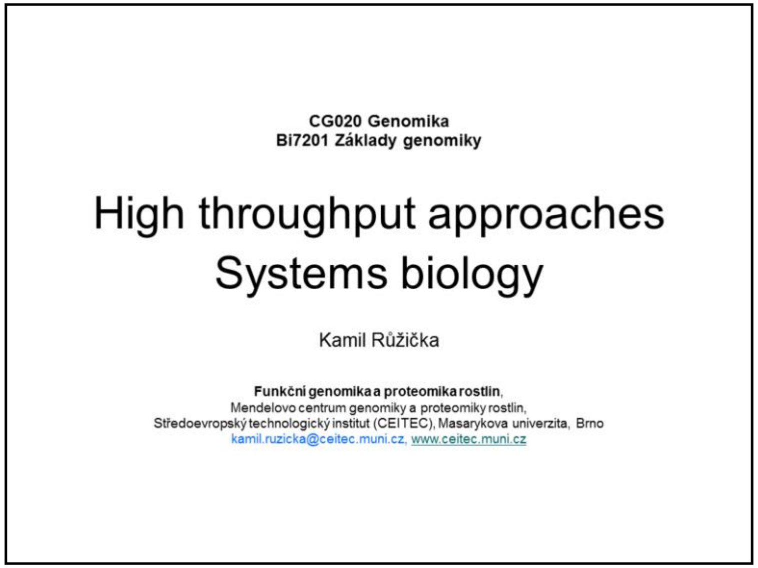


CG020 Genomika
Bi7201 Základy genomiky

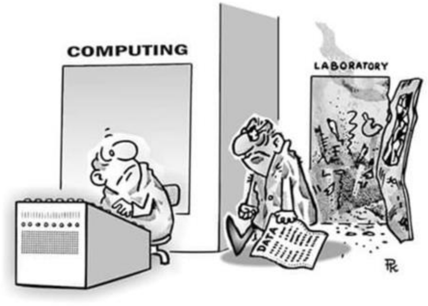
High throughput approaches
Systems biology

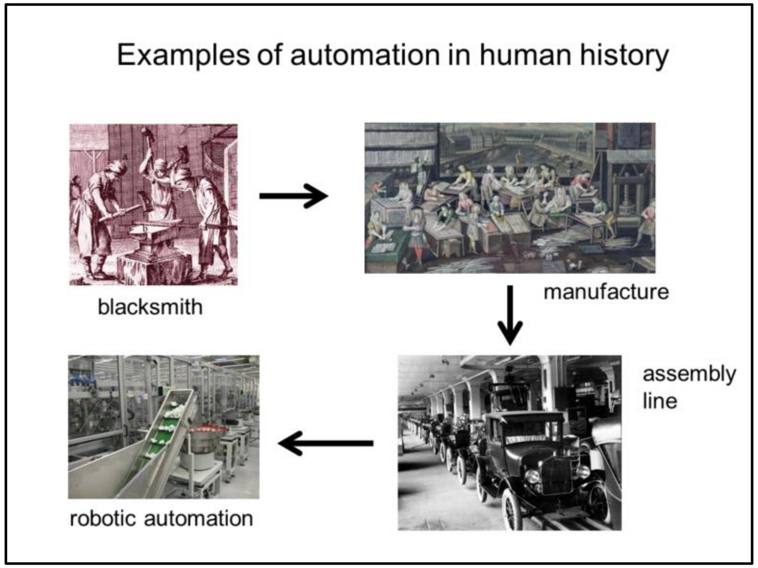
Kamil Růžička

Funkční genomika a proteomika rostlin,
Mendelovo centrum genomiky a proteomiky rostlin,
Středoevropský technologický institut (CEITEC), Masarykova univerzita, Brno
kamil.ruzicka@ceitec.muni.cz, www.ceitec.muni.cz

Přehled

- High throughput biology
  - Automation
  - Omics
  - Transcriptomics and high throughput transcriptomics
  - High throughput interactomics and how to read it
  - High throughput of anything
  - 1000(+1) genomes, GWAS
  - ENCODE
  - Epigenenome and epitranscriptome
- Little about Systems biology
  - Omics
  - Holism and modules
  - Gene regulation in E. coli


COMPUTING
LABORATORY
DATA



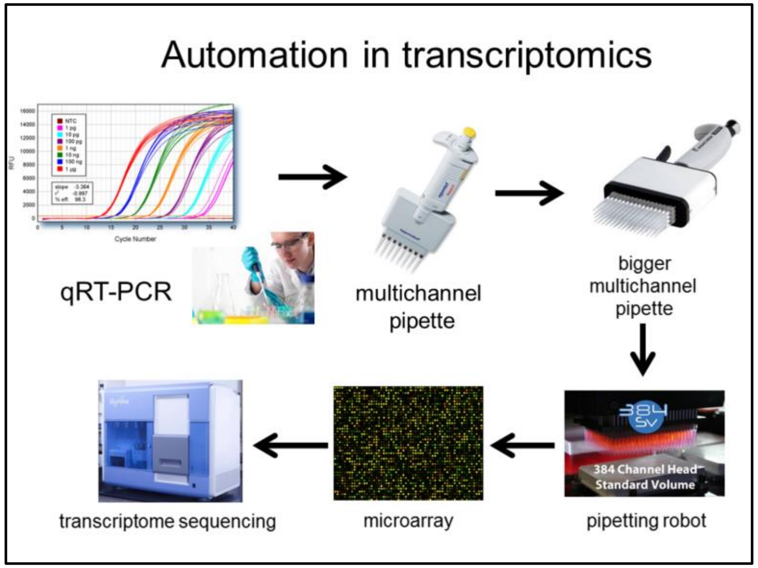
Examples of automation in human history

blacksmith

manufacture

assembly line

robotic automation



Automation in transcriptomics

qRT-PCR

multichannel pipette

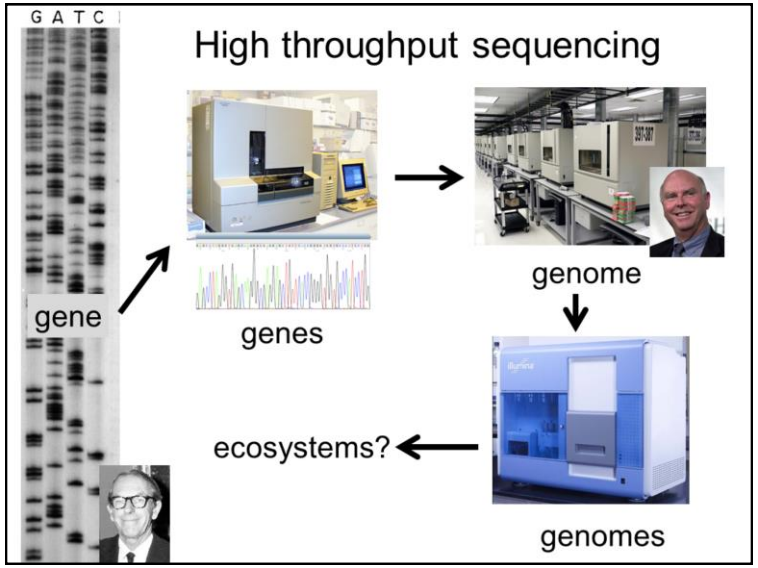
bigger multichannel pipette

pipetting robot

microarray

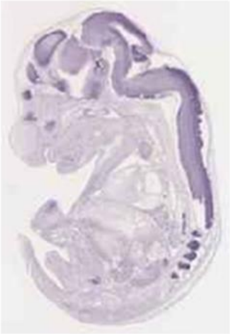
transcriptome sequencing

In short, the construction of single methods shown in other lecture(s) can be automatized and we can find plenty of resources online, premade. The most widely used model organism, the better.

