

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

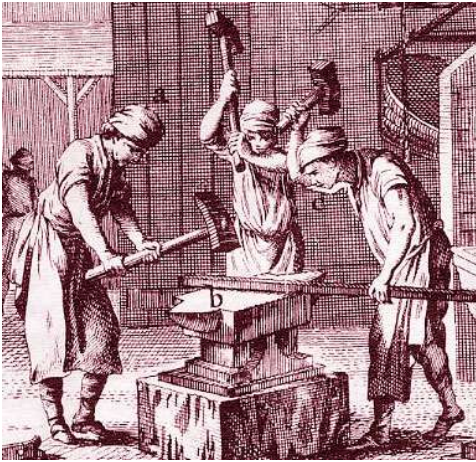
Přehled

- High throughput biology
 - Automation
 - High throughput of anything
 - 1000(+1) genomes, natural variation, GWAS
 - Epigenome and epitranscriptome
 - ENCODE
- Little about Systems biology
 - Omics
 - Holism and modules



Pavel Kantorek - RIP

Examples of automation in human history



blacksmith



manufacture

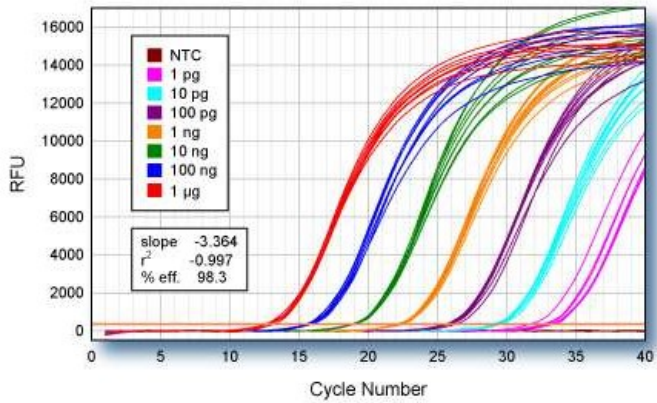


robotic automation



assembly
line

Automation in transcriptomics



qRT-PCR



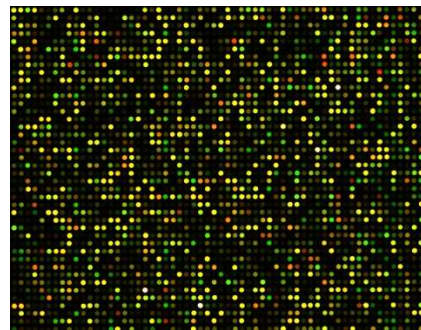
multichannel pipette



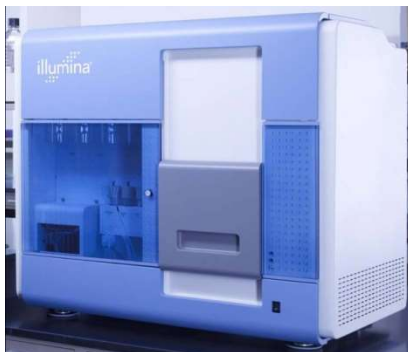
bigger multichannel pipette



pipetting robot



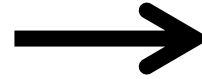
microarray



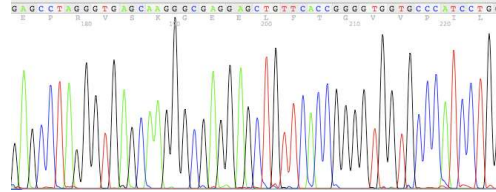
transcriptome sequencing

G A T C

High throughput sequencing

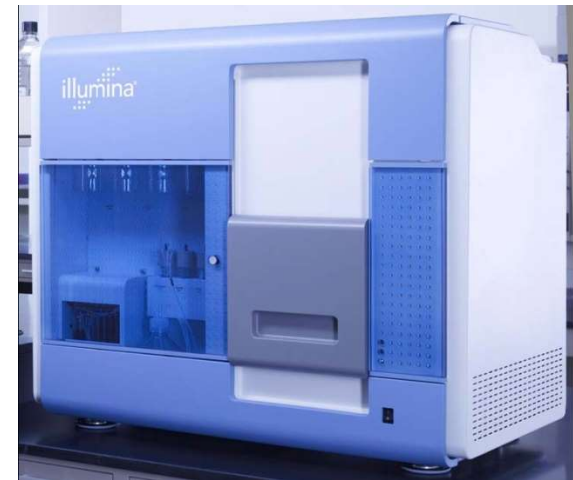


gene



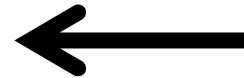
genes

genome



genomes

ecosystems
transcriptomes
epigenomes
etc.



In situ gene expression atlases -
find expression pattern of your gene

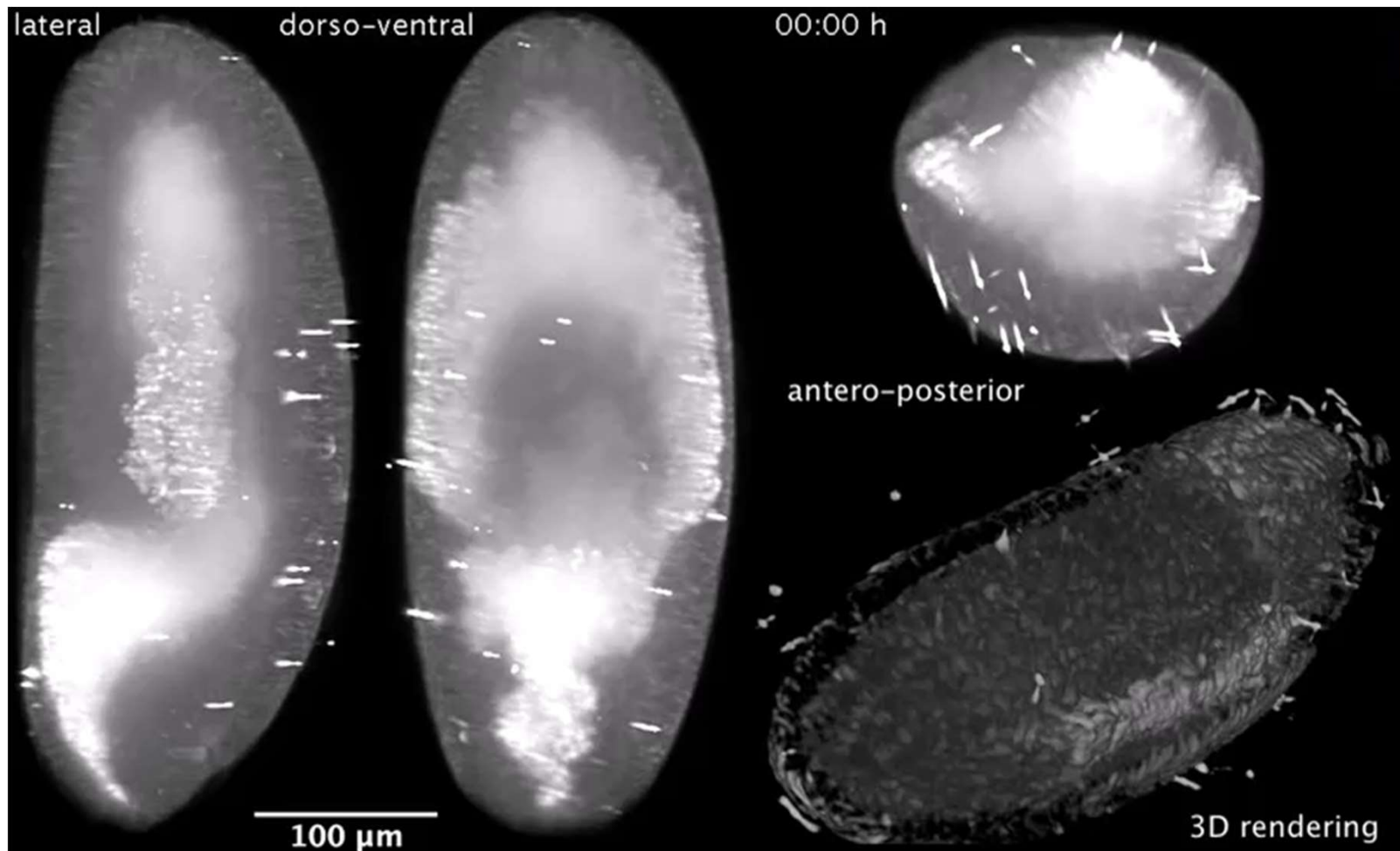


FI(2)D in *Drosophila*

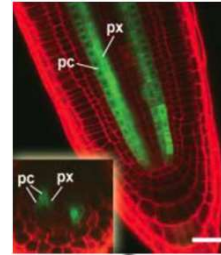
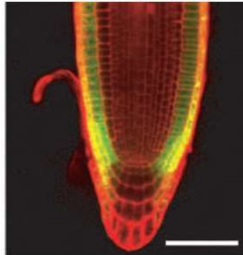
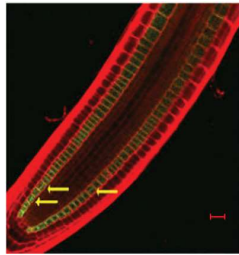


KIAA1841 in mouse
expressed in neurons

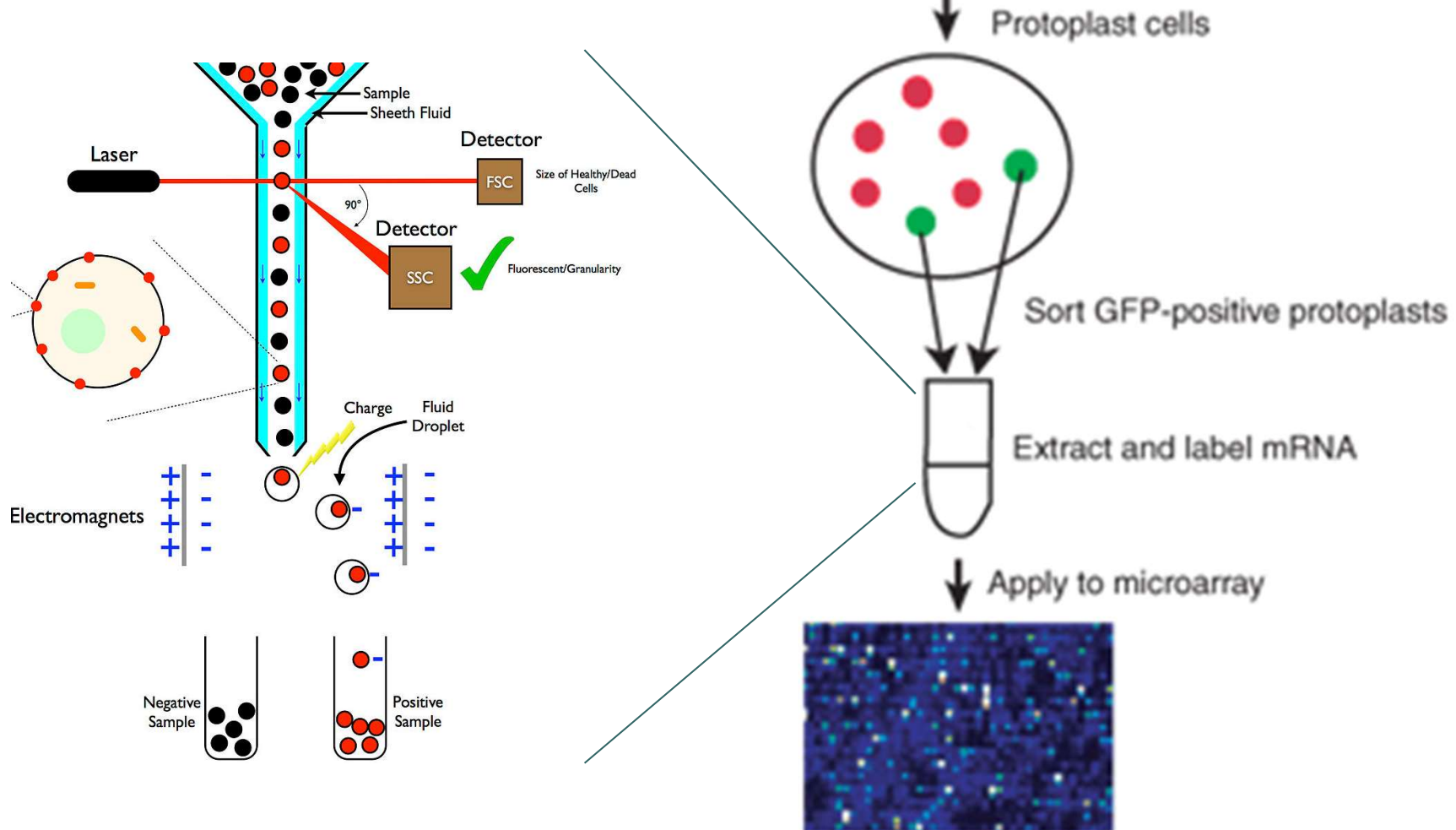
Light sheet microscopy – high throughput



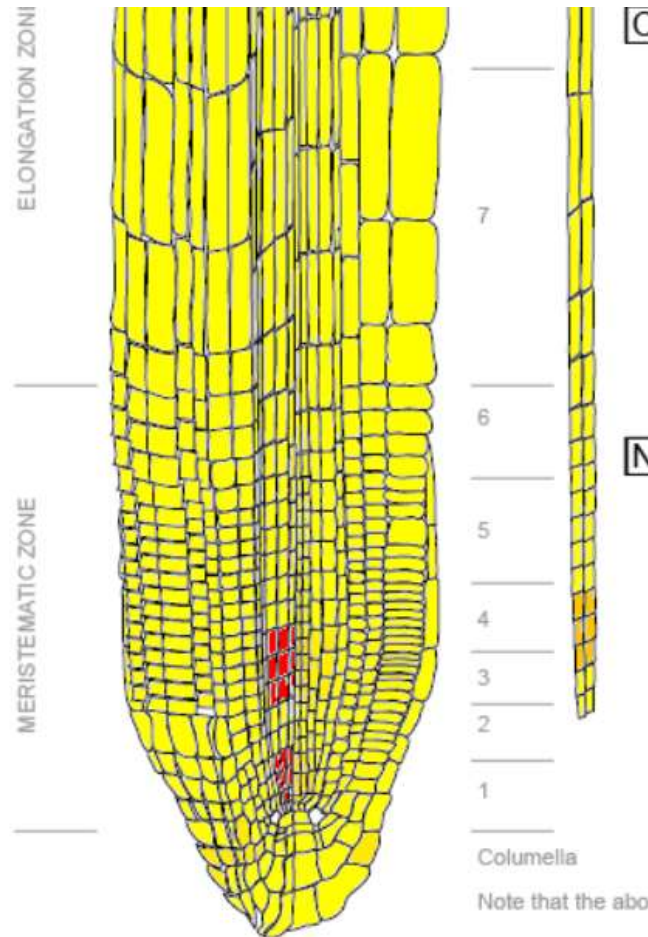
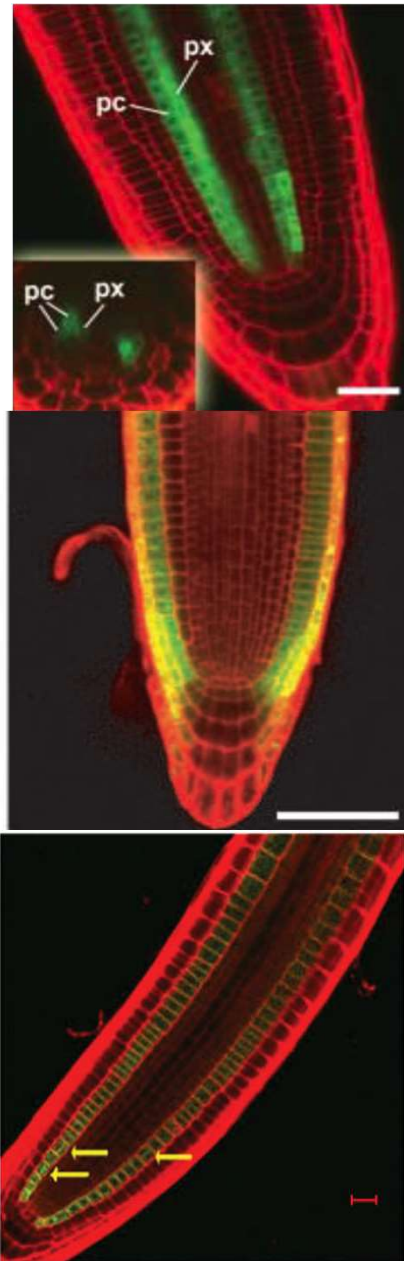
Protoplasting/cell sorting



