functional genomics

Astrid Lægreid

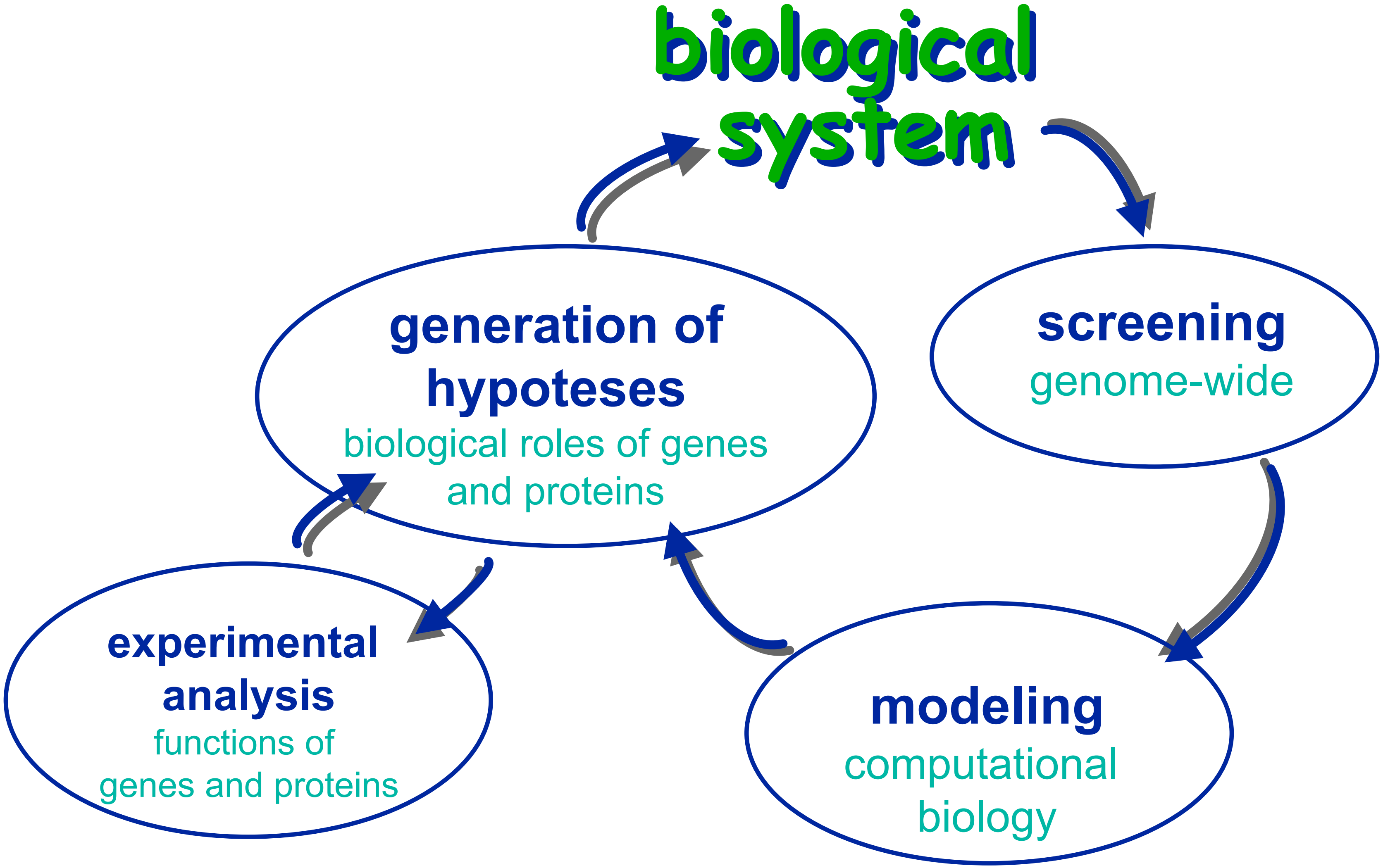
Department of Cancer Research and Molecular Medicine
Norwegian University of Science and Technology

Microarray Core Facility
Norwegian Microarray Consortium

outline

- functional genomics
- gene expression
- predicting gene function
- challenges

functional genomics



biological system

screening
genome-wide

modeling
computational biology

generation of hypoteses
biological roles of genes and proteins

experimental analysis
functions of genes and proteins

genome sequencing

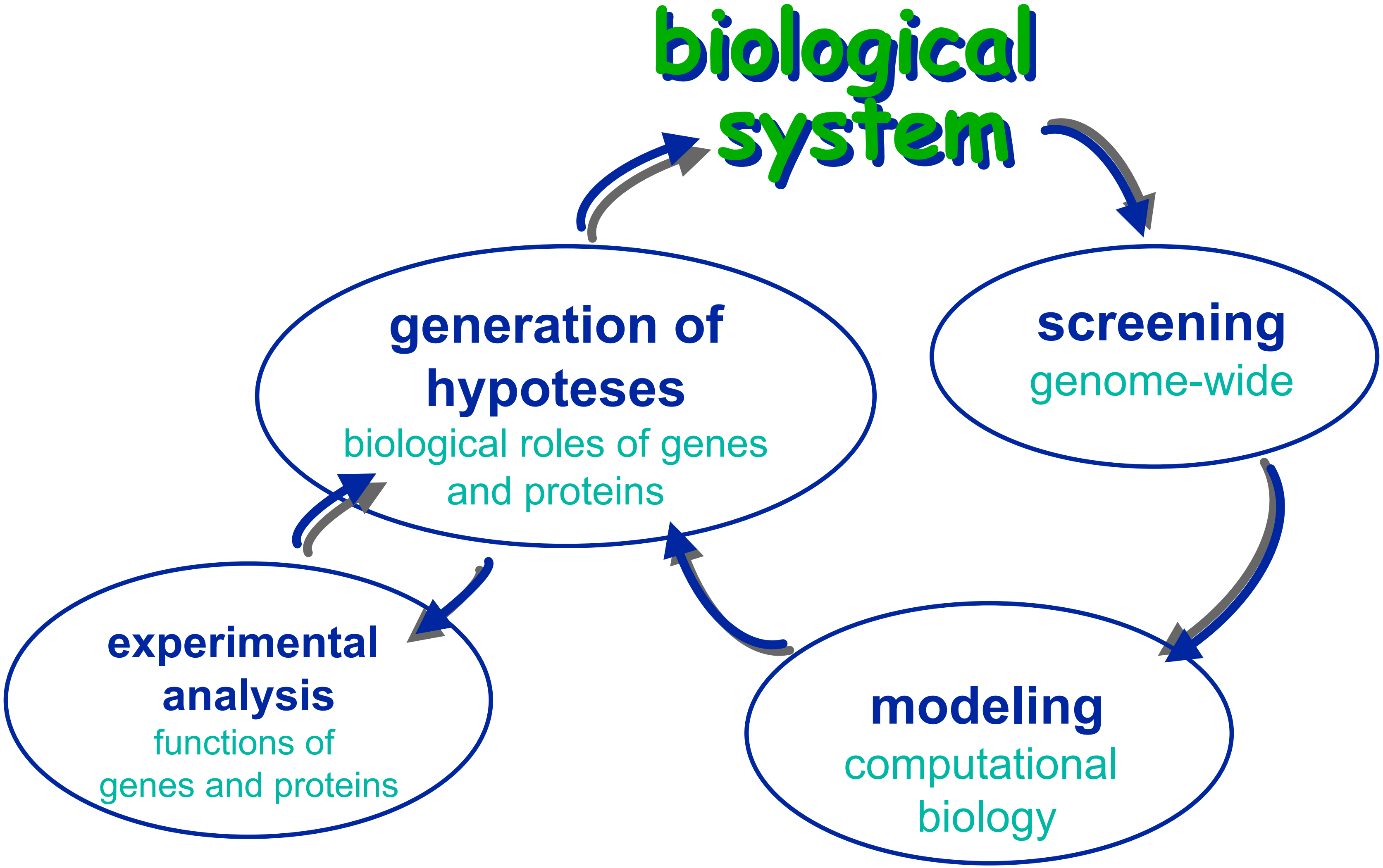
human genome

- 3×10^9 basepairs
- ~ 35.000 genes
- > 100.000 splice variants

genome-wide screening

- **how?** high-throughput - HTP
- **what?** gene expression, gene-dosage, gene-variation (SNP), protein
- **with?** microarray, mass spectrometry, 2D-gel electrophoresis

functional genomics



biological system

screening
genome-wide

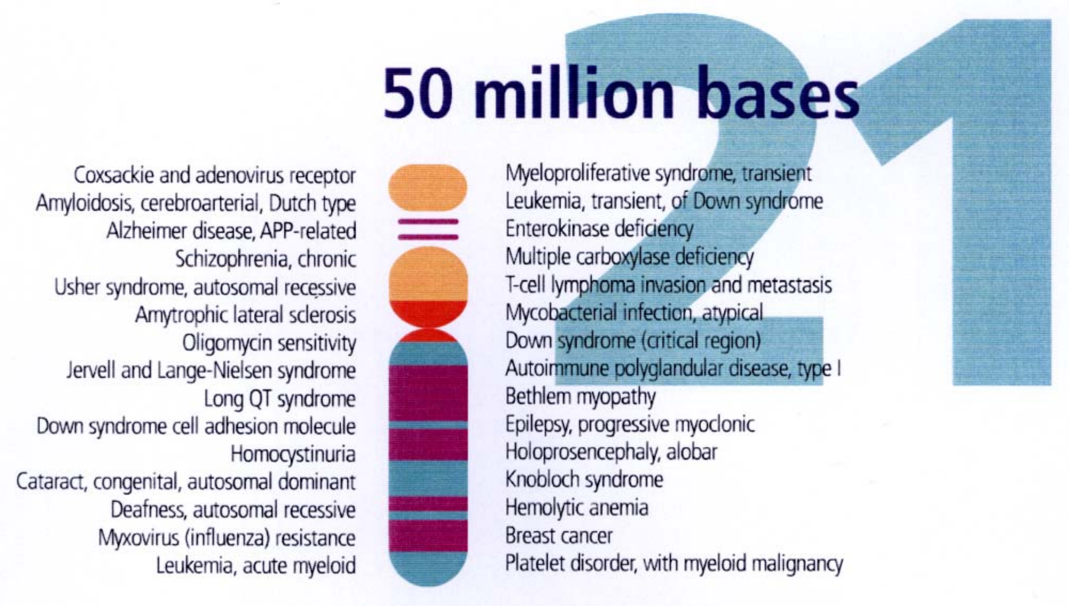
modeling
computational biology

generation of hypoteses
biological roles of genes and proteins

experimental analysis
functions of genes and proteins

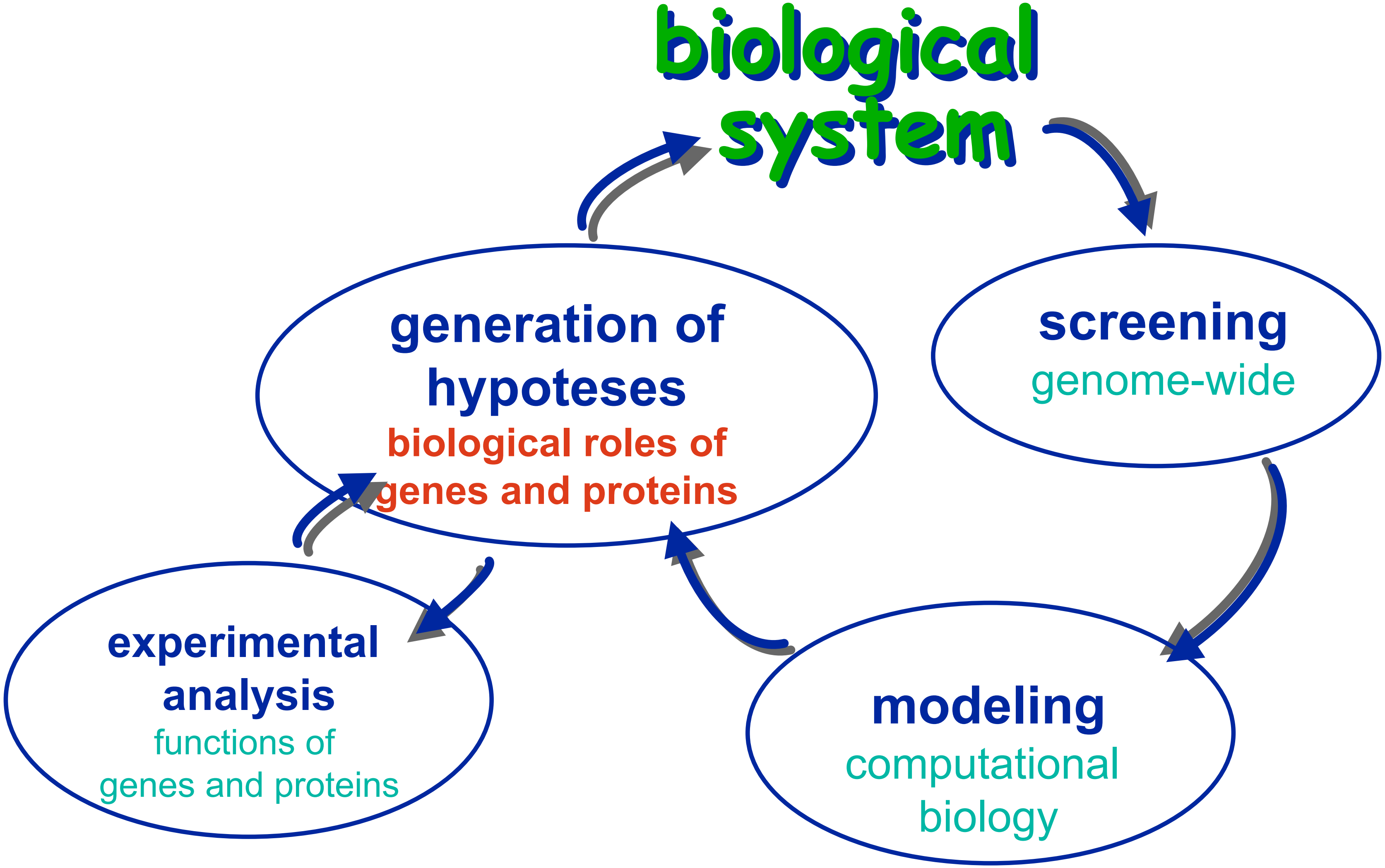
chromosome 21

- 127 known genes
- 98 unknown genes




164 million bases
X
Short stature, idiopathic familial
Leri-Weill dyschondrosteosis
Langer mesomelic dysplasia
Leukemia, acute myeloid, M2 type
Chondrodysplasia punctata
Kallmann syndrome
Ocular albinism, Nettleship-Falls type
Oral-facial-digital syndrome
Nance-Horan cataract-dental syndrome
Heterocellular hereditary persistence of fetal hemoglobin
Pyruvate dehydrogenase deficiency
Glycogen storage disease
Coffin-Lowry syndrome
Mental retardation
Spondyloepiphyseal dysplasia tarda
Paroxysmal nocturnal hemoglobinuria
Infantile spasm syndrome
Aicardi syndrome
Deafness, sensorineural
Simpson-Golabi-Behmel syndrome, type 2
Adrenal hypoplasia, congenital
Dosage-sensitive sex reversal
Deafness, congenital sensorineural
Retinitis pigmentosa
Wilson-Turner syndrome
Cone dystrophy
Aland island eye disease (ocular albinism)
Optic atrophy
Night blindness, congenital stationary, type 1
Erythroid-potentiating activity
Arthrogryposis multiplex congenita
Night blindness, congenital stationary, type 2
Brunner syndrome
Wiskott-Aldrich syndrome
Thrombocytopenia
Dent disease
Nephrolithiasis, type I
Hypophosphatemia, type III
Proteinuria
Anemia, sideroblastic/hypochromic
Cerebellar ataxia
Renal cell carcinoma, papillary
Diabetes mellitus, insulin-dependent
Sutherland-Haan syndrome
Cognitive function, social
Mental retardation, nonspecific
Menkes disease
Occipital horn syndrome
Cutis laxa, neonatal
FG syndrome
Immunodeficiency, moderate and severe
Miles-Carpenter syndrome
Charcot-Marie-Tooth neuropathy, dominant
Mental retardation
X-inactivation center
Premature ovarian failure
Arts syndrome
Cleft palate and/or ankyloglossia
Megalocornea
Epilepsy (Juberg-Hellman syndrome)
Pelizaeus-Merzbacher disease
Spastic paraplegia
Alport syndrome
Cowchock syndrome
Hypertrichosis, congenital generalized
Ptosis, hereditary congenital
Apoptosis inhibitor
Panhypopituitarism
Thoracoabdominal syndrome
Simpson-Golabi-Behmel syndrome, type 1
Split hand/foot malformation, type 2
Hypoparathyroidism
Mental retardation, Shashi type
Lesch-Nyhan syndrome
HPRT-related gout
Lowe syndrome
Borjeson-Forssman-Lehmann syndrome
Testicular germ cell tumor
Hemophilia B
Warfarin sensitivity
Osseous dysplasia (male lethal), digital
Adrenoleukodystrophy
Adrenomyeloneuropathy
Colorblindness, blue monochromatic
Cardiac valvular dysplasia
Emery-Dreifuss muscular dystrophy
Heterotopia, periventricular
Favism
Hemolytic anemia
Colorblindness, green cone pigment
Incontinentia pigmenti, type II
Hydrocephalus
MASA syndrome
Spastic paraplegia
Rett syndrome
Mature T-cell proliferation
Myopia (Bornholm eye disease)
Mental retardation with psychosis
Endocardial fibroelastosis
Hodgkin disease susceptibility, pseudoautosomal
Ichthyosis
Microphthalmia, dermal aplasia, and sclerocornea
Episodic muscle weakness
Mental retardation
Ocular albinism and sensorineural deafness
Amelogenesis imperfecta
Charcot-Marie-Tooth disease, recessive
Keratosis follicularis spinulosa decalvans
Hypophosphatemia, hereditary
Partington syndrome
Retinoschisis
Gonadal dysgenesis, XY female type
Mental retardation, non-dysmorphic
Agammaglobulinemia, type 2
Craniofrontonasal dysplasia
Opitz G syndrome, type I
Pigment disorder, reticulate
Melanoma
Duchenne muscular dystrophy
Becker muscular dystrophy
Cardiomyopathy, dilated
Chronic granulomatous disease
Snyder-Robinson mental retardation
Norrie disease
Exudative vitreoretinopathy
Coats disease
Renpenning syndrome
Retinitis pigmentosa, recessive
Mental retardation, nonspecific and syndromic
Dyserythropoietic anemia with thrombocytopenia
Chondrodysplasia punctata, dominant
Autoimmunity-immunodeficiency syndrome
Renal cell carcinoma, papillary
Faciogenital dysplasia (Aarskog-Scott syndrome)
Choreoathetosis with mental retardation
Sarcoma, synovial
Prieto syndrome
Spinal muscular atrophy, lethal infantile
Migraine, familial typical
Androgen insensitivity
Spinal and bulbar muscular atrophy
Prostate cancer
Perineal hypospadias
Breast cancer, male, with Reifenstein syndrome
Ectodermal dysplasia, anhidrotic
Alpha-thalassemia/mental retardation
Juberg-Marsidi syndrome
Sutherland-Haan syndrome
Smith-Fineman-Myers syndrome
Hemolytic anemia
Myoglobinuria/hemolysis
Wieacker-Wolff syndrome
Torsion dystonia-parkinsonism, Filipino type
Leukemia, myeloid/lymphoid or mixed-lineage
Anemia, sideroblastic, with ataxia
Allan-Herndon syndrome
Deafness
Choroideremia
Agammaglobulinemia
Fabry disease
Mohr-Tranebjaerg syndrom
Jensen syndrome
Lissencephaly
Bazex syndrome
Mental retardation with growth hormone deficiency
Mental retardation, South African type
Lymphoproliferative syndrome
X inactivation, familial skewed
Pettigrew syndrome
Gustavson mental retardation syndrome
Immunodeficiency, with hyper-IgM
Retinitis pigmentosa
SRY (sex determining region Y)
Wood neuroimmunologic syndrome
Heterotaxy, visceral
Albinism-deafness syndrome
Cone dystrophy, progressive
Prostate cancer susceptibility
Fragile X mental retardation
Epidermolysis bullosa, macular type
Diabetes insipidus, nephrogenic
Cancer/testis antigen
Dyskeratosis
Hemophilia A
Hunter syndrome
Mucopolysaccharidosis
Intestinal pseudoobstruction, neuronal
Melanoma antigens
Mental retardation-skeletal dysplasia
Myotubular myopathy
Otopalatodigital syndrome, type I
Colorblindness, red cone pigment
Goeminne TKCR syndrome
Waisman parkinsonism-mental retardation
Barth syndrome
Cardiomyopathy, dilated
Noncompaction of left ventricular myocardium
Von Hippel-Lindau binding protein

functional genomics



biological system

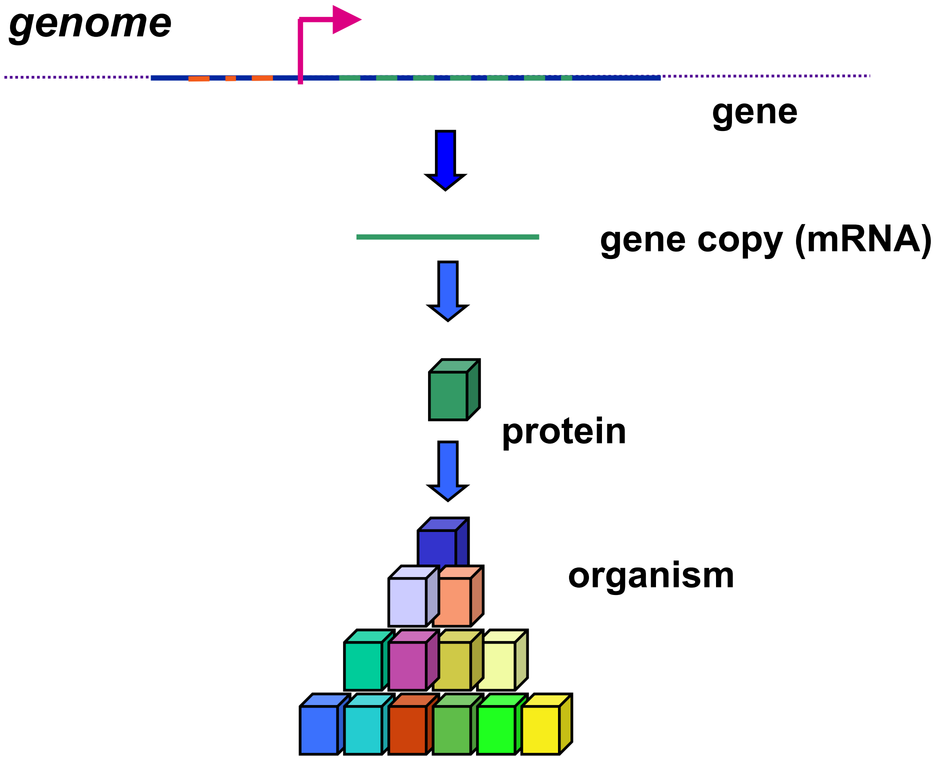
screening
genome-wide

modeling
computational biology

generation of hypoteses
biological roles of genes and proteins

experimental analysis
functions of genes and proteins

from genome to organism



genome

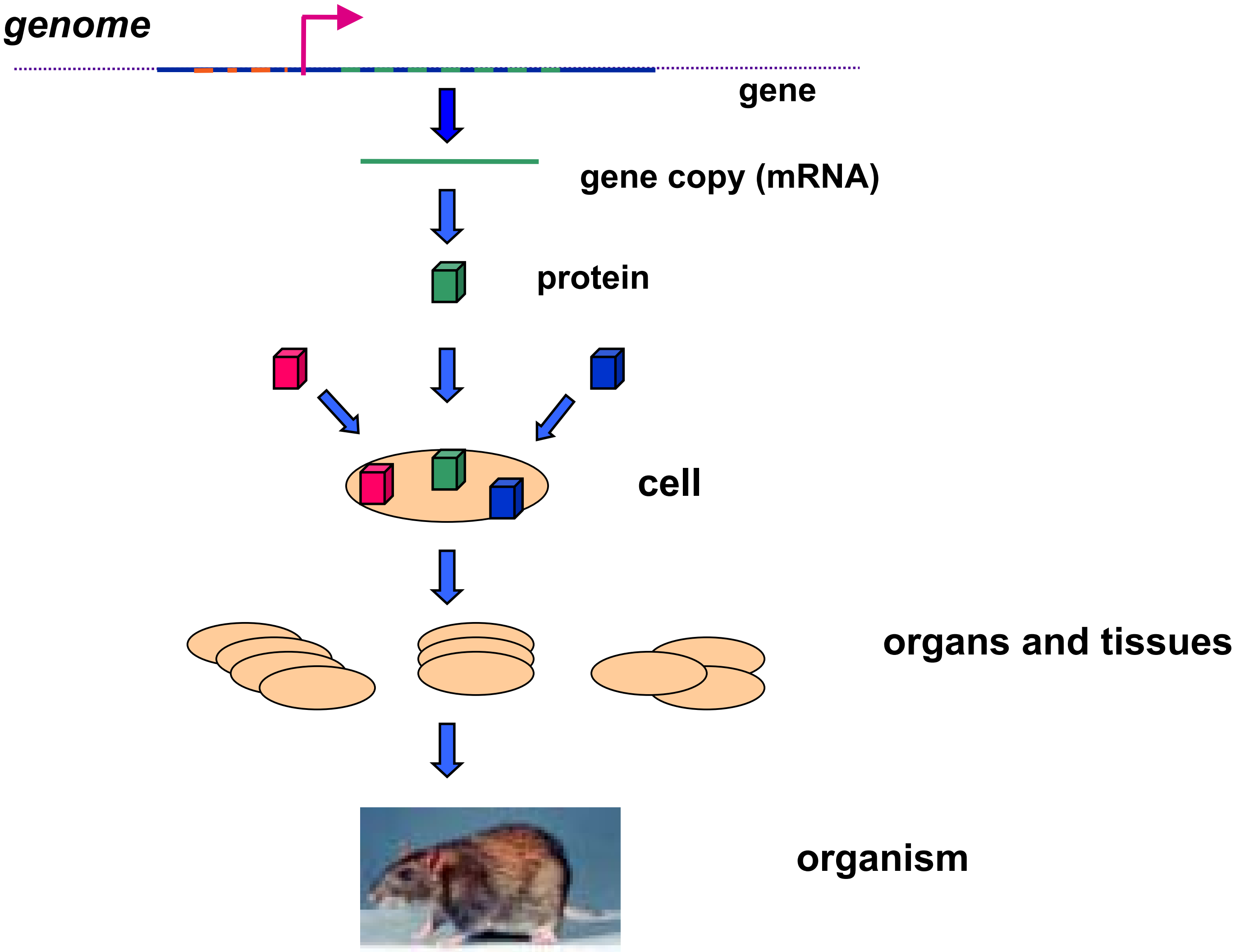
gene

gene copy (mRNA)

protein

organism

from genome to organism



genome

gene

gene copy (mRNA)

