

From raw data to publication

...what happens in between?

Anne-Kristin Stavrum





Summary: Steps in microarray analysis

- Preprocessing/QC
 - Normalisation
 - Transformation
- Explorative analysis
 - Hierarchical cluster analysis
 - Principal component analysis
- Hypothesis testing
 - Differentially expressed genes
 - Predictive models
- Interpretation
 - Gene ontology
 - KEGG Pathways
 - Protein atlas
- Validation / confirmation
 - PCR of almost significant genes
 - PCR on new samples
 - Protein studies
 - Functional verification



Data analysis

(as presented with rat brain)

Explorative;

get a quick overview of the data and how gene expression in the different samples is related

Differential expression;

find genes enriched in each brain region

Functional analysis;

Find functional characteristics of genes enriched in different regions; are these similar or distinct?

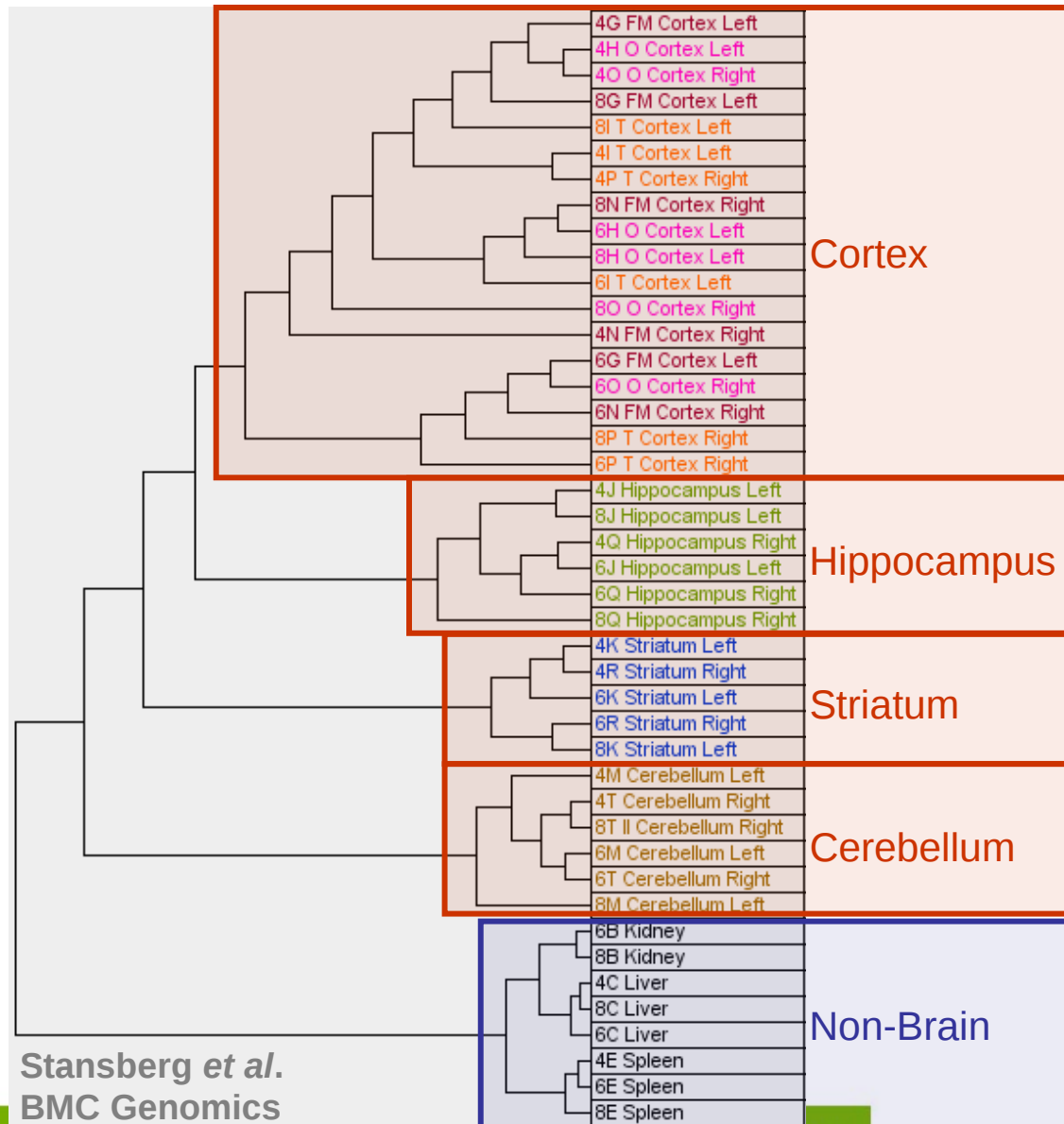
Could be enough, but interpretation, validation and follow up studies of selected genes or processes/pathways would be required for publication in most cases

And of course – did you find anything worth telling the world about?