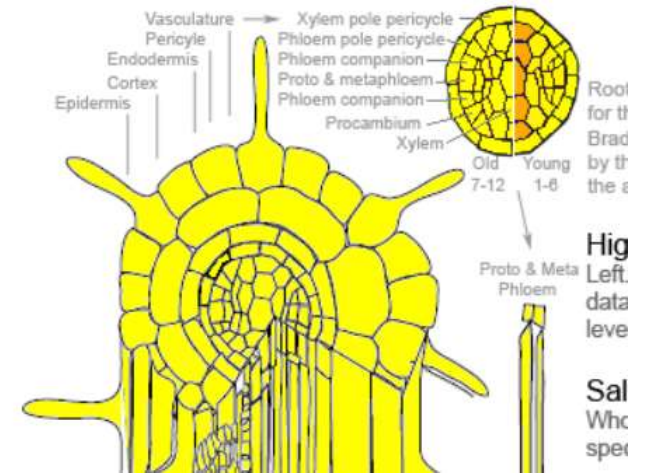
Cell-specific GFP line



eFP browser



At1g80100 262041_at AHP6



Single cell transcriptomics takes over

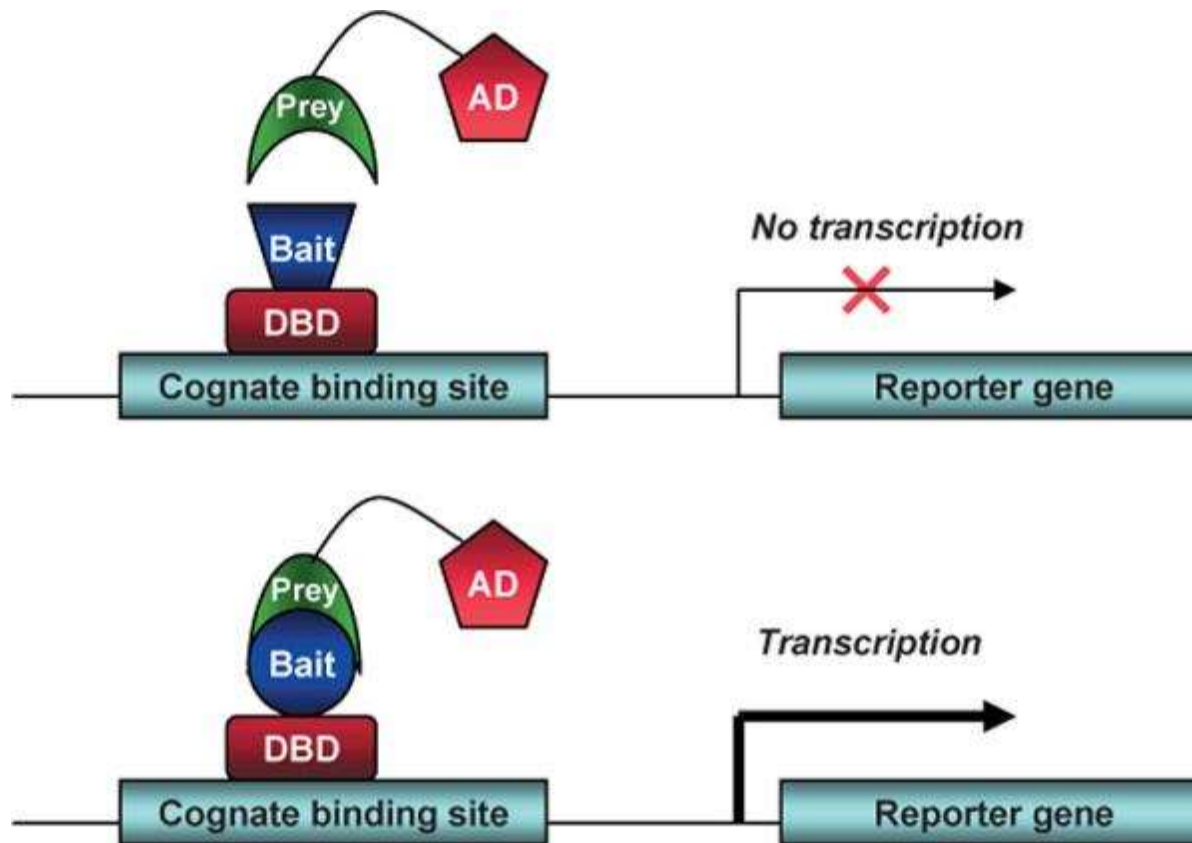
SINGLE-CELL GENOMICS

Comprehensive single-cell transcriptional profiling of a multicellular organism

To resolve cellular heterogeneity, we developed a combinatorial indexing strategy to profile the transcriptomes of single cells or nuclei, termed sci-RNA-seq (single-cell combinatorial indexing RNA sequencing). We applied sci-RNA-seq to profile nearly 50,000 cells from the nematode *Caenorhabditis elegans* at the L2 larval stage, which provided >50-fold “shotgun” cellular coverage of its somatic cell composition. From these data, we defined consensus expression profiles for 27 cell types and recovered rare neuronal cell types corresponding to as few as one or two cells in the L2 worm. We integrated these profiles with whole-animal chromatin immunoprecipitation sequencing data to deconvolve the cell type-specific effects of transcription factors. The data generated by sci-RNA-seq constitute a powerful resource for nematode biology and foreshadow similar atlases for other organisms.

Yeast two-hybrid (Y2H) summary

protein-protein interaction hunt



High throughput yeast two hybrid for various organisms

articles

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

(2000)

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[‡]

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
V. Furtak,¹ R. Renzulli,¹ N. Aanensen,¹ S. Carroll,¹
S. Sickelhaupt,¹ Y. Lazovatsky,¹ A. DaSilva,¹ J. Zhong,²
S. Antonyan,² R. L. Finley Jr.,² K. P. White,³ M. Braverman,¹
S. Arrieve,¹ S. Gold,¹ M. Leach,¹ J. Knight,¹ R. A. Shimkets,¹
M. P. McKenna,¹ J. Chant,^{1†} J. M. Rothberg¹

Evidence for Network Evolution in an *Arabidopsis* Interactome Map

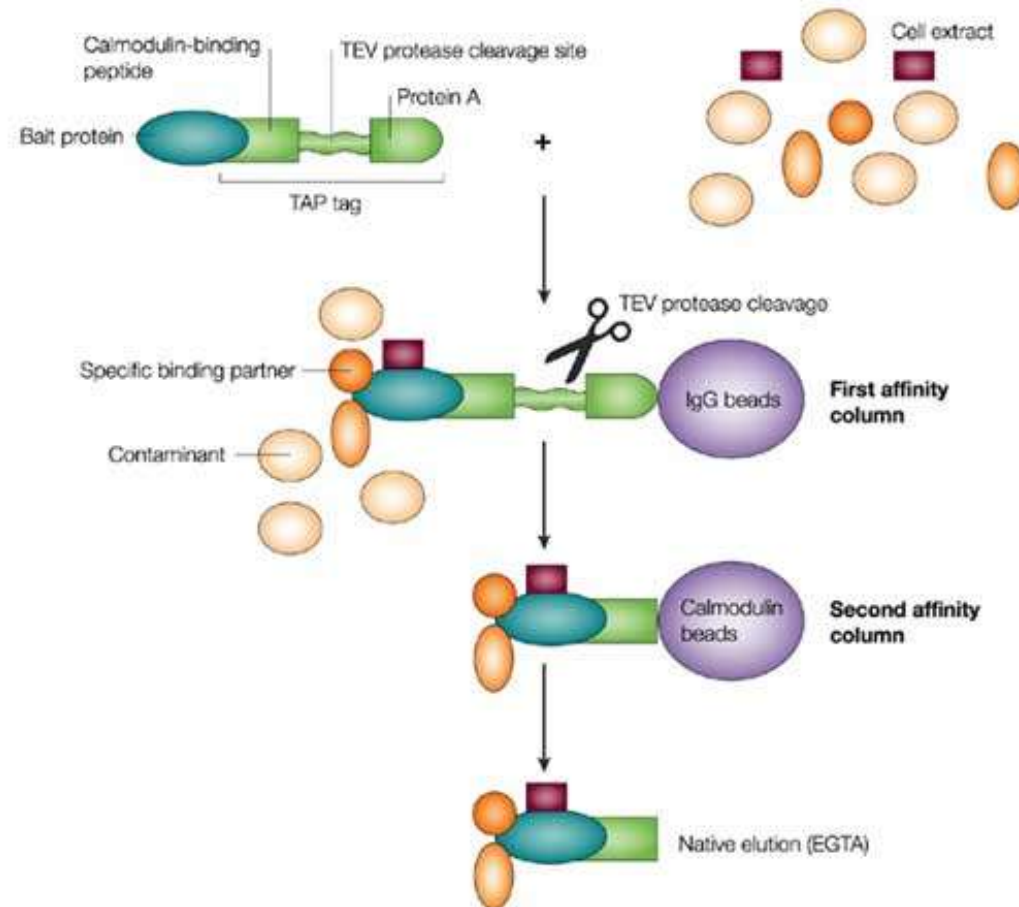
Arabidopsis Interactome Mapping Consortium^{*†}

(2009)

(2005)

TAP purification

affinity purification interaction hunt



Nature Reviews | Molecular Cell Biology



MALDI-TOF

High throughput affinity purification

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*†

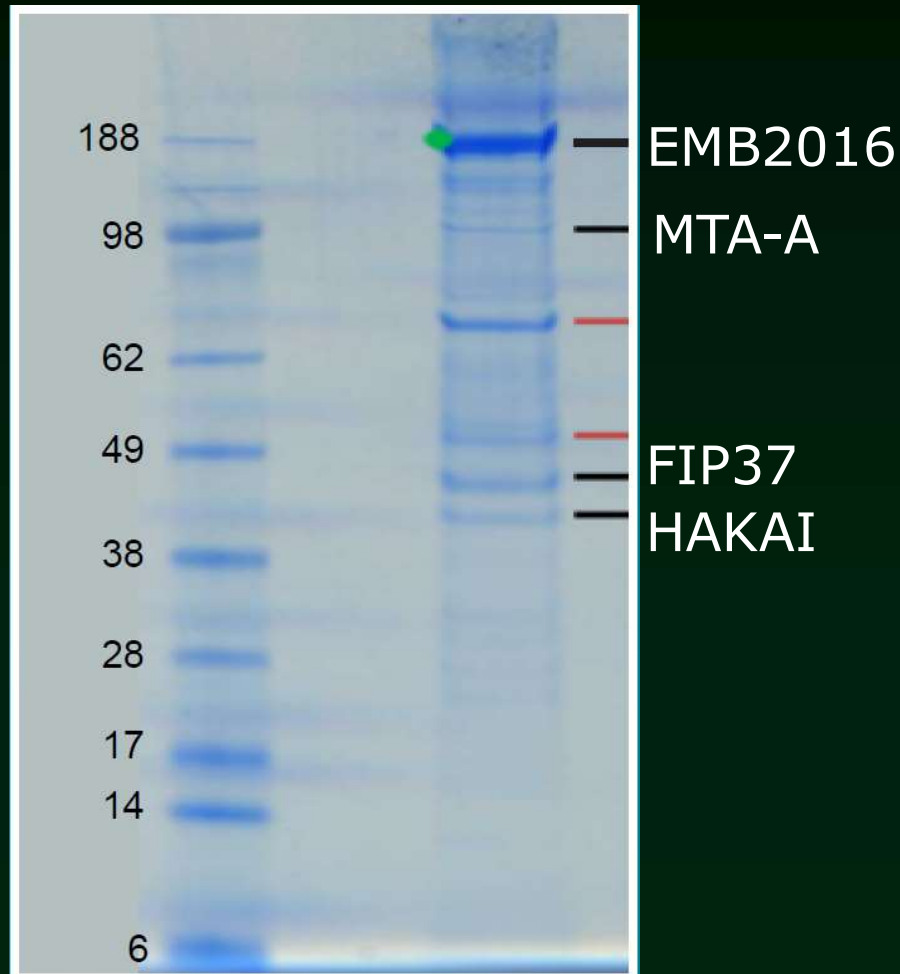
A Protein Complex Network of *Drosophila melanogaster* (2011)