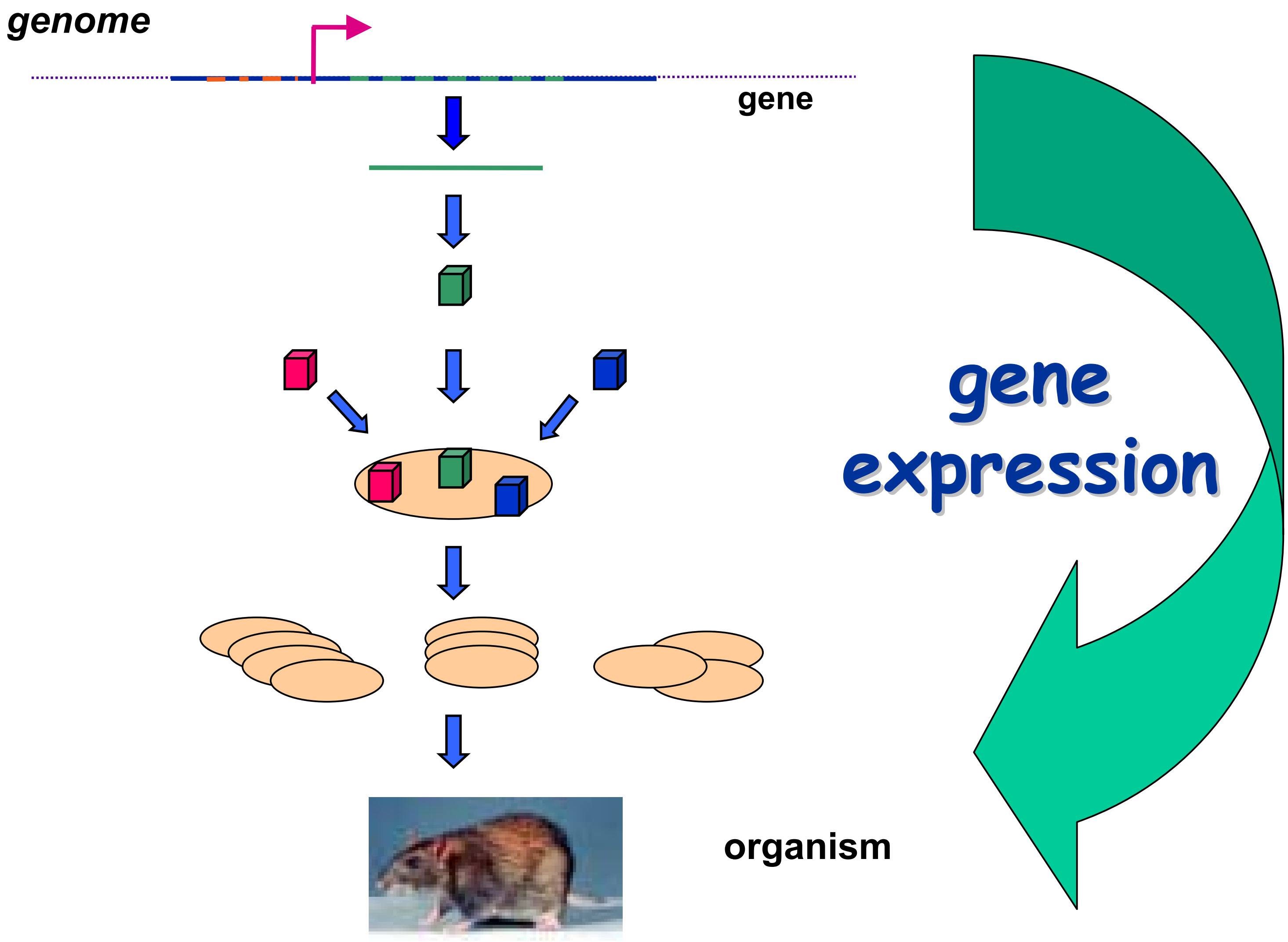
protein

cell

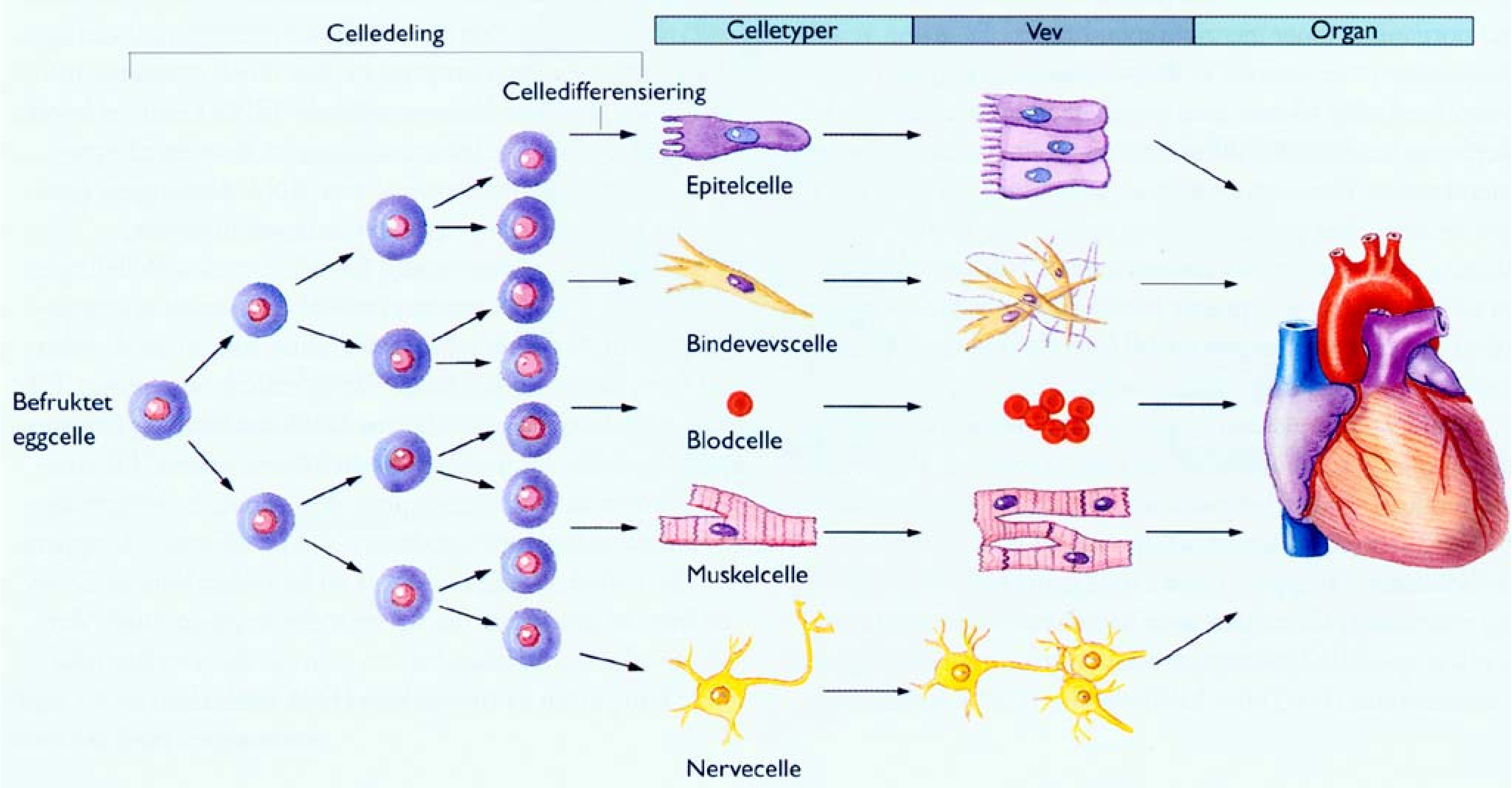
organs and tissues

organism

from genome to organism

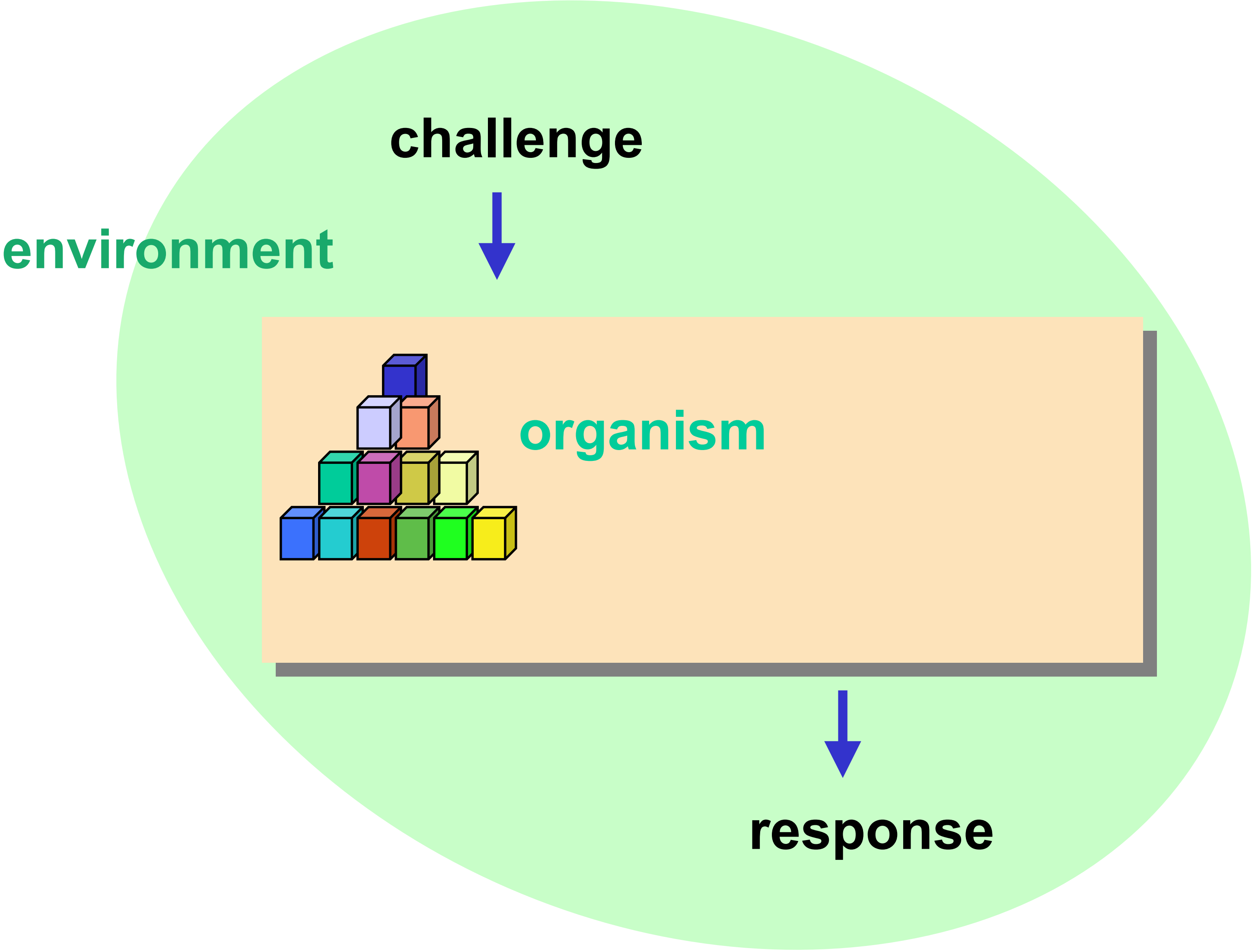


all cell types contain the same genome
..but differ in gene expression patterns....



gene expression determines what you are....

to live is
to interact with the environment



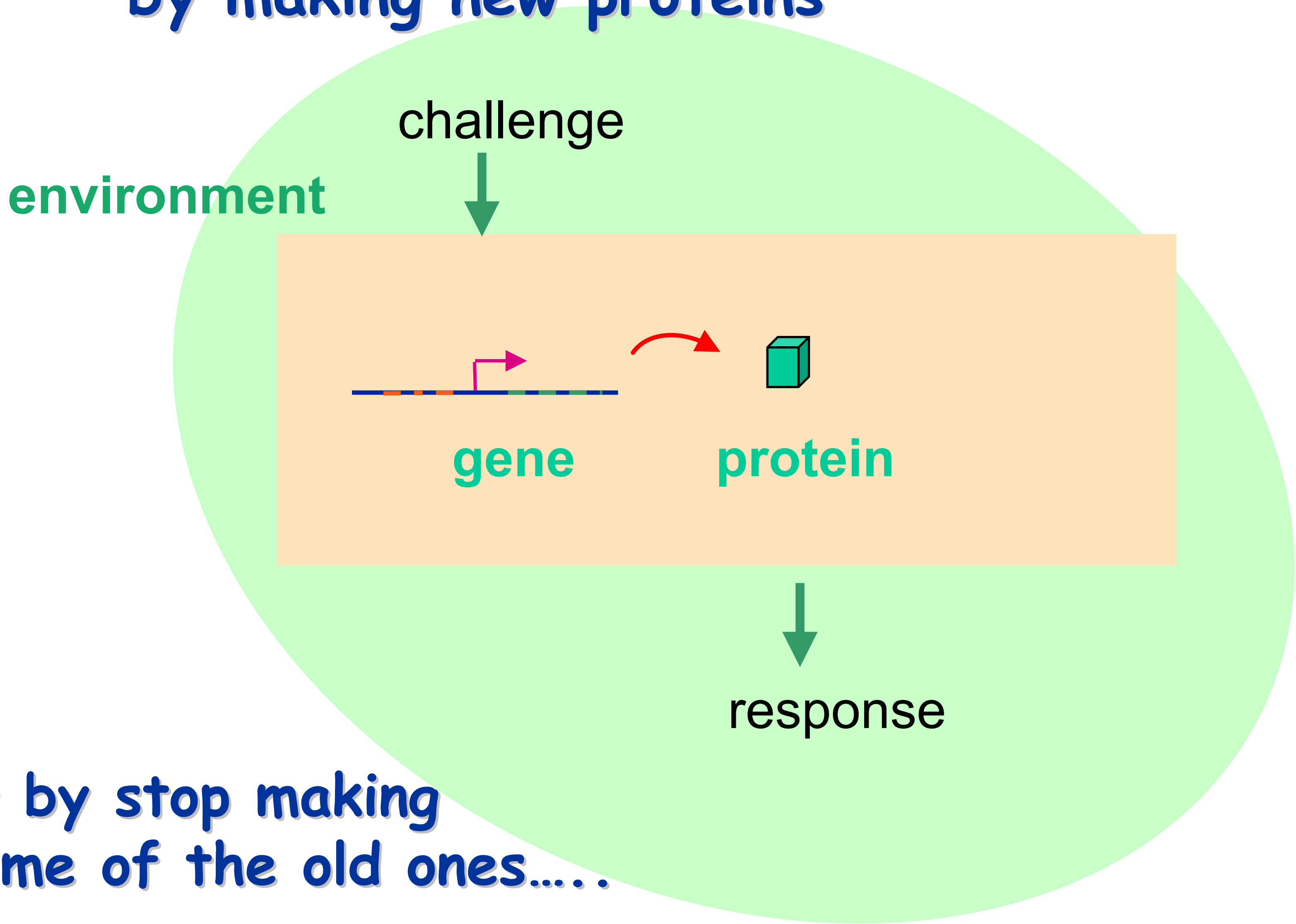
challenge

environment

organism

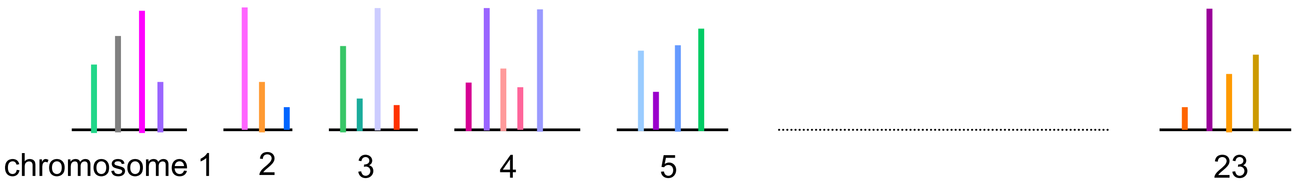
response

the organism responds by making new proteins



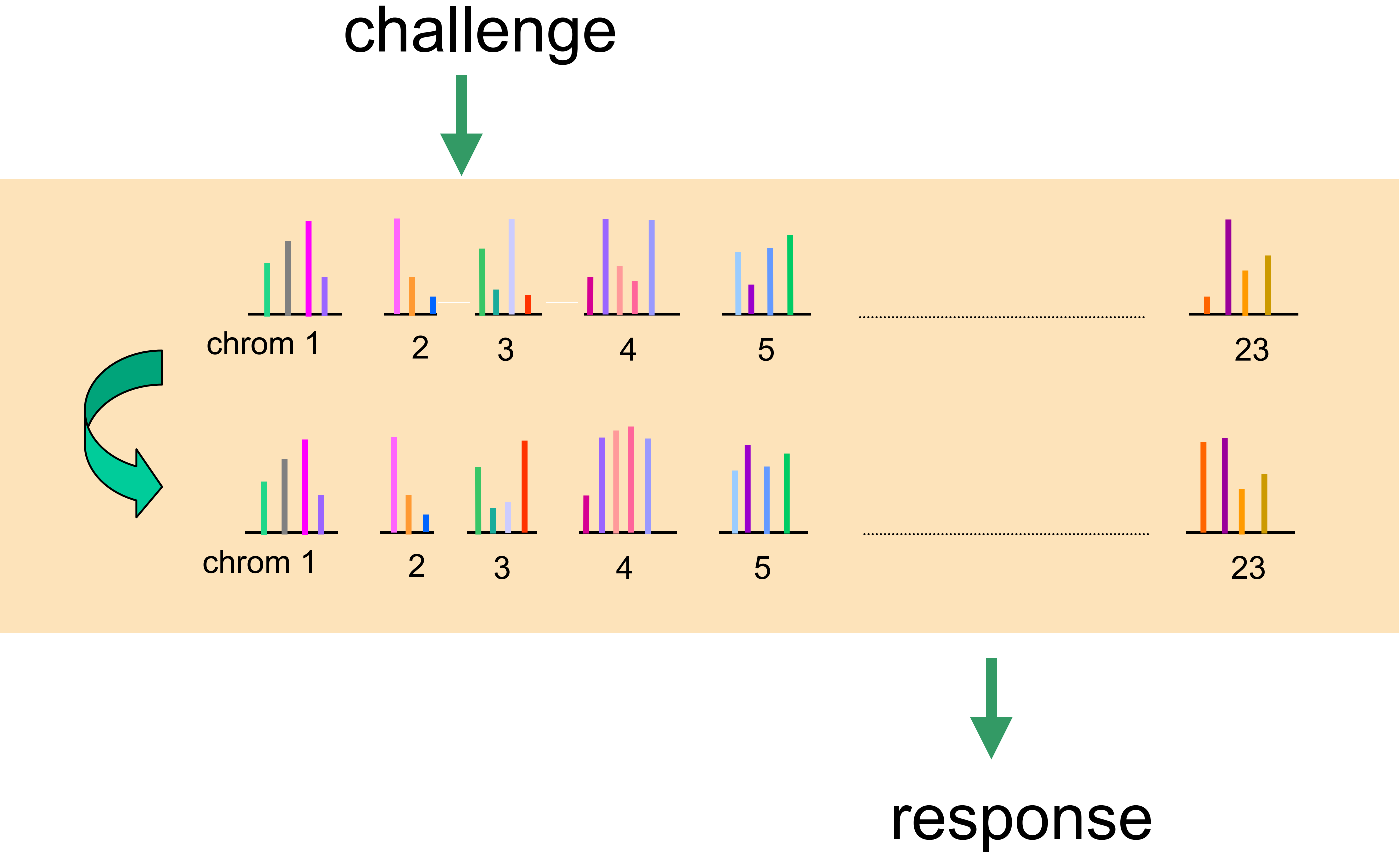
or by stop making some of the old ones…..

we can learn more by measuring expression of **all** genes

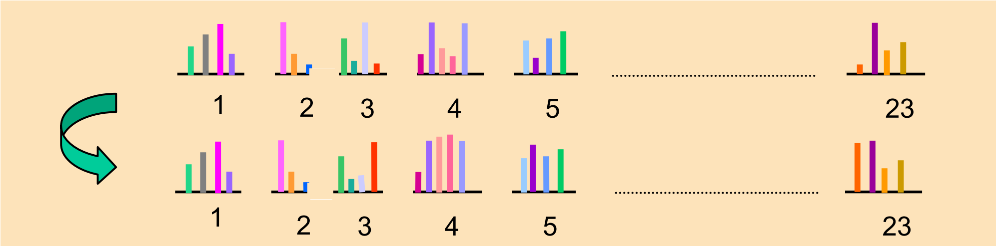


gene expression depends on cell type and cell state

by measuring changes in gene expression we can discover genes participating in a given biological response