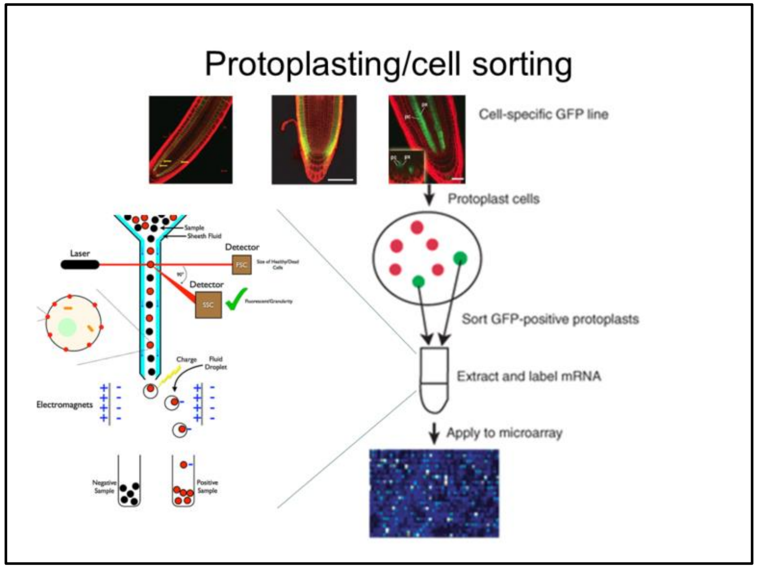


High throughput sequencing

KIAA1841 in mouse expressed in neurons

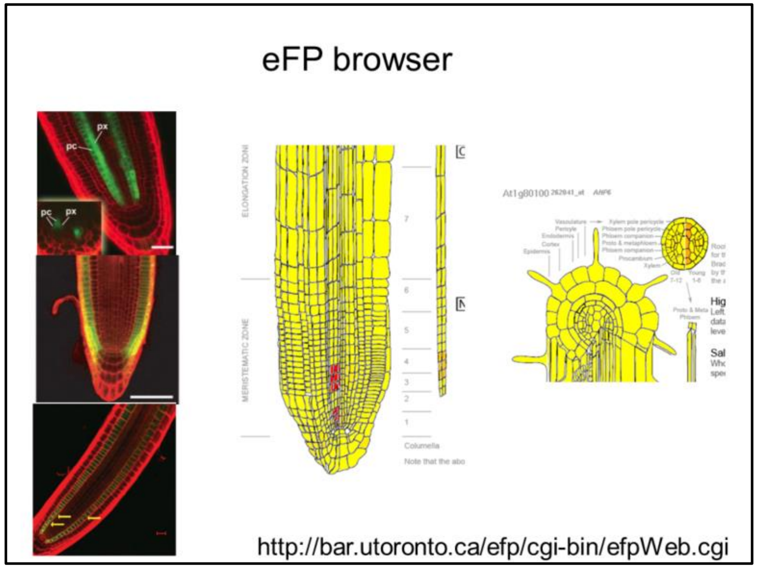


emouseatlas.org



Protoplasting/cell sorting

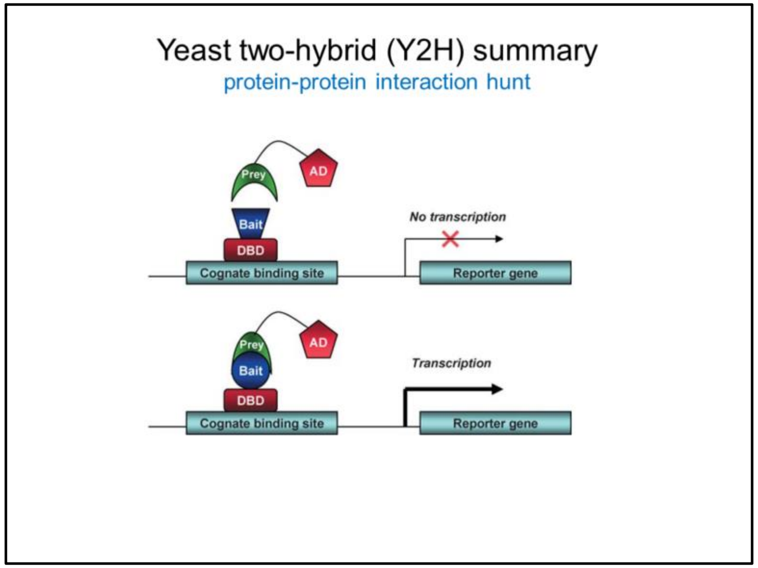
Plants have cell walls which is removed by cell wall degrading enzymes. The cell-wall free cell – protoplast – is released. If that contains fluorescently labeled signal (originating from a tissue specific marker) it can be separated by flow cytometer. From the tube of fluorescent cells collected, we can extract RNA and put to microarray or RNAseq to check the expression levels.



eFP browser

http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi

Some people did such experiments and one can online check the expression of one’s favorite gene (unfortunately, slightly prone of artifacts).

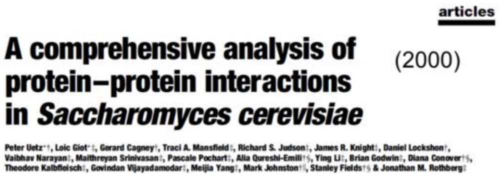


Yeast two-hybrid (Y2H) summary

protein-protein interaction hunt

In simple words: Y2H allows us identifying interacting partner of our favorite gene.

High throughput yeast two hybrid for various organisms

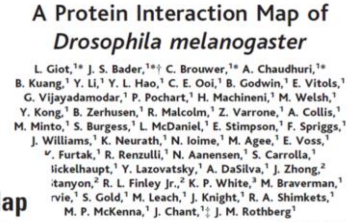


articles

A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*

Peter Uetz*†, Loic Giot*‡, Gerard Cagney†, Traci A. Mansfield‡, Richard S. Judson‡, James R. Knight‡, Daniel Lockshon†, Vaibhav Narayan‡, Maithreyan Srinivasan‡, Pascale Pochart‡, Alia Qureshi-Emili†§, Ying Li‡, Brian Godwin‡, Diana Conover†§, Theodore Kalbfleisch‡, Govindan Vijayadamodar‡, Meijia Yang‡, Mark Johnston†‖, Stanley Fields†§ & Jonathan M. Rothberg‡

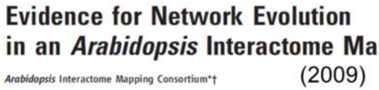
(2000)



A Protein Interaction Map of *Drosophila melanogaster*

L. Giot,[1*] J. S. Bader,[1*†] C. Brouwer,[1*] A. Chaudhuri,[1*]
B. Kuang,[1] Y. Li,[1] Y. L. Hao,[1] C. E. Ooi,[1] B. Godwin,[1] E. Vitols,[1]
G. Vijayadamodar,[1] P. Pochart,[1] H. Machineni,[1] M. Welsh,[1]
Y. Kong,[1] B. Zerhusen,[1] R. Malcolm,[1] Z. Varrone,[1] A. Collis,[1]
M. Minto,[1] S. Burgess,[1] L. McDaniel,[1] E. Stimpson,[1] F. Spriggs,[1]
J. Williams,[1] K. Neurath,[1] N. Ioime,[1] M. Agee,[1] E. Voss,[1]
. Furtak,[1] R. Renzulli,[1] N. Aanensen,[1] S. Carrolla,[1]
ickelhaupt,[1] Y. Lazovatsky,[1] A. DaSilva,[1] J. Zhong,[2]
tanyon,[2] R. L. Finley Jr.,[2] K. P. White,[3] M. Braverman,[1]
rvie,[1] S. Gold,[1] M. Leach,[1] J. Knight,[1] R. A. Shimkets,[1]
M. P. McKenna,[1] J. Chant,[1‡] J. M. Rothberg[1]

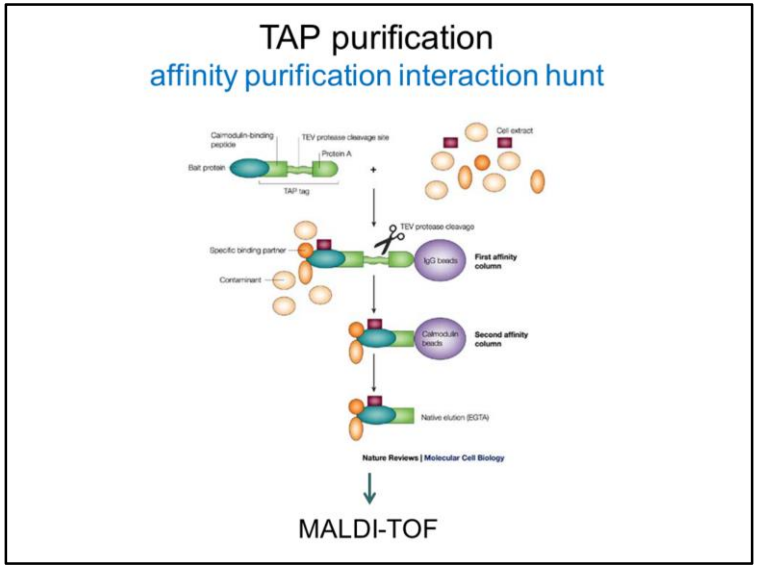
(2005)



Evidence for Network Evolution in an *Arabidopsis* Interactome Map

Arabidopsis Interactome Mapping Consortium*†

(2009)



TAP purification

affinity purification interaction hunt

Nature Reviews | Molecular Cell Biology

↓

MALDI-TOF

Tandem affinity purification is better and becomes more and more popular: identifies whole complexes.

So, far high throughput affinity purification approach slightly less popular

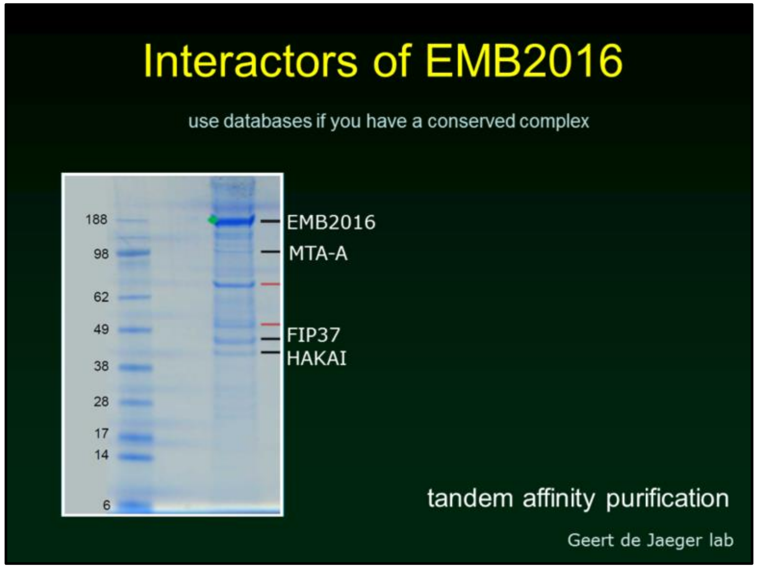
Functional organization of the yeast proteome by systematic analysis of protein complexes (2002)

Anne-Claude Gavin[*], Markus Bösche[*], Roland Krause[*], Paola Grandi[*], Martina Marzioch[*], Andreas Bauer[*], Jörg Schultz[*], Jens M. Rick[*], Anne-Marie Michon[*], Cristina-Maria Cruciat[*], Marita Remor[*], Christian Höfert[*], Malgorzata Schelder[*], Miro Brajenovic[*], Heinz Ruffner[*], Alejandro Merino[*], Karin Klein[*], Manuela Hudak[*], David Dickson[*], Tatjana Rudi[*], Volker Gnau[*], Angela Bauch[*], Sonja Bastuck[*], Bettina Huhse[*], Christina Leutwein[*], Marie-Anne Heurtier[*], Richard R. Copley[†], Angela Edelmann[*], Erich Querfurth[*], Vladimir Rybin[*], Gerard Drewes[*], Manfred Raida[*], Tewis Bouwmeester[*], Peer Bork[†], Bertrand Seraphin[†‡], Bernhard Kuster[*], Gitte Neubauer[*] & Giulio Superti-Furga[*†]

A Protein Complex Network of *Drosophila melanogaster* (2011)

K.G. Guruharsha,[1,4] Jean-François Rual,[1,4] Bo Zhai,[1,4] Julian Mintseris,[1,4] Pujita Vaidya,[1] Namita Vaidya,[1] Chapman Beekman,[1] Christina Wong,[1] David Y. Rhee,[1] Odise Cenaj,[1] Emily McKillip,[1] Saumini Shah,[1] Mark Stapleton,[2] Kenneth H. Wan,[2] Charles Yu,[2] Bayan Parsa,[2] Joseph W. Carlson,[2] Xiao Chen,[2] Bhaveen Kapadia,[2] K. VijayRaghavan,[3] Steven P. Gygi,[1] Susan E. Celniker,[2] Robert A. Obar,[1,*] and Spyros Artavanis-Tsakonas[1,*]

thebiogrid.org - highly relevant for searching for interactors, but look also elsewhere!