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 – very nice tool

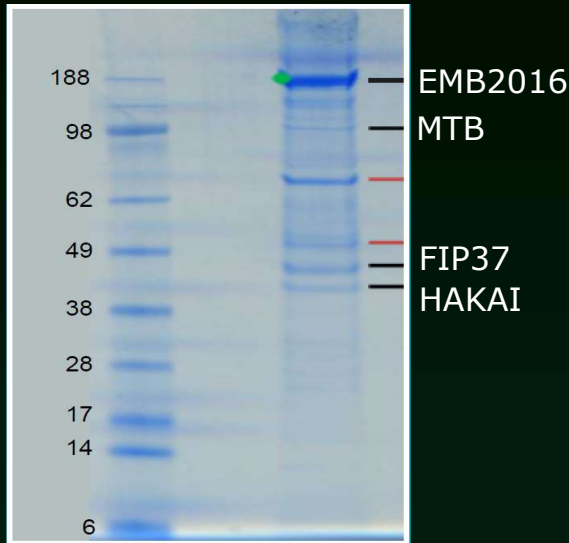
Interactors of EMB2016

use databases if you have a conserved complex

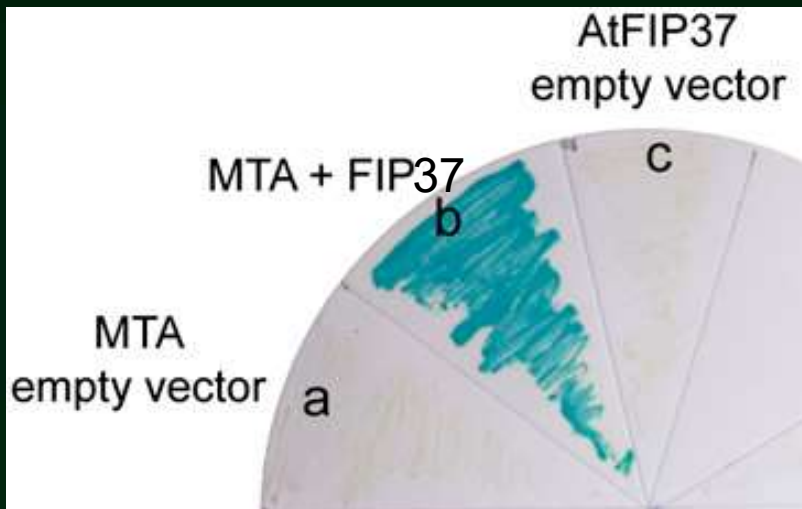


tandem affinity purification

EMB2016 interactors – RNA methylase

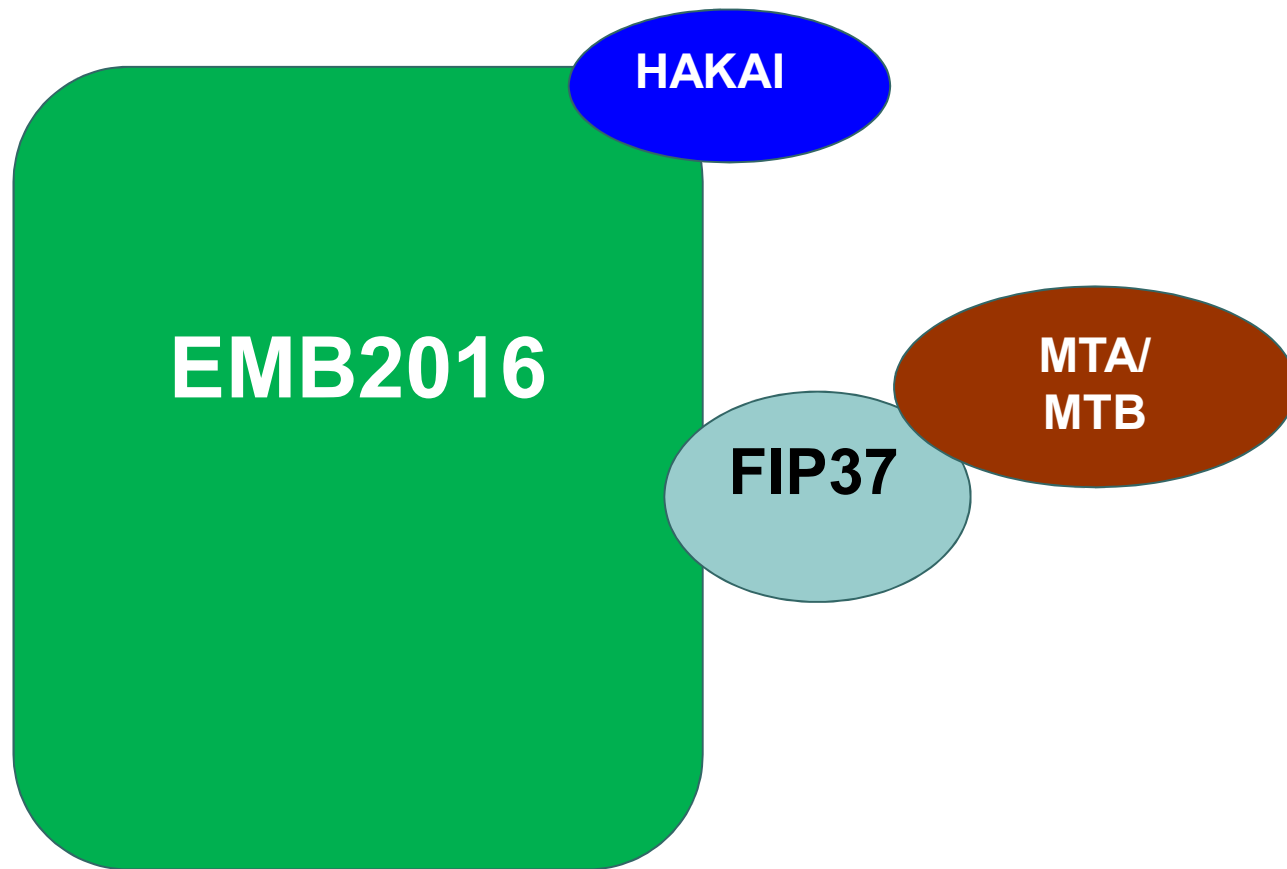


RING finger/HAKAI was also shown to associate with splicing factors (human)

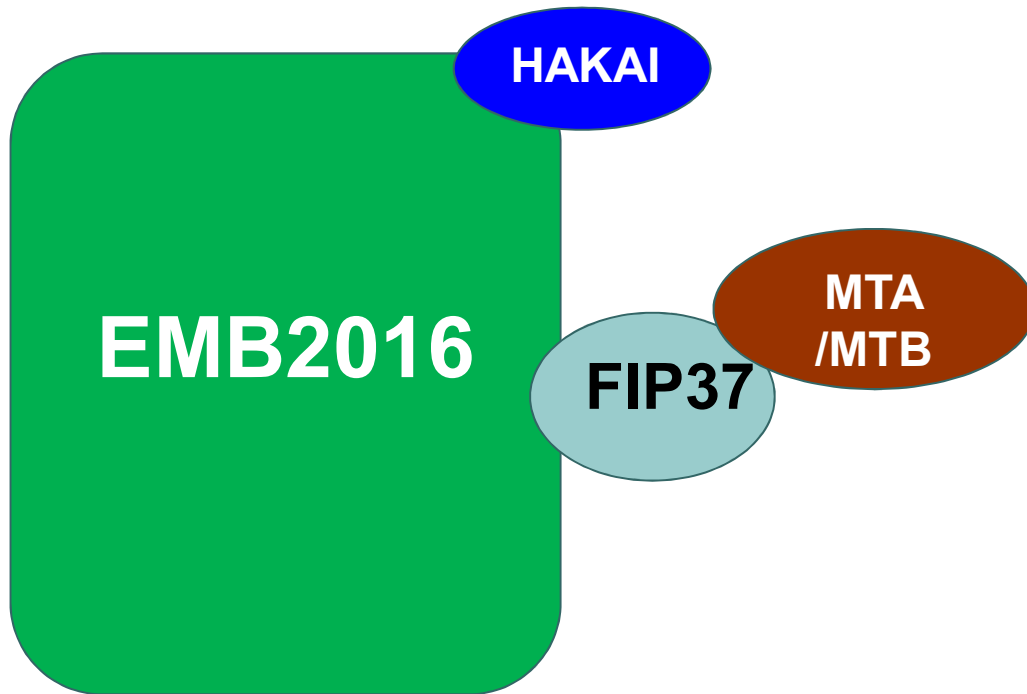


MTA-A – homolog of MTA

Do other proteins specifically bind?

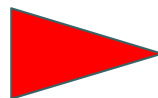


Advantage of conserved proteins in high throughput data

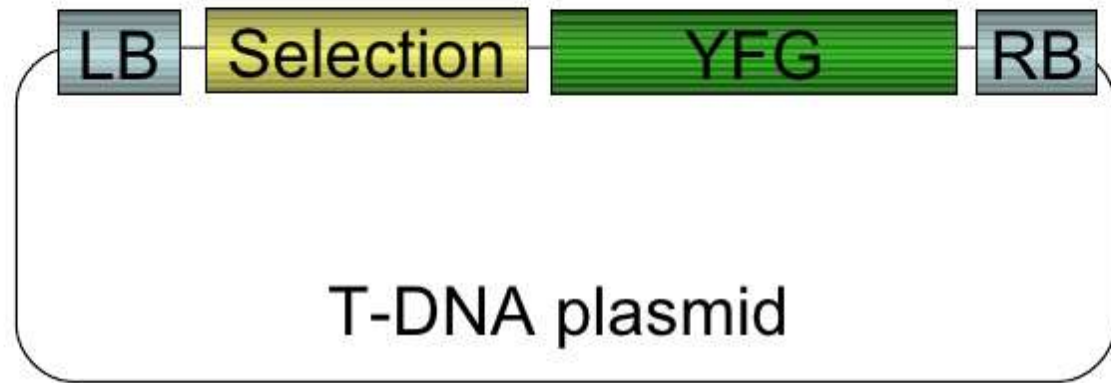
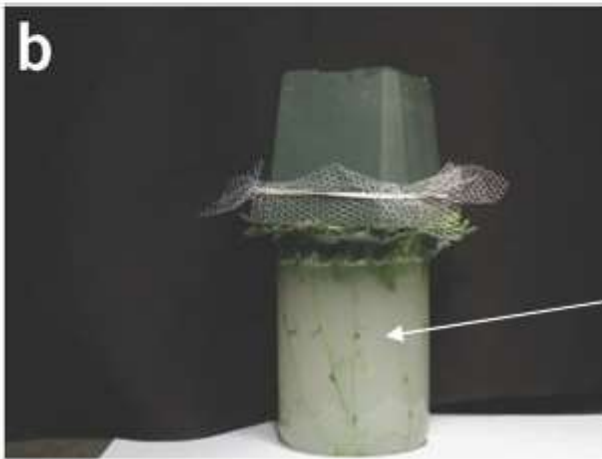


Flybase: EMB2016 interacts with HAKAI (no data on Biogrid)

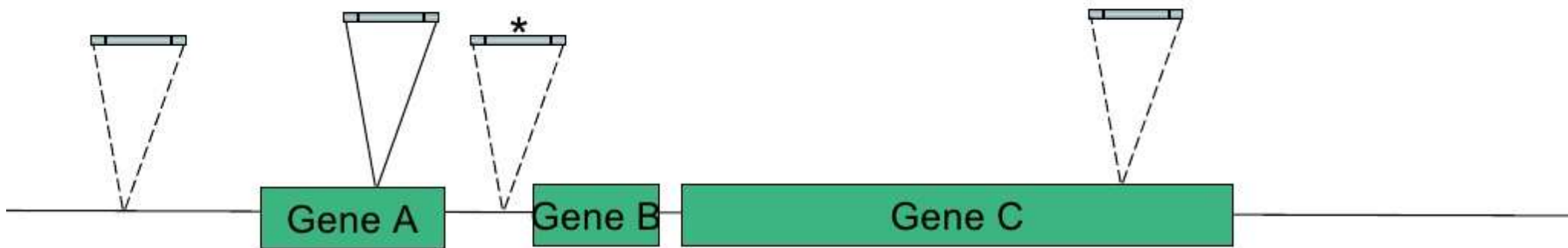
Summary of Physical Interactions		
RNA-protein		
Interacting group	Assay	References
vir - stau	anti bait coimmunoprecipitation, partial dna sequence identification by hybridization	(Laver et al., 2013)
protein-protein		
Interacting group	Assay	References
vir - CG7358	experimental knowledge based	(Guruharsha 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



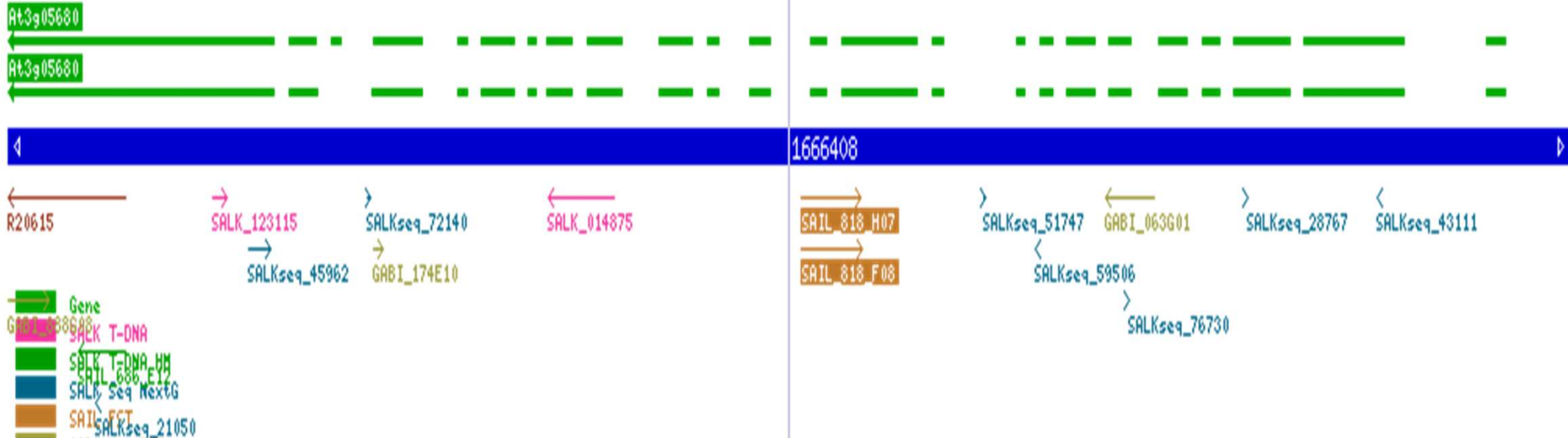
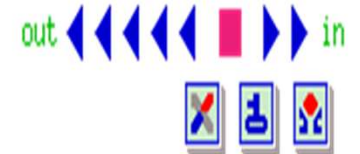
Examples of possible insertions:



You can order your mutant from the stock center

Arabidopsis thaliana [TAIR V10]

chr3 1661408 - 1671408



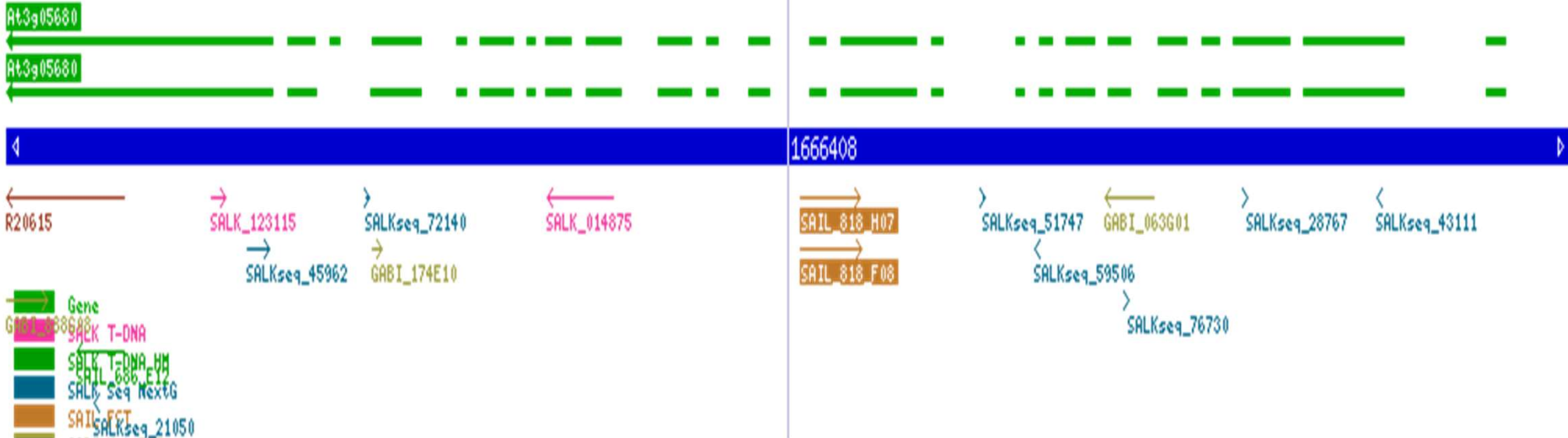
the same for Drosophila,
mouse, worm etc.

signal.salk.edu

You can order your mutant from the stock center

Arabidopsis thaliana [TAIR V10]

chr3 1661408 - 1671408



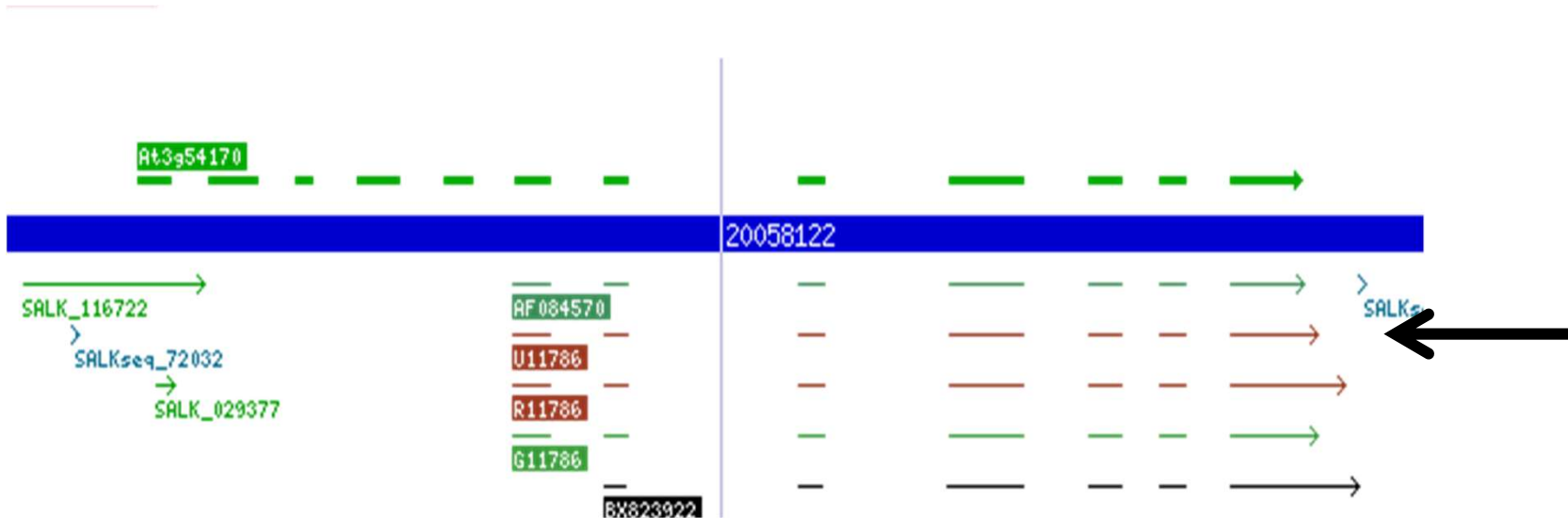
What to do if you cannot find insertion line for your gene?