Not all studies have interesting findings...



Explorative overview of data



microarray.no

Stansberg *et al.*
BMC Genomics

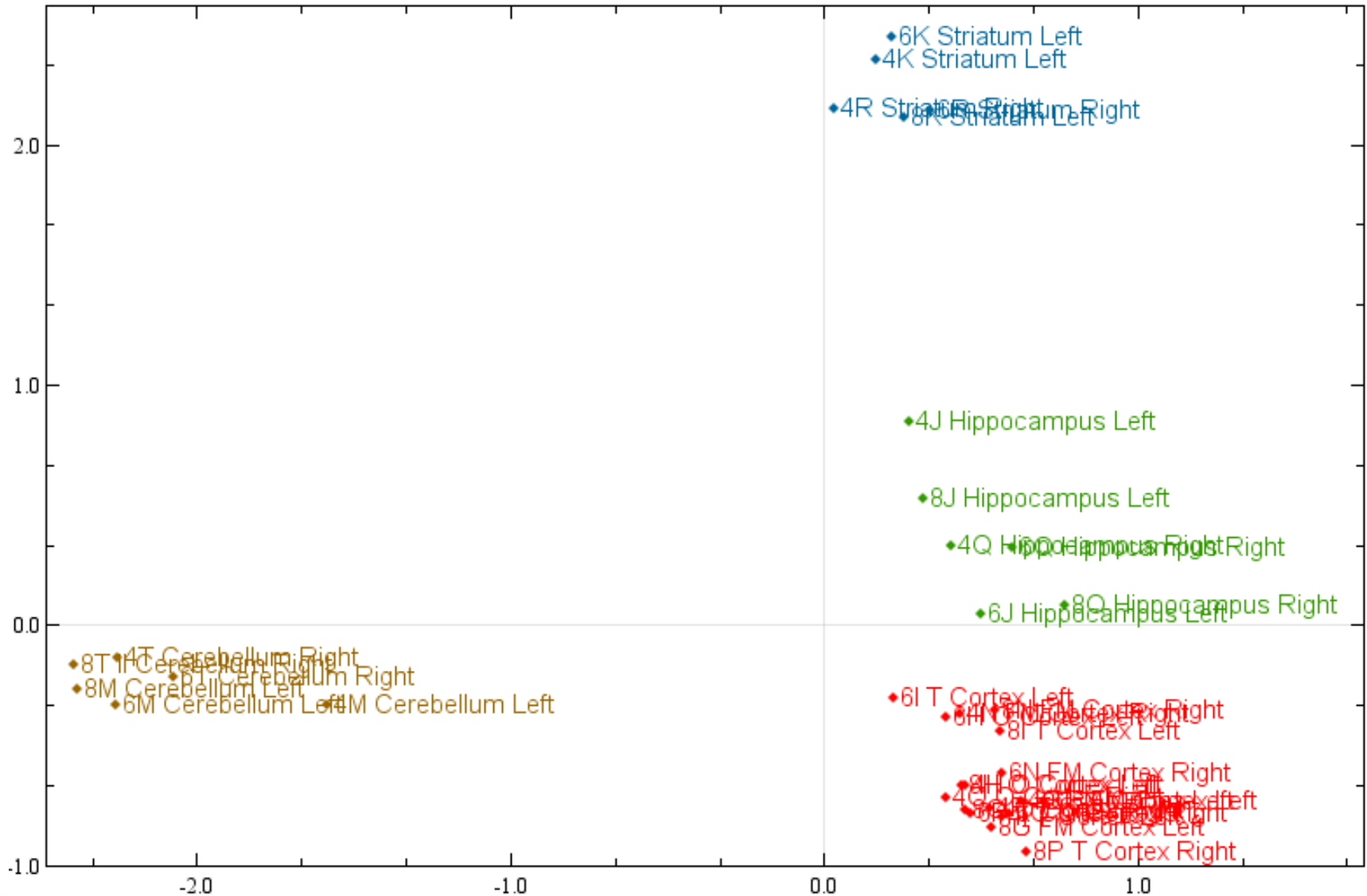
2007

www.uib.no





Explorative overview of data



microarray.no



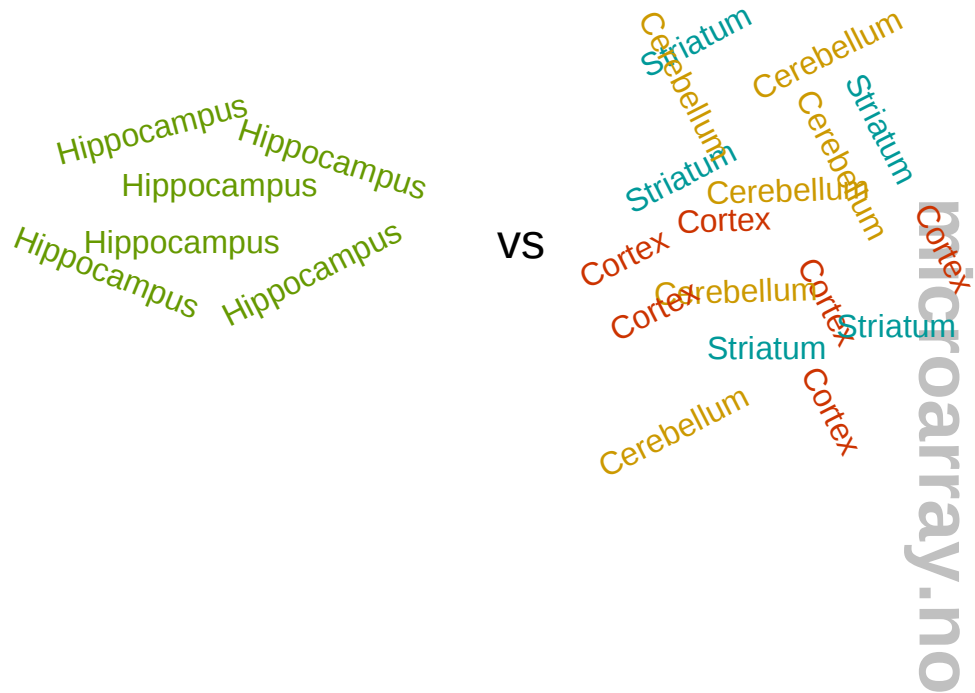


Finding regional genes

Differential expression

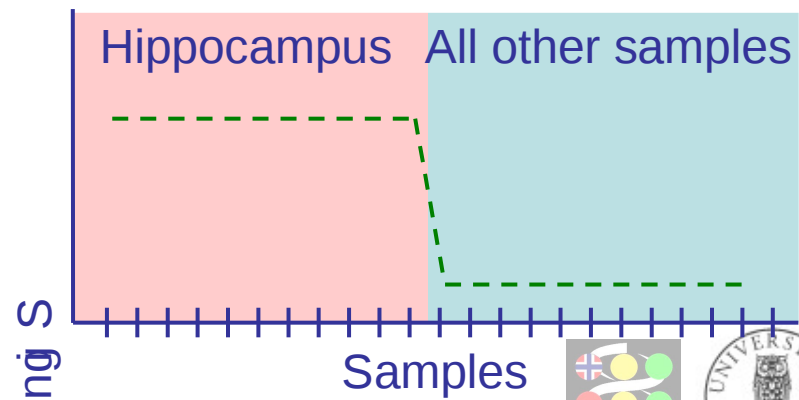


For each gene,
compare expression levels of all
samples in one region to all other
samples



i.e.

For each region,
find genes that show a higher
expression level in that region
than in all other samples