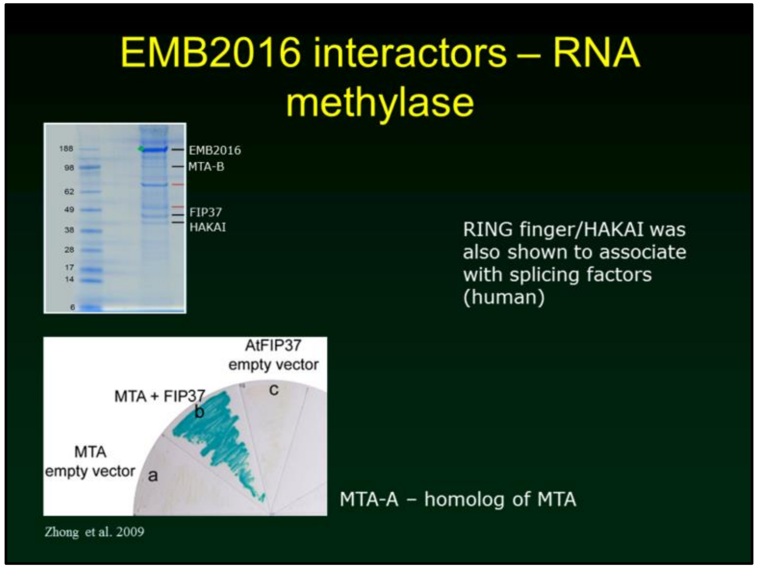
Interactors of EMB2016

use databases if you have a conserved complex

tandem affinity purification

Geert de Jaeger lab

Example how to work with databases.



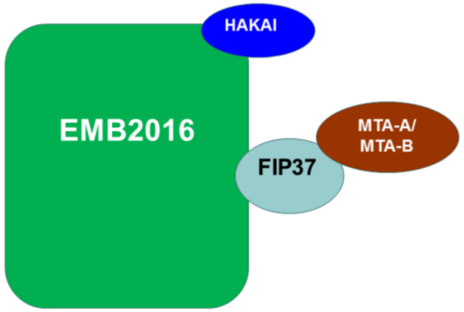
EMB2016 interactors – RNA methylase

RING finger/HAKAI was also shown to associate with splicing factors (human)

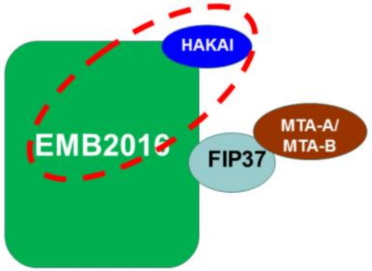
MTA-A – homolog of MTA

Zhong et al. 2009

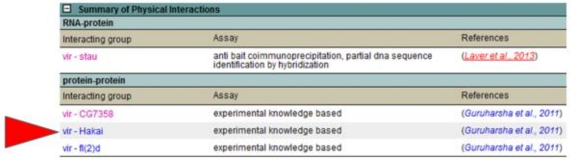
Inferred protein complex



Inferred protein complex



Flybase: EMB2016 interacts with HAKAI (no data on Biogrid)



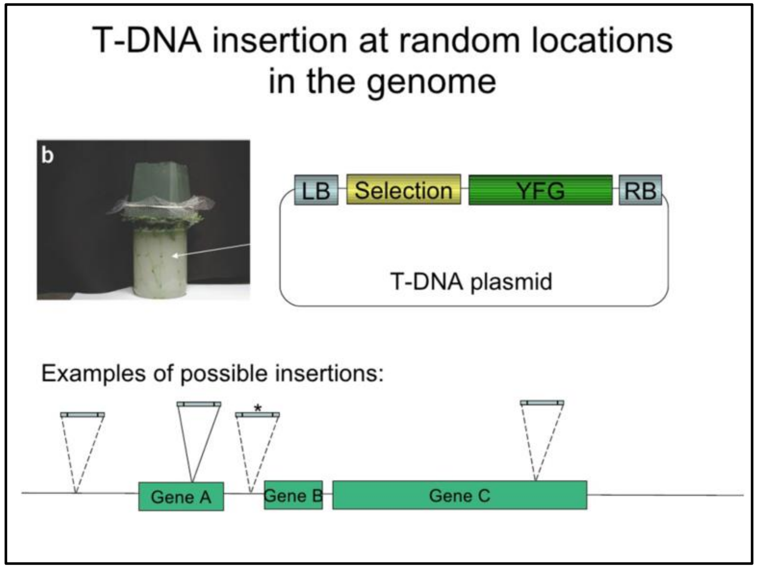
Summary of Physical Interactions

RNA-protein

Interacting group	Assay	References
vir - stau	anti bait coimmunoprecipitation, partial dna sequence identification by hybridization	(*Laver et al., 2013*)

protein-protein

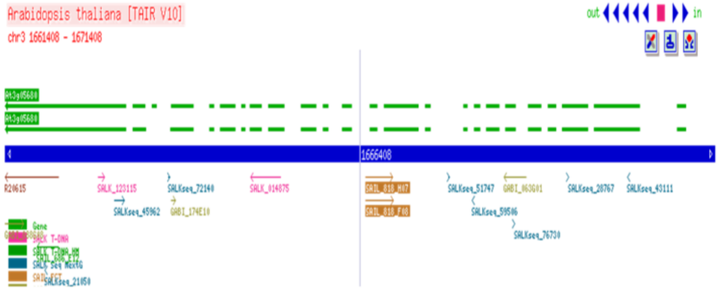
Interacting group	Assay	References
vir - CG7358	experimental knowledge based	(*Guruharsha et al., 2011*)
vir - Hakai	experimental knowledge based	(*Guruharsha et al., 2011*)
vir - fl(2)d	experimental knowledge based	(*Guruharsha et al., 2011*)



T-DNA insertion at random locations in the genome

Examples of possible insertions:

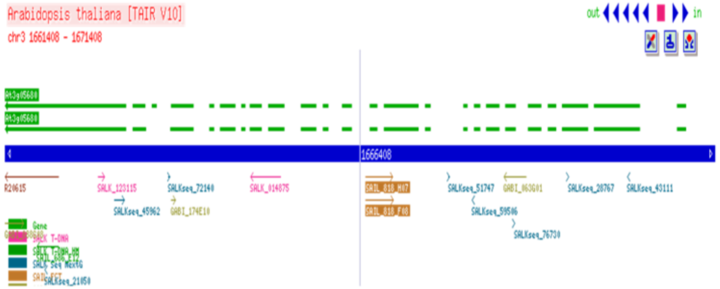
You can order your mutant from the stock center



the same for Drosophila,
mouse, worm etc.

signal.salk.edu

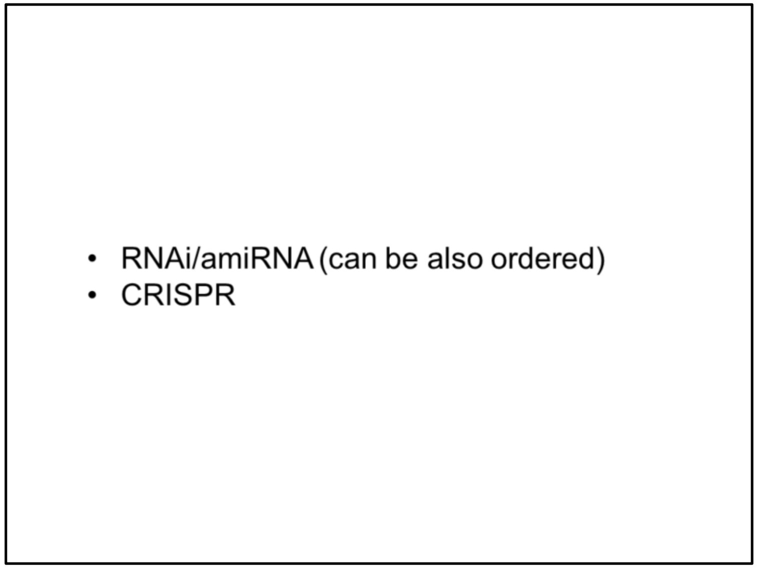
You can order your mutant from the stock center



the same for Drosophila, mouse, worm etc.

What to do if you cannot find insertion line for your gene?

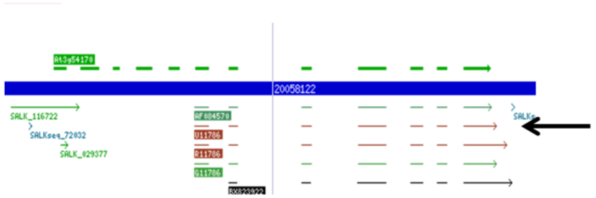
signal.salk.edu



- RNAi/amiRNA (can be also ordered)
- CRISPR

CRISPR – an elegant and rapidly developing technology which allows specific editing of your gene. Everybody lieks CRISPR nowadays.

You can order even various constructs regarding your gene from stock centers



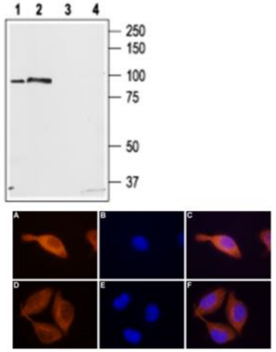
even basic fusions (GFP, myc, TAP etc.) often ready for you

signal.salk.edu

You can order antibodies against your protein

several human proteins providers:
http://www.scbt.com/
www.acris-antibodies.com/
etc.