the same for Drosophila, mouse, worm etc.

signal.salk.edu

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

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



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

You can order antibodies against your protein

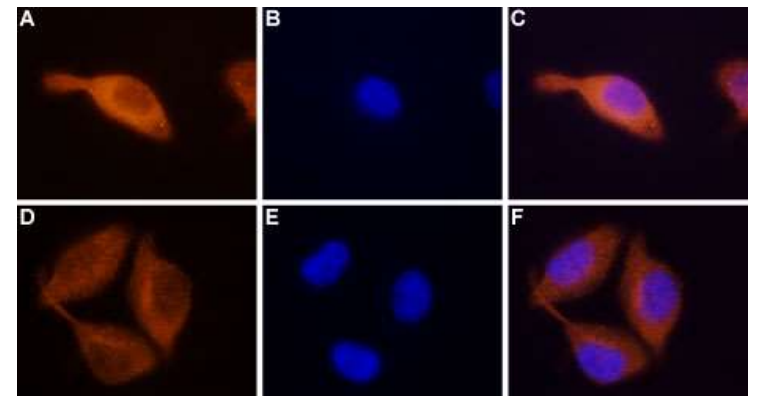
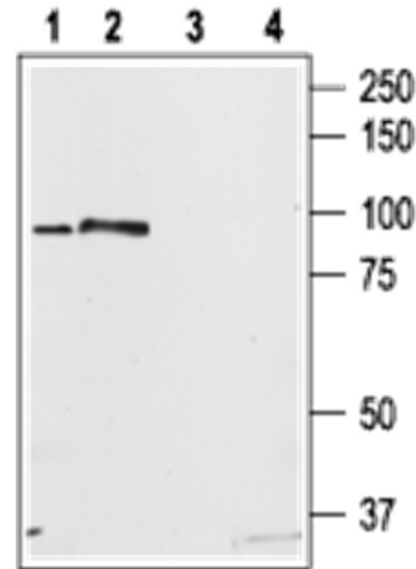
several human proteins
providers (mostly commercial):

<http://www.scbt.com/>

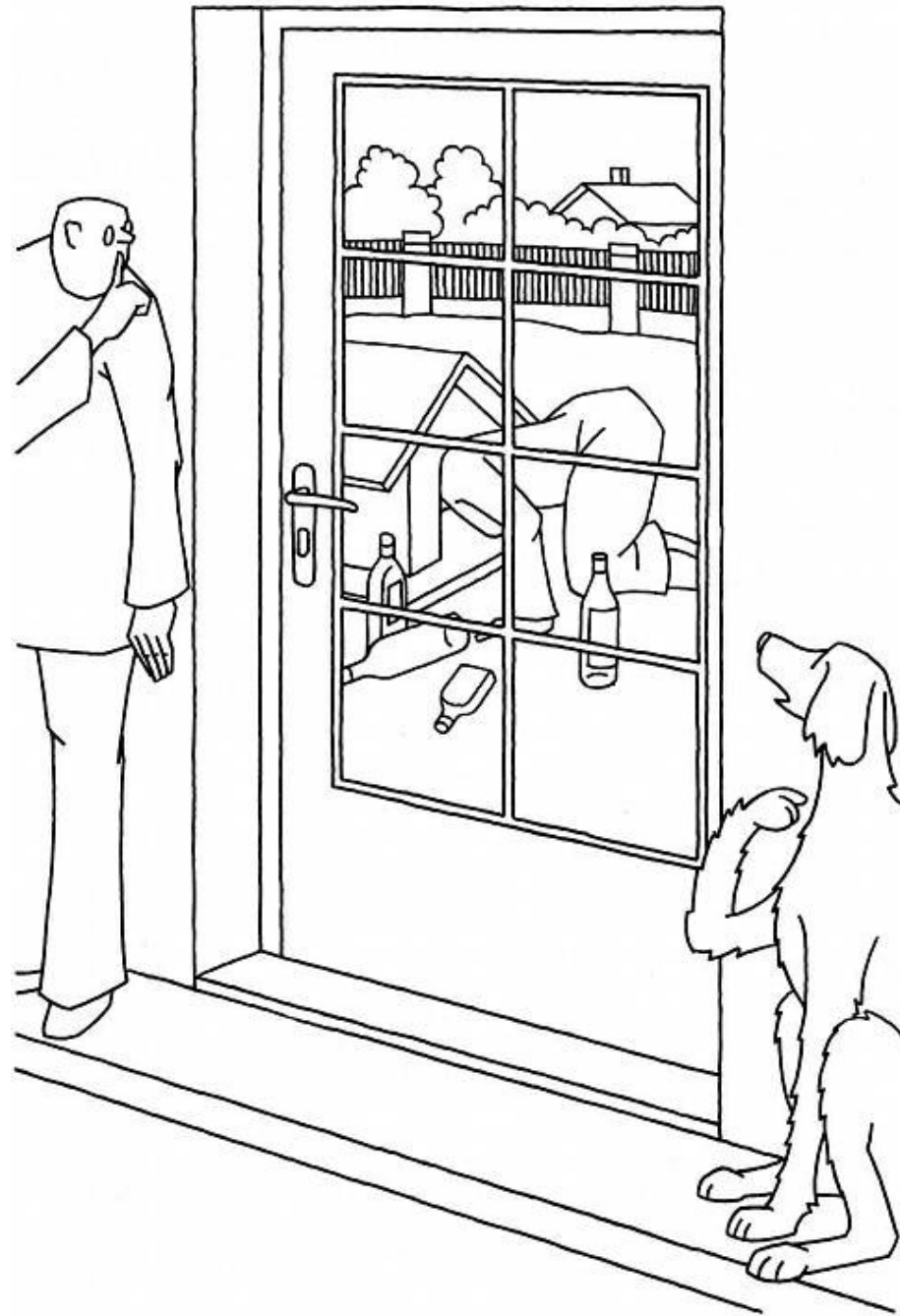
www.acris-antibodies.com/

etc.

- even get western and
immunocytochemistry in
advance

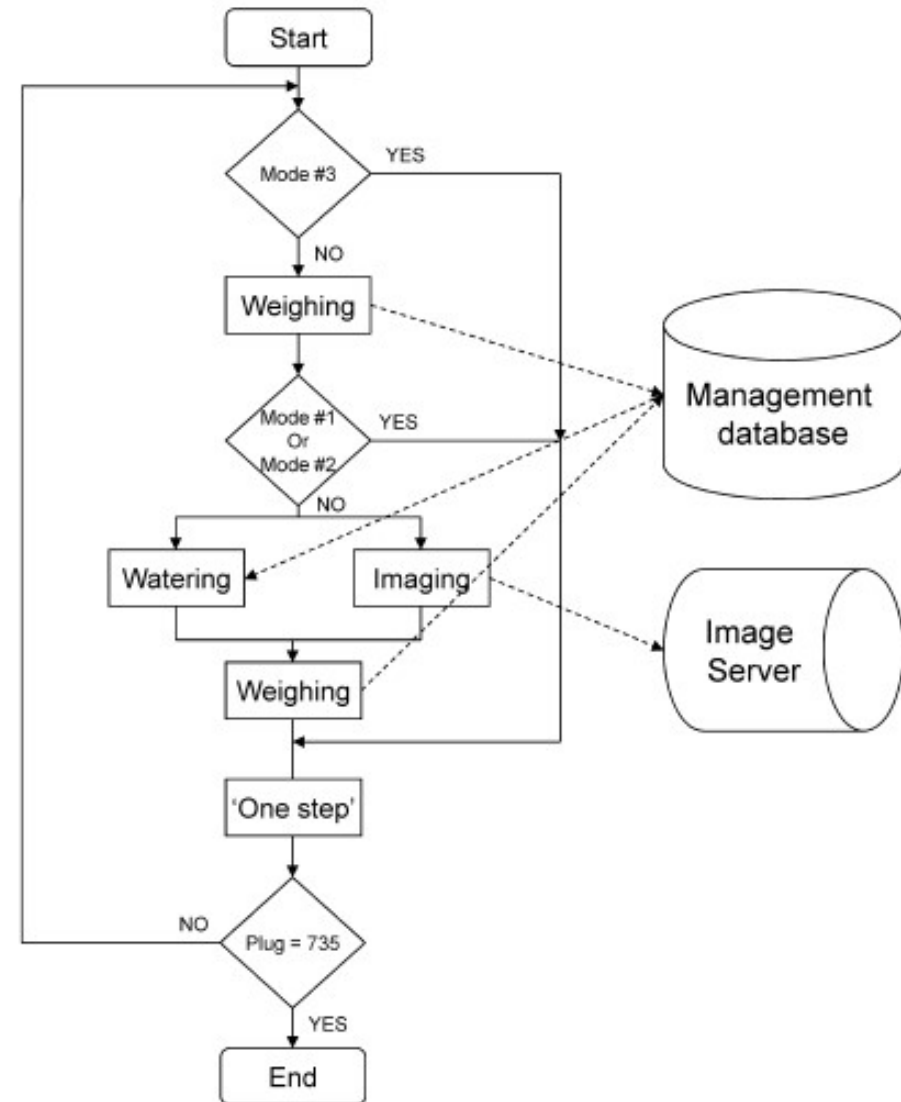
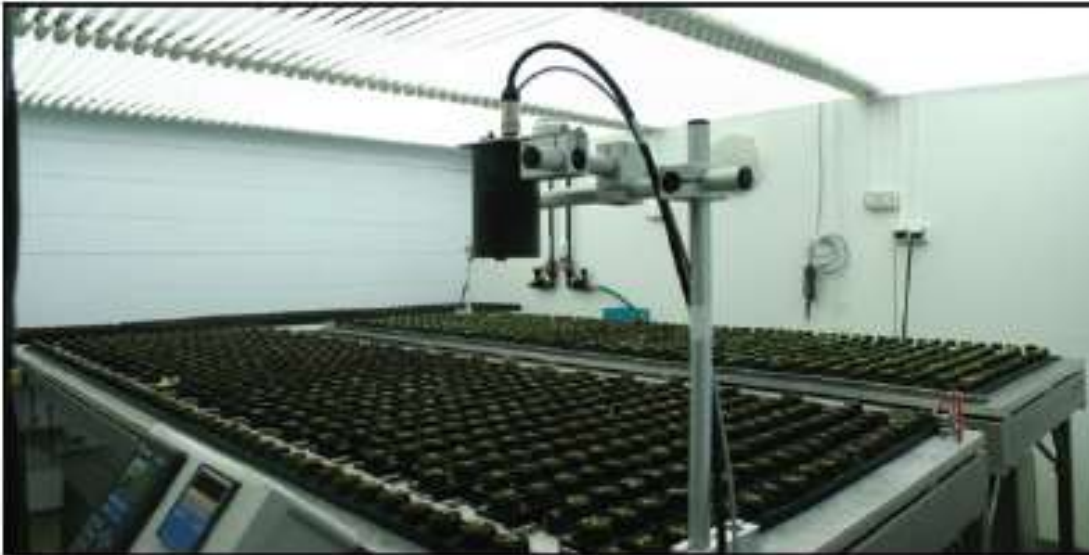


Plants – underdeveloped (Agrisera)



Phenoscope

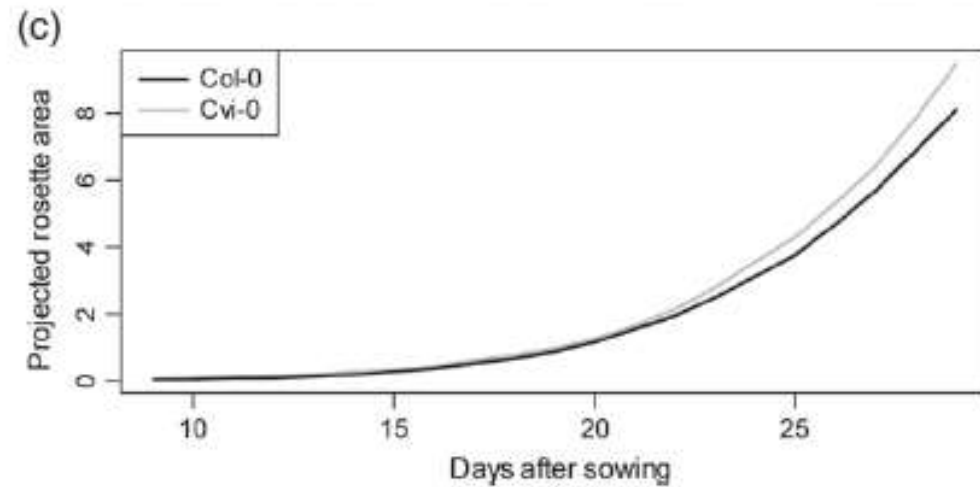
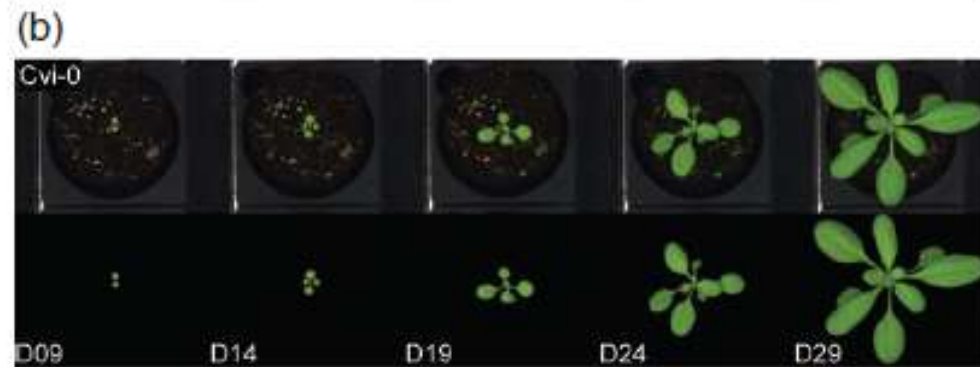
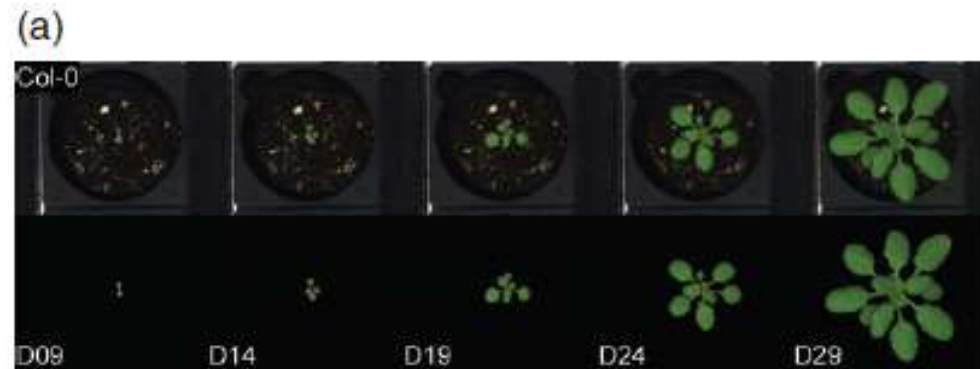
(a)