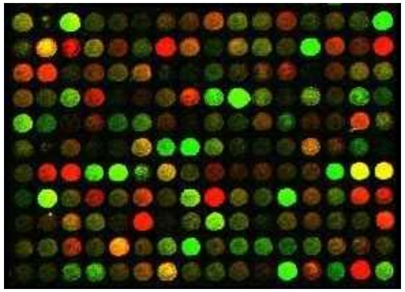


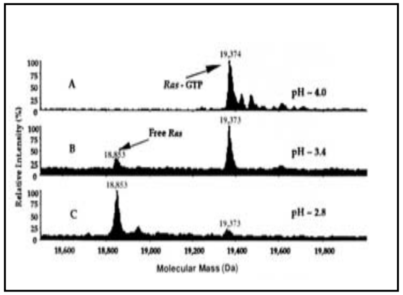
measure thousands of genes and proteins in high throughput analyses



mRNA profiling

protein profiling

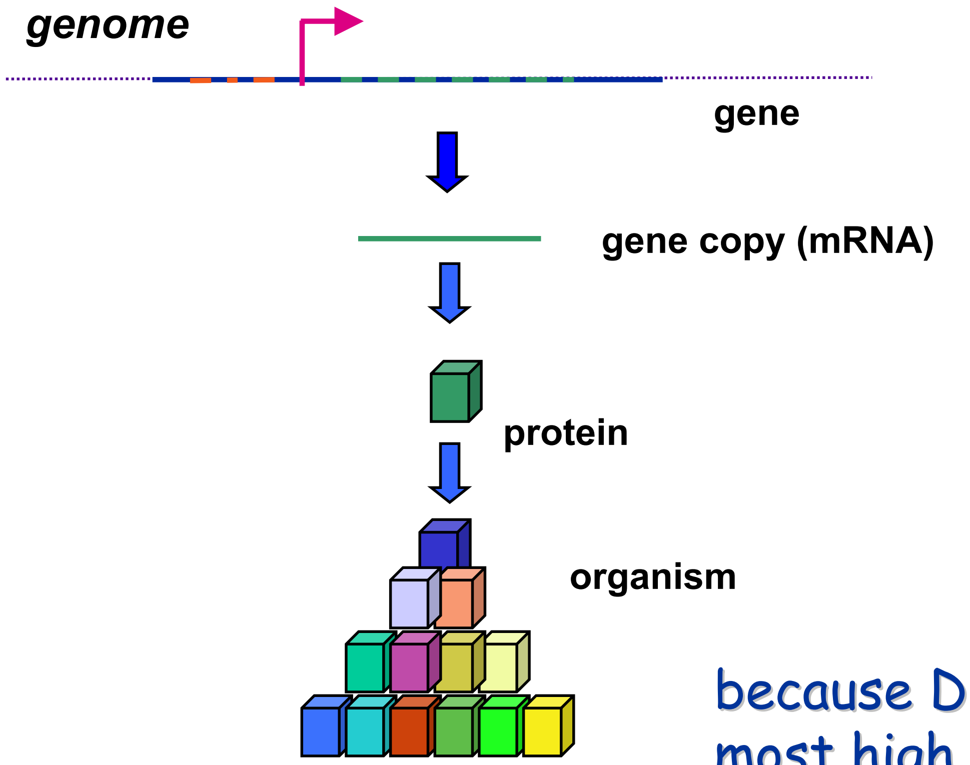




microarray

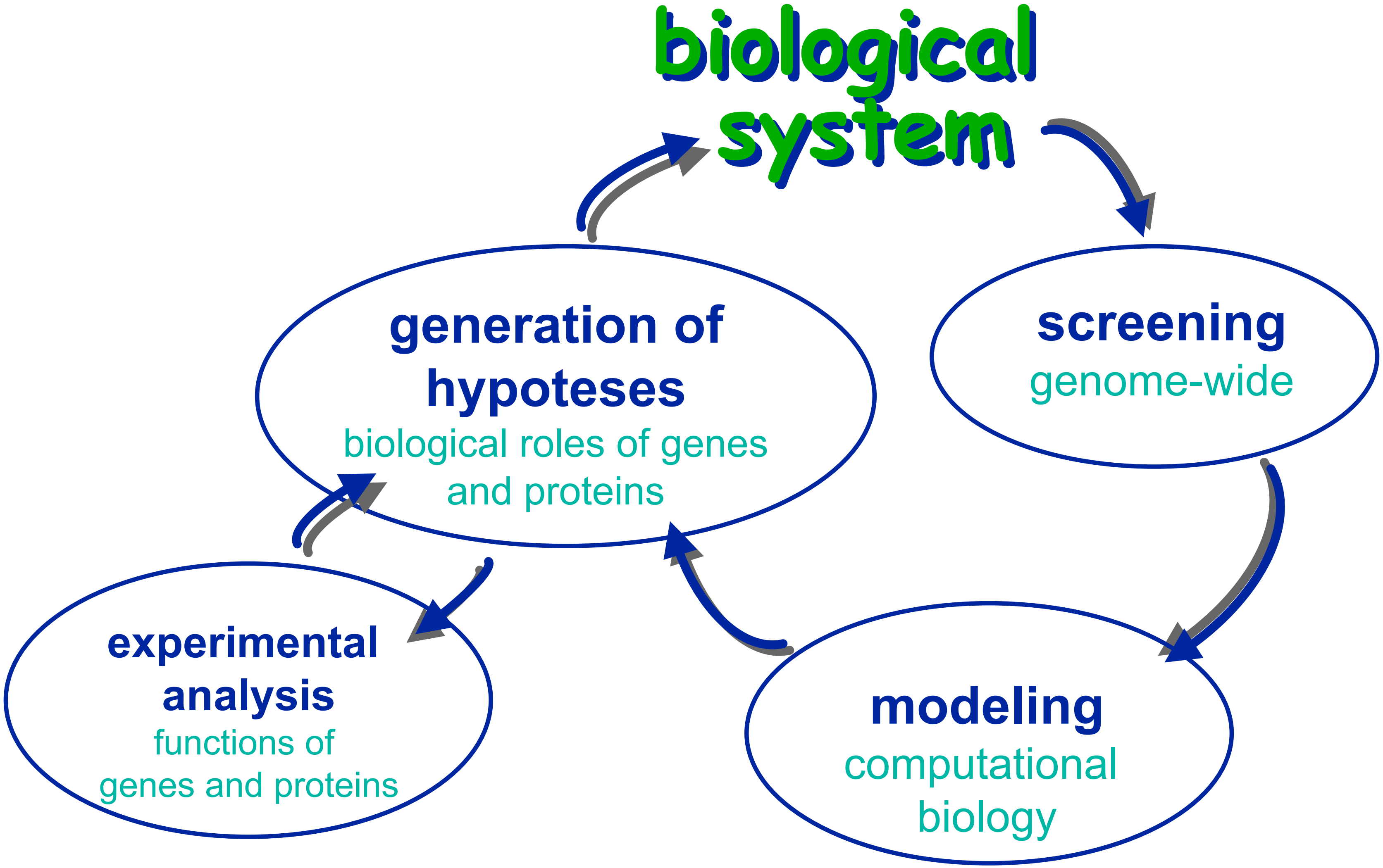
2D gel electrophoresis
mass spectrometry

why measure mRNA?



because DNA microarray is the most high throughput method that can measure
gene expression with high sensitivity and specificity

functional genomics



biological system

generation of hypoteses

biological roles of genes and proteins

screening

genome-wide

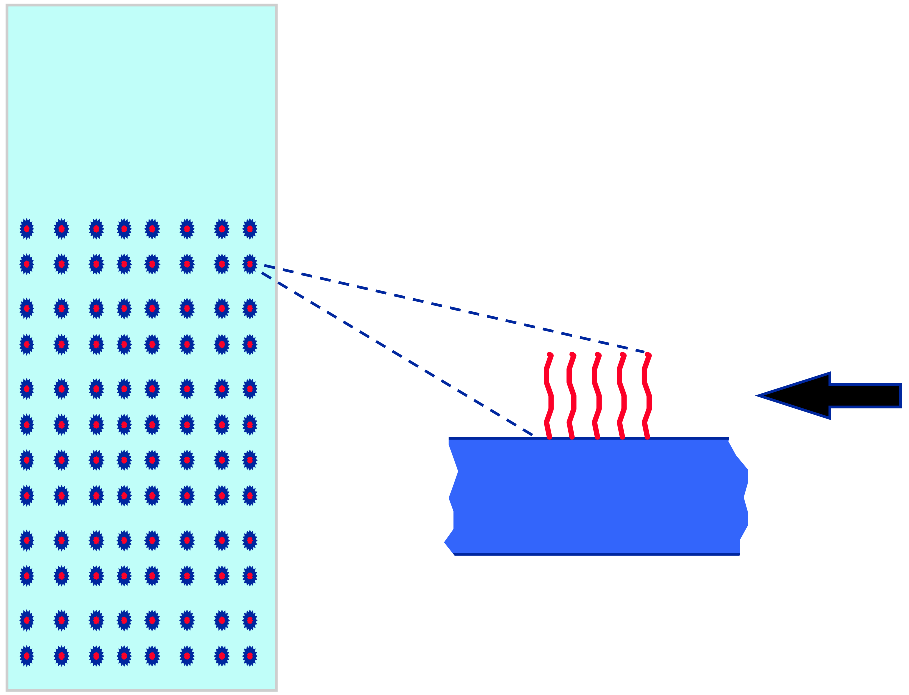
modeling

computational biology

experimental analysis

functions of genes and proteins

DNA microarray



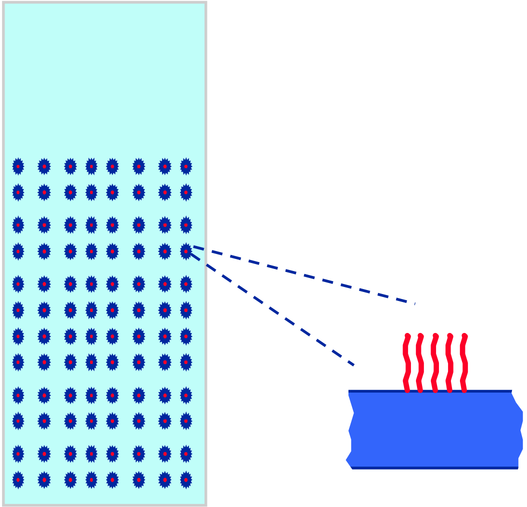
DNA molecules

= *specific probe*

microscopic slide

5.000 - 80.000 probes pr. array

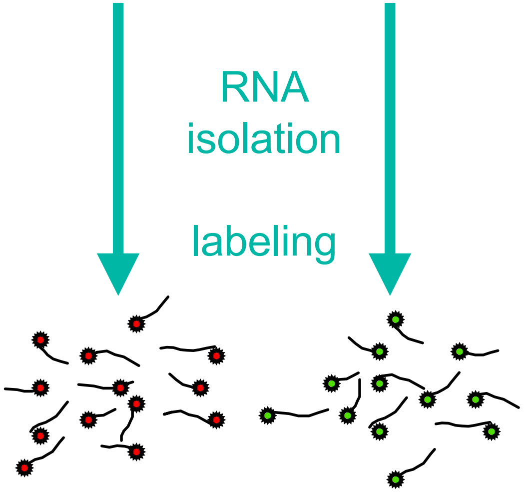
microarray formats



- cDNA (500-1500 bp)
- long oligonuleotides (40-70-mers)
- short oligonucleotides (20-25-mers)

microarray analysis

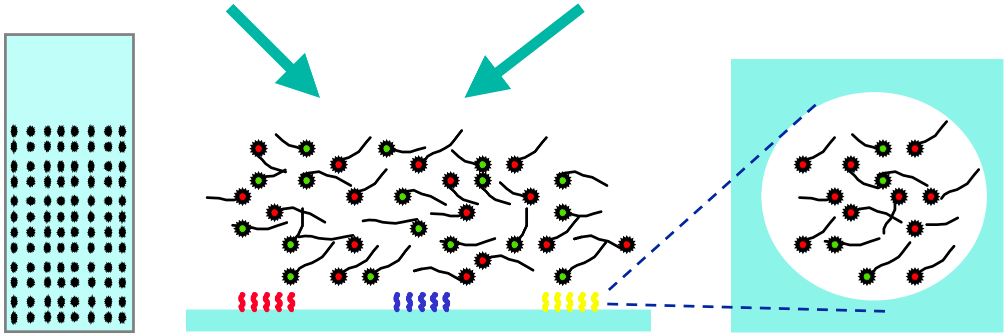
1 sample / control



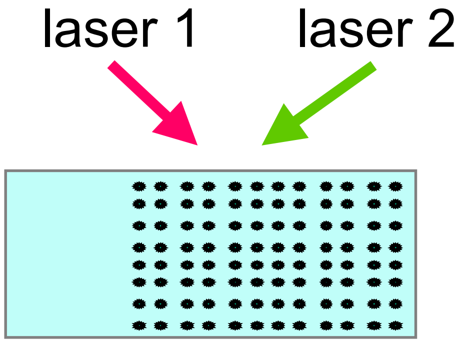
RNA
isolation

labeling

2 hybridization



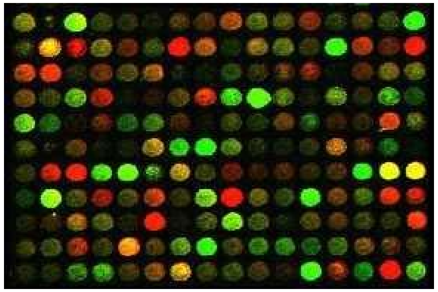
3 scanning



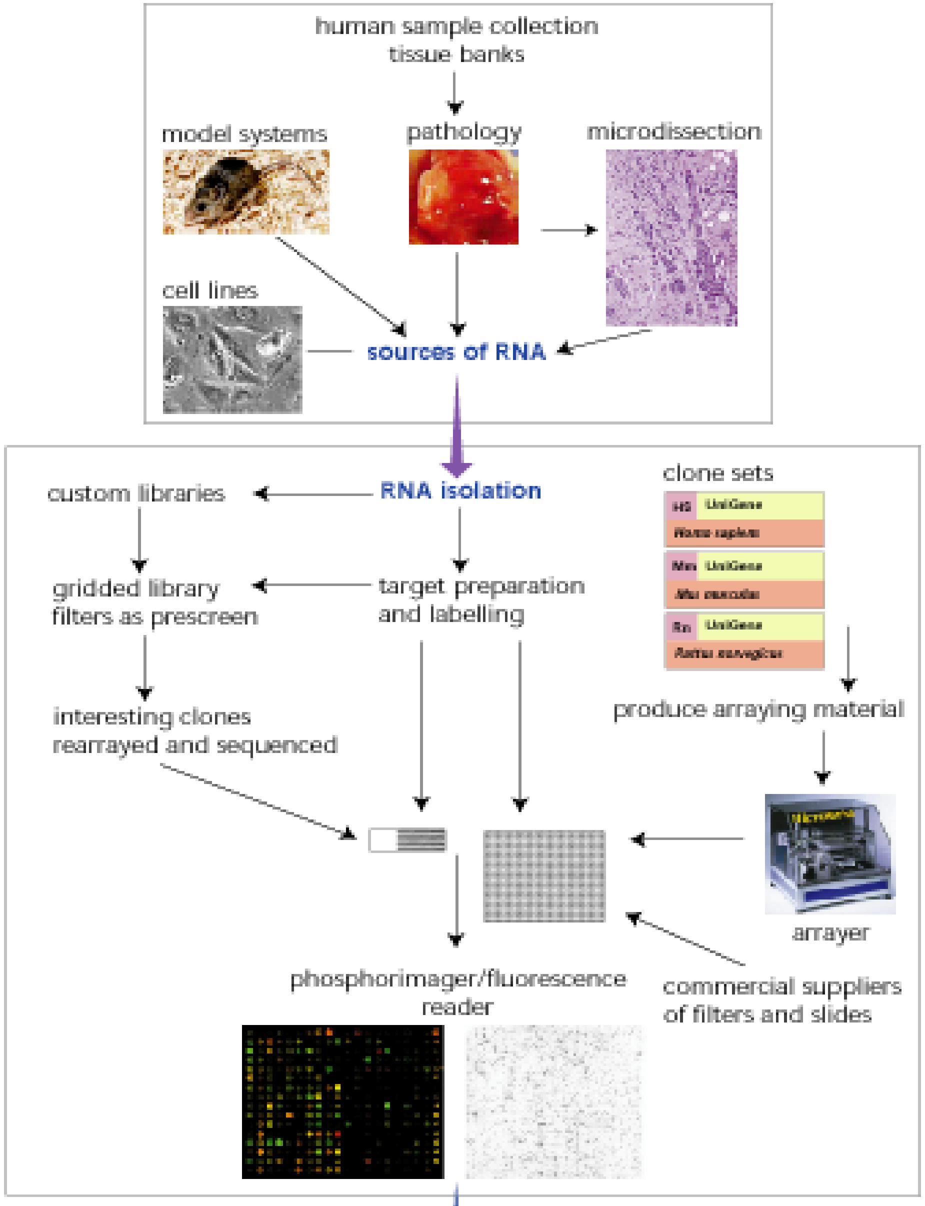
laser 1

laser 2

red = ”up”
green = ”down”

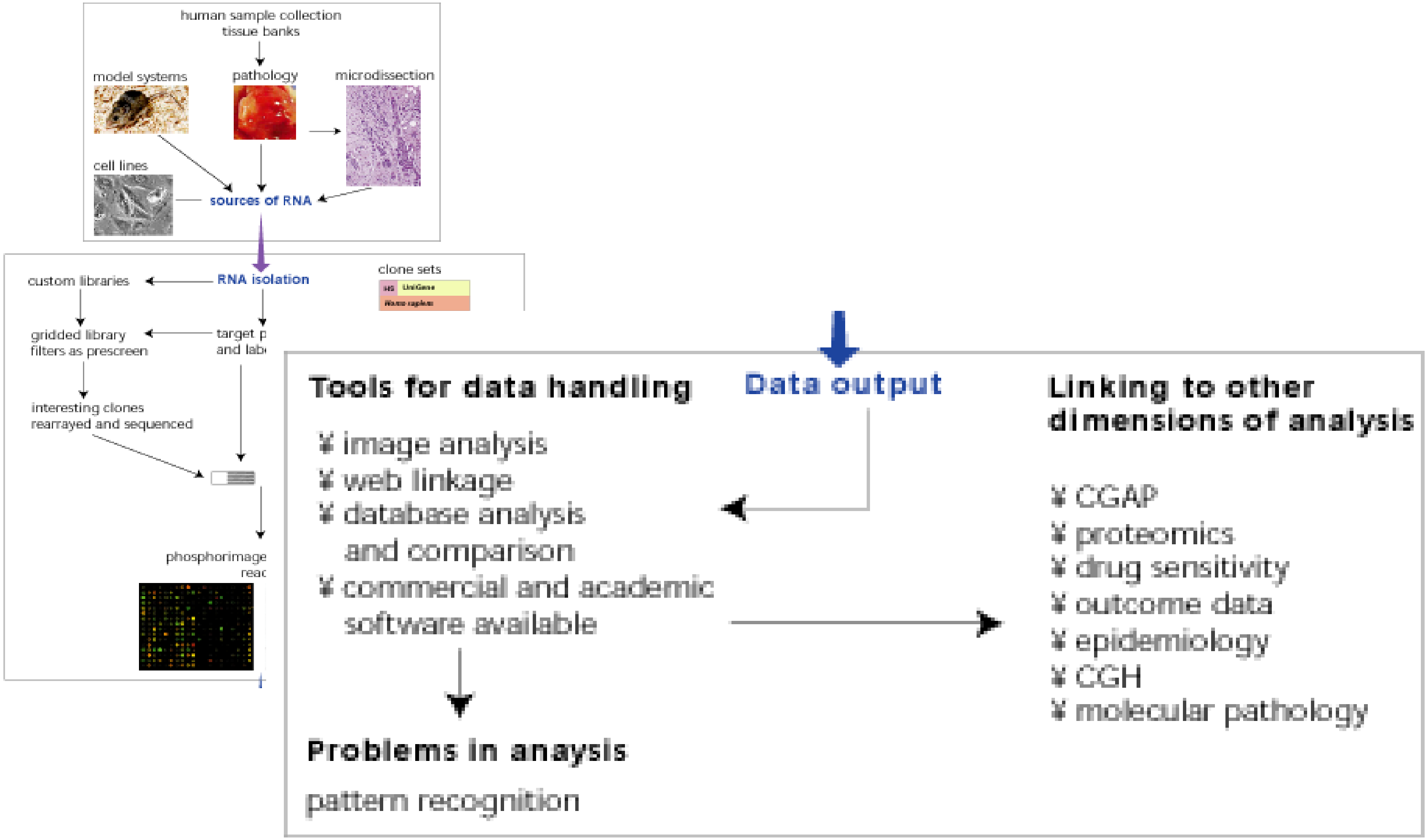


microarray analysis



Bowtell, Nature Genetics, Supplement, 21:25, 1999

microarray analysis



Data output

Tools for data handling

¥ image analysis
¥ web linkage
¥ database analysis
and comparison
¥ commercial and academic
software available

Problems in anaysis

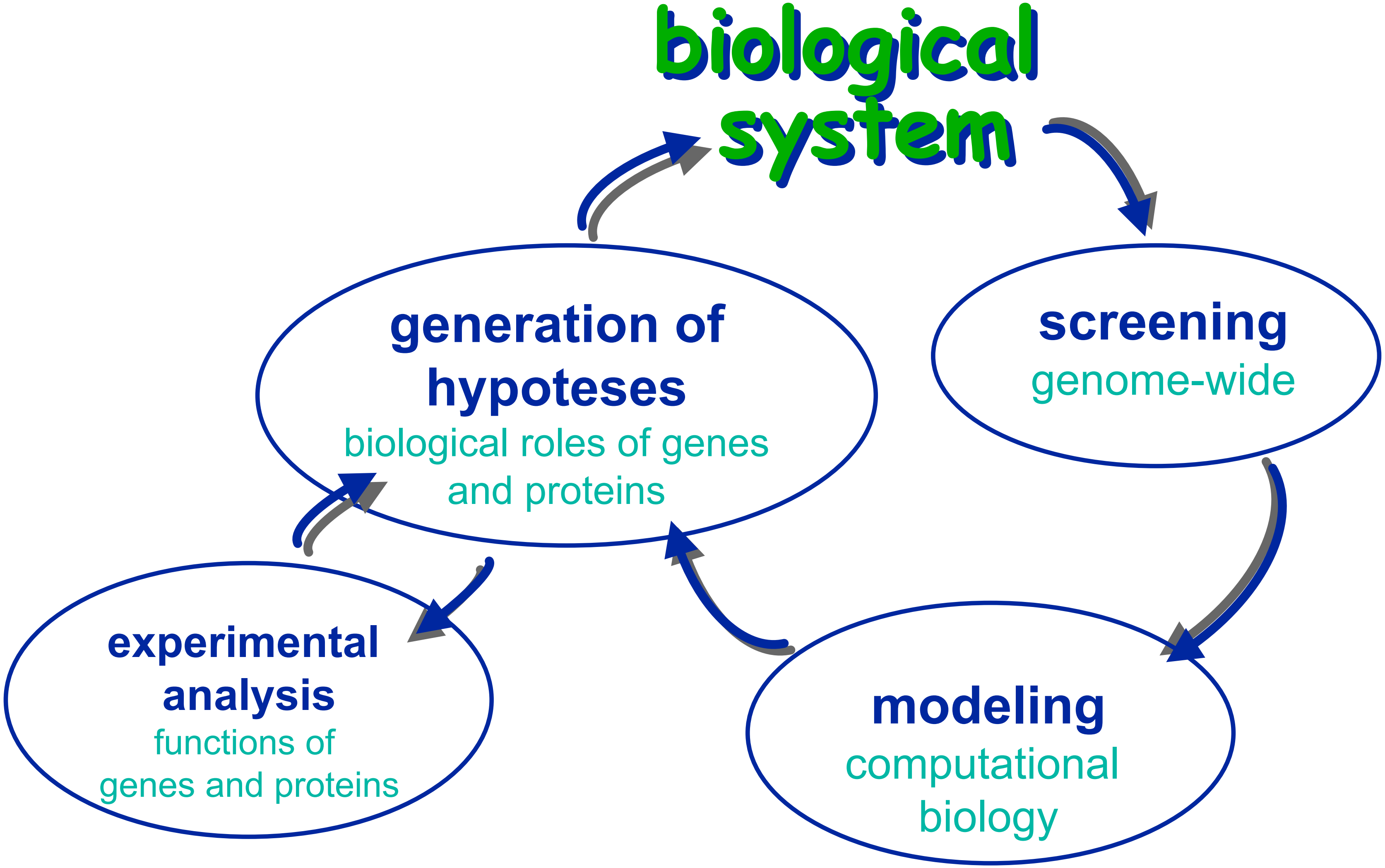
pattern recognition

Linking to other dimensions of analysis

¥ CGAP
¥ proteomics
¥ drug sensitivity
¥ outcome data
¥ epidemiology
¥ CGH
¥ molecular pathology