- even get western and immunocytochemistry in advance

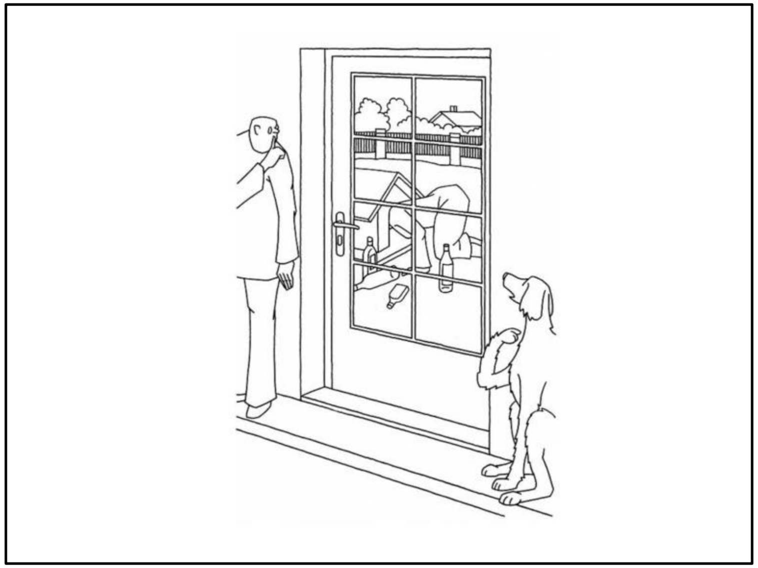


Plants so far lagging behind – agrisera.com perhaps little bit. Rather commercial service.

Light sheet microscopy – high throughput

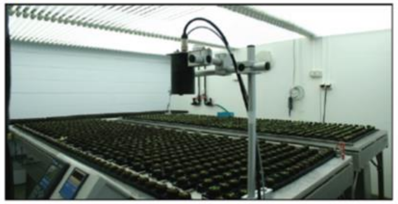
…perhaps video will work

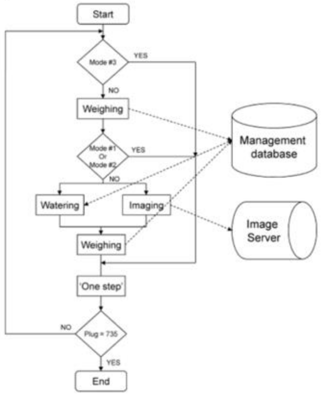
Tomancak lab, MPI Dresden



Phenoscope

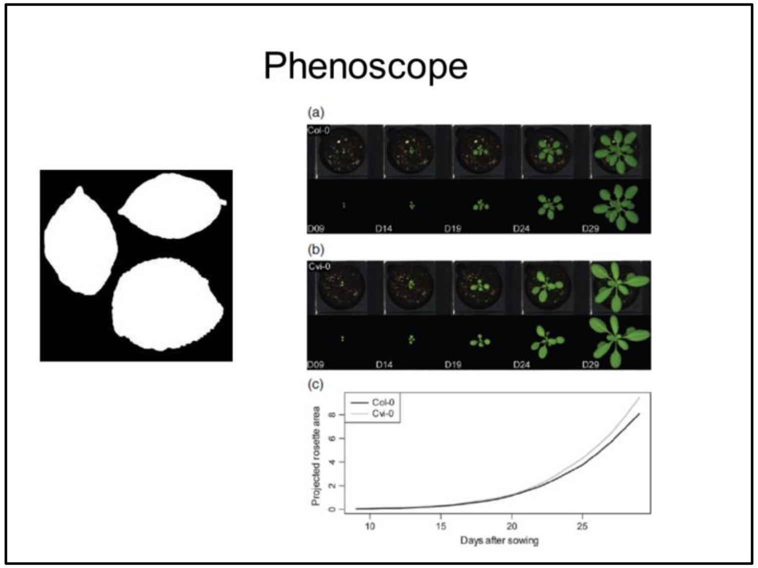
(a)





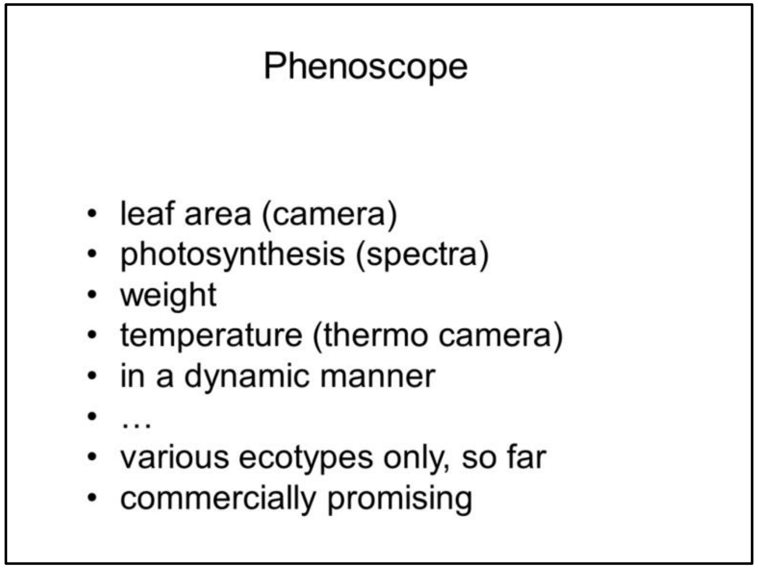
PHENOSCOPE: an automated large-scale phenotyping platform

Thisne et al. 2013



Phenoscope

Automatized detection of leaf shape area and its quatification.



Phenoscope

- leaf area (camera)
- photosynthesis (spectra)
- weight
- temperature (thermo camera)
- in a dynamic manner
- …
- various ecotypes only, so far
- commercially promising

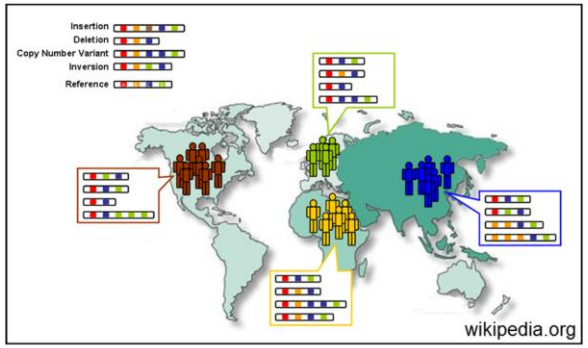
Check your phenotype online

seedgenes.org
– database of plant embryonic mutants (in-dept)

http://rarge.psc.riken.jp/phenome/
- RIKEN Arabidopsis Phenome Information Database (kind of attempt on adult plant)

1000 genomes

1000 human genomes sequenced over the world

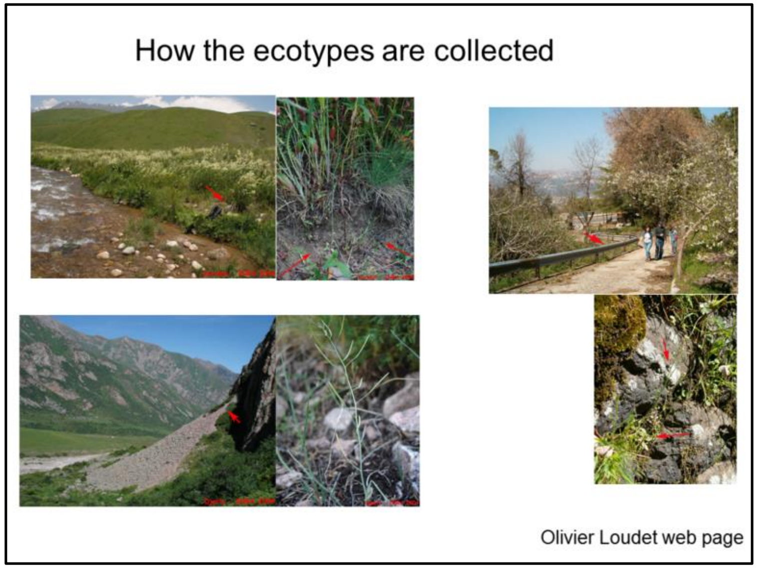


wikipedia.org

1001 genomes - Arabidopsis

http://1001genomes.org/

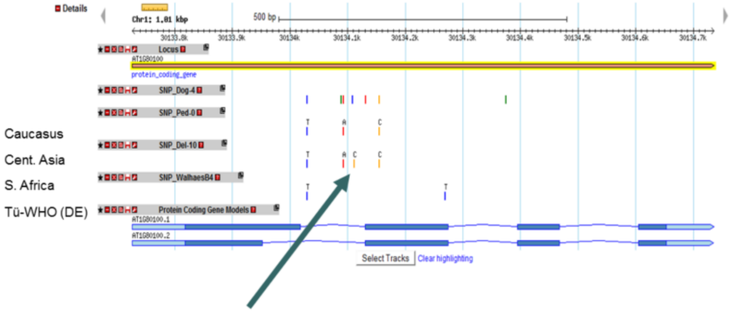
in both cases, much more lines already sequenced



