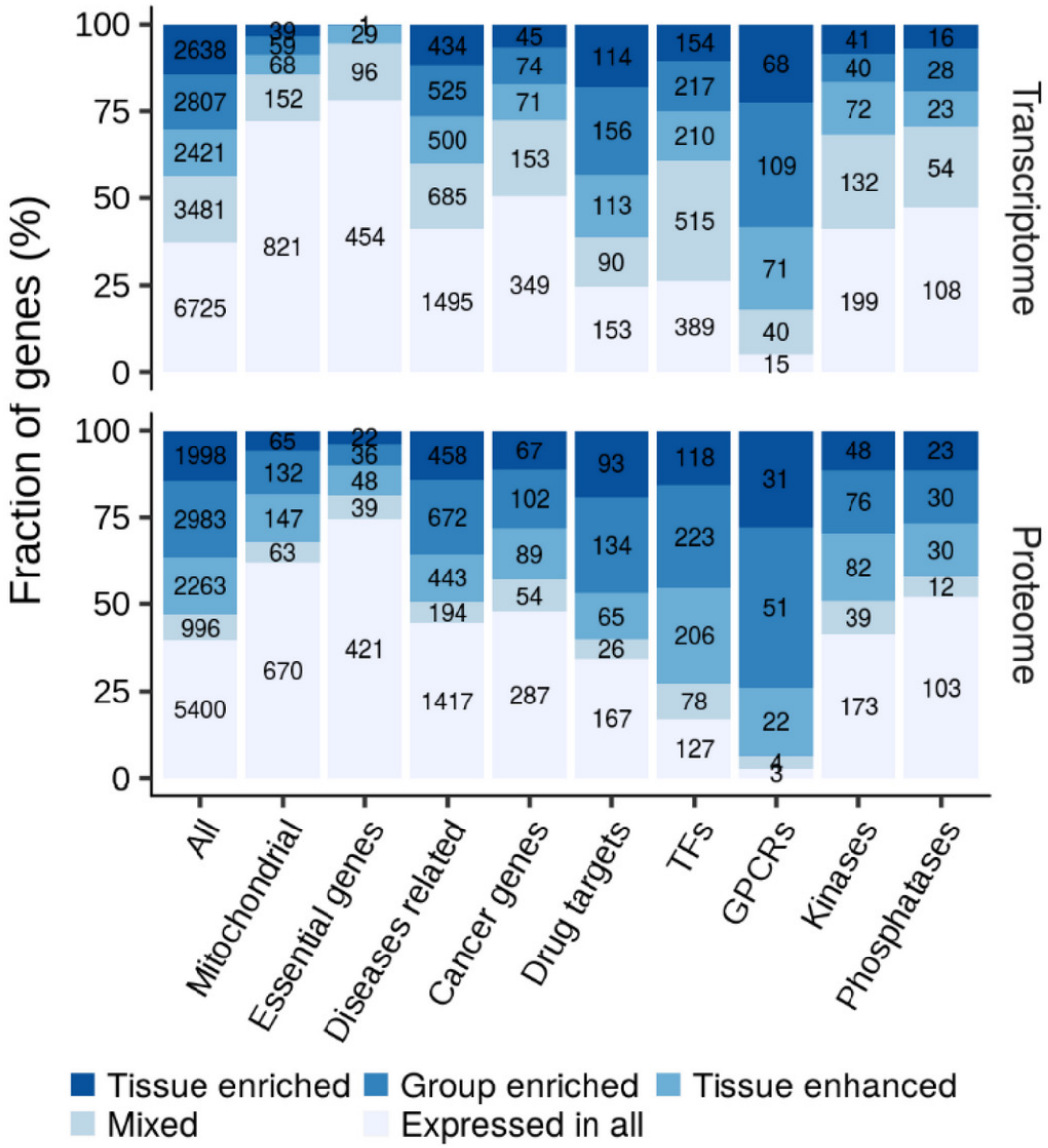
Blue: fraction of detected proteins

Orange: fraction of transcripts for which no protein was detected.
Ca. 1/3 of these transcripts were found in testis.

Interpretation: the mRNA of not detected proteins shows - on average - smaller levels of mRNA expression.
However, even some highly expressed mRNAs were missing as proteins.

Wang et al. Mol. Syst. Biol. 15, e8503 (2019)

Transcriptome / Proteome



The tissue distribution of expression of disease-associated genes followed that of all genes,

However, the expression of drug targets in general and **GPCRs** in particular was much more tissue restricted.

This suggests that proteins may make for better drug targets if they are not ubiquitously expressed.

Wang et al. Mol. Syst. Biol. 15, e8503 (2019)

Summary

Transcription and translation are tightly regulated processes in cells because the cells need
(a) to make sure that the **right mRNAs** and **proteins** are being synthesized which are needed for the **particular cell state** or cell fate, and
(b) to make sure that **no unnecessary molecules** are synthesized which would be costly in terms of resources.

How transcription and translation processes are **regulated** is still subject of intense research.

Recently, the SILAC method and the **ribosome profiling method** (where processing ribosomes are stalled by application of small-molecule inhibitors, and the mRNA sequences the ribosomes bind to get sequenced) have enabled researchers to pinpoint the precise kinetics of expressing individual genes and of translating individual mRNAs.