Bowtell, Nature Genetics,
Supplement, 21:25, 1999

functional genomics



biological system

generation of hypoteses

biological roles of genes and proteins

screening

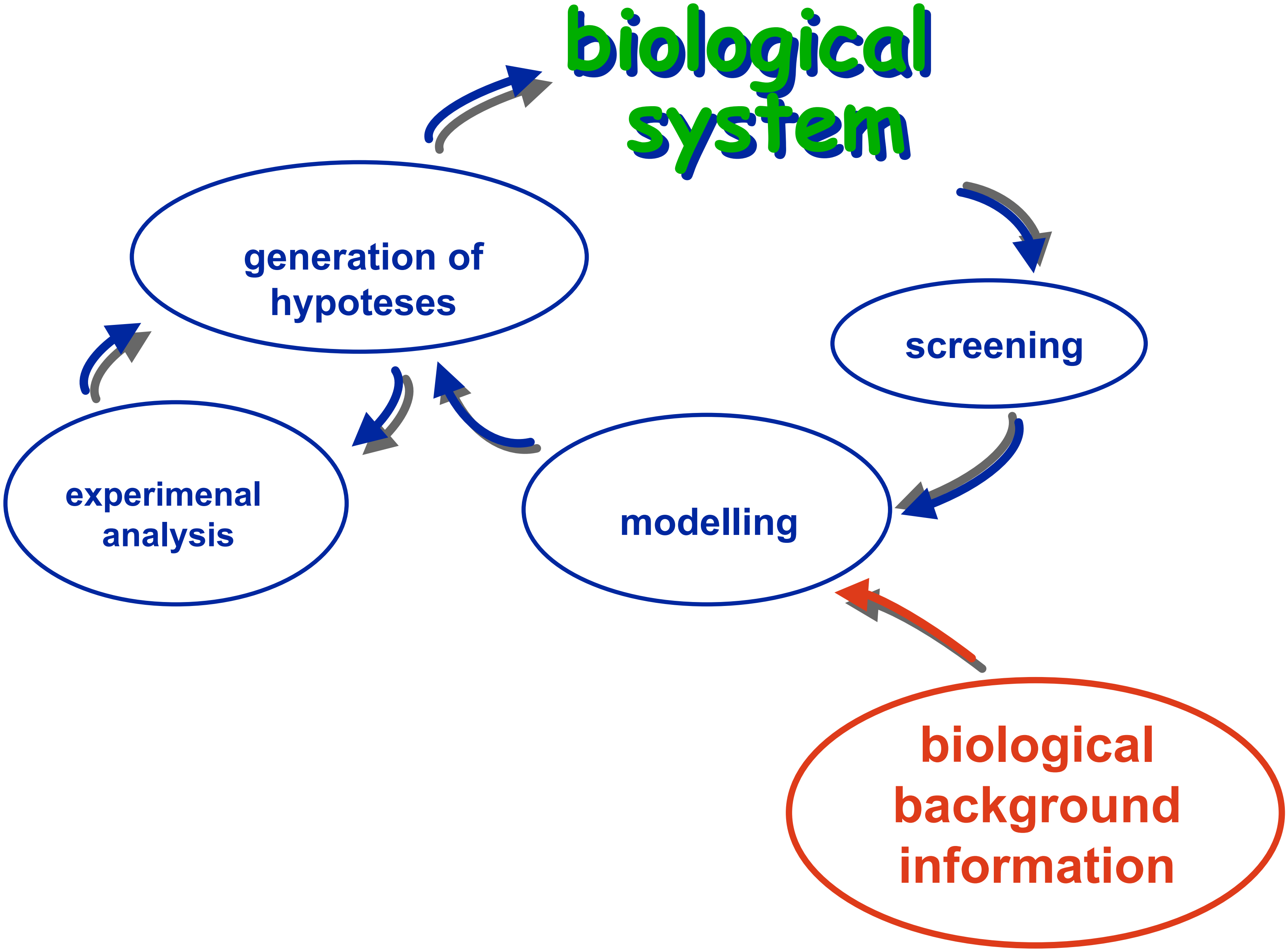
genome-wide

modeling

computational biology

experimental analysis

functions of genes and proteins



biological system

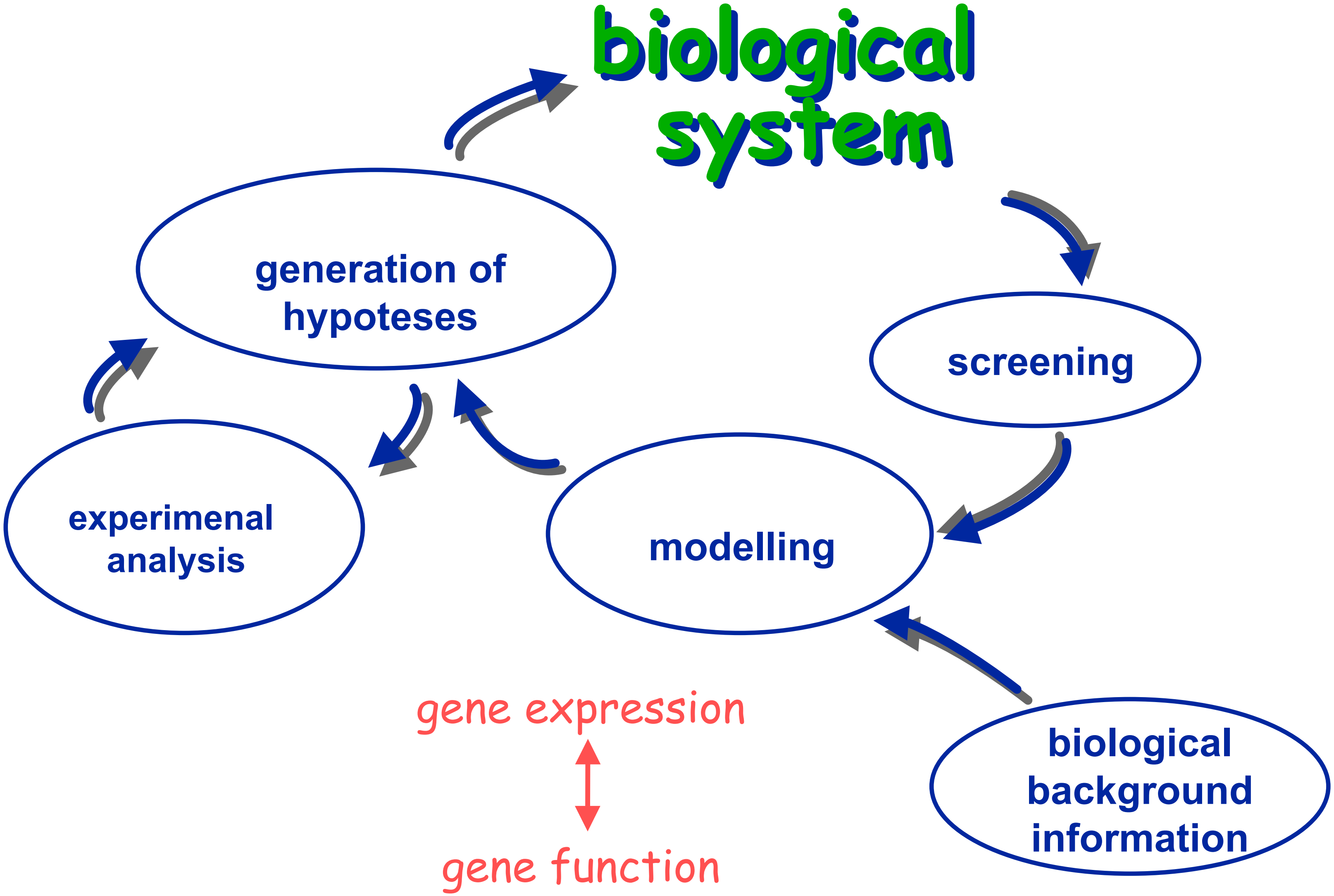
generation of hypoteses

screening

experimental analysis

modelling

biological background information



biological system

generation of hypoteses

screening

modelling

experimenal analysis

biological background information

gene expression

↕

gene function

functional classification of genes from time profiles

Astrid Lægreid[1] and Jan Komorowski[2]

[1]Department of Cancer Research and Molecular Medicine

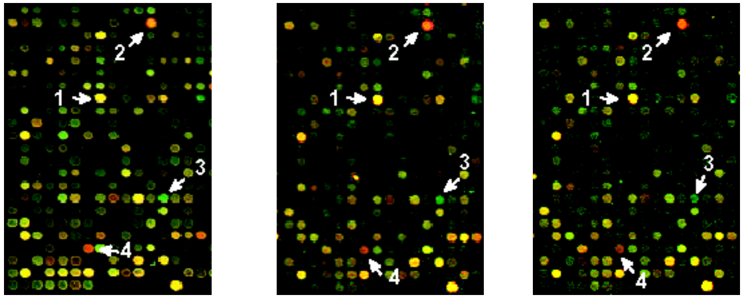
Norwegian University of Science and Technology

[2]The Linnaeus Centre for Bioinformatics,

Uppsala

The Transcriptional Program in the Response of Human Fibroblasts to Serum

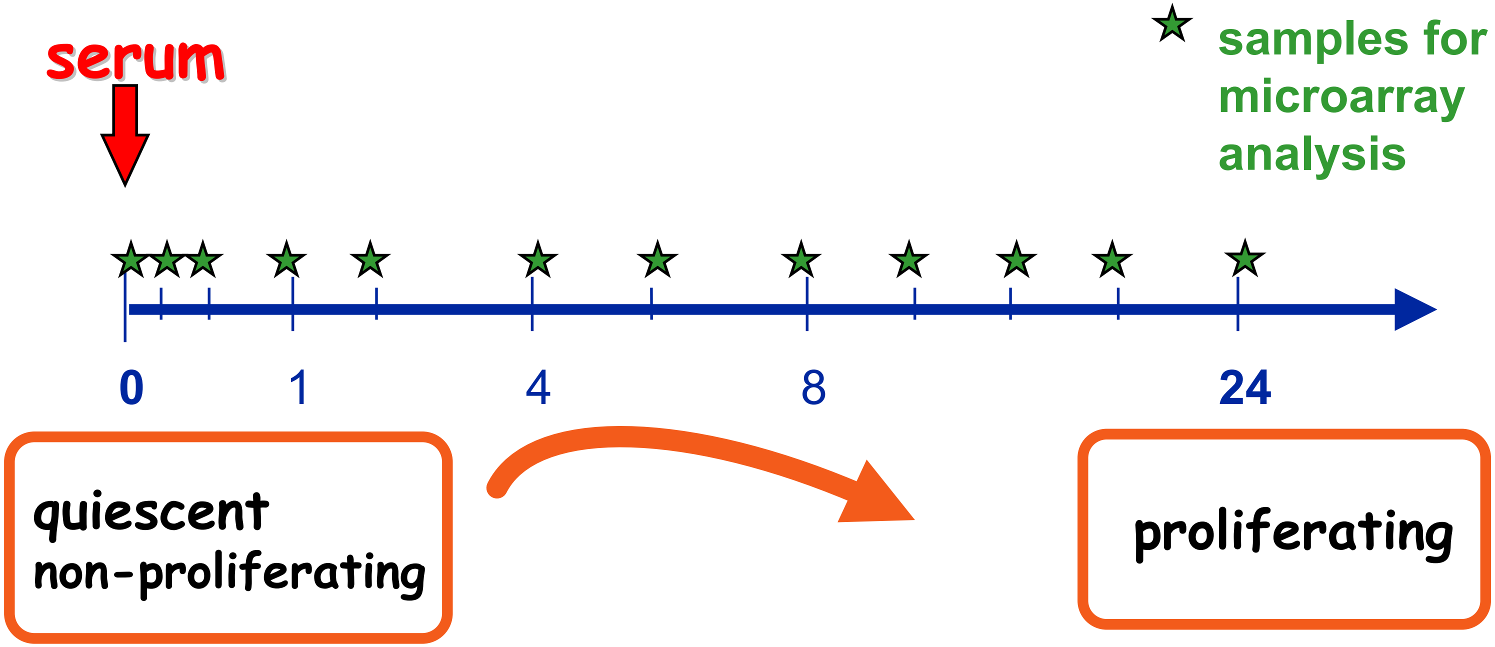
Iyer et al, Science, 283: 83, 1999



8 hours serum treatment

1, protein disulfide isomerase-related protein
2, IL-8 precursor
3, EST AA057170
4, vascular endothelial growth factor

fibroblast - 24 h serum response


serum

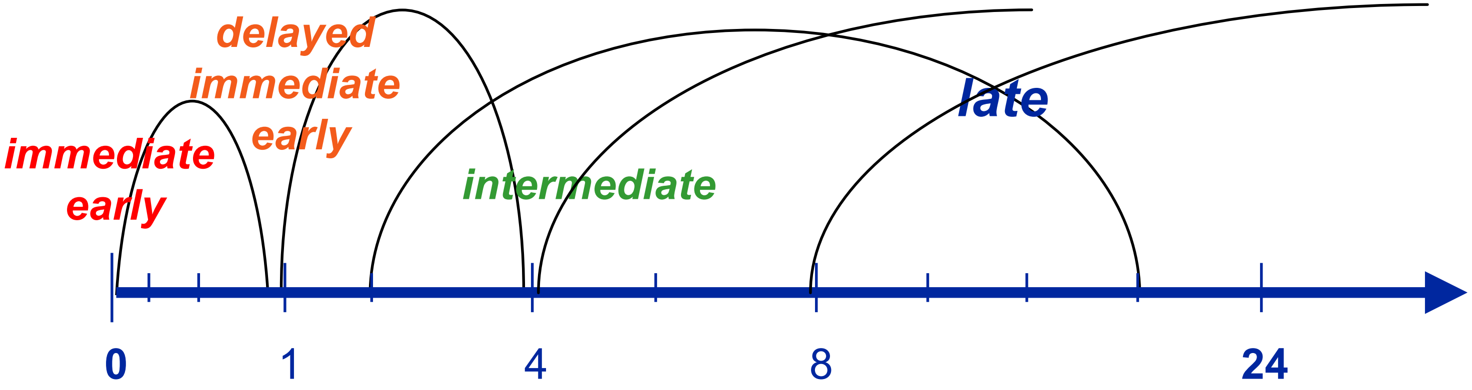
★ samples for microarray analysis

0
1
4
8
24

quiescent non-proliferating

proliferating

dynamic processes


immediate early

delayed immediate early

intermediate

late

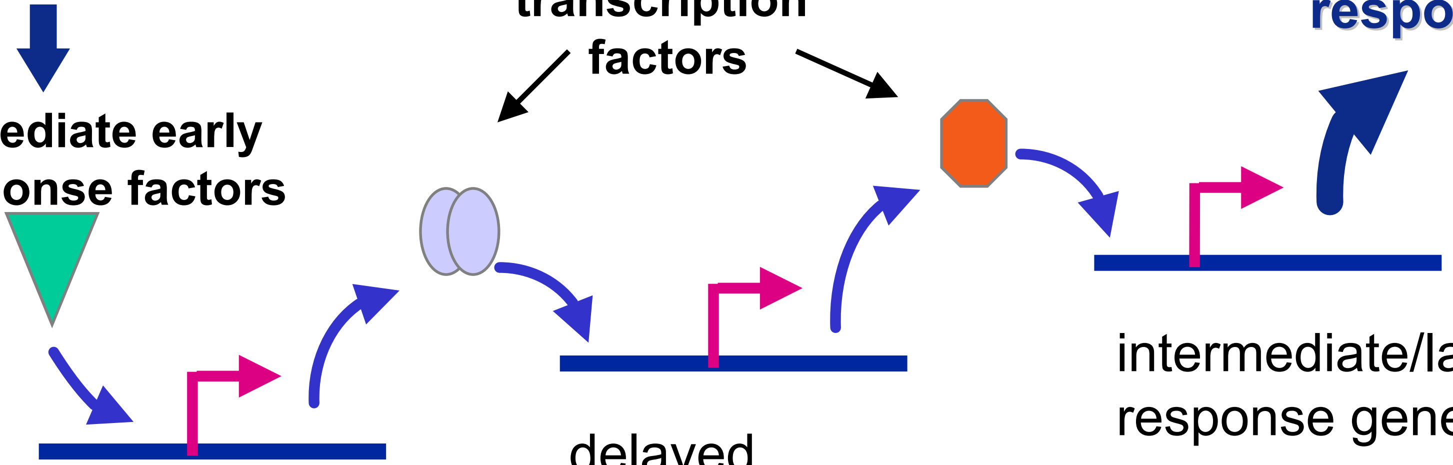
0 1 4 8 24

primary **secondary** **tertiary**

quiescent
non-proliferating

proliferating

molecular mechanisms of transcriptional response



serum

= signal

immediate early response factors

secondary transcription factors

effectors

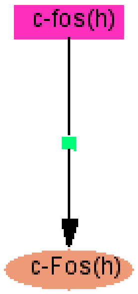
= cellular response

immediate early response genes

delayed immediate early response genes

intermediate/late response genes

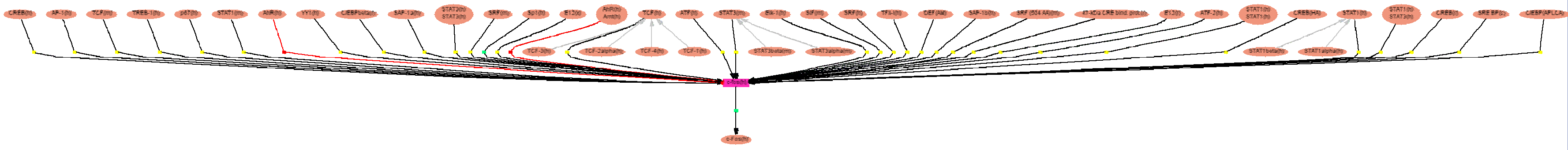
fos - immediate early transcription factor



Transpath; biobase.de

fos - immediate early transcription factor

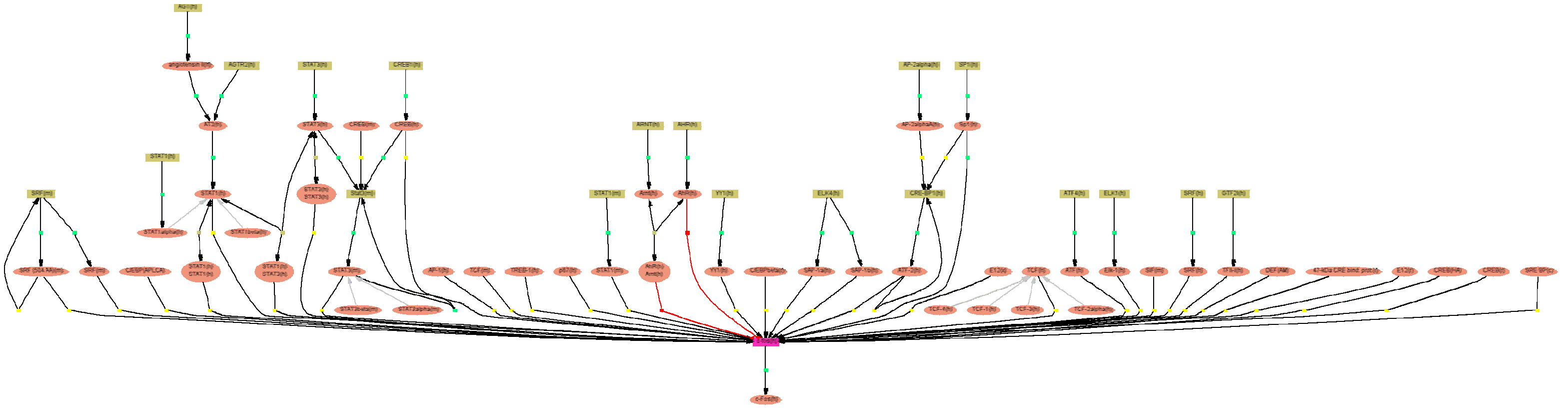
upstream factors



Transpath; biobase.de

fos - immediate early transcription factor

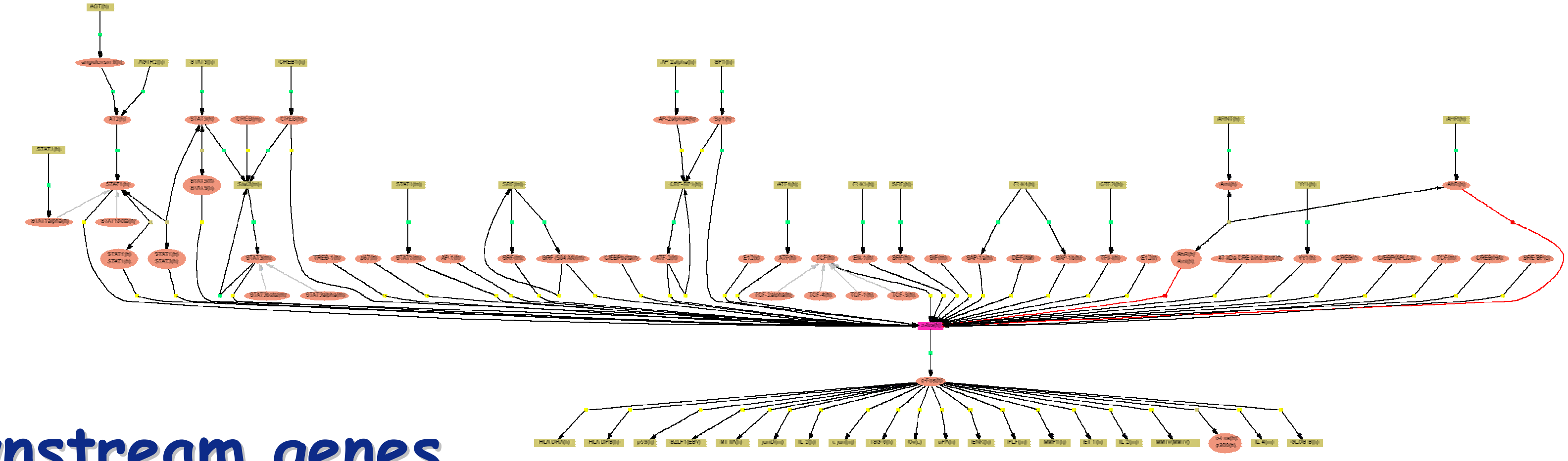
upstream factors



Transpath; biobase.de

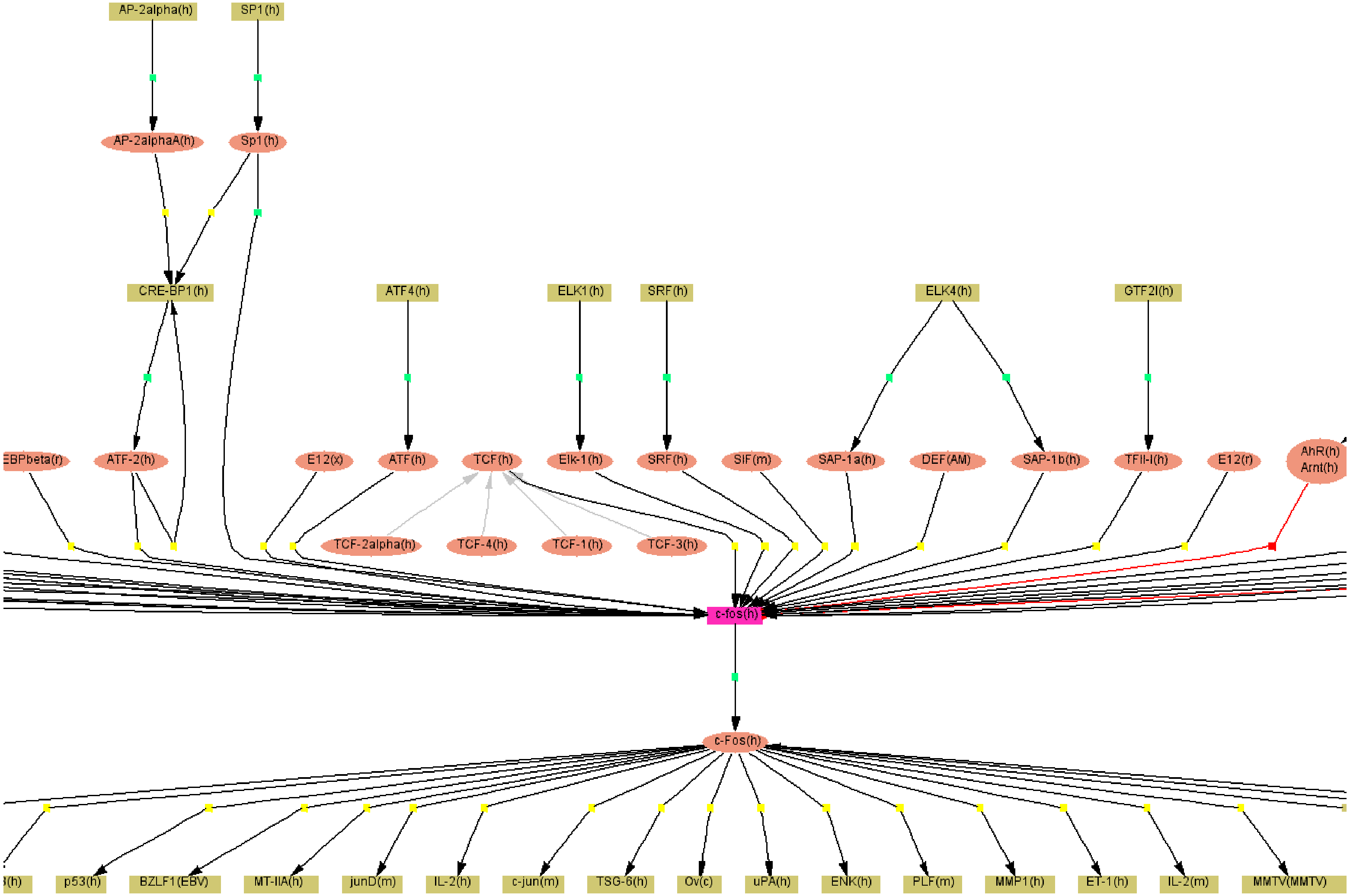
fos - immediate early transcription factor