PHENOSCOPE: an automated large-scale phenotyping platform

Thisne et al. 2013

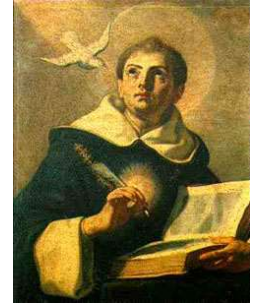
Phenoscope



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

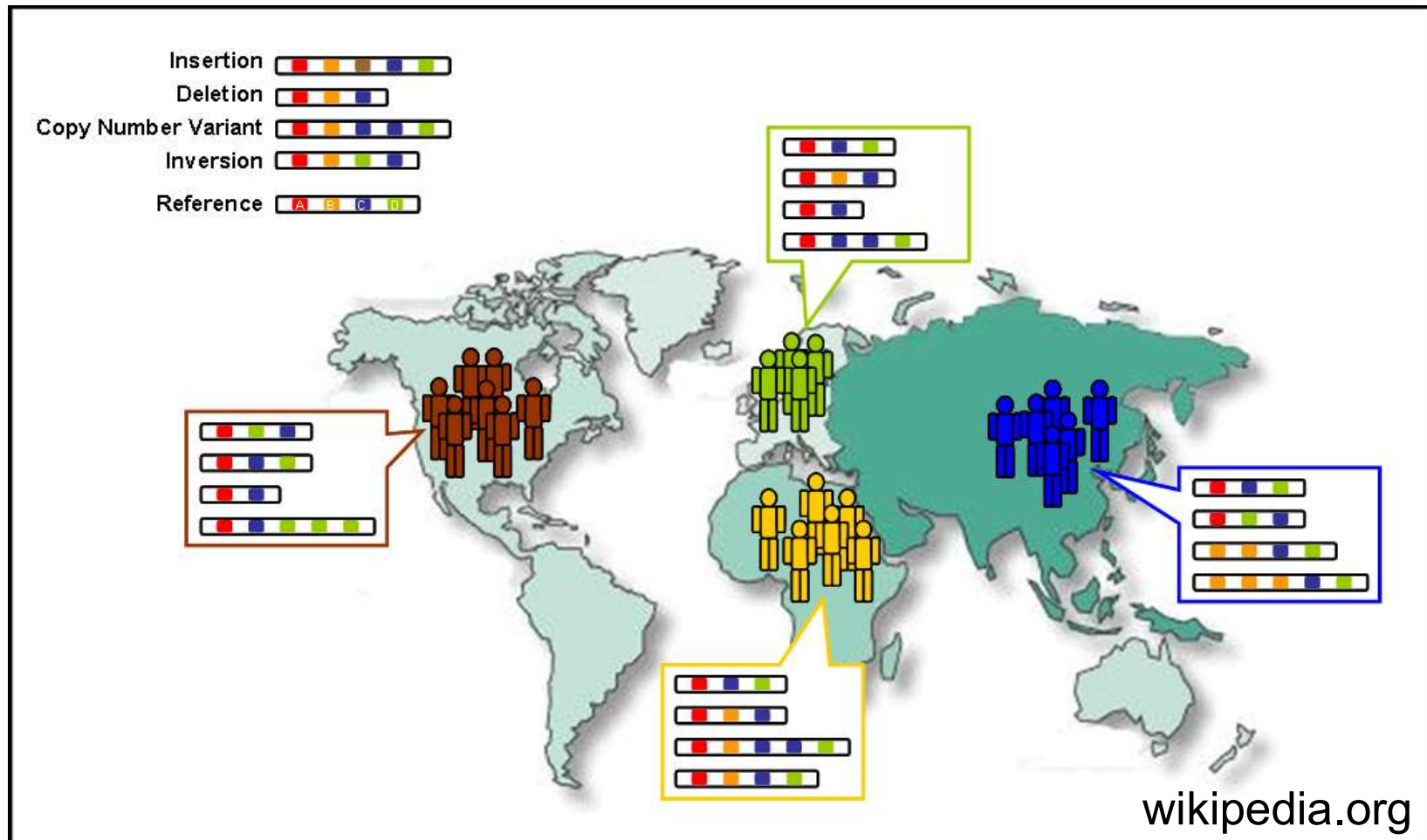
– database of plant embryonic mutants (in-depth)

<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
(already history)



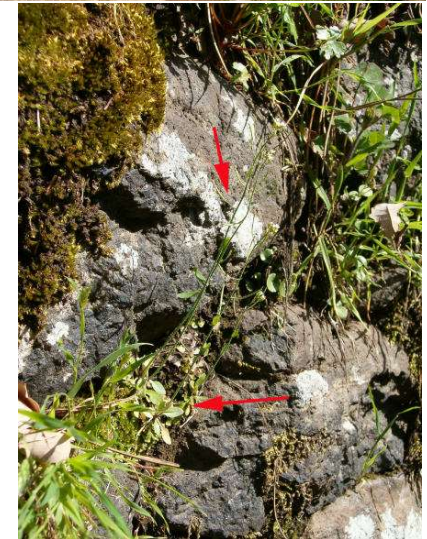
1001 genomes - Arabidopsis



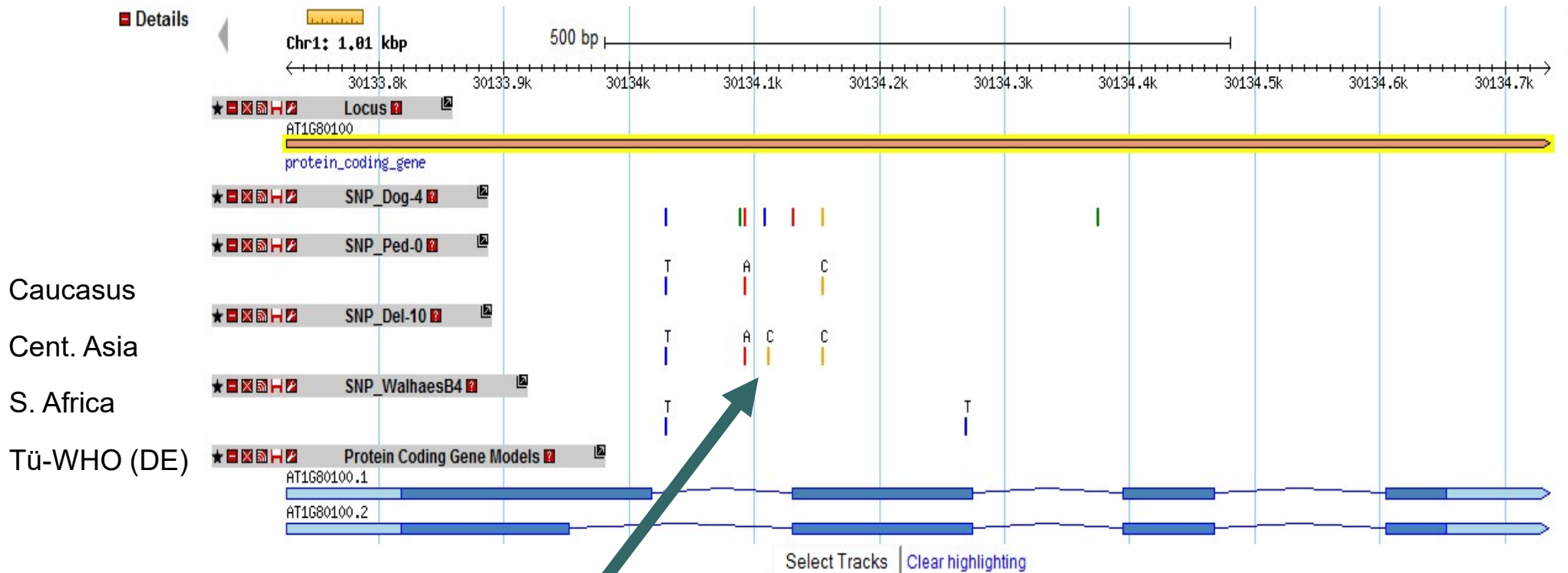
<http://1001genomes.org/>

in both cases, much more lines already sequenced

How the ecotypes are collected



1001 genomes user interface



several single nucleotide polymorphisms (**SNP**)
in the selected gene

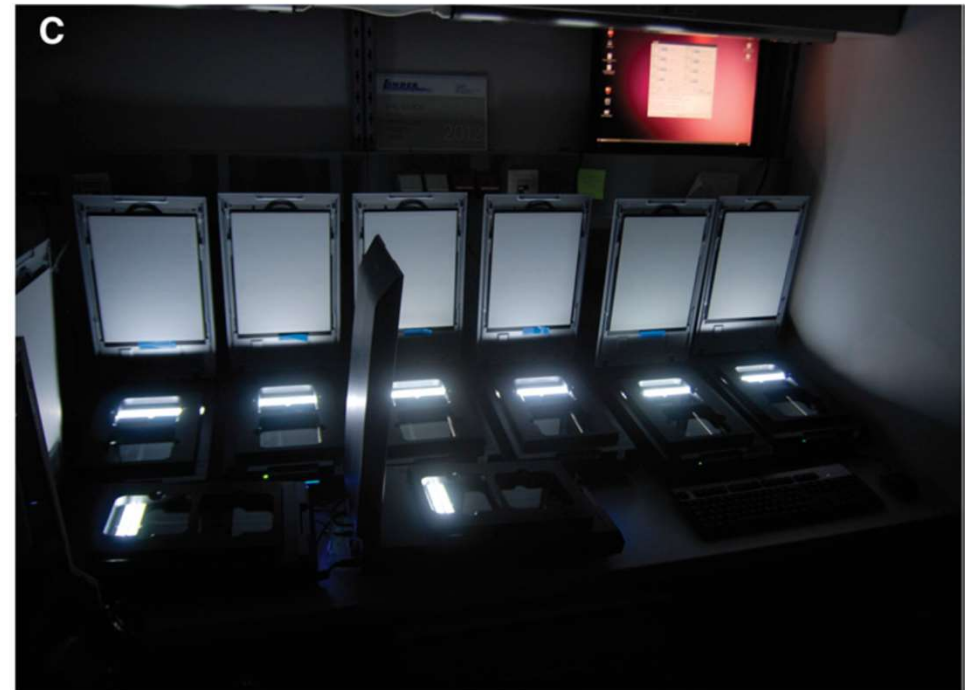
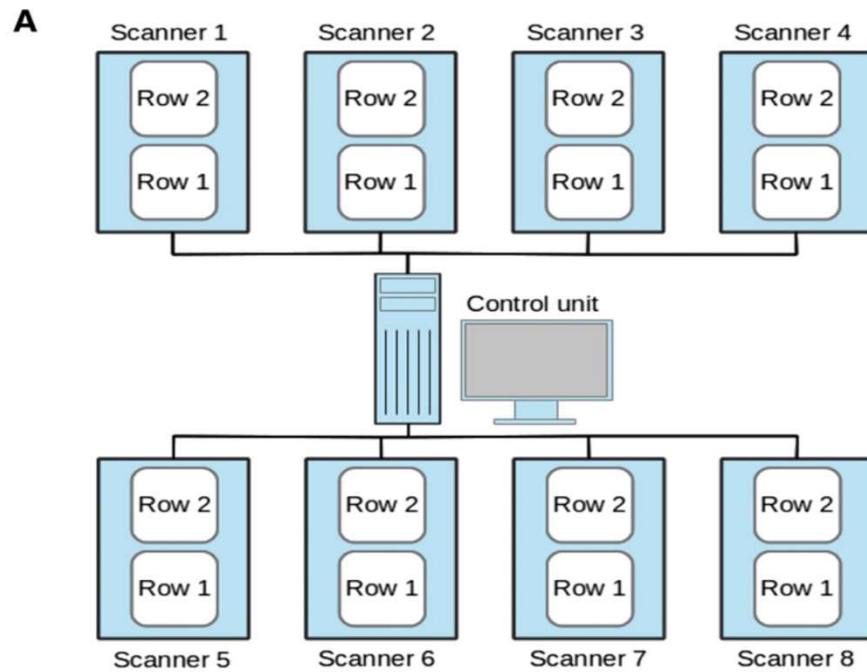
What could be natural variation good for?

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)



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!)

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

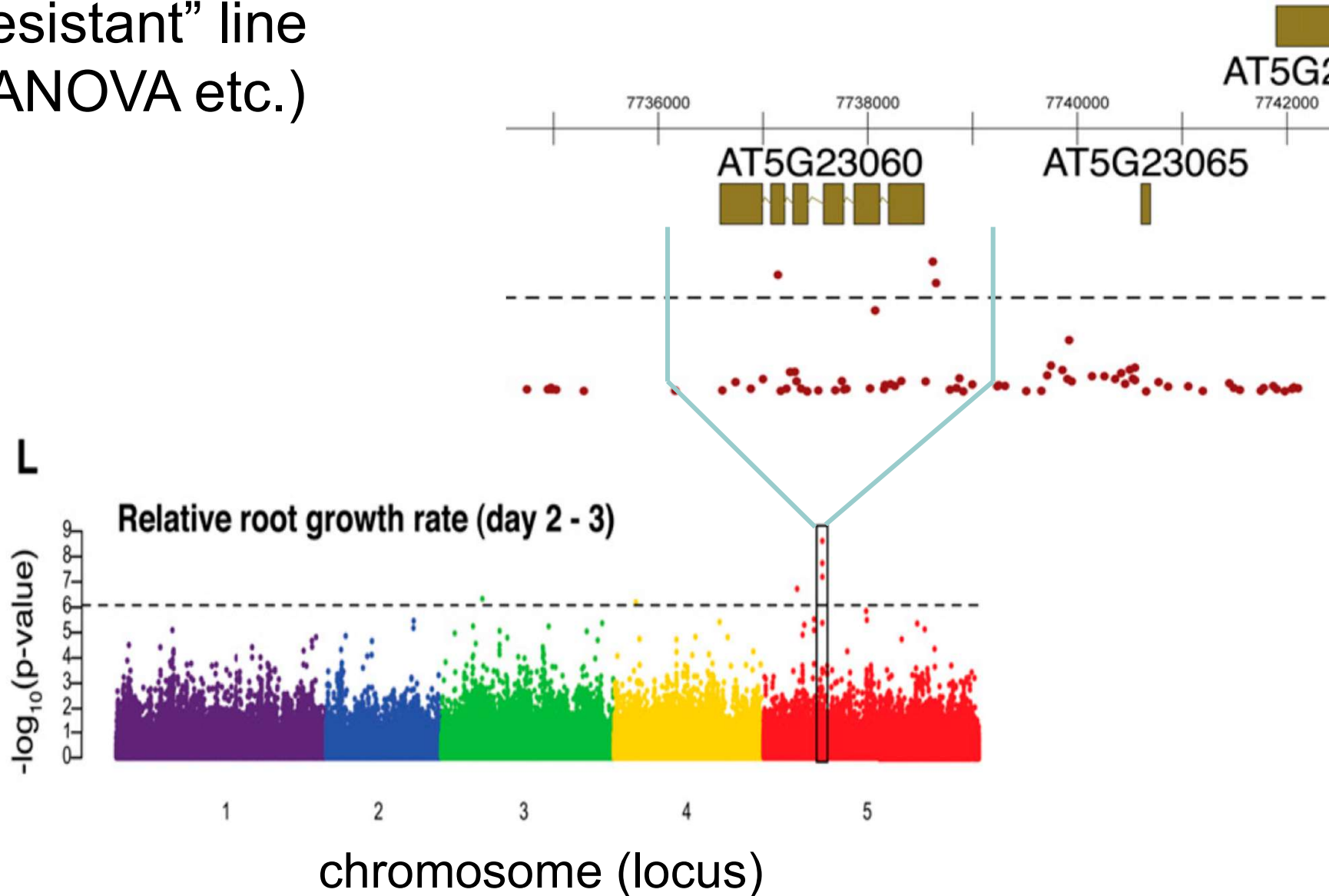
163 accessions (ecotypes),
several replicates (8 x 3)



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

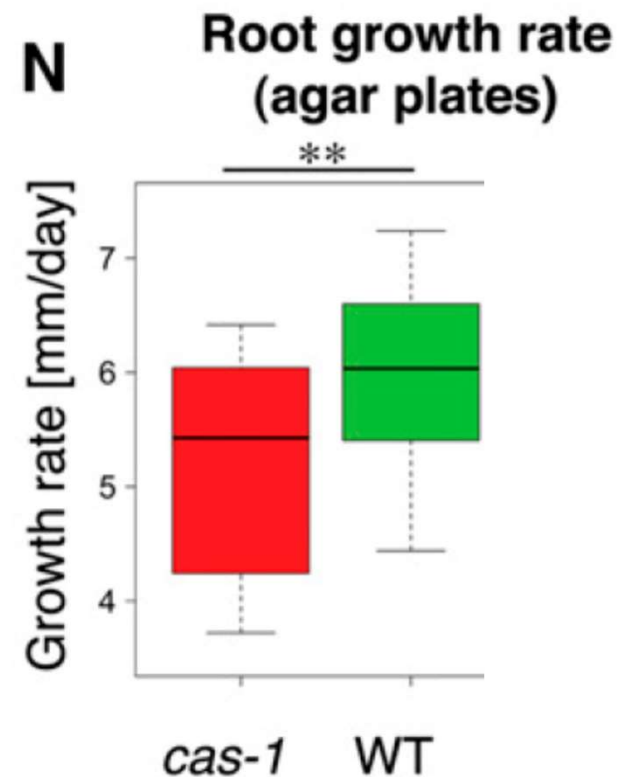
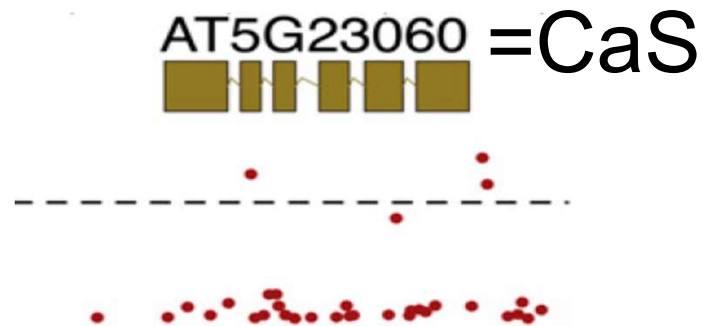
Genome wide association studies (GWAS)