How the ecotypes are collected

Olivier Loudet web page

1001 genomes user interface



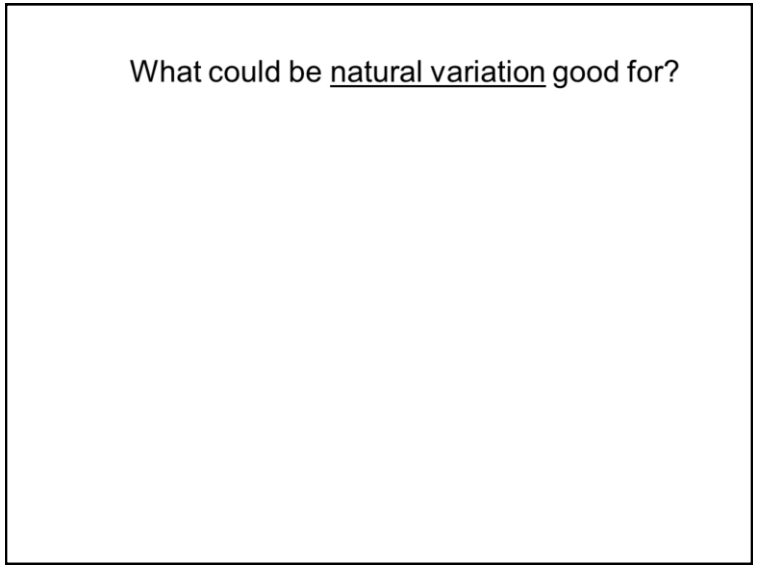
Caucasus

Cent. Asia

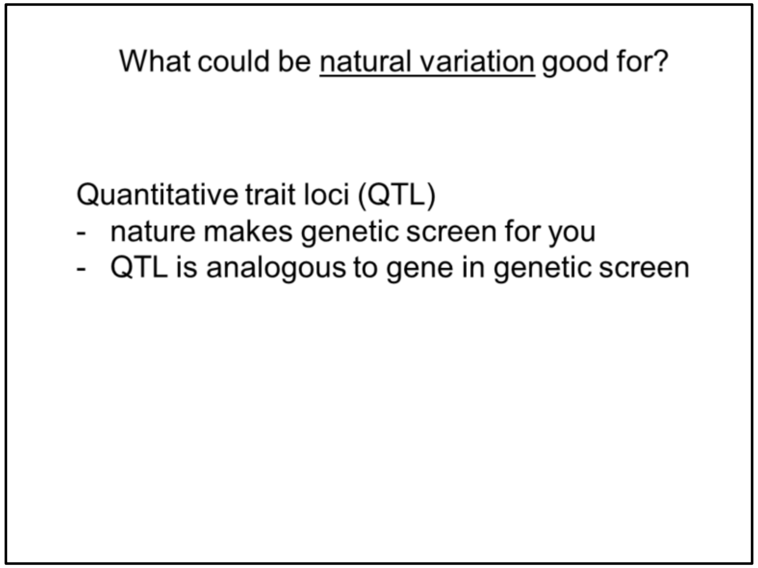
S. Africa

Tü-WHO (DE)

several single nucleotide polymorphisms (**<u>SNP</u>**) in the selected gene

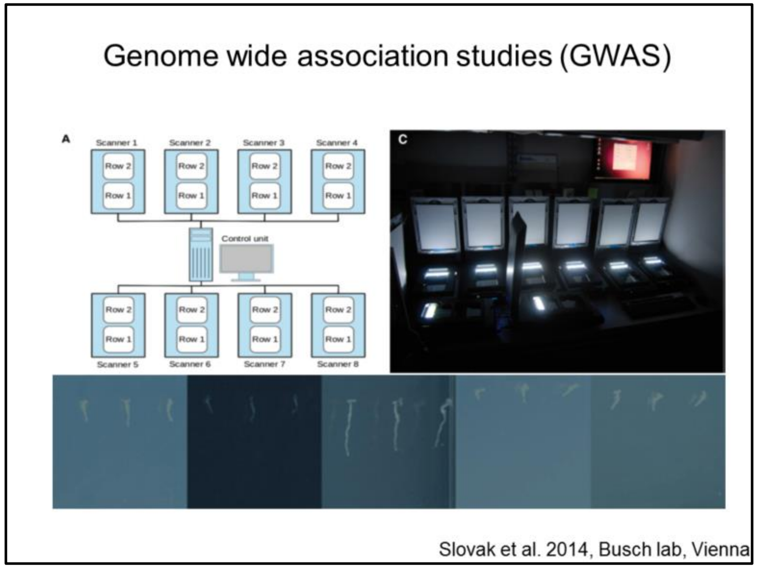


What could be natural variation good for?



What could be natural variation good for?

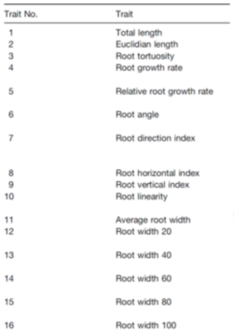
Quantitative trait loci (QTL)
- nature makes genetic screen for you
- QTL is analogous to gene in genetic screen



Genome wide association studies (GWAS)

Slovak et al. 2014, Busch lab, Vienna

Genome wide association studies (GWAS)



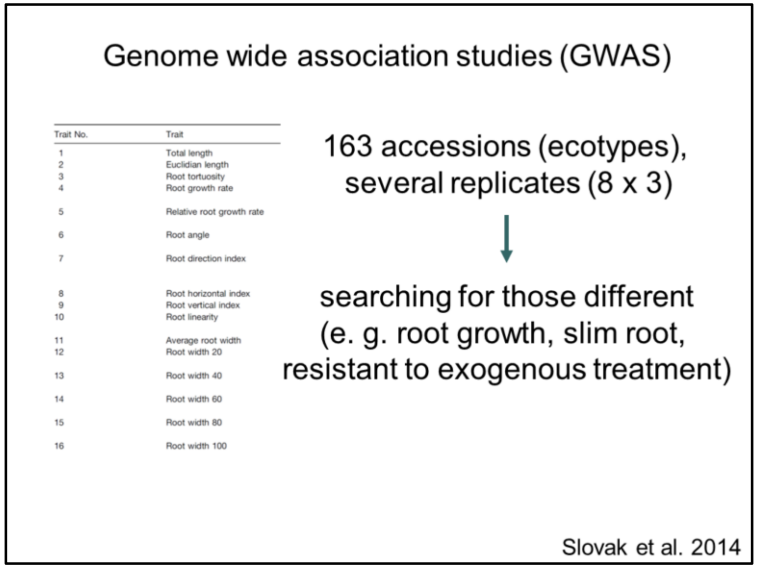
Trait No.	Trait
1	Total length
2	Euclidian length
3	Root tortuosity
4	Root growth rate
5	Relative root growth rate
6	Root angle
7	Root direction index
8	Root horizontal index
9	Root vertical index
10	Root linearity
11	Average root width
12	Root width 20
13	Root width 40
14	Root width 60
15	Root width 80
16	Root width 100

163 accessions (ecotypes),
several replicates (8 x 3)

↓

searching for those different
(say how different they might be!)

Slovak et al. 2014



Genome wide association studies (GWAS)

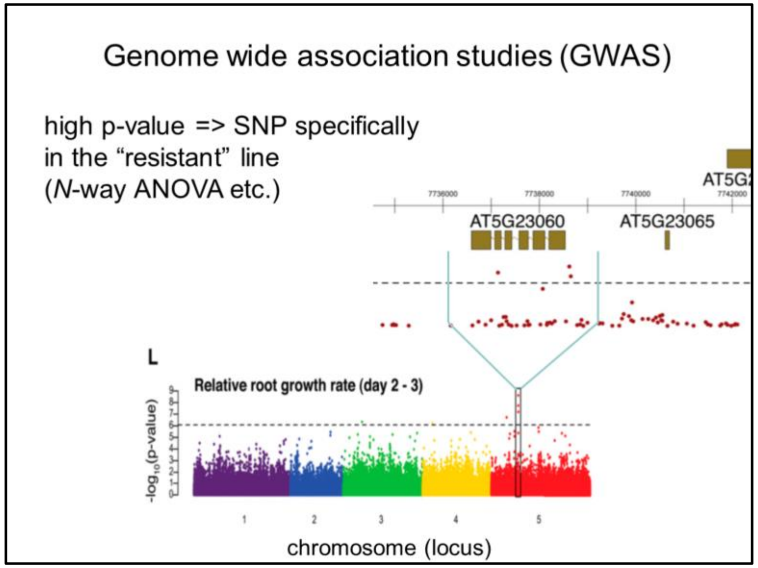
Trait No.	Trait
1	Total length
2	Euclidian length
3	Root tortuosity
4	Root growth rate
5	Relative root growth rate
6	Root angle
7	Root direction index
8	Root horizontal index
9	Root vertical index
10	Root linearity
11	Average root width
12	Root width 20
13	Root width 40
14	Root width 60
15	Root width 80
16	Root width 100

163 accessions (ecotypes),
several replicates (8 x 3)

↓

searching for those different
(e. g. root growth, slim root,
resistant to exogenous treatment)

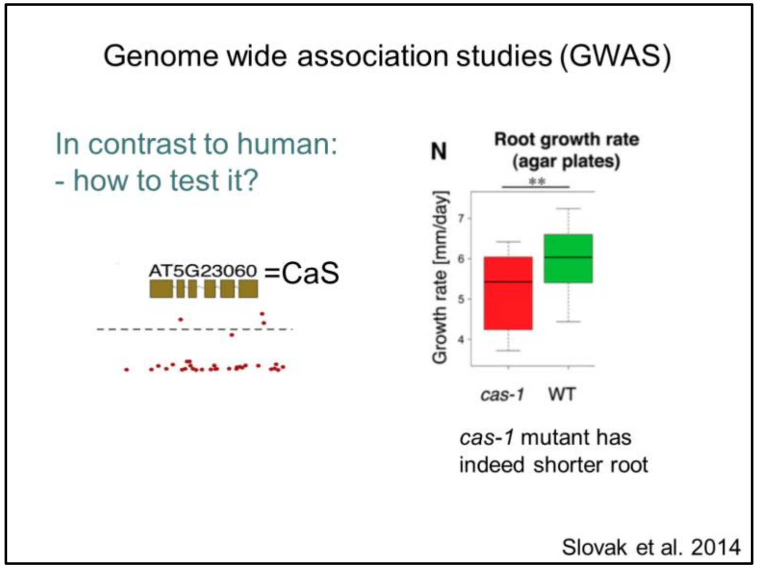
Slovak et al. 2014



Genome wide association studies (GWAS)

high p-value => SNP specifically
in the “resistant” line
(*N*-way ANOVA etc.)

The phenotypic variation was 3 times associated with ecotypes having polymorphisms (“mutations”) in gene AT5G23060 above given threshold, that’s too unlikely to be an accident.



Genome wide association studies (GWAS)

In contrast to human:
- how to test it?