upstream factors



+ downstream genes

Transpath; biobase.de


AP-2alpha(h)
SP1(h)
AP-2alphaA(h)
Sp1(h)
CRE-BP1(h)
ATF4(h)
ELK1(h)
SRF(h)
ELK4(h)
GTF2I(h)
EBPbeta(r)
ATF-2(h)
E12(x)
ATF(h)
TCF(h)
Elk-1(h)
SRF(h)
SIF(m)
SAP-1a(h)
DEF(AM)
SAP-1b(h)
TFII-I(h)
E12(r)
AhR(h)
Arnt(h)
TCF-2alpha(h)
TCF-4(h)
TCF-1(h)
TCF-3(h)
c-fos(h)
c-Fos(h)
3(h)
p53(h)
BZLF1(EBV)
MT-IIA(h)
junD(m)
IL-2(h)
c-jun(m)
TSG-6(h)
Ov(c)
uPA(h)
ENK(h)
PLF(m)
MMP1(h)
ET-1(h)
IL-2(m)
MMTV(MMTV)

co-regulation of genes

coding for proteins in a network

in fibroblast serum-response

pro-endothelin → active endothelin → inactive endothelin

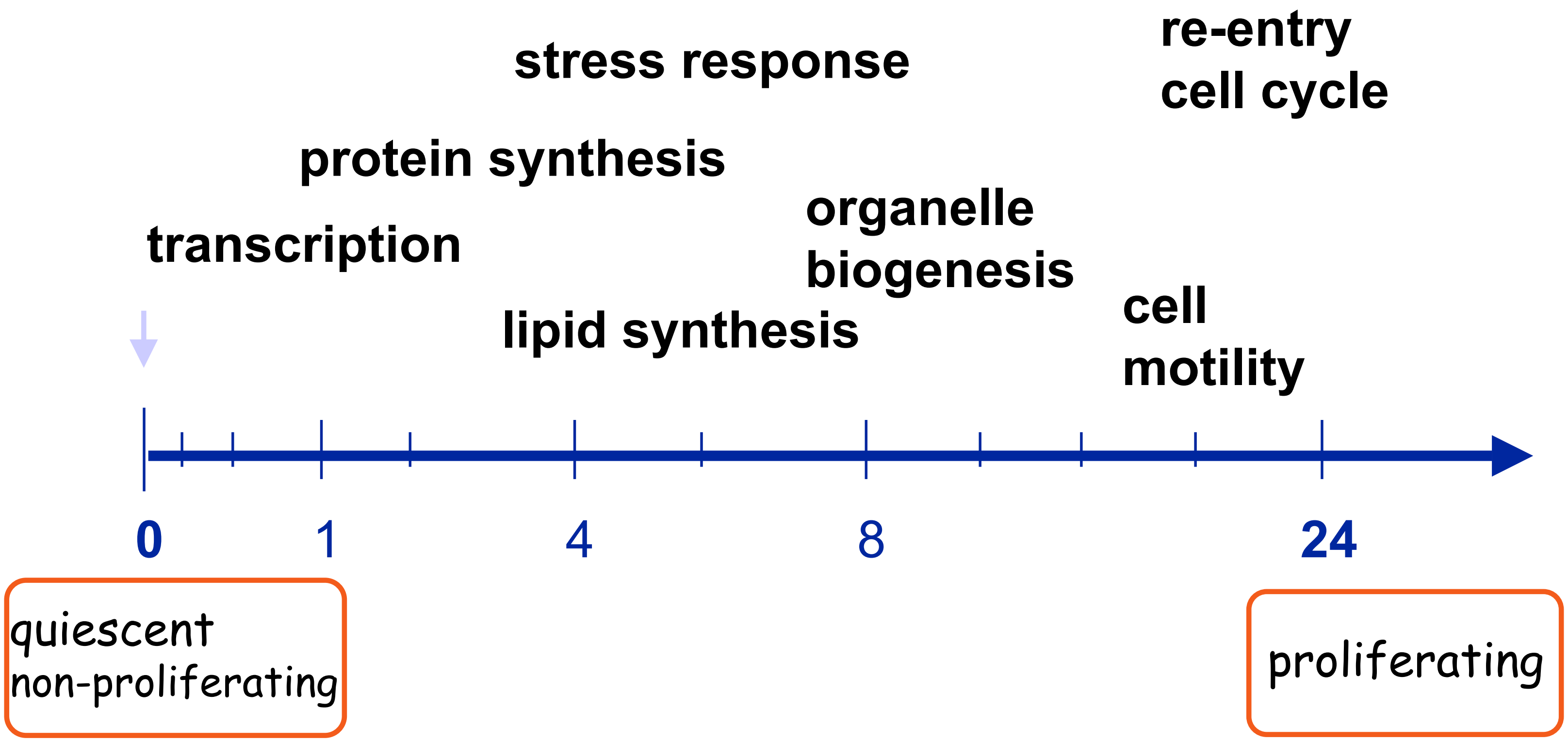
co-regulation of genes
coding for proteins in a network
in fibroblast serum-response

⊕ furin

⊖ CALLA/CD10

⊕ pro-endothelin → active endothelin → inactive endothelin

cellular processes



stress response

re-entry
cell cycle

protein synthesis

organelle
biogenesis

transcription

cell
motility

lipid synthesis

quiescent
non-proliferating

proliferating

fibroblast serum-response transcriptional program

517 gen-probes differential gene expression

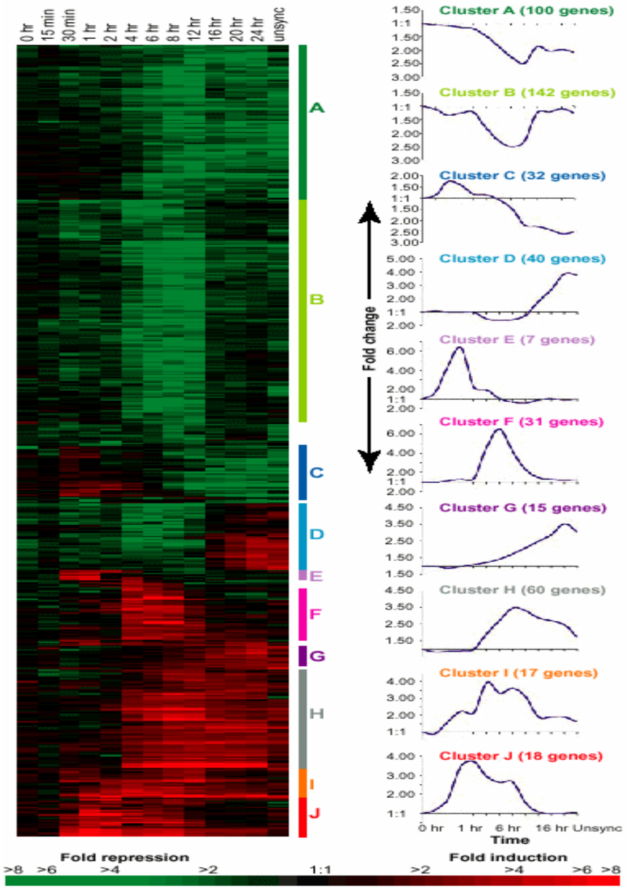
497 unique genes

284 known genes

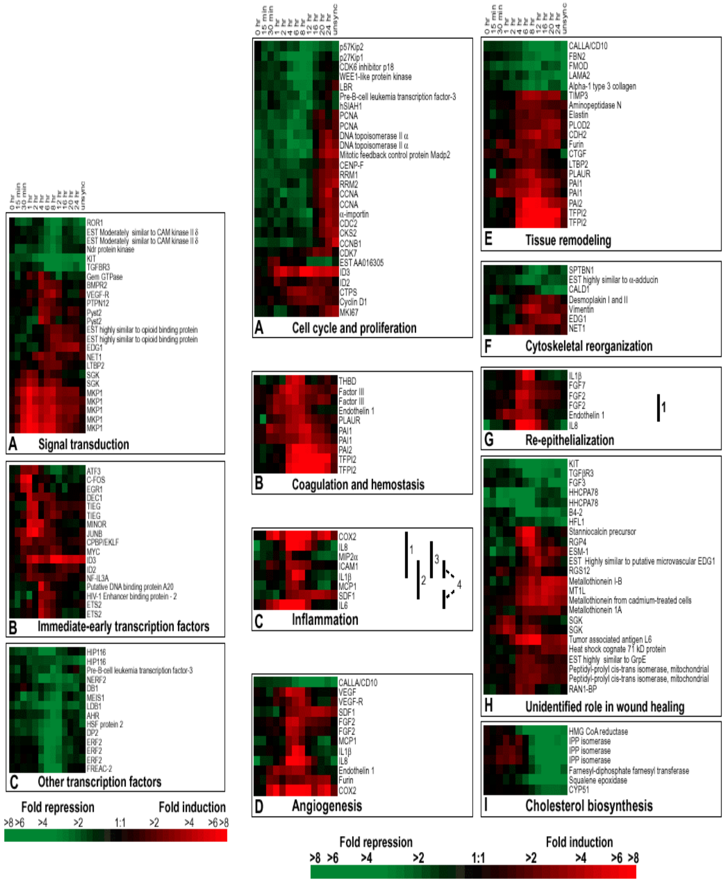
213 unknown genes

Iyer's analysis of transcriptional fibroblast serum response

Expression clusters



Functional clusters



our aim

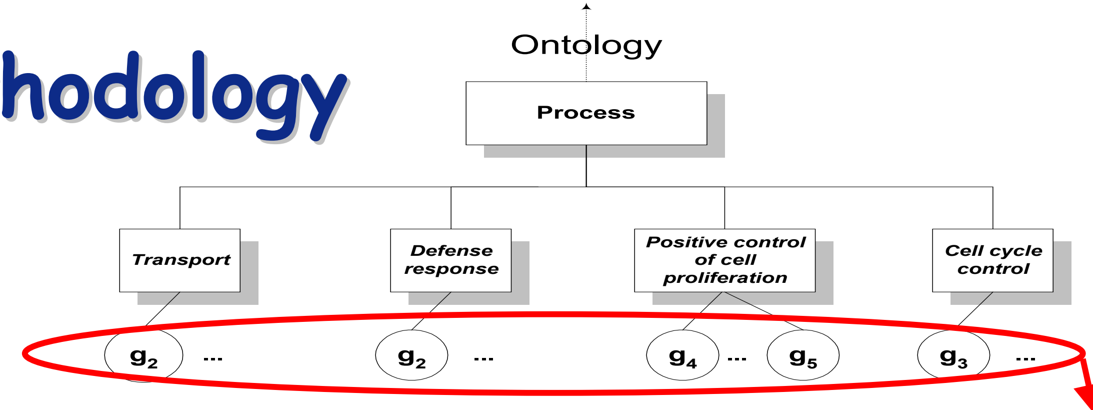
find relationship between

gene function - gene expression profile

selected challenges in gene-expression analysis

- function similarity corresponds to expression similarity but:
  - functionally correlated genes may be expression-wise dissimilar (e.g. anti-coregulated)
  - genes usually have multiple function
  - measurements may be *approximate* and *contradictory*
- can we obtain clusters of biologically related genes?
- can we build models that classify unknown genes to functional classes, that are human legible, and that handle *approximate* and often *contradictory* data?
- how can we re-use biological knowledge?

methodology



Ontology

Process

Transport

Defense response

Positive control of cell proliferation

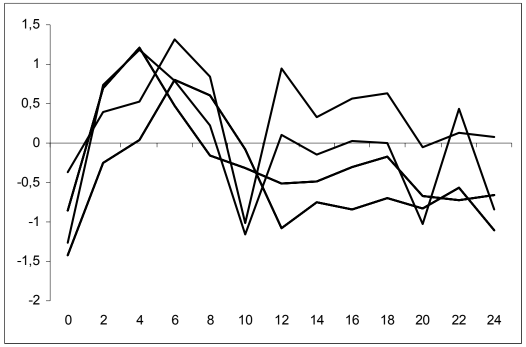
Cell cycle control

g_2 … g_2 … g_4 … g_5 g_3 …

1. Mining functional classes from an ontology

Gene	0HR	15MIN	30MIN	1HR	2HR	4HR	6HR	8HR	12HR	16HR	20HR	24HR	Process
g_1	0.00	-0.47	-3.32	-0.81	0.11	-0.60	-1.36	-1.03	-1.84	-1.00	-0.60	-0.94	Unknown
g_2	0.00	0.66	0.07	0.20	0.29	-0.89	-0.45	-0.29	-0.29	-0.15	-0.45	-0.42	Transport and defense response
g_3	0.00	0.14	-0.04	0.00	-0.15	-0.58	-0.30	-0.18	-0.38	-0.49	-0.81	-1.12	Cell cycle control
g_4	0.00	-0.04	0.00	-0.23	-0.25	-0.47	-0.60	-0.56	-1.09	-0.71	-0.76	-0.62	Positive control of cell proliferation
g_5	0.00	0.28	0.37	0.11	-0.17	-0.18	-0.60	-0.23	-0.58	-0.79	-0.29	-0.74	Positive control of cell proliferation
…	…	…	…	…	…	…	…	…	…	…	…	…	…

2. Extracting features for learning



3. Inducing minimal decision rules using rough sets

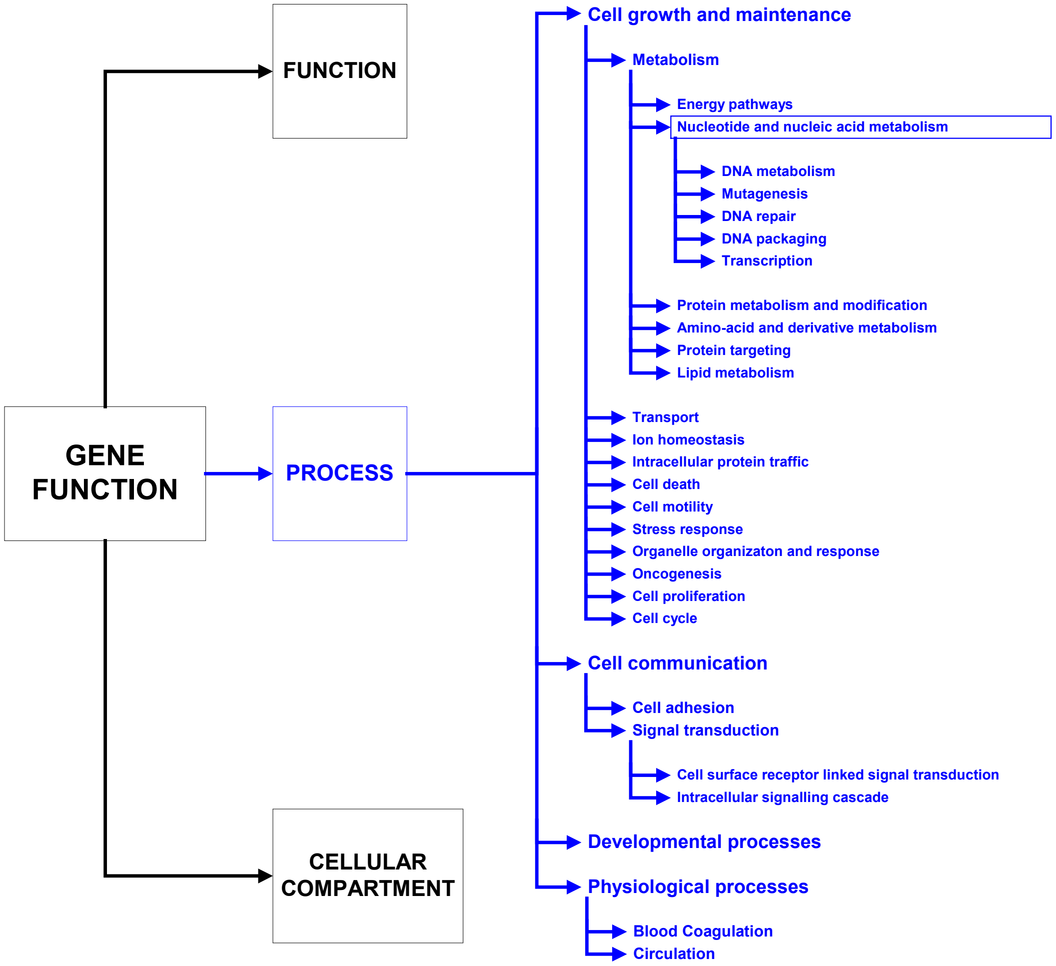
0 - 4 (Increasing) AND 6 - 10 (Decreasing) AND 14 - 18 (Constant) => GO (cell proliferation)

4. The function of unknown genes is predicted using the rules

!

Lægreid A, Hvidsten T, Midelfart H, Komorowski J, Sandvik AK.
Genome Research. 13: 965-979, 2003

Gene Ontology



- GENE FUNCTION
  - FUNCTION
  - PROCESS
    - Cell growth and maintenance
      - Metabolism
        - Energy pathways
        - Nucleotide and nucleic acid metabolism
          - DNA metabolism
          - Mutagenesis
          - DNA repair
          - DNA packaging
          - Transcription
        - Protein metabolism and modification
        - Amino-acid and derivative metabolism
        - Protein targeting
        - Lipid metabolism
      - Transport
      - Ion homeostasis
      - Intracellular protein traffic
      - Cell death
      - Cell motility
      - Stress response
      - Organelle organizaton and response
      - Oncogenesis
      - Cell proliferation
      - Cell cycle
    - Cell communication
      - Cell adhesion
      - Signal transduction
        - Cell surface receptor linked signal transduction
        - Intracellular signalling cascade
    - Developmental processes
    - Physiological processes
      - Blood Coagulation
      - Circulation
  - CELLULAR COMPARTMENT

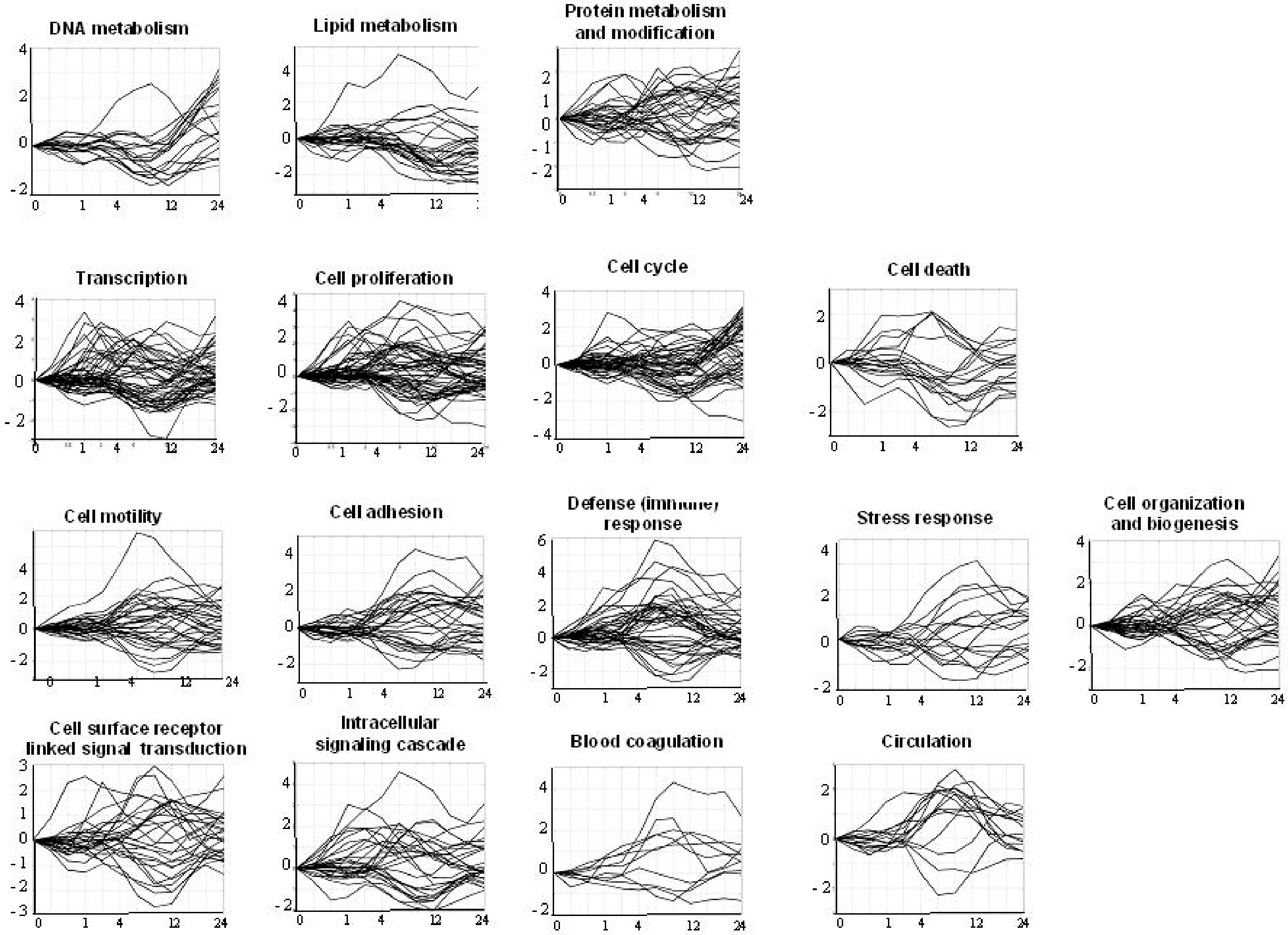
* The homepage of Ashburner's Gene Ontology: http://genome-www.standford.edu/GO/