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



Genome wide association studies (GWAS)

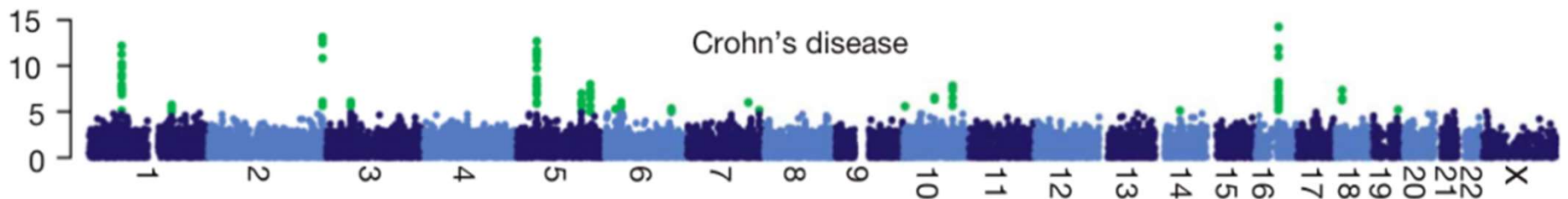
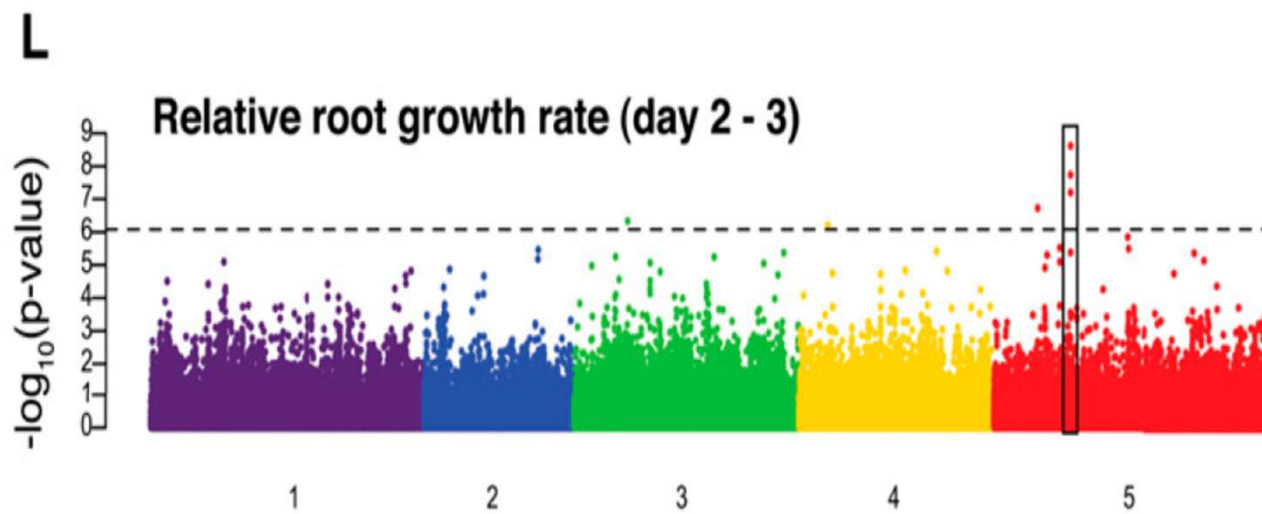
In contrast to human:
- how to test it?



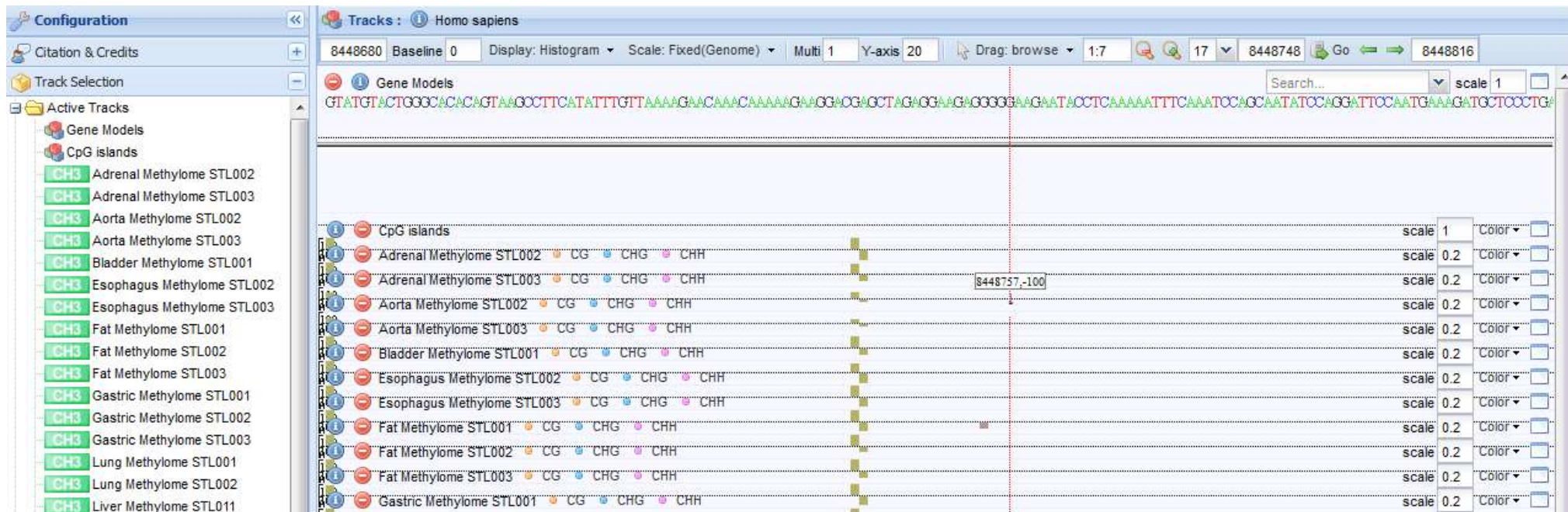
cas-1 mutant has indeed shorter root

Genome wide association studies (GWAS)

Manhattan plot by human



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



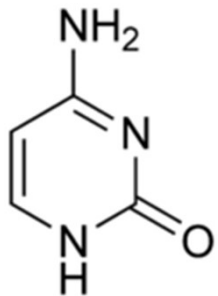
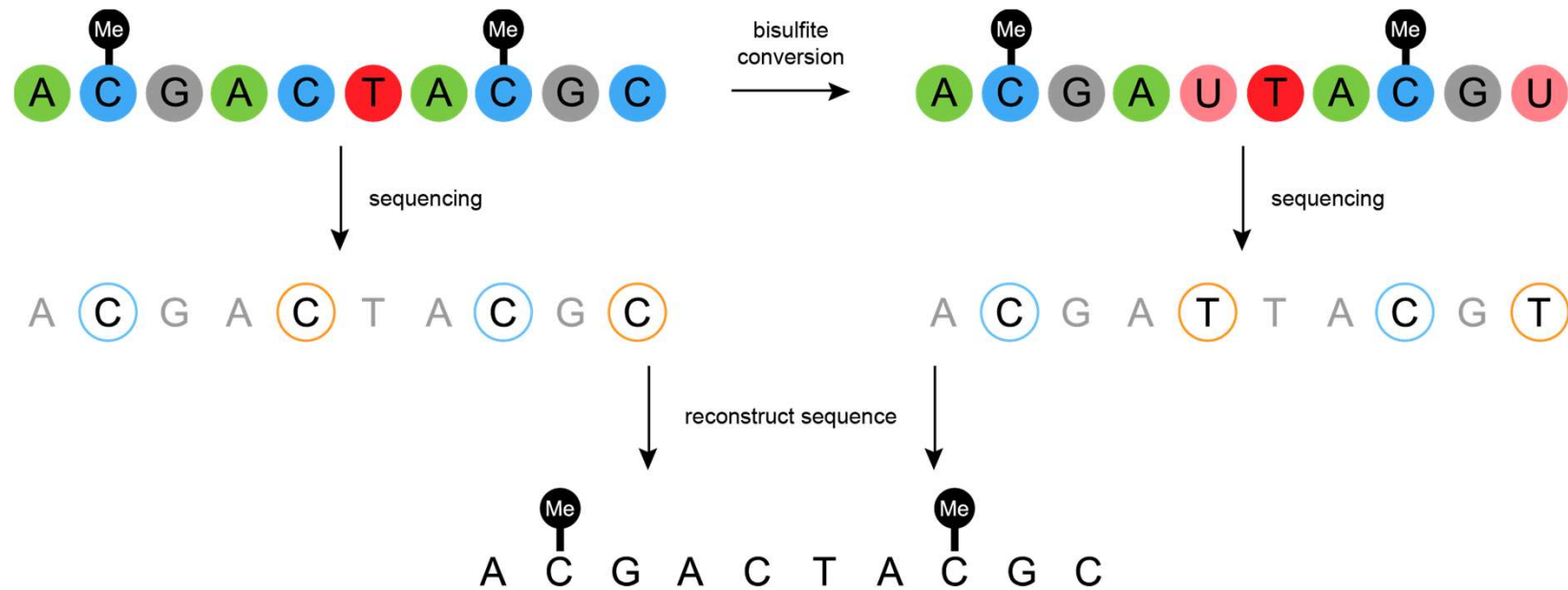
Epigenetic modifications

How to find methylated bases in genome?

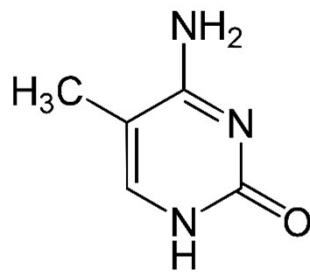
Which bases are methylated?

How to sequence methylation of genome?

bisulfite sequencing

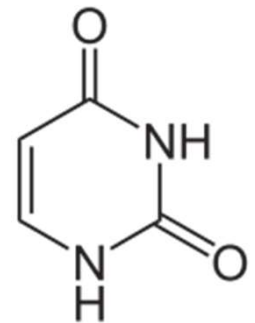


cytosine



5-methylcytosine

non-methylated
cytosine
is converted
to uracil



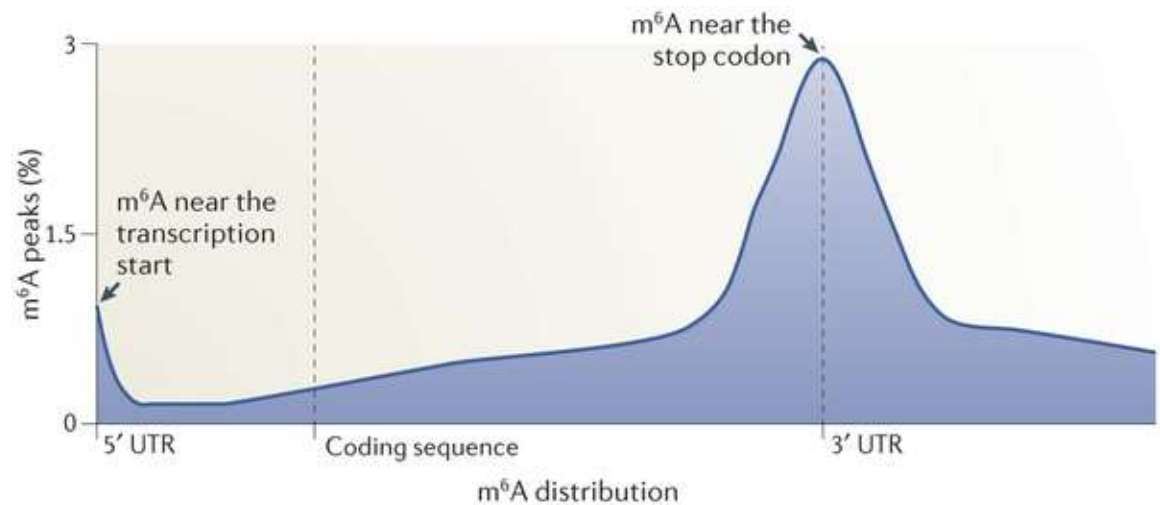
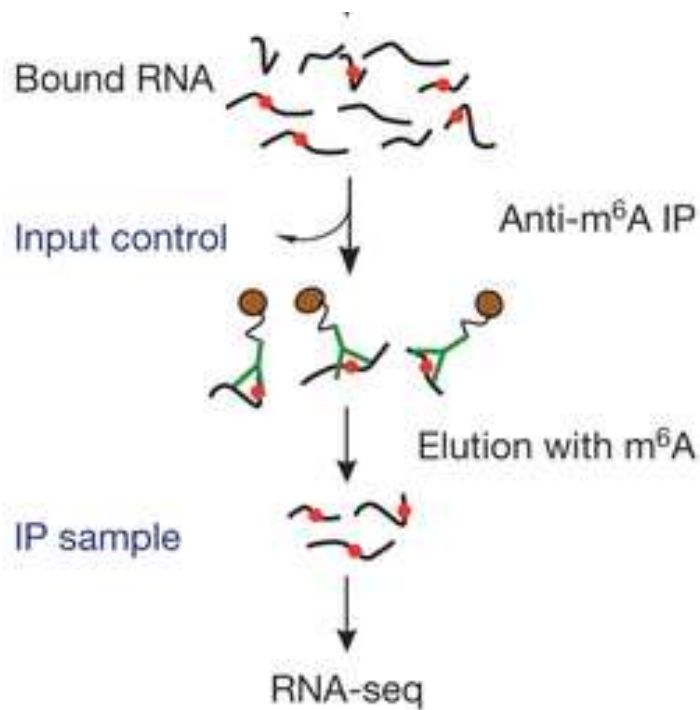
uracil

What is methylation of cytosine good for?