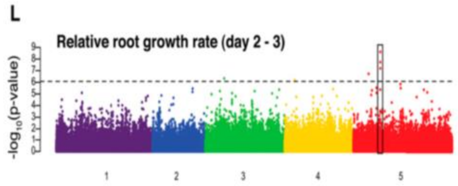
cas-1 mutant has indeed shorter root

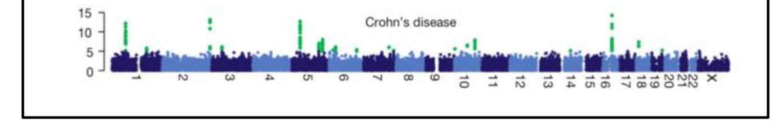
Slovak et al. 2014

You can test it by complementation (transforming with wild type DNA)..

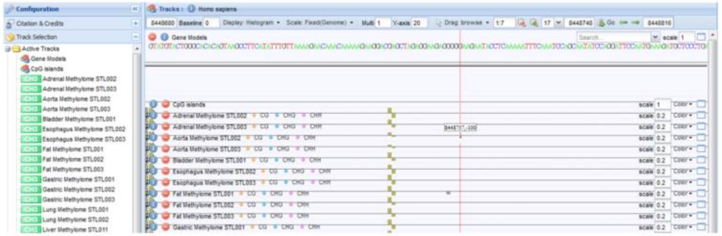
Genome wide association studies (GWAS)

Manhattan plot by human





Status of cytosine methylations in various tissues can be explored in various tissues (human)



8448680 Baseline 0 Display: Histogram Scale: Fixed(Genome) Multi 1 Y-axis 20 Drag: browse 1:7 17 8448748 8448816

scale 1

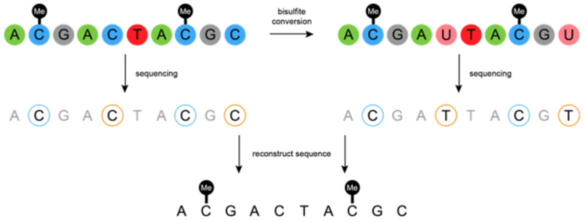
scale 0.2

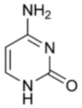
Epigenetic modifications

How to find methylated bases in genome?

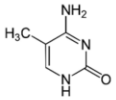
Which bases are methylated?

How to sequence methylation of genome?



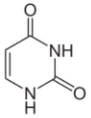


cytosine



5-methylcytosine

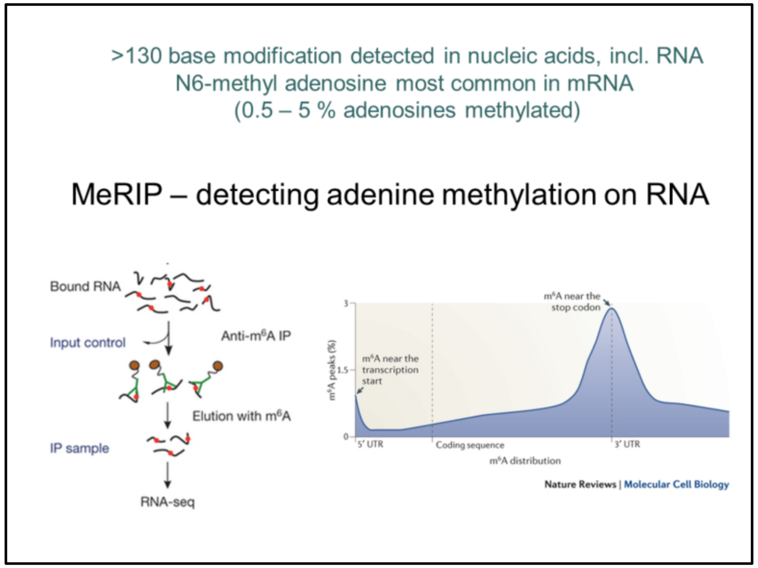
bisulfite sequencing



uracil

What is methylation of cytosine good for?

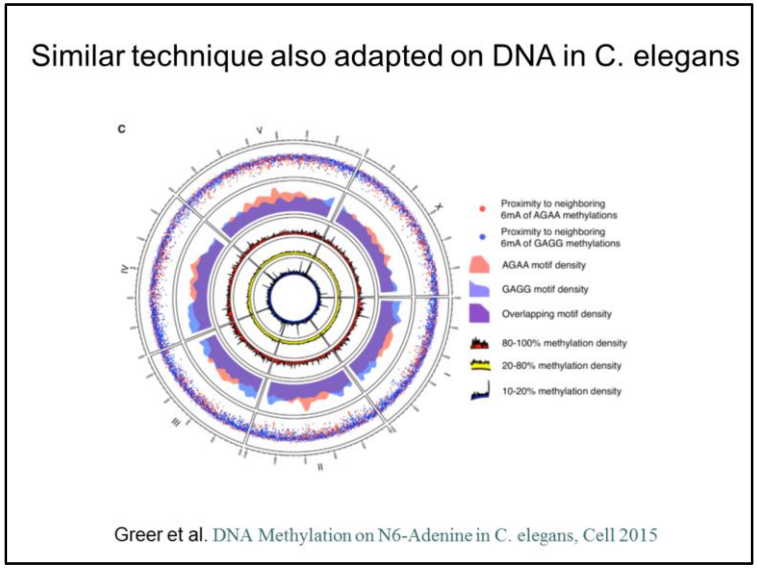
Are there other covalent modifications?



>130 base modification detected in nucleic acids, incl. RNA
N6-methyl adenosine most common in mRNA
(0.5 – 5 % adenosines methylated)

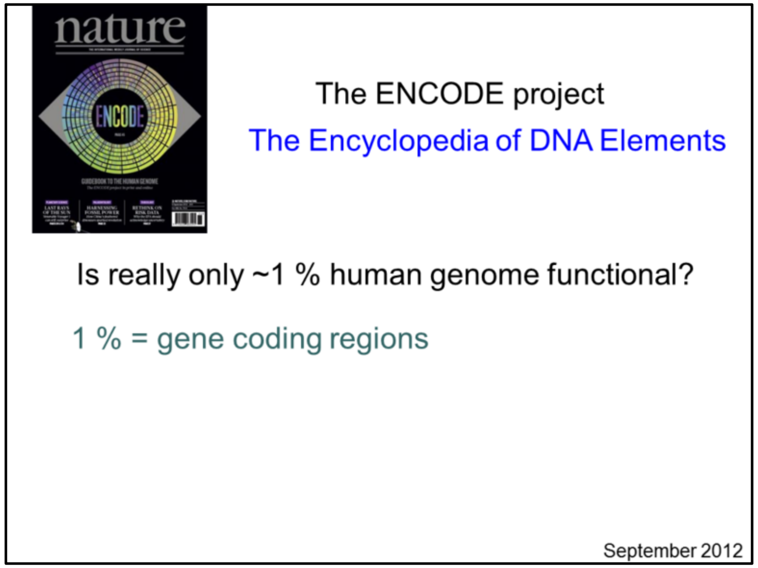
MeRIP – detecting adenine methylation on RNA

Nature Reviews | Molecular Cell Biology



Similar technique also adapted on DNA in C. elegans

Greer et al. DNA Methylation on N6-Adenine in C. elegans, Cell 2015



The ENCODE project

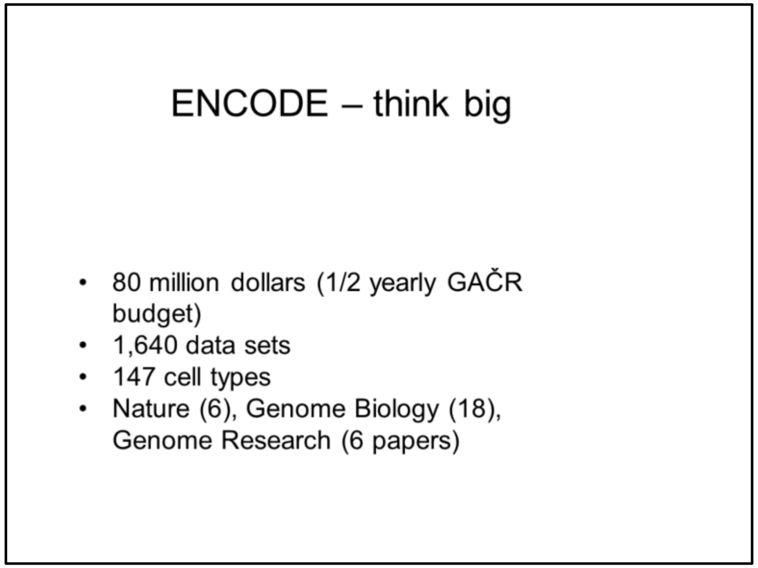
The Encyclopedia of DNA Elements

Is really only ~1 % human genome functional?

1 % = gene coding regions

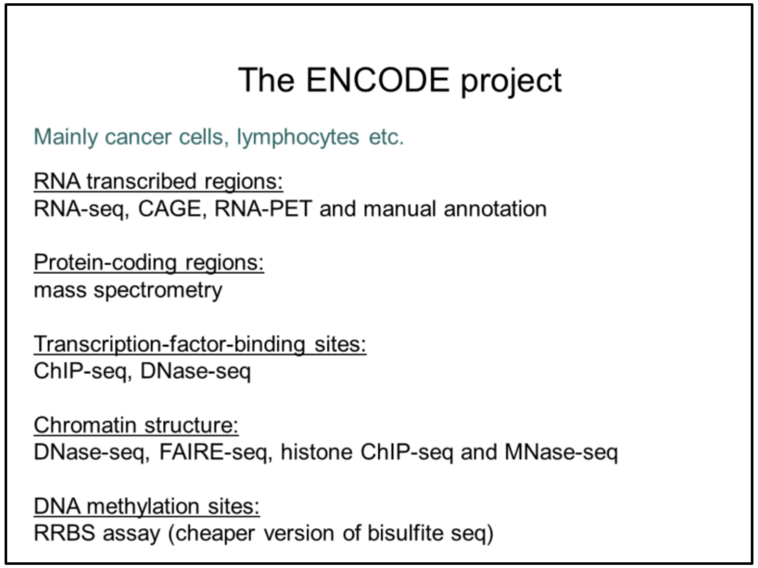
September 2012

About an ambitious high throughput initiative in human.