annotations

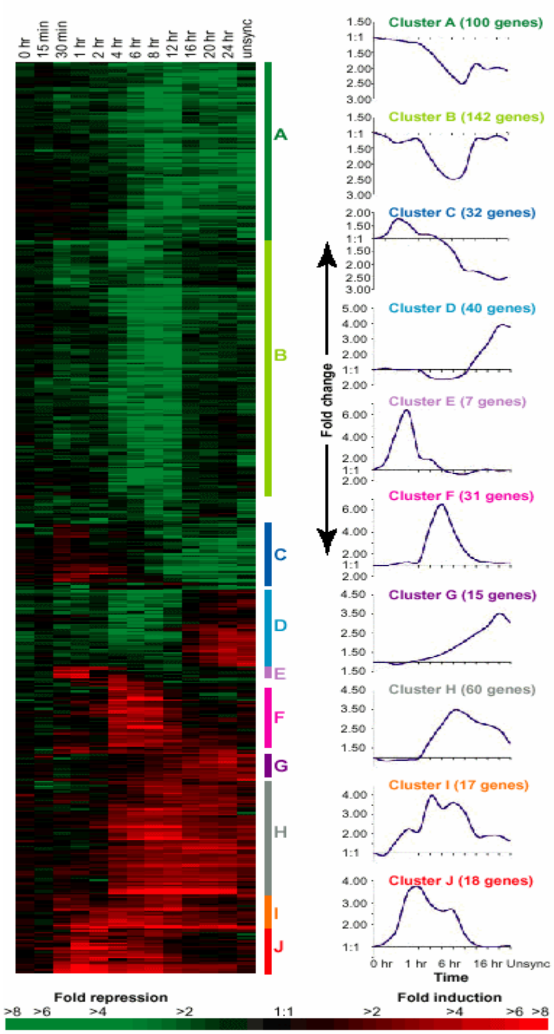
Annotation of Known Genes

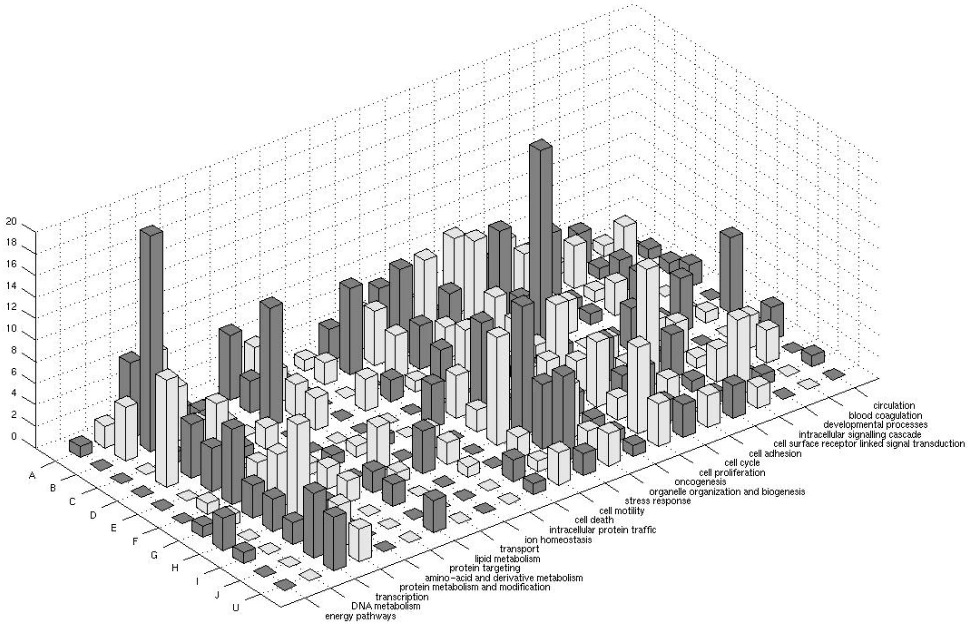
GENE SYMBOL	GENE NAME	GENEBANK ACCESSION NUMBER	ANNOTATIONS AT THE MOST SPECIFIC LEVEL OF GO	ANNOTATIONS TO THE 23 BROAD CELLULAR PROCESSES USED FOR LEARNING
SEPP1	selenoprotein P, plasma, 1	AA045003	oxidative stress response(GO:0006979), metal ion transport(GO:0006823)	stress response(GO:0006950), transport(GO:0006810)
EPB41L2	erythrocyte membrane protein band 4.1-like 2	W88572	positive control of cell proliferation(GO:0008284)	cell proliferation(GO:0008283)
OA48-18	acid-inducible phosphoprotein	AA029909	cell proliferation(GO:0008283)	cell proliferation(GO:0008283)
CTSK	cathepsin K (pycnodysostosis)	AA044619	proteolysis and peptidolysis(GO:0006508)	protein metabolism and modification(GO:0006411)
CPT1B	carnitine palmitoyltransferase I, muscle	W89012	fatty acid beta-oxidation(GO:0006635)	lipid metabolism(GO:0006629)
CLDN11	claudin 11 (oligodendrocyte transmembrane protein)	N22392	cell adhesion(GO:0007155), substrate-bound cell migration(GO:0006929), cell proliferation(GO:0008283), developmental processes(GO:0007275)	cell adhesion(GO:0007155), cell motility(GO:0006928), cell proliferation(GO:0008283), developmental processes(GO:0007275)
RPL5	ribosomal protein L5	AA027277	protein biosynthesis(GO:0006412), ribosomal large subunit assembly and maintenance(GO:0000027)	protein metabolism and modification(GO:0006411), cell organization and biogenesis(GO:0006996)
	Homo sapiens clone 23785 mRNA sequence	N32247	calcium-independent cell-cell matrix adhesion(GO:0007161)	cell adhesion(GO:0007155)

time profiles of selected processes

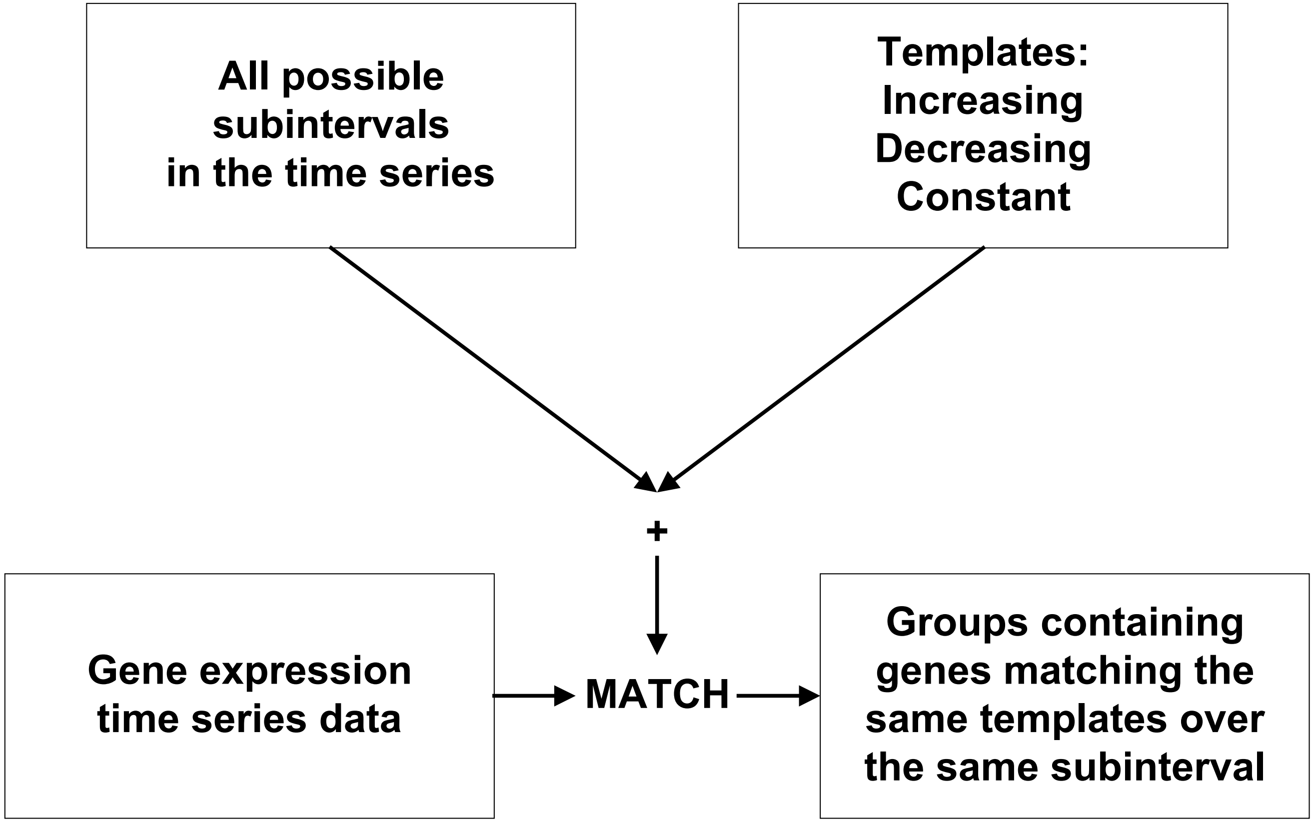


Gene Ontology vs. clusters





template-based feature synthesis



All possible subintervals in the time series

Templates:
Increasing
Decreasing
Constant

+

Gene expression time series data

MATCH

Groups containing genes matching the same templates over the same subinterval

12 measurement points, 55 possible intervals of length >2

cross validation estimates

PROCESS	AUC	SE
Ion homeostasis	1.00	0.00
Protein targeting	0.99	0.03
Blood coagulation	0.96	0.08
DNA metabolism	0.94	0.09
Intracellular signaling cascade	0.94	0.06
Energy pathways	0.93	0.12
Cell cycle	0.93	0.04
Oncogenesis	0.92	0.11
Circulation	0.91	0.11
Cell death	0.90	0.10
Developmental processes	0.90	0.07
Transcription	0.88	0.11
Defense (immune) response	0.88	0.05
Cell adhesion	0.87	0.09
Stress response	0.86	0.15
Protein metabolism and modification	0.85	0.10
Cell motility	0.84	0.11
Cell surface rec linked signal transd	0.82	0.15
Lipid metabolism	0.81	0.14
Transport	0.79	0.17
Cell organization and biogenesis	0.79	0.11
Cell proliferation	0.79	0.06
Amino acid and derivative metabolism	0.69	0.06
AVERAGE	**0.88**	**0.09**

A:
Coverage: 84%
Precision: 50%

B:
Coverage: 71%
Precision: 60%

C:
Coverage: 39%
Precision: 90%

Coverage = TP/(TP+FN)
Precision = TP/(TP+FP)

the model

Annotations, Rules and Classifications

Annotated genes		273
within the 23 broad classes of GO biological process		
Gene probes		284
associated with the 273 genes within the 23 broad biological process classes		
Training examples		
annotations associated with the genes in the 23 broad biological process classes		549
co-annotations associated with the genes in the 23 broad biological process classes		444
Rules generated from the training examples		18064
Estimated quality of classifications of unknown genes (cross-validation estimates)		
Sensitivity		84%
Specificity		91%
Fraction of classifications that are correct		49%
Classifications for unknown (uncharacterized) genes		548
classifications were obtained for 211 of the 213 unknown genes		
(Re-)Classifications for training examples		728
True positive classifications	519	
True positive co-classifications	356	
False positive classifications	219	
False negative (missing) classifications	30	

→ For 272 of the 273 training examples at least one correct (re-)classification was obtained

conclusions

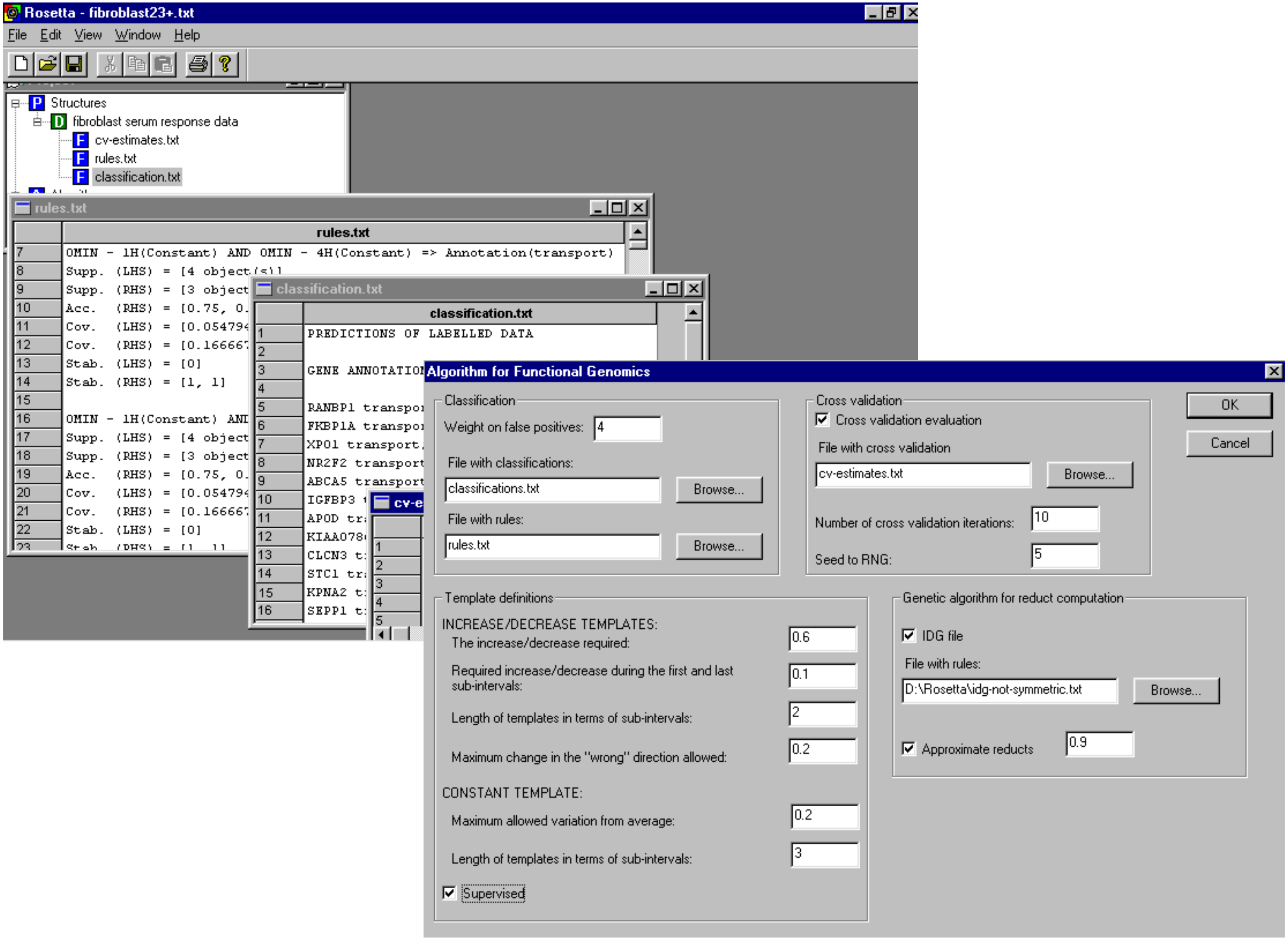
- our methodology
  - incorporates background biological knowledge
  - handles well the noise and incompleteness in the microarray data
  - can be objectively evaluated
  - predicts multiple functions per gene
  - can re-classify known genes and provide possible new functions of the known genes
  - can provide hypotheses about the function of unknown genes
- experimental work needs to be done to confirm our predictions

Lægreid A, Hvidsten T, Midelfart H, Komorowski J, Sandvik AK.
Predicting Gene Ontology Biological Process from Temporal Gene Expression Patterns.
Genome Research. 13: 965-979, 2003

Hvidsten TR, Lægreid A, Komorowski J.
Learning rule-based models from gene expression time profiles annotated using Gene Ontology.
Bioinformatics, 19:1116-23, 2003

Genomic ROSETTA:

http://www.idi.ntnu.no/~aleks/rosetta



rules.txt

```
7  OMIN - 1H(Constant) AND OMIN - 4H(Constant) => Annotation(transport)
8  Supp.  (LHS) = [4 object(s)]
9  Supp.  (RHS) = [3 object
10 Acc.   (RHS) = [0.75, 0.
11 Cov.   (LHS) = [0.054794
12 Cov.   (RHS) = [0.166667
13 Stab.  (LHS) = [0]
14 Stab.  (RHS) = [1, 1]
15
16 OMIN - 1H(Constant) AND
17 Supp.  (LHS) = [4 object
18 Supp.  (RHS) = [3 object
19 Acc.   (RHS) = [0.75, 0.
20 Cov.   (LHS) = [0.054794
21 Cov.   (RHS) = [0.166667
22 Stab.  (LHS) = [0]
```

classification.txt

```
1  PREDICTIONS OF LABELLED DATA
2
3  GENE ANNOTATION
4
5  RANBP1 transpo
6  FKBP1A transpo
7  XPO1 transport,
8  NR2F2 transport
9  ABCA5 transport
10 IGFBP3
11 APOD tr
12 KIAA078
13 CLCN3 t
14 STC1 tr
15 KPNA2 t
16 SEPP1 t
```

Algorithm for Functional Genomics

Classification
- Weight on false positives: 4
- File with classifications: classifications.txt
- File with rules: rules.txt

Cross validation
- ☑ Cross validation evaluation
- File with cross validation: cv-estimates.txt
- Number of cross validation iterations: 10
- Seed to RNG: 5

Template definitions

INCREASE/DECREASE TEMPLATES:
- The increase/decrease required: 0.6
- Required increase/decrease during the first and last sub-intervals: 0.1
- Length of templates in terms of sub-intervals: 2
- Maximum change in the "wrong" direction allowed: 0.2

CONSTANT TEMPLATE:
- Maximum allowed variation from average: 0.2
- Length of templates in terms of sub-intervals: 3

☑ Supervised

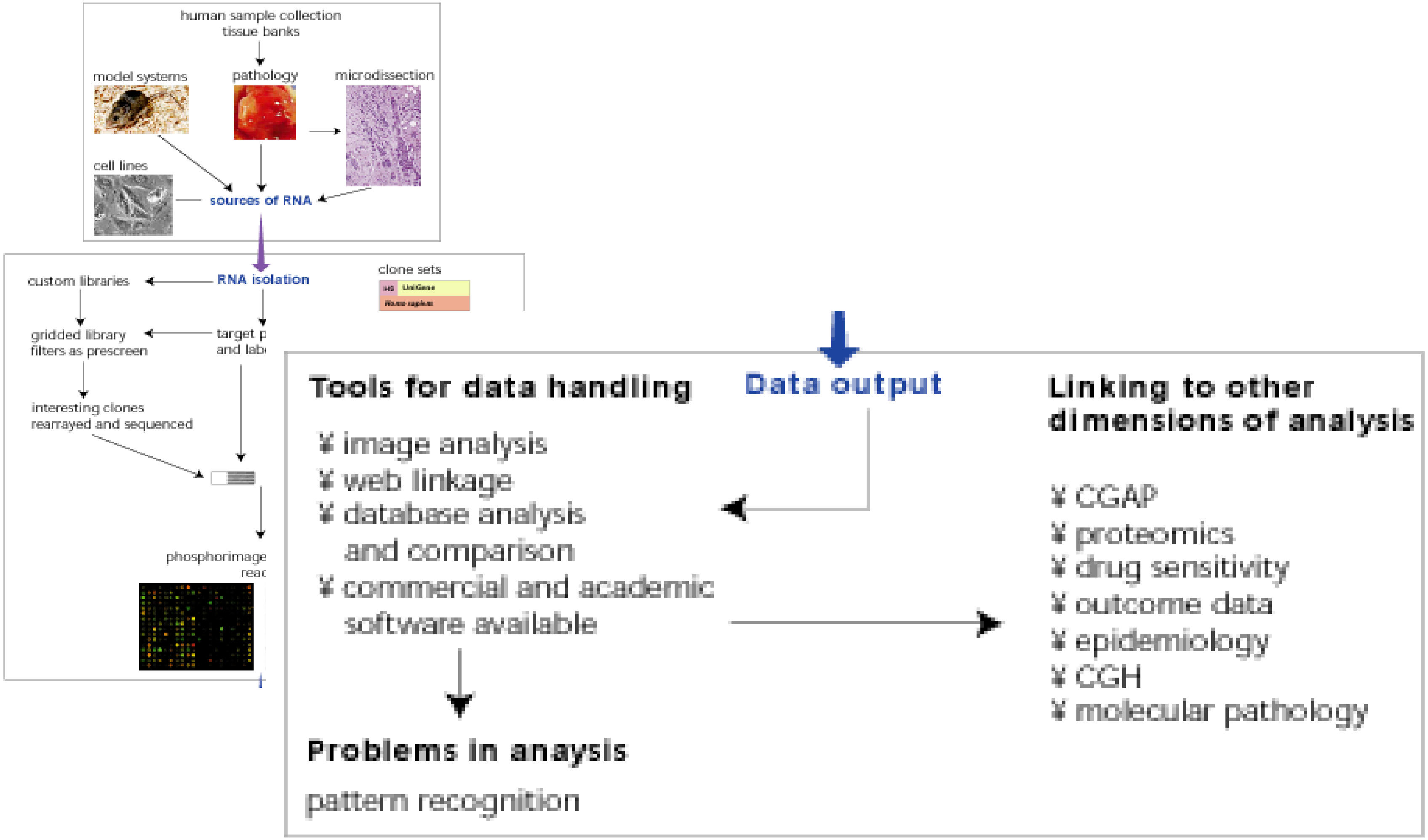
Genetic algorithm for reduct computation
- ☑ IDG file
- File with rules: D:\Rosetta\idg-not-symmetric.txt
- ☑ Approximate reducts 0.9

how to improve models for prediction of biological roles of genes/proteins?