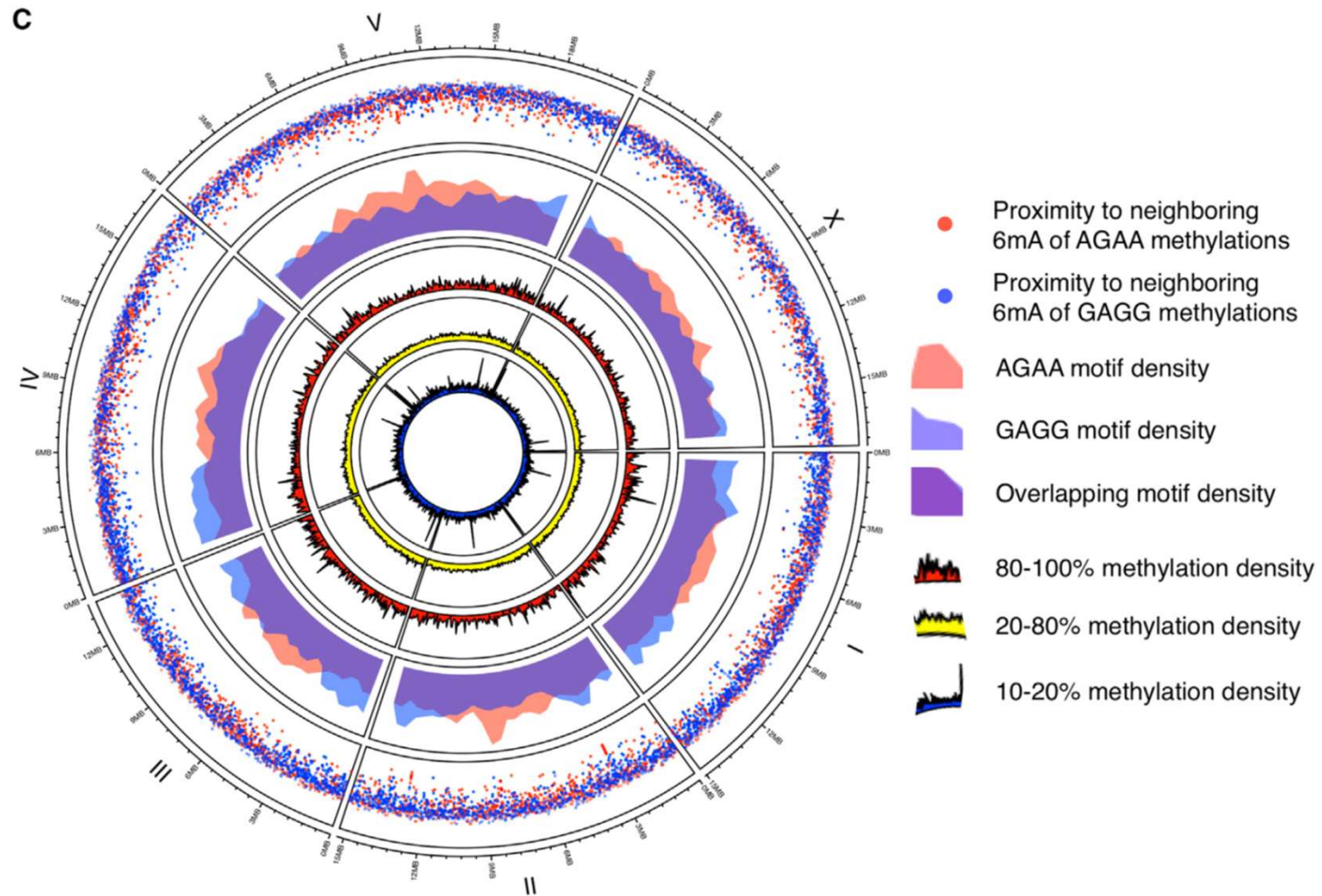
Are there other covalent modifications?

>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 (m6A)



Similar technique also adapted on DNA in *C. elegans* (6mA)



HOT! Novel avenues in m6A sequencing



Subscriber access provided by MASARYK UNIV BRNO

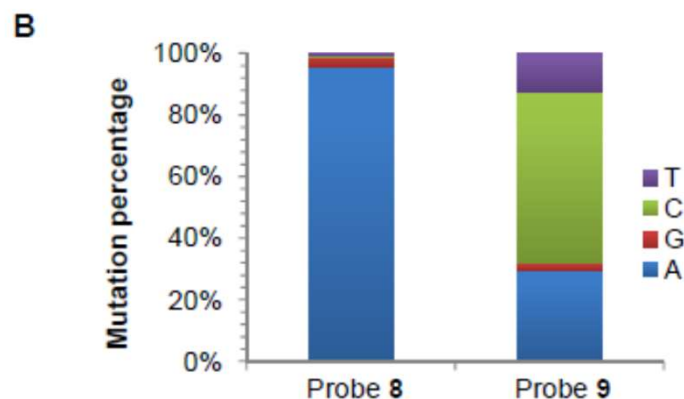
Communication

N6-Allyladenosine: a New Small Molecule for RNA Labeling Identified by Mutation Assay

Xiao Shu, Qing Dai, Tong Wu, Ian R. Bothwell, Yanan Yue, Zezhou Zhang, Jie Cao, Qili Fei, Minkui Luo, Chuan He, and Jianzhao Liu

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.7b06837 • Publication Date (Web): 08 Nov 2017

Downloaded from <http://pubs.acs.org> on November 14, 2017



HOT! Novel avenues in m⁶A sequencing even better one

Angewandte Chemie

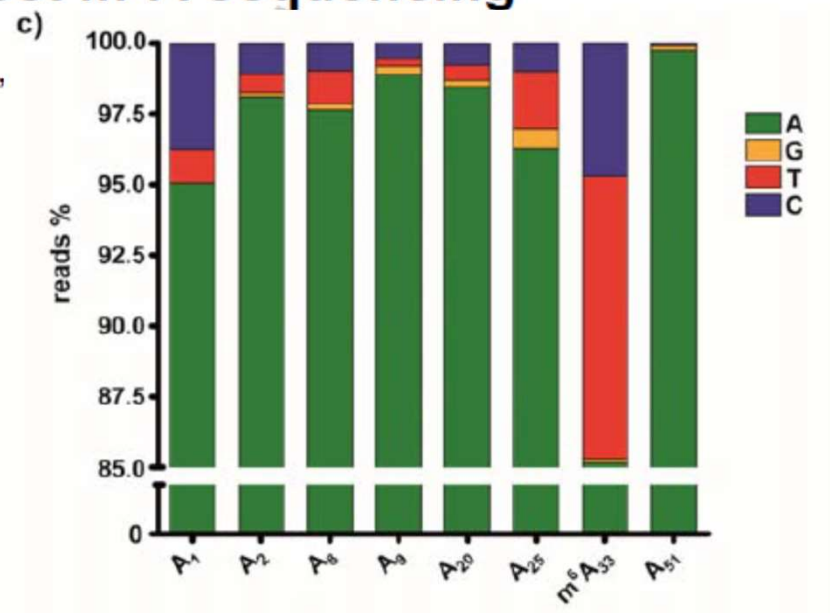
10.1002/ange.201710209

WILEY-VCH

COMMUNICATION

Engineering of a DNA polymerase for direct m⁶A sequencing

Joos Aschenbrenner^{†[a]}, Stephan Werner^{†[b]}, Virginie Marchand^[c],
Helm^[b] and Andreas Marx^{*[a]}



HOT! Novel avenues in m6A sequencing

nanopore 6mA sequencing

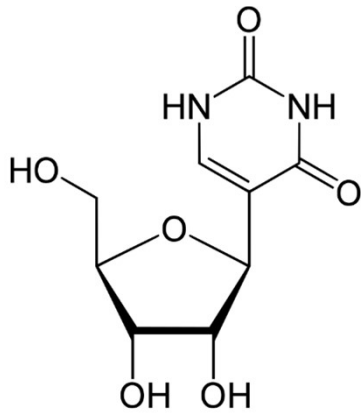
bioRxiv preprint first posted online Apr. 13, 2017; doi: <http://dx.doi.org/10.1101/127100>. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a [CC-BY-NC 4.0 International license](#).

Nanopore detection of bacterial DNA base modifications

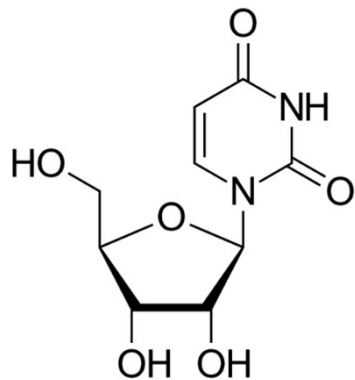
Alexa B.R. McIntyre^{1,2}, Noah Alexander¹, Aaron S. Burton³, Sarah Castro-Wallace⁴, Charles Y. Chiu^{5,6}, Kristen K. John⁷, Sarah E. Stahl⁸, Sheng Li^{9,10,11}, Christopher E. Mason^{1,12,13*}

<https://www.biorxiv.org/content/early/2017/04/13/127100>

Presence of pseudouridine in mammalian mRNA highly dependent on method, lab, conditions...



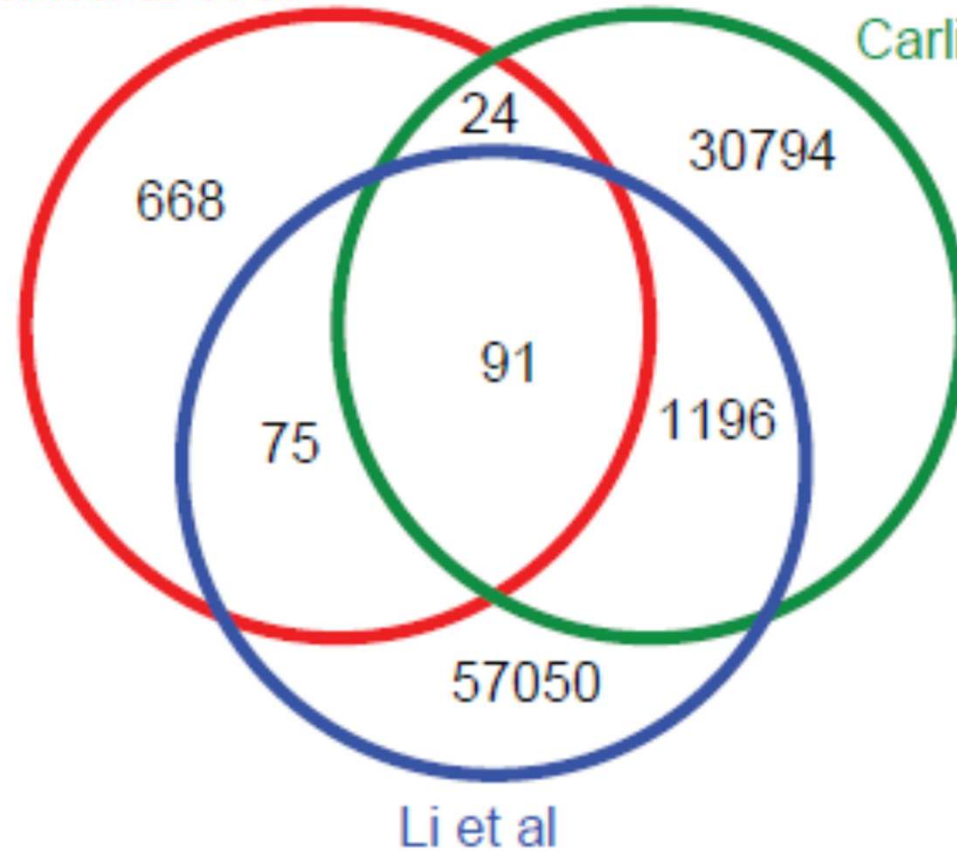
uridine

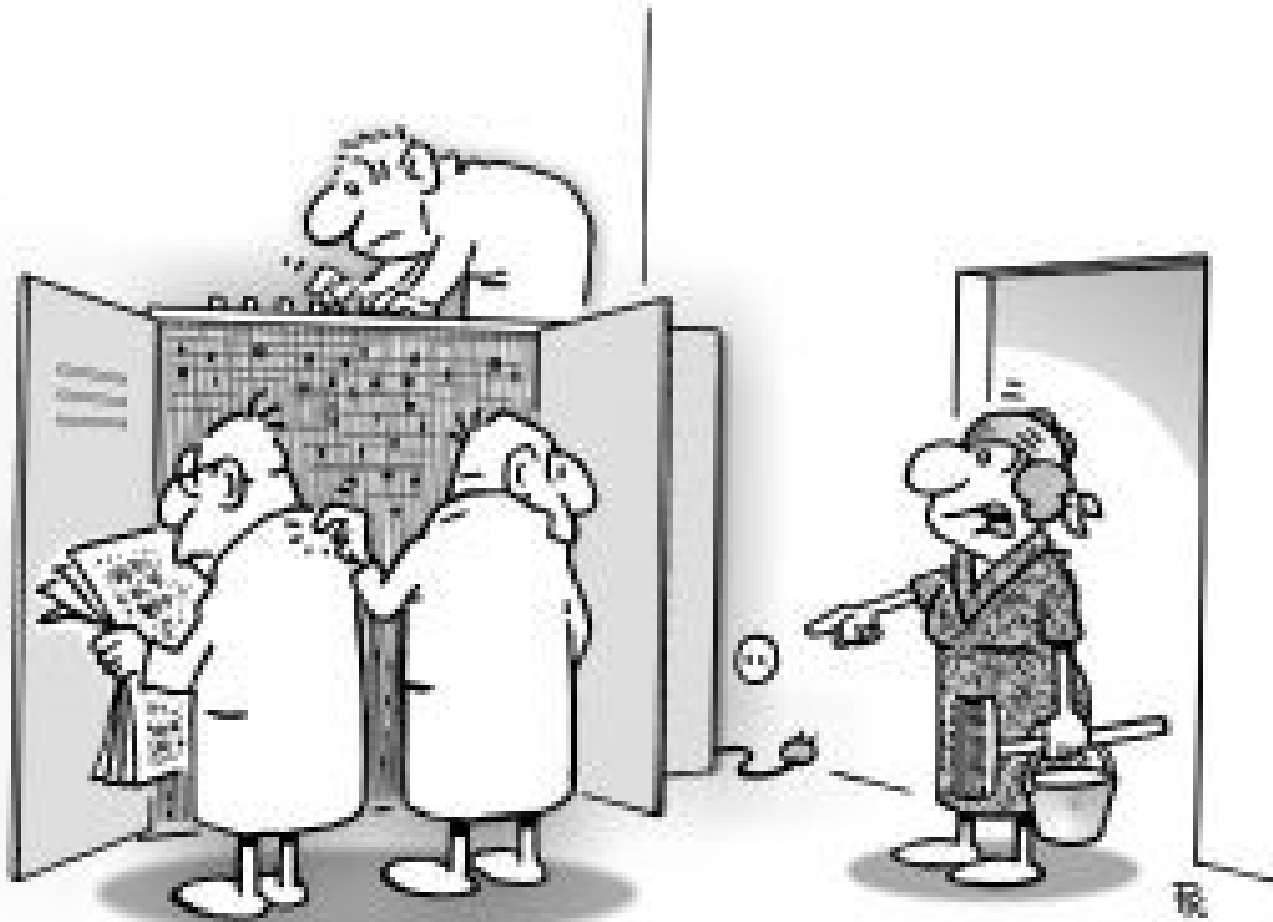


pseudouridine

Schwartz et al

Carlile et al







The ENCODE project

The Encyclopedia of DNA Elements

Is really only ~1 % human genome functional?

1 % = gene coding regions

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)

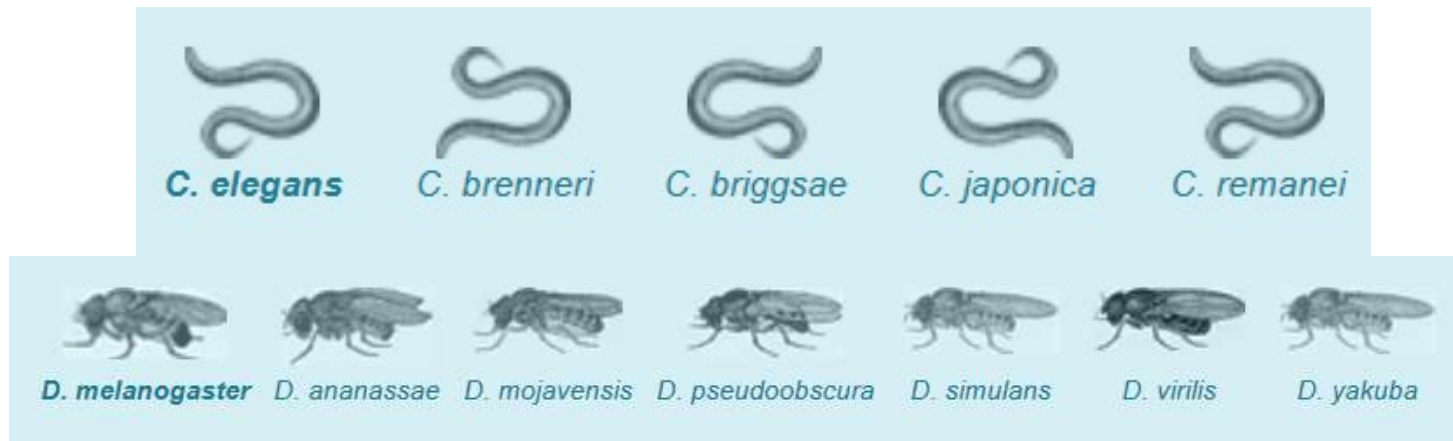
ENCODE - summary

~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



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.