ENCODE – think big

- 80 million dollars (1/2 yearly GAČR budget)
- 1,640 data sets
- 147 cell types
- Nature (6), Genome Biology (18), Genome Research (6 papers)



The ENCODE project

Mainly cancer cells, lymphocytes etc.

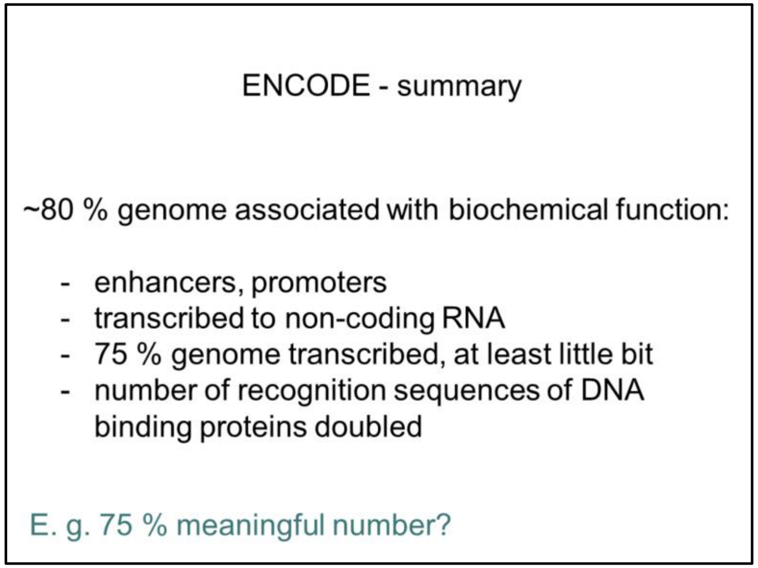
RNA transcribed regions:
RNA-seq, CAGE, RNA-PET and manual annotation

Protein-coding regions:
mass spectrometry

Transcription-factor-binding sites:
ChIP-seq, DNase-seq

Chromatin structure:
DNase-seq, FAIRE-seq, histone ChIP-seq and MNase-seq

DNA methylation sites:
RRBS assay (cheaper version of bisulfite seq)



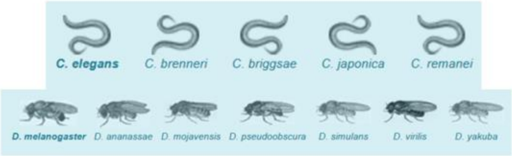
ENCODE - summary

~80 % genome associated with biochemical function:

- enhancers, promoters
- transcribed to non-coding RNA
- 75 % genome transcribed, at least little bit
- number of recognition sequences of DNA binding proteins doubled

E. g. 75 % meaningful number?

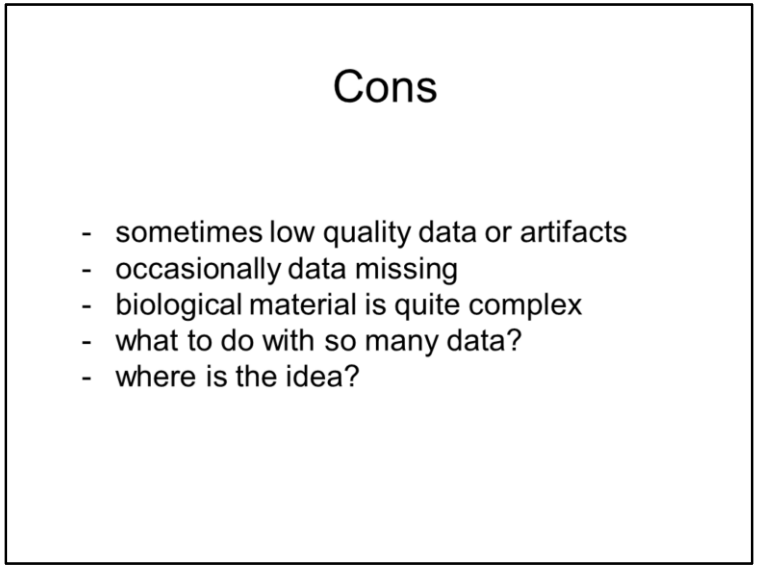
ModENCODE on the way



Drosophila tissue sources:
- Adult eclosion + several days
- Adult female
- Adult male
- Embryos 0-1, 0-2, 0-12, 10-12 hr etc
- Larvae in various instars
- Pupae in various stages
- Mated males or females
- etc.

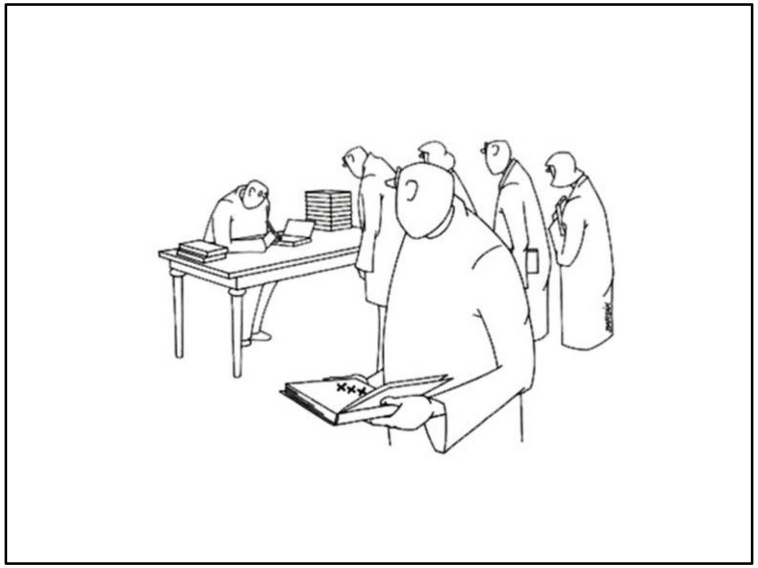
http://www.modencode.org/

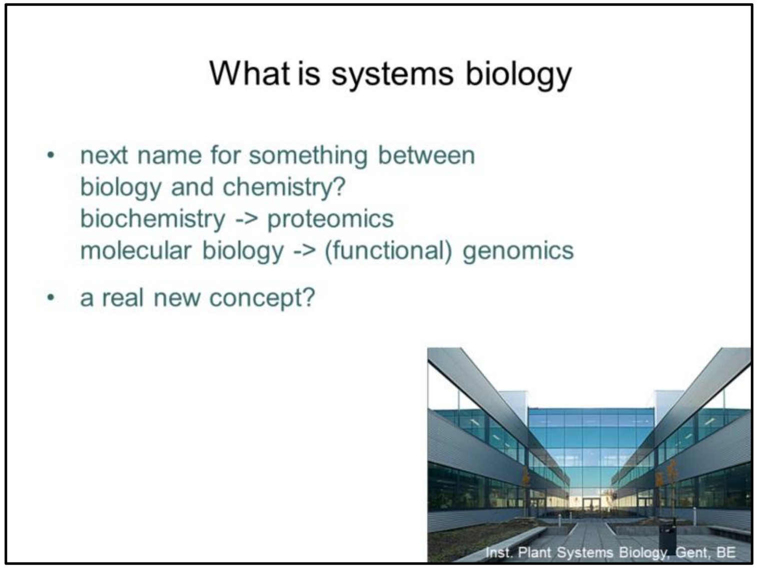
Question: where do you see the limits of high throughput biology?



Cons

- sometimes low quality data or artifacts
- occasionally data missing
- biological material is quite complex
- what to do with so many data?
- where is the idea?



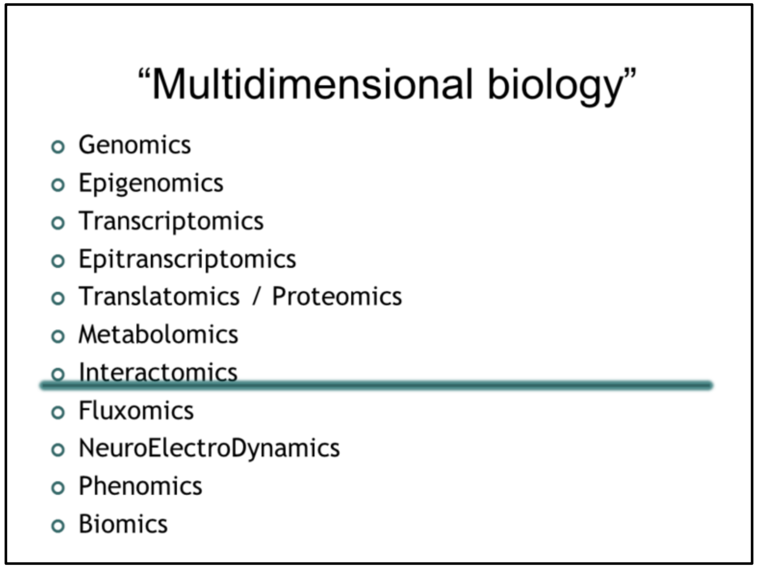


What is systems biology

- next name for something between
  biology and chemistry?
  biochemistry -> proteomics
  molecular biology -> (functional) genomics
- a real new concept?

Inst. Plant Systems Biology, Gent, BE

Few words about systems biology. An effort to use advanced mathematics to model – “understanding everything in organism at once’’.



“Multidimensional biology”

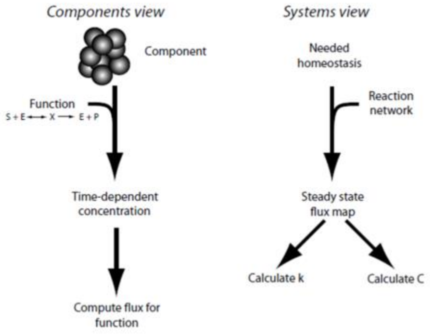
- Genomics
- Epigenomics
- Transcriptomics
- Epitranscriptomics
- Translatomics / Proteomics
- Metabolomics
- Interactomics
- Fluxomics
- NeuroElectroDynamics
- Phenomics
- Biomics

Systems theory

Forget about **reductionism**, think **holistically**.