- improved computational methods
- more training examples
    - more genes/proteins
    - more measurements per gene/protein (time points, cell types, tissues, states,...)
    - more annotations

      (GO, sequence, protein structure, cell biology, physiology, pathology,…)

many levels of information



Data output

Tools for data handling

¥ image analysis
¥ web linkage
¥ database analysis
and comparison
¥ commercial and academic
software available

Linking to other dimensions of analysis

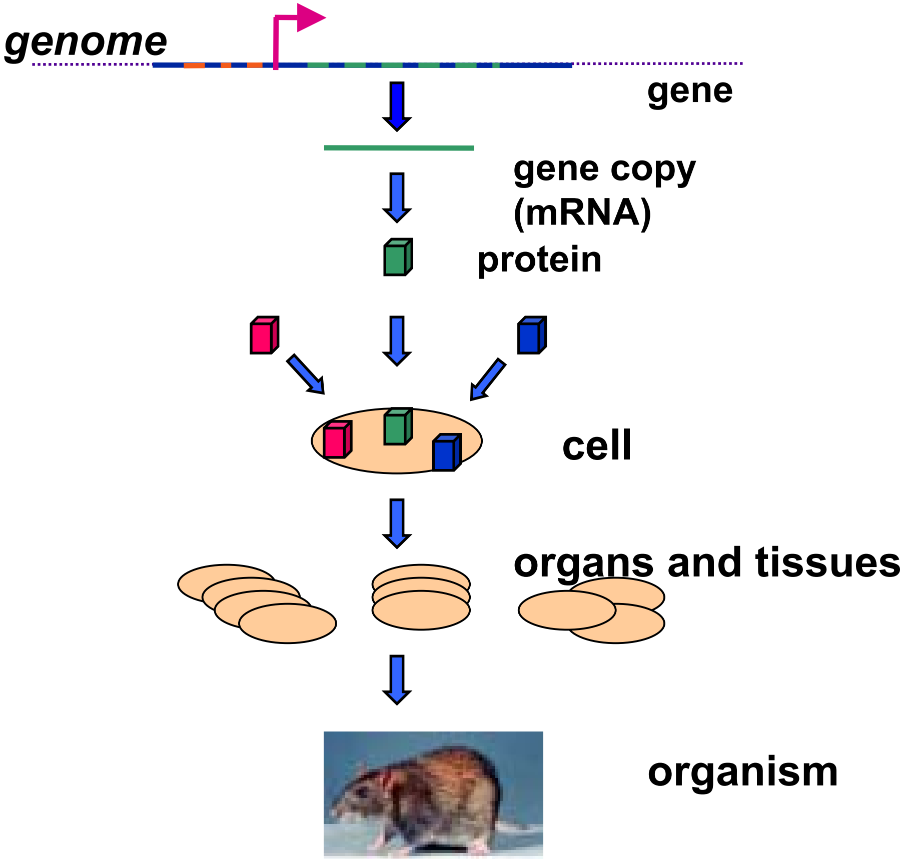
¥ CGAP
¥ proteomics
¥ drug sensitivity
¥ outcome data
¥ epidemiology
¥ CGH
¥ molecular pathology

Problems in anaysis

pattern recognition

Bowtell, Nature Genetics,
Supplement, 21:25, 1999

high complexity



~35.000 genes

> 100.000 gene (splice) products

> 100.000 proteins

> 200.000 protein states

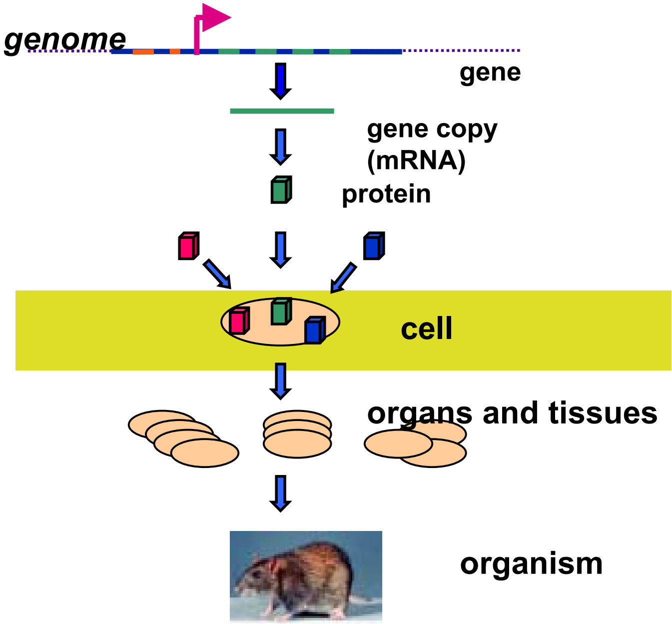
each cell expresses
5.- 15.000 genes
40.-60.000 proteins

several hundred cell types

many different states per cell

tissues and organs are composed of many different cell types

molecular networks within cells



genome

gene

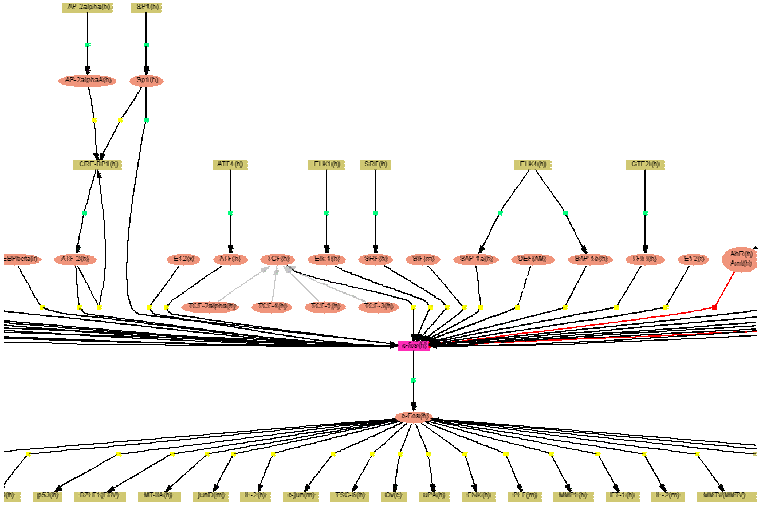
gene copy
(mRNA)

protein

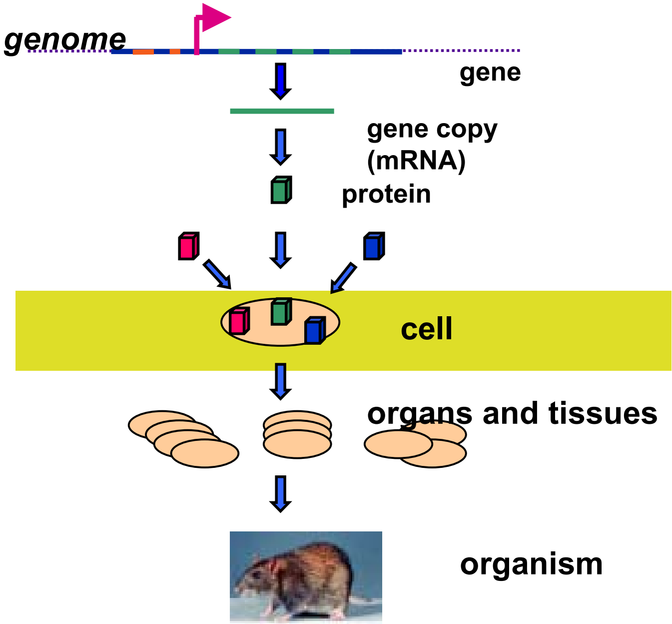
cell

organs and tissues

organism



molecular networks within cells



genome

gene

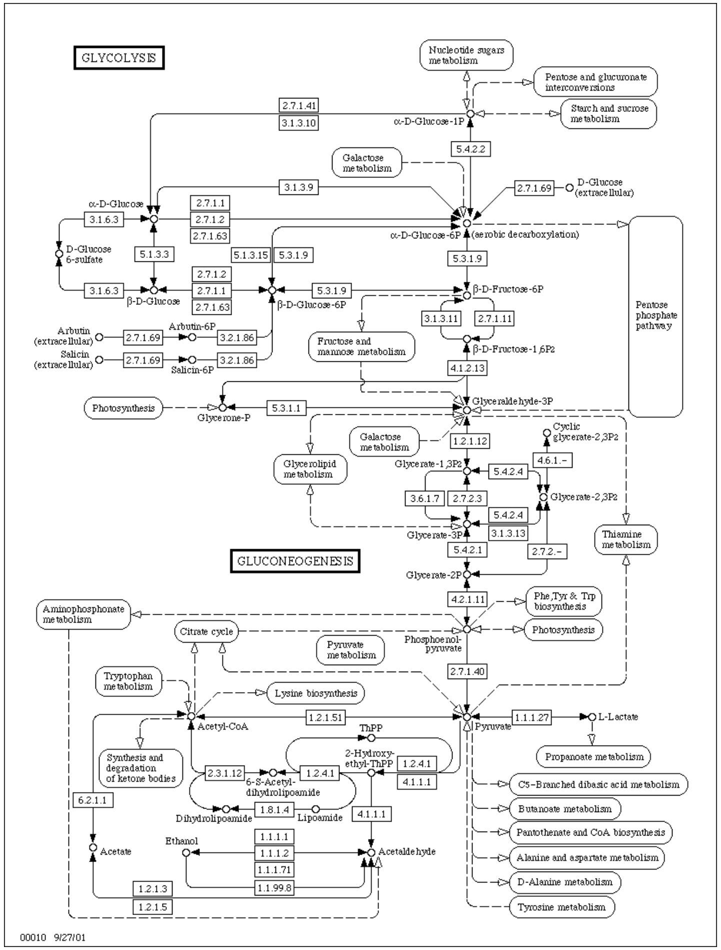
gene copy (mRNA)

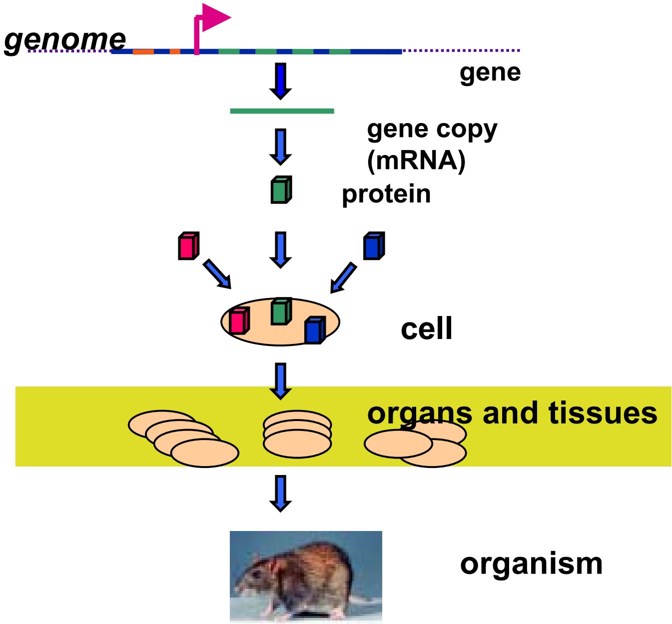
protein

cell

organs and tissues

organism

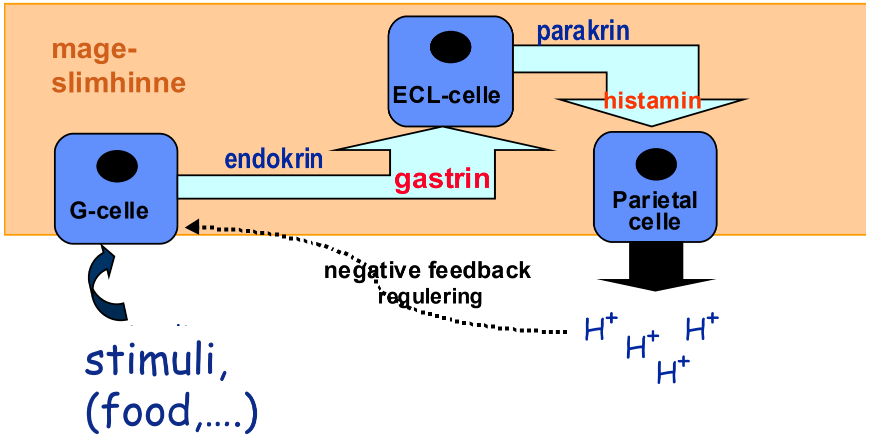


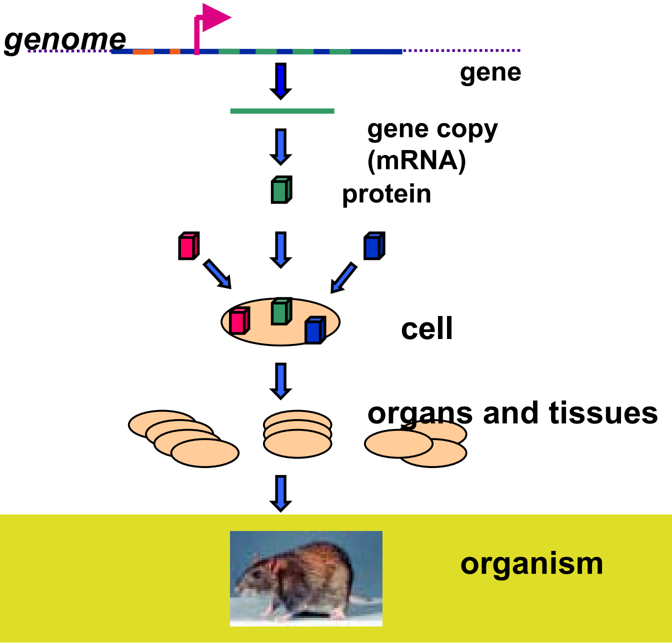


different cell types interact within organs and tissues

different cell types interact during gastric acid secretion

stomach mucosa





genome

gene

gene copy
(mRNA)

protein

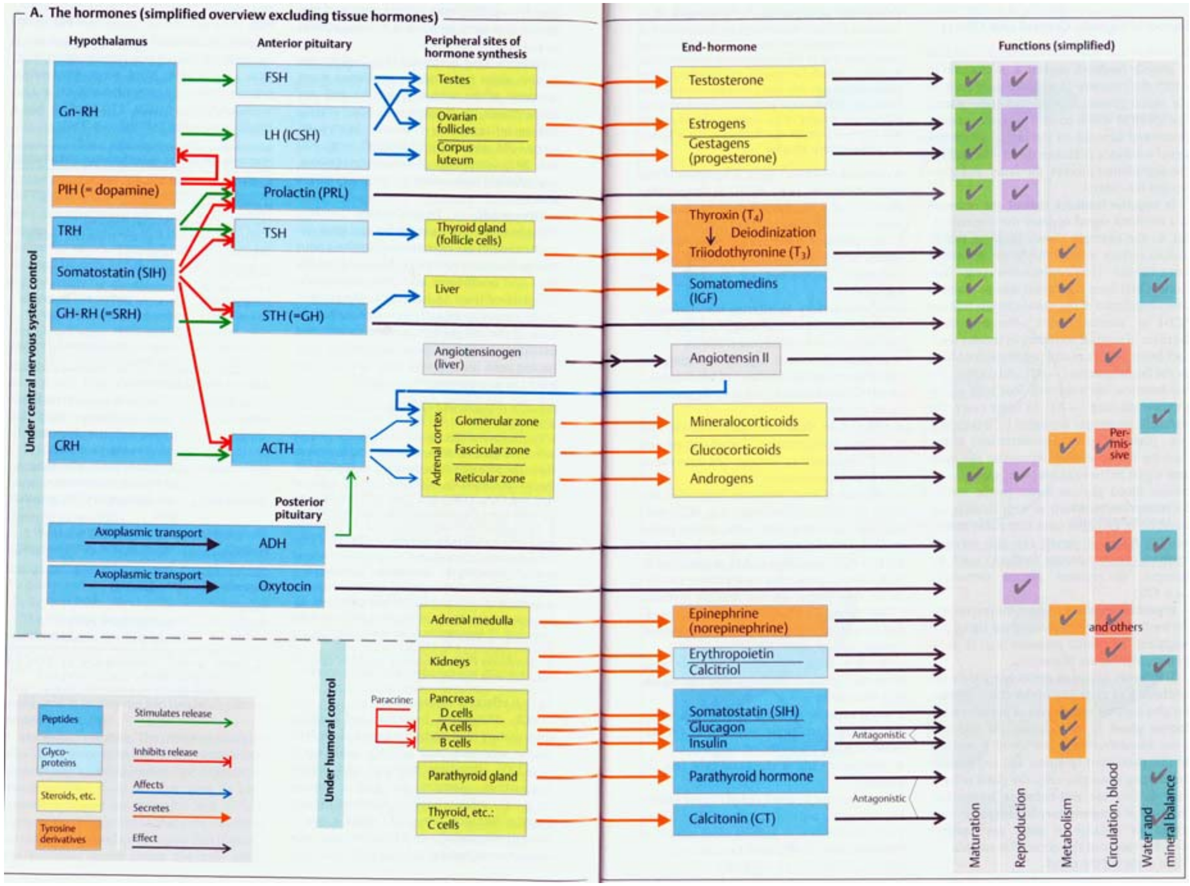
cell

organs and tissues

organism

interconnection within organism

hormones regulate interactions between organs and tissues



expression profiling in biology

determine molecular mechanisms underlying

- cell function related to cell type and state
- physiological functions of organims

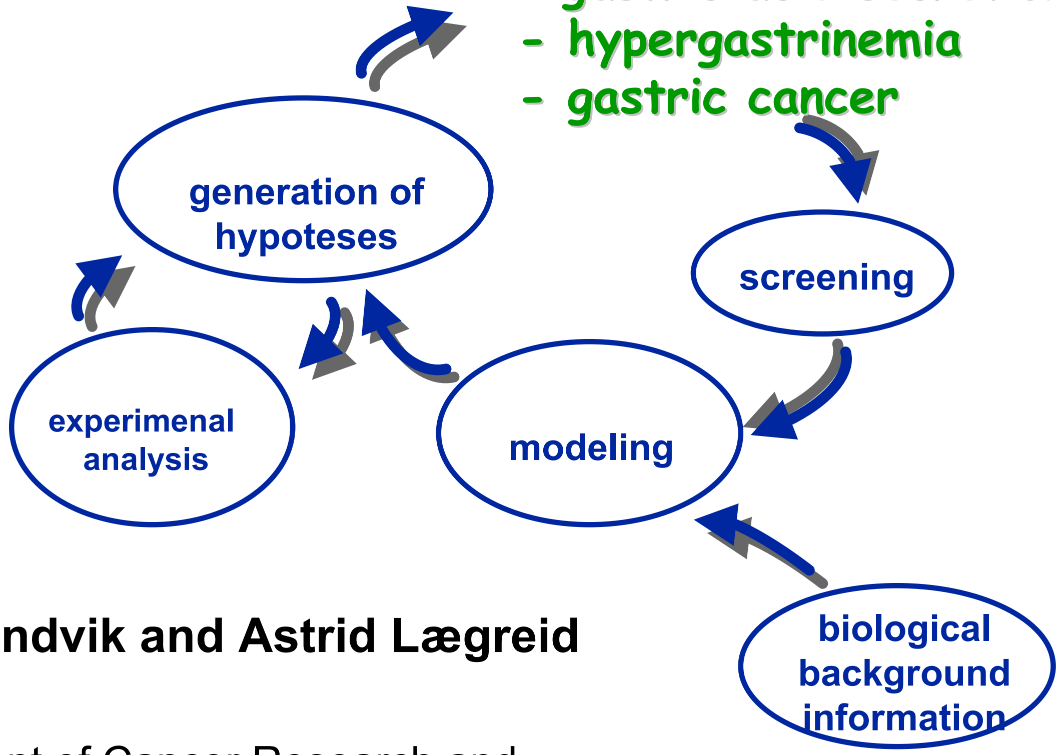
expression profiling in disease managment

- dicover disease subtypes
- improve disease diagnostics
- improve prognostics/choice of treatment
- discover new drug targets

our focus:

Molecular Mechansisms of the Normal and Diseased Gastrointestinal System

- gastric acid secretion
- hypergastrinemia
- gastric cancer



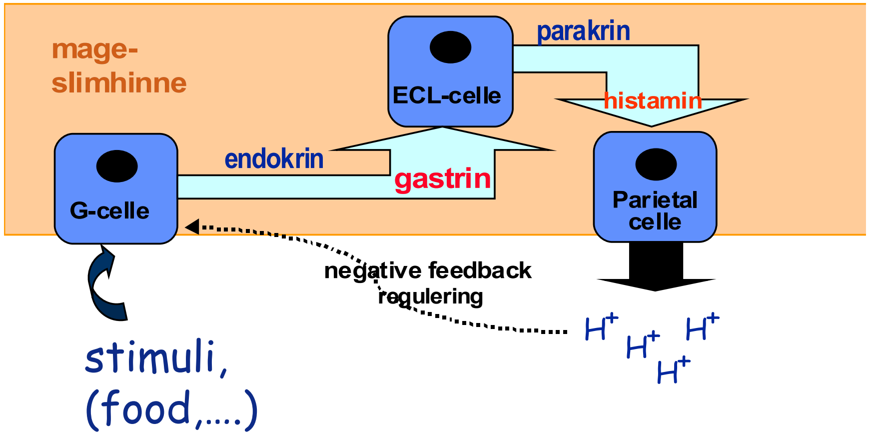
Arne Sandvik and Astrid Lægreid

Department of Cancer Research and
Molecular Medicine, NTNU

gastrointestinal physiology and pathophysiology

- gastric acid secretion

stomach mucosa



- molecular mechanisms?
- regulators, effectors?

gastrointestinal physiology and pathophysiology

- hypergatrinemia

↑ gastrin

↑ proliferation gastric mucosa

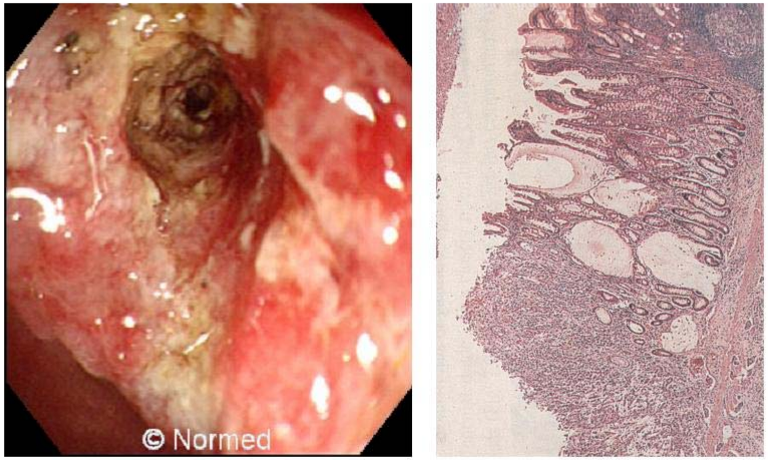
↑ ECL-cells

↑ cancer

- molecular mechanisms?
- regulators, effectors?

gastrointestinal physiology and pathophysiology

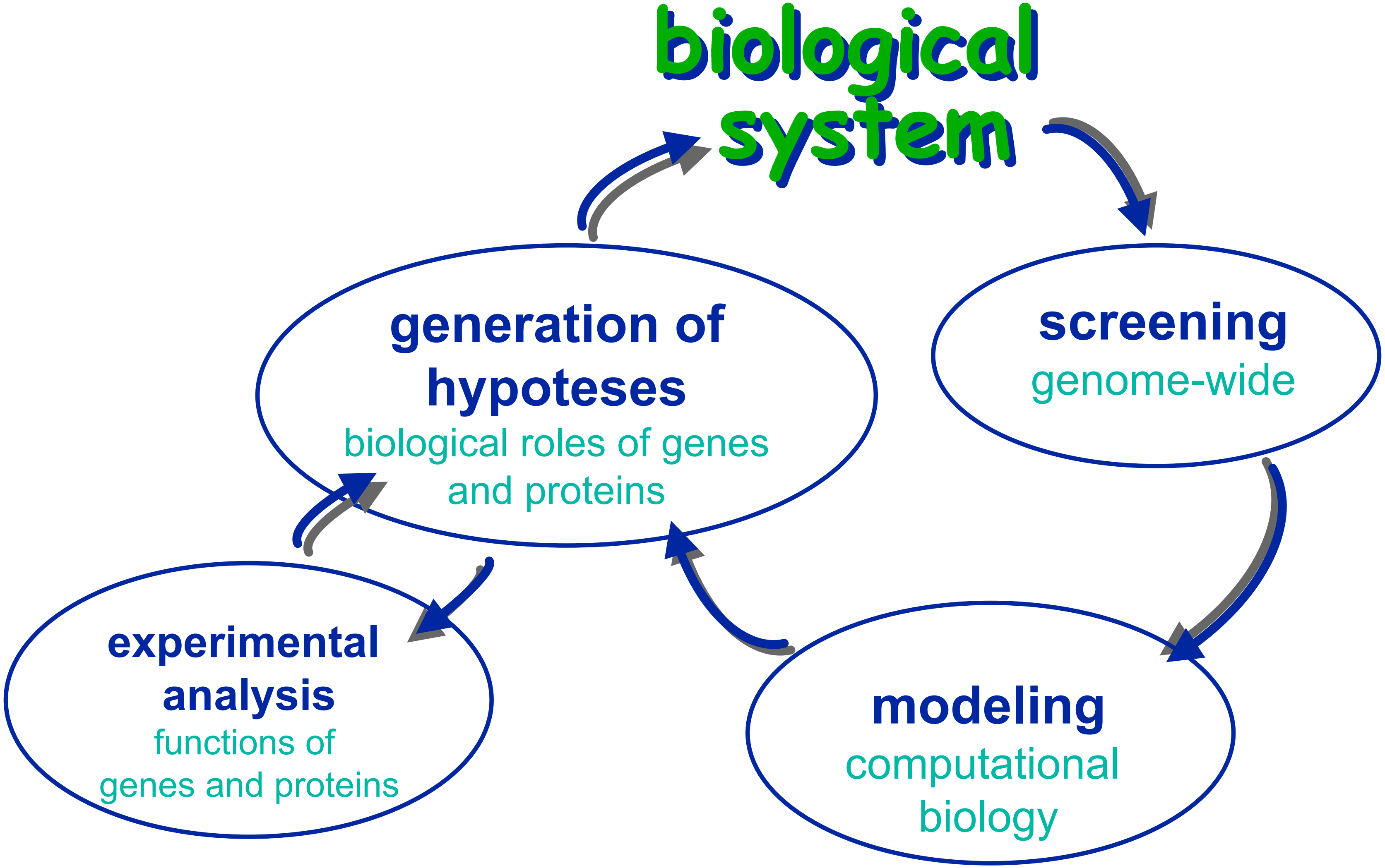
- gastric cancer



- classification & prediction
  subtype diagnostics
  prognostics
  optimal treatment
  early diagnostics

- molecular mechanisms?
- regulators, effectors?

functional genomics



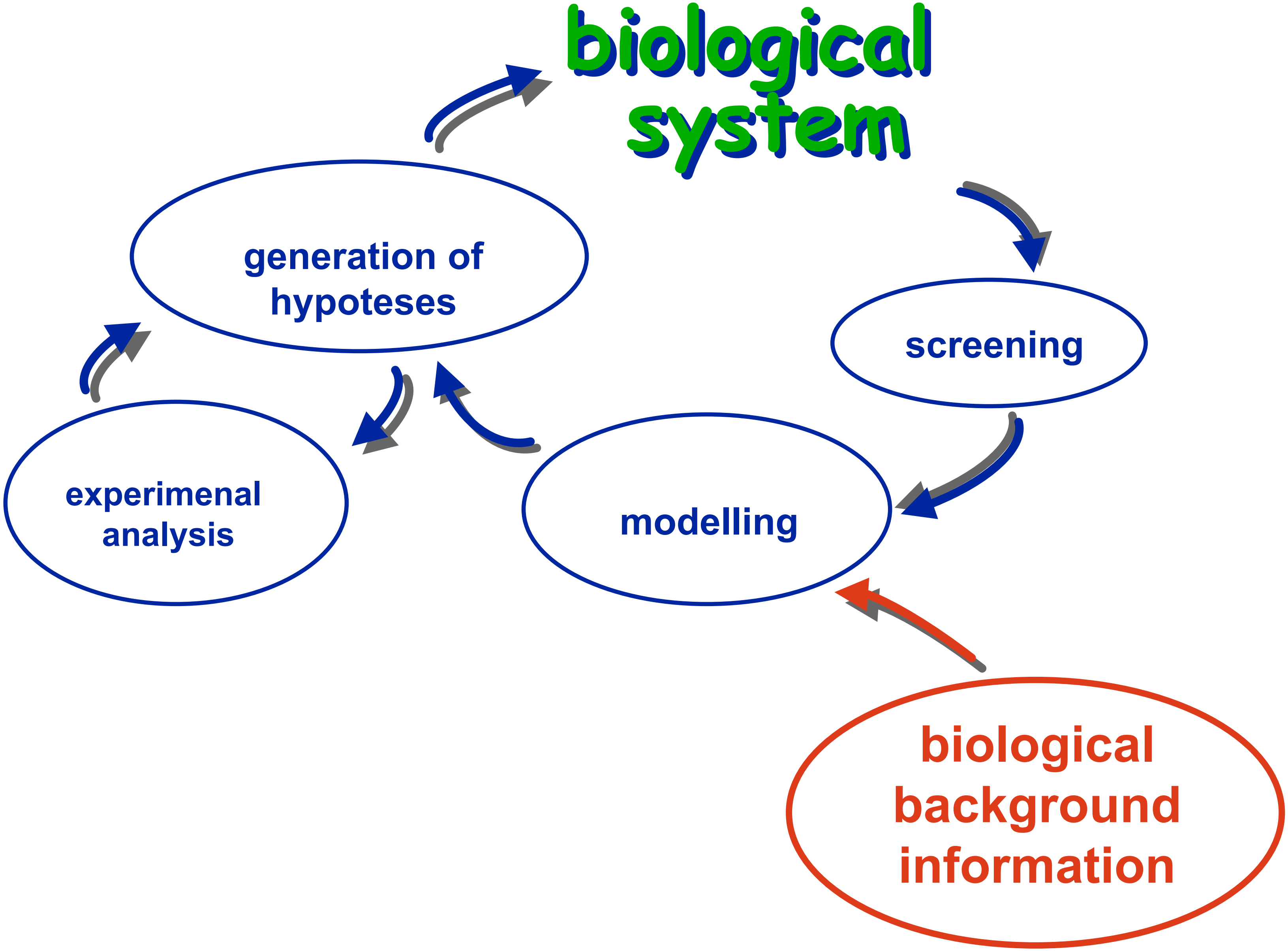
biological system

generation of hypoteses
biological roles of genes and proteins

screening
genome-wide

modeling
computational biology

experimental analysis
functions of genes and proteins



biological system

generation of hypoteses

screening

modelling

experimenal analysis

biological background information

challenges….