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?



“Multidimensional biology”

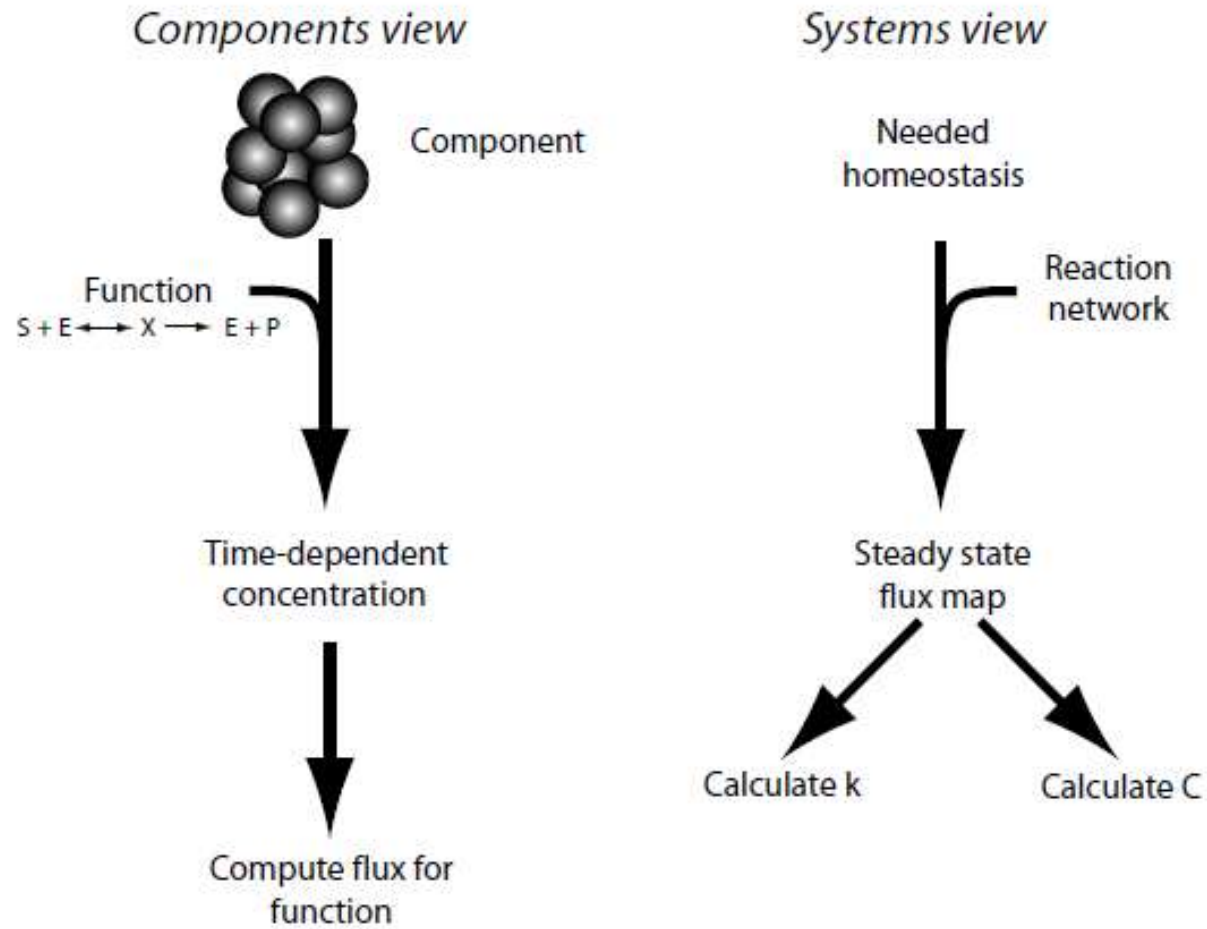
- Genomics
 - Epigenomics
 - Transcriptomics
 - Epitranscriptomics
 - Translatomics / Proteomics
 - Metabolomics
 - Interactomics
 - Fluxomics
 - NeuroElectroDynamics
 - Phenomics
 - Biomics
-

Systems theory

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

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

Reductionism vs. holism



Ludwig von Bertalanffy

(1901-1972)

Copyrighted material; sample page 22 of 22

\$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

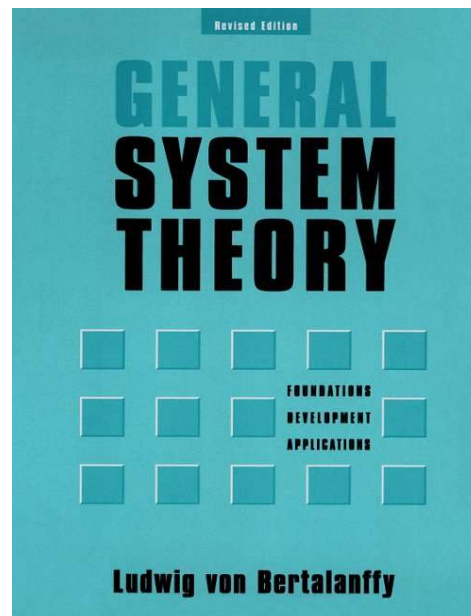
Hierarchy Theory

ISBN 0-8076-0674-X, hb, \$7.95

GEORGE BRAZILLER, INC.

171 Madison Avenue
New York, NY 10016

ISBN 0-8076-0453-4

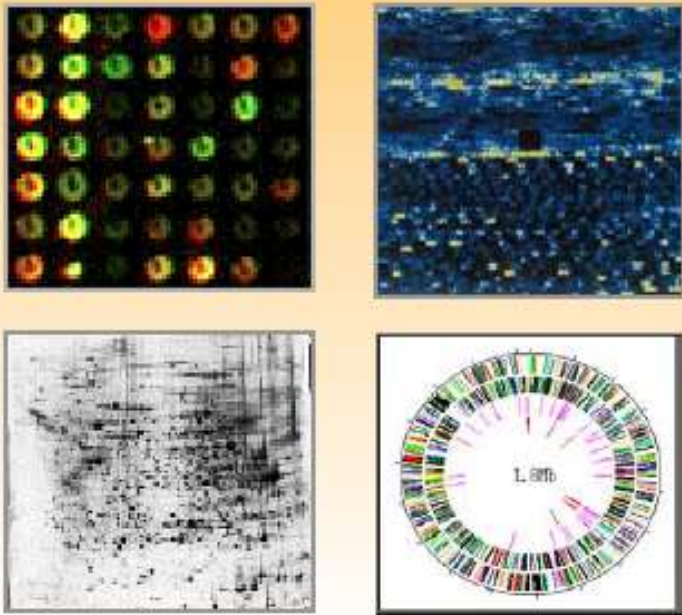


Copyrighted material; sample page 3 of 22

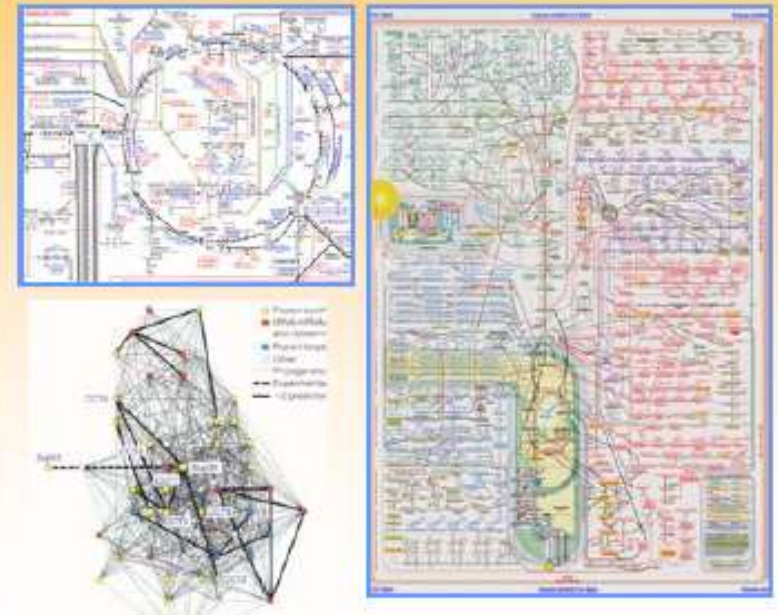
4	Advances in General System Theory	87
	Approaches and Aims in Systems Science	87
	Methods in General Systems Research	94
	Advances of General System Theory	99
5	The Organism Considered as Physical System	120
	The Organism as Open System	120
	General Characteristics of Open	
	Chemical Systems	124
	Equifinality	131
	Biological Applications	134
6	The Model of Open System	139
	The Living Machine and Its Limitations	139
	Some Characteristics of Open Systems	141
	Open Systems in Biology	145
	Open Systems and Cybernetics	149
	Unsolved Problems	151
	Conclusion	153
7	Some Aspects of System Theory in Biology	155
	Open Systems and Steady States	156
	Feedback and Homeostasis	160
	Allometry and the Surface Rule	163
	Theory of Animal Growth	171
	Summary	184
8	The System Concept in the Sciences of Man	186
	The Organismic Revolution	186
	The Image of Man in Contemporary Thought	188
	System-Theoretical Re-orientation	192
	Systems in the Social Sciences	194
	A System-Theoretical Concept of History	197
	The Future in System-Theoretical Aspect	203
9	General System Theory in Psychology and Psychiatry	205
	The Quandary of Modern Psychology	205
	System Concepts in Psychopathology	208
	Conclusion	220
10	The Relativity of Categories	222
	The Whorfian Hypothesis	222
	The Biological Relativity of Categories	227

Omic-revolution shifts paradigm to large systems

High Throughput Data

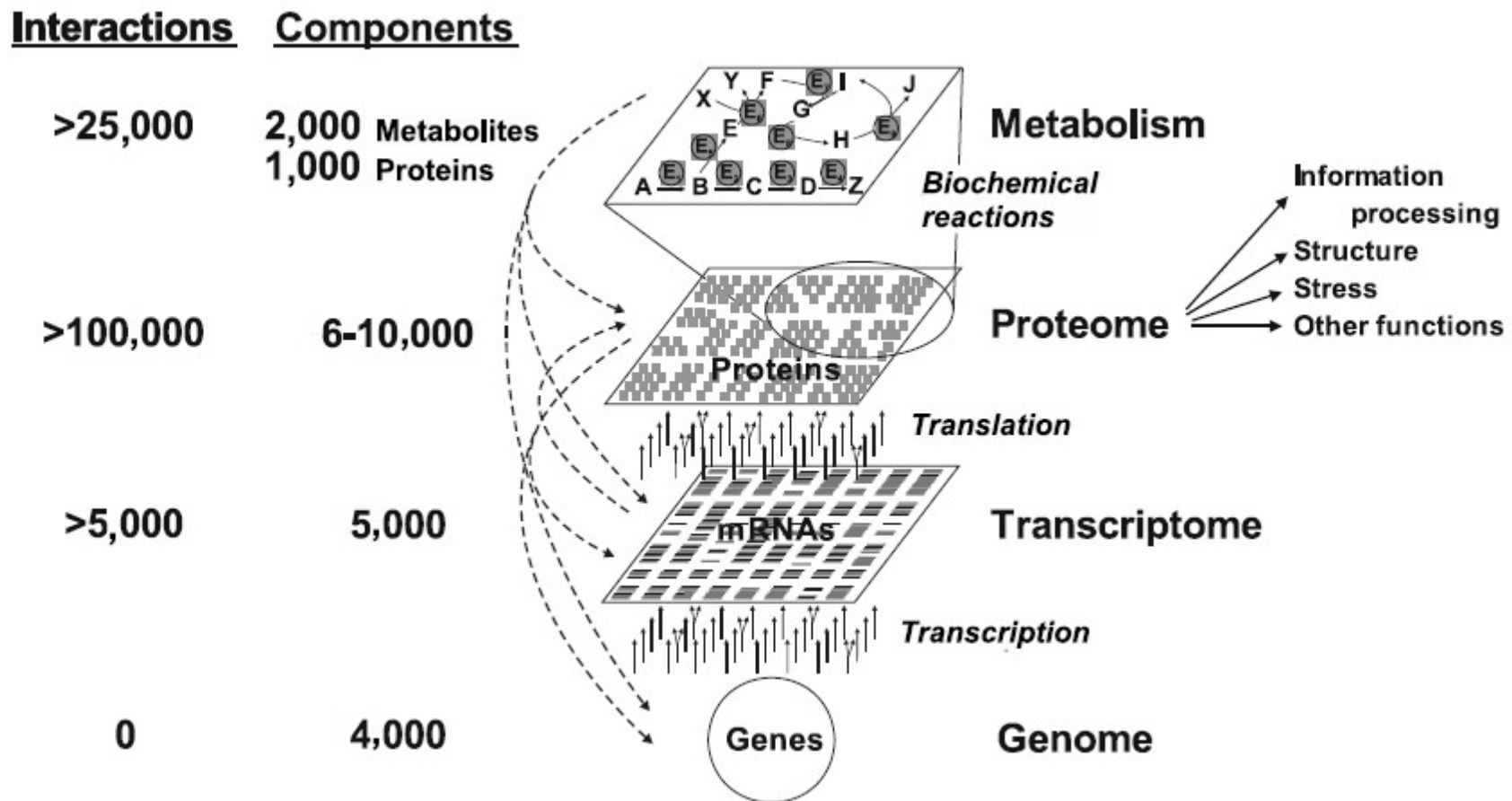


Cellular Complexity

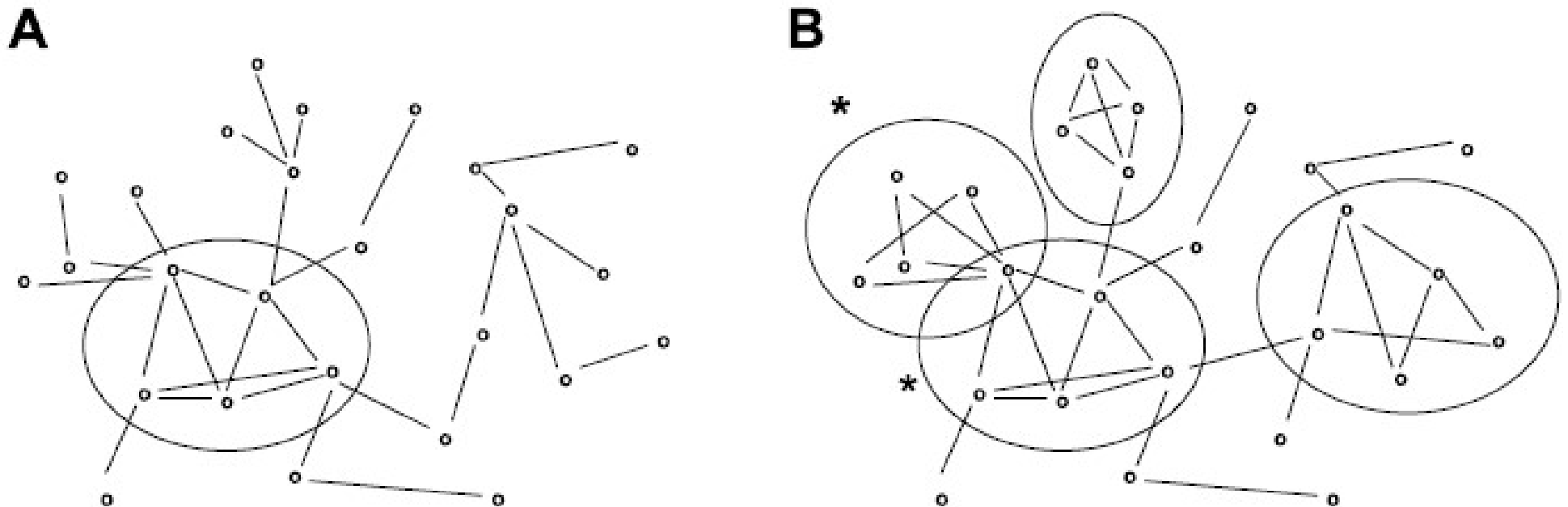


- Integrative bioinformatics
- (Network) modeling

E. coli genome and proteome is small



Reductionism within holism



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

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

<https://www.araport.org/>

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 pavčině 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