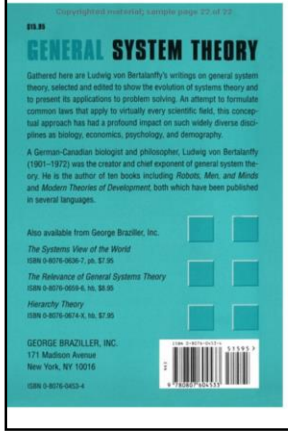
ὅλος [hol'-os] – greek. all, the whole, entire, complete

Reductionism vs. holism



Ludwig von Bertalanffy

(1901-1972)



$15.95

GENERAL SYSTEM THEORY

Gathered here are Ludwig von Bertalanffy's writings on general system theory, selected and edited to show the evolution of systems theory and to present its applications to problem solving. An attempt to formulate common laws that apply to virtually every scientific field, this conceptual approach has had a profound impact on such widely diverse disciplines as biology, economics, psychology, and demography.

A German-Canadian biologist and philosopher, Ludwig von Bertalanffy (1901–1972) was the creator and chief exponent of general system theory. He is the author of ten books including *Robots, Men, and Minds* and *Modern Theories of Development*, both which have been published in several languages.

Also available from George Braziller, Inc.

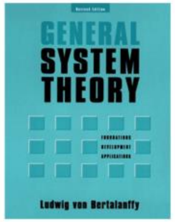
The Systems View of the World
ISBN 0-8076-0636-7, pb, $7.95

The Relevance of General Systems Theory
ISBN 0-8076-0659-6, hb, $8.95

Hierarchy Theory
ISBN 0-8076-0674-X, hb, $7.95

GEORGE BRAZILLER, INC.
171 Madison Avenue
New York, NY 10016

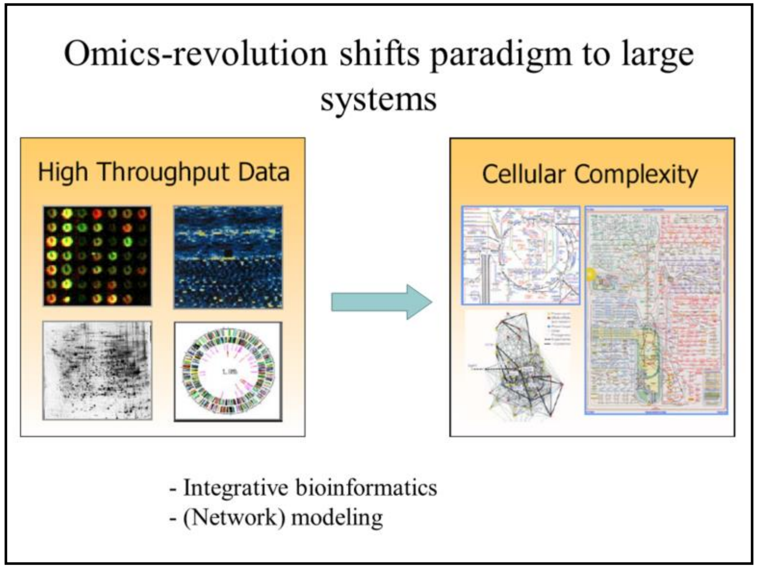
ISBN 0-8076-0453-4



GENERAL SYSTEM THEORY

Ludwig von Bertalanffy

xiv



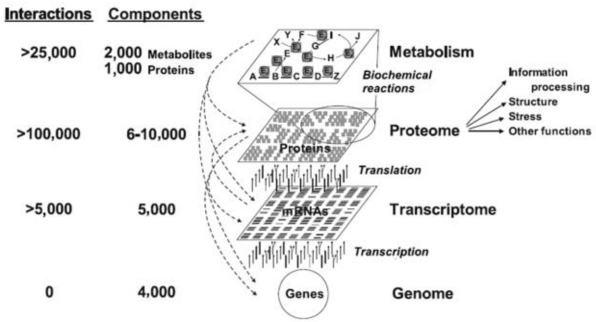
Omics-revolution shifts paradigm to large systems

High Throughput Data

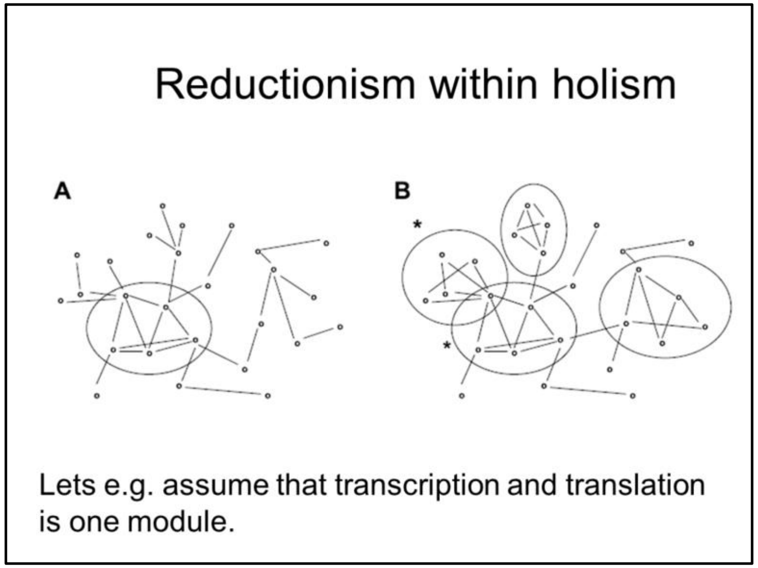
Cellular Complexity

- Integrative bioinformatics
- (Network) modeling

E. coli genome and proteome is small



Interactions	Components	
>25,000	2,000 Metabolites 1,000 Proteins	Metabolism
>100,000	6-10,000	Proteome
>5,000	5,000	Transcriptome
0	4,000	Genome



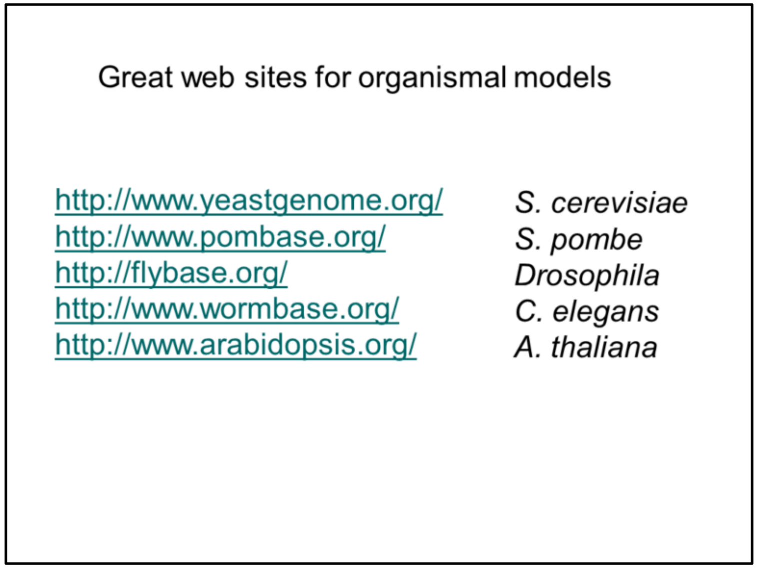
Reductionism within holism

Lets e.g. assume that transcription and translation is one module.

Example of module – simplifying the models.

Conclusions – systems biology

- computing capacities allow handling large data sets
- fashionable
- modelling whole cell processes in silico?
- story frequently missing, there will be always question marks

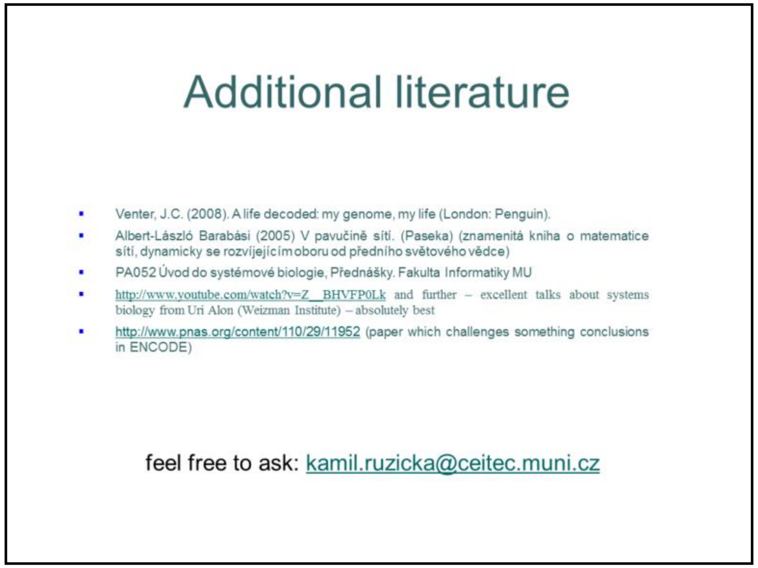


Great web sites for organismal models

http://www.yeastgenome.org/	*S. cerevisiae*
http://www.pombase.org/	*S. pombe*
http://flybase.org/	*Drosophila*
http://www.wormbase.org/	*C. elegans*
http://www.arabidopsis.org/	*A. thaliana*

Also nice web sites

http://encodeproject.org/
http://www.thebiogrid.org/
http://www.genemania.org/
http://string-db.org/
…and many others
…pay attention, if they are kept alive and curated



Additional literature

- Venter, J.C. (2008). A life decoded: my genome, my life (London: Penguin).
- Albert-László Barabási (2005) V pavučině sítí. (Paseka) (znamenitá kniha o matematice sítí, dynamicky se rozvíjejícím oboru od předního světového vědce)
- PA052 Úvod do systémové biologie, Přednášky. Fakulta Informatiky MU
- http://www.youtube.com/watch?v=Z__BHVFP0Lk and further – excellent talks about systems biology from Uri Alon (Weizman Institute) – absolutely best
- http://www.pnas.org/content/110/29/11952 (paper which challenges something conclusions in ENCODE)

feel free to ask: kamil.ruzicka@ceitec.muni.cz