Information Bases/Derived-Data Databases

Experimental/Clinical Data

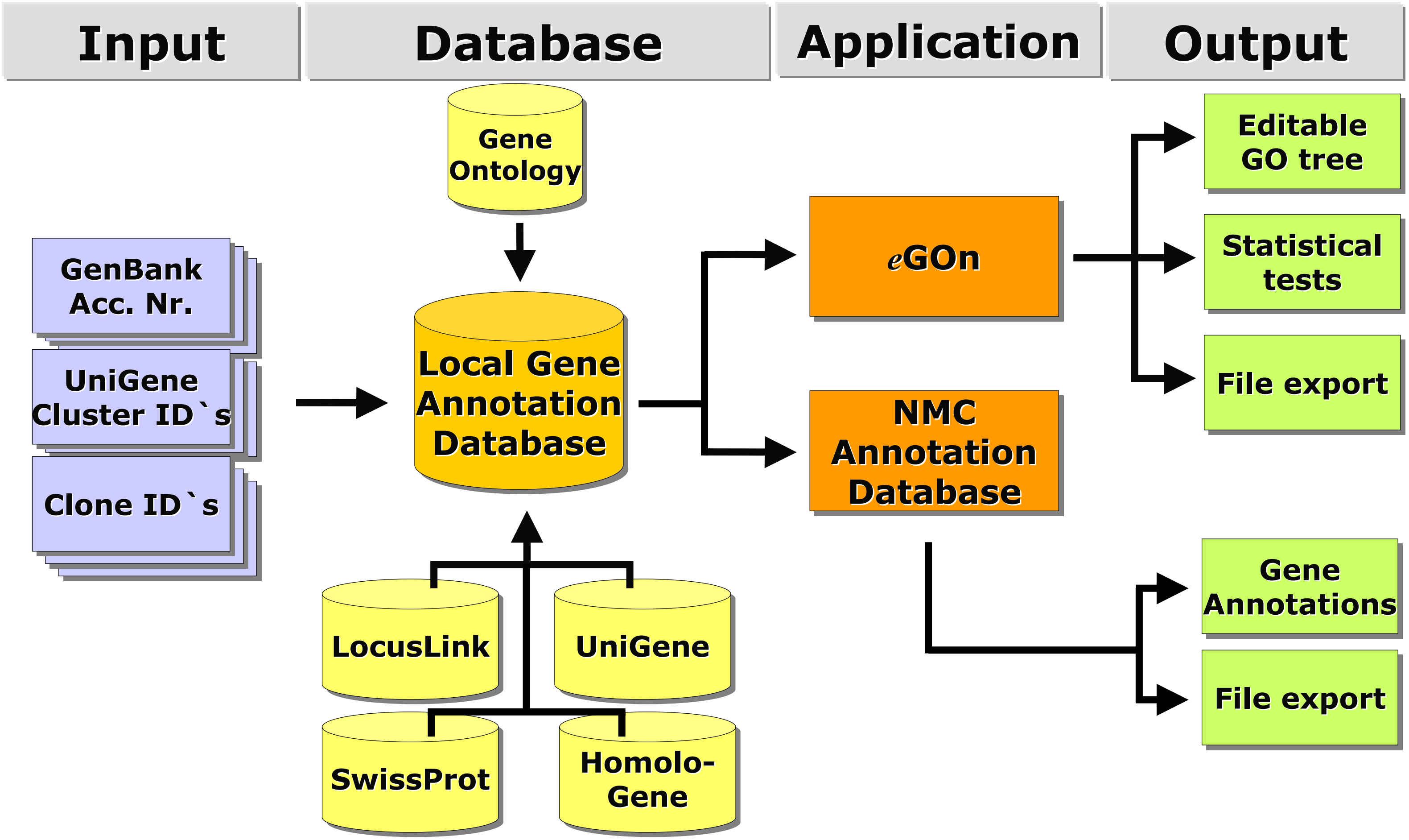
challenges….

Information Bases/Derived-Data Databases

Experimental/Clinical Data

link information from various sources
in a relevant way

relational database & tools at NTNU



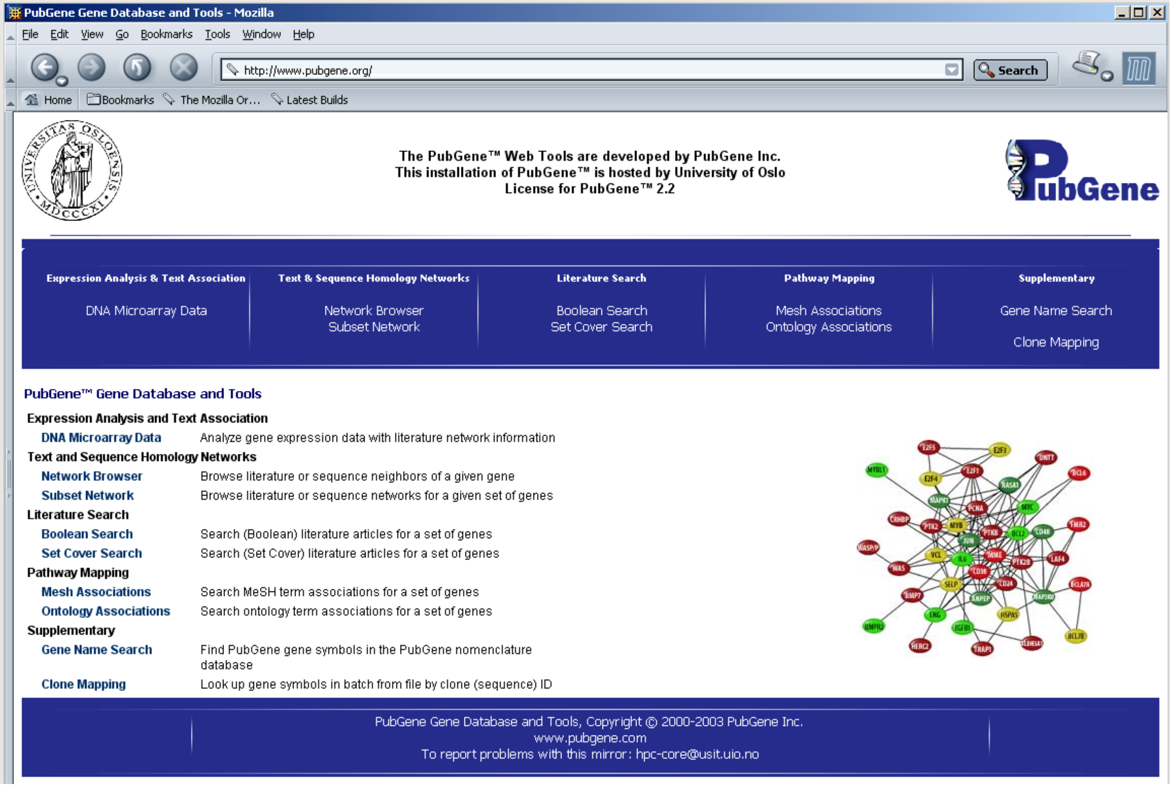
challenges….

Information Bases/Derived-Data Databases

Experimental/Clinical Data

mine information from unstructured information sources

mining the literature



The PubGene™ Web Tools are developed by PubGene Inc.
This installation of PubGene™ is hosted by University of Oslo
License for PubGene™ 2.2

Expression Analysis & Text Association	Text & Sequence Homology Networks	Literature Search	Pathway Mapping	Supplementary
DNA Microarray Data	Network Browser Subset Network	Boolean Search Set Cover Search	Mesh Associations Ontology Associations	Gene Name Search Clone Mapping

PubGene™ Gene Database and Tools

Expression Analysis and Text Association	
DNA Microarray Data	Analyze gene expression data with literature network information
Text and Sequence Homology Networks	
Network Browser	Browse literature or sequence neighbors of a given gene
Subset Network	Browse literature or sequence networks for a given set of genes
Literature Search	
Boolean Search	Search (Boolean) literature articles for a set of genes
Set Cover Search	Search (Set Cover) literature articles for a set of genes
Pathway Mapping	
Mesh Associations	Search MeSH term associations for a set of genes
Ontology Associations	Search ontology term associations for a set of genes
Supplementary	
Gene Name Search	Find PubGene gene symbols in the PubGene nomenclature database
Clone Mapping	Look up gene symbols in batch from file by clone (sequence) ID

PubGene Gene Database and Tools, Copyright © 2000-2003 PubGene Inc.
www.pubgene.com
To report problems with this mirror: hpc-core@usit.uio.no

Tor-Kristian Jenssen, Astrid Lægreid, Jan Komorowski, Eivind Hovig.
A literature network of human genes for high throughput gene-expression analysis.
Nature Genetics, 28: 21-28

mining the literature

(at NTNU)

statistical methods

machine learning

natural language processing

challenges….

Information Bases/Derived-Data Databases

Experimental/Clinical Data

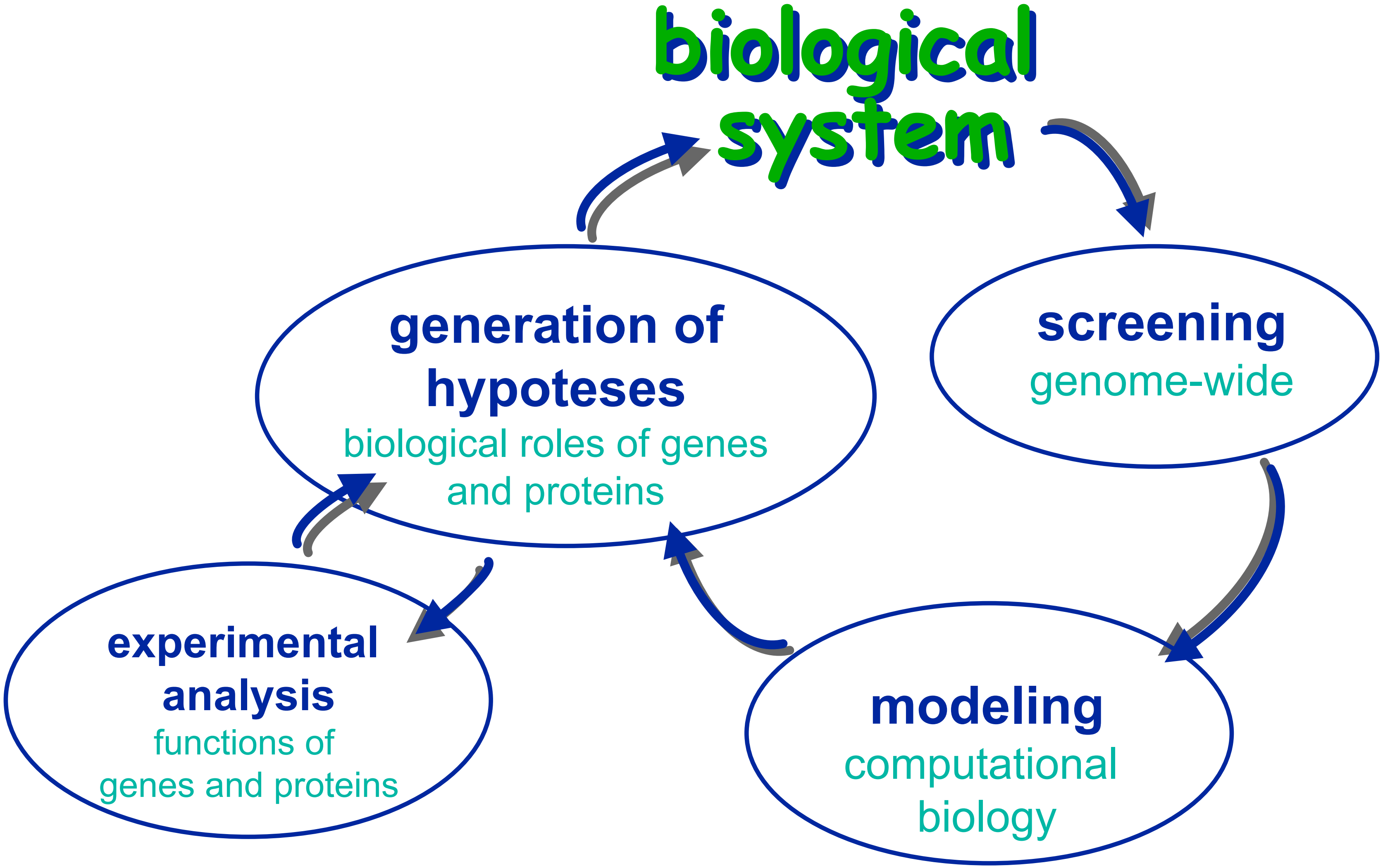
develop improved methods for modeling

modeling

data driven

first principles

functional genomics



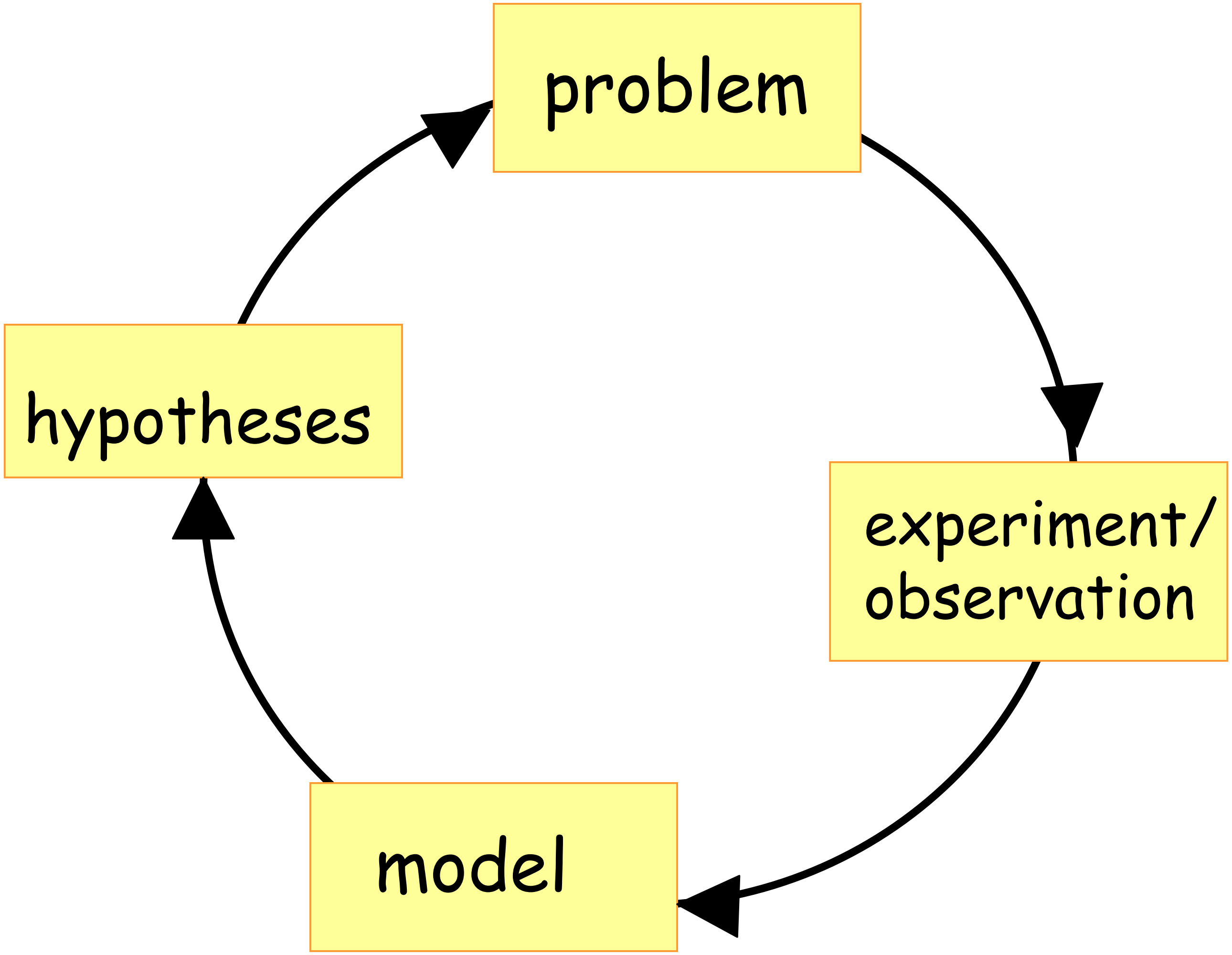
biological system

screening
genome-wide

modeling
computational biology

generation of hypoteses
biological roles of genes and proteins

experimental analysis
functions of genes and proteins



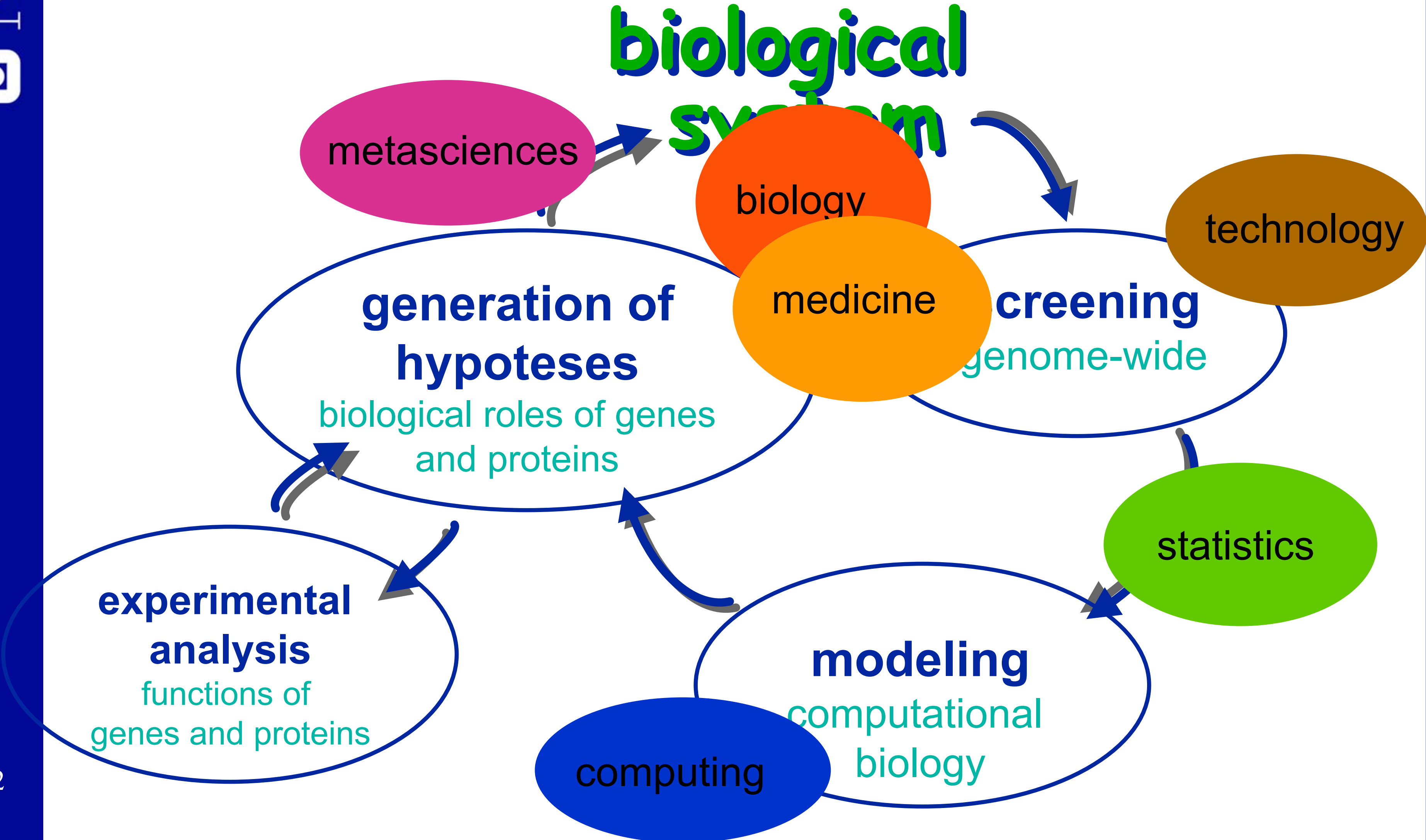
problem

experiment/
observation

model

hypotheses

multi-disciplinary effort



biological system

metasciences

biology

medicine

technology

generation of hypoteses

biological roles of genes and proteins

screening

genome-wide

statistics

experimental analysis

functions of genes and proteins

modeling

computational biology

computing

Arne K. Sandvik
Helge L. Waldum

Fekadu Yadetie
Kristin Nørsett
Vidar Beisvåg
Berit Doseth Eitrem
Hallgeir Bergum
Frode Jünge
Torunn Bruland
Ola Ween
Liv Thommesen
Kristine Misund
Tonje Strømmen

Mette Langaas

Raymond Dingledine

Agnar Aamodt
Waclaw Kusnierczyk
Pauline Haddow
Gunnar Tufte
Kjell Bratbergsengen
Heri Ramampiaro
Tore Amble
Rune Sætre
Bjørn Alsberg
Arnar Flatberg
Lars Giskehaug
Torulf Mollestad
Henrik Tveit

Jan Komorowski
Torgeir Hvidsten
Herman Midelfart
Vladimir Yankovski

Norwegian Microarray Consortium: www.mikromatrise.no