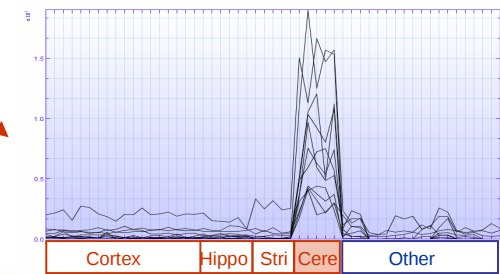
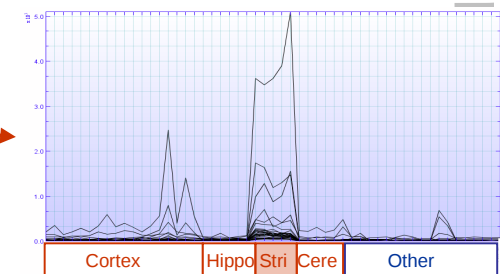
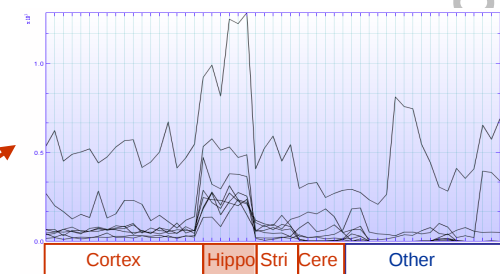
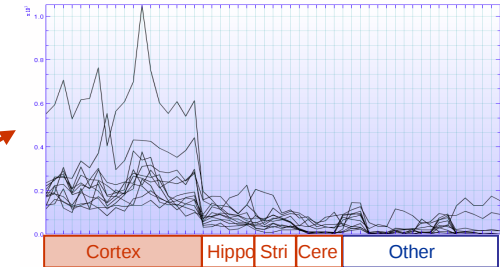
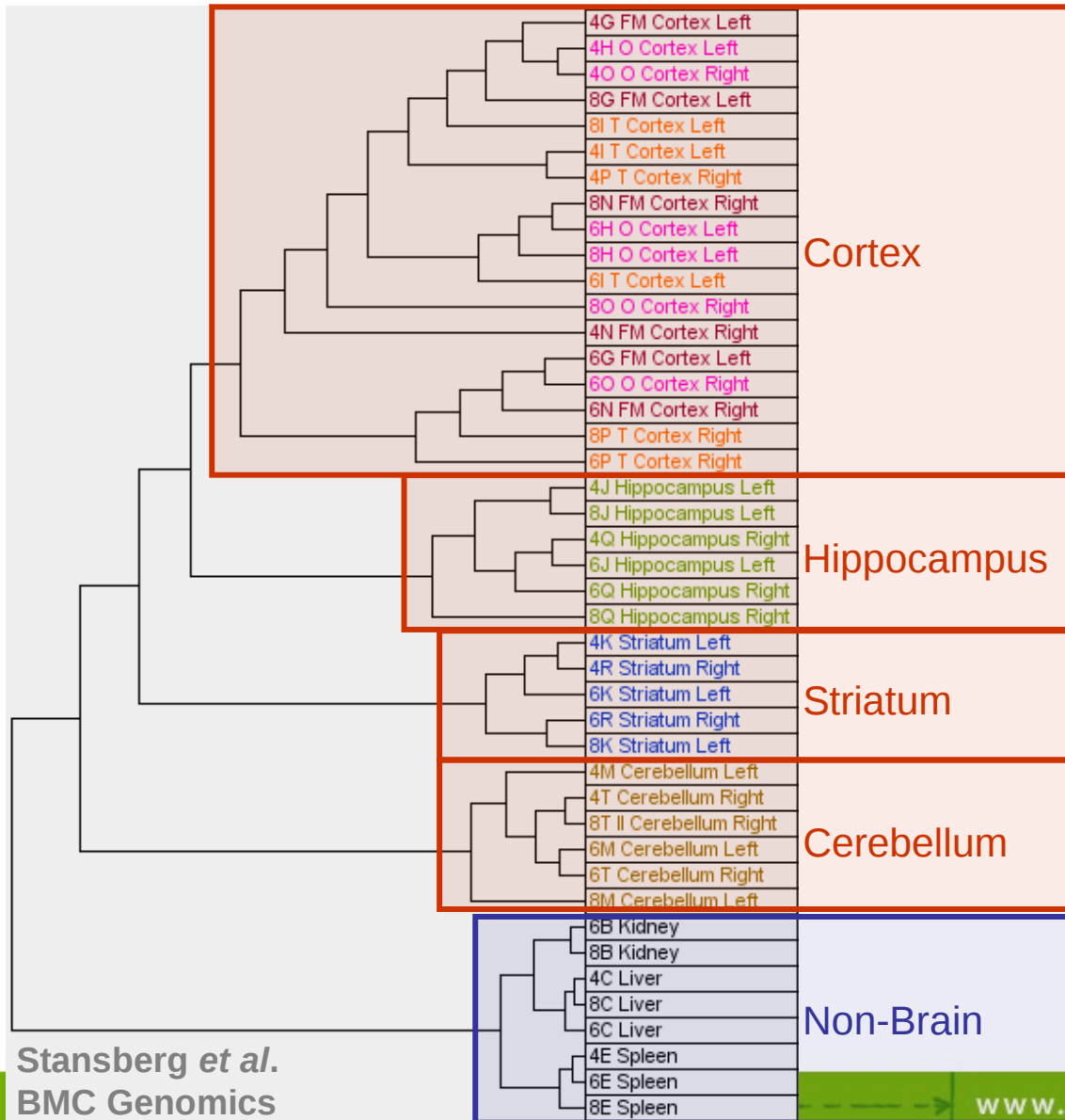
log₂ ang S

Samples





Finding regional genes





Functional analysis

Do genes enriched in one region have something in common, that may distinguish them from genes enriched in the other regions?

i.e.

do we see a correlation between regional gene expression patterns in the rat brain and specialised brain functions?

YES!





Regional patterns reveal functional specialisation



Ref (25170)	Cx (353)	HIF (189)	Str (314)	Cb (627)	Panther Biological Process
49.2	39.1	33.9	40.4	42.7	Biological process unclassified
4.8	7.9	9.5	8.9	6.9	Transport
2.3	6.5	7.9	7.3	4.8	Ion transport
2.4	9.3	8.5	10.2	5.6	Neuronal activities
1.1	3.4	3.2	3.2	3.2	Synaptic transmission
0.1	0.3	0.0	1.3	0.3	Action potential propagation
13.2	30.0	31.2	24.2	21.5	Signal transduction
6.6	14.4	13.2	12.7	7.3	Cell surface receptor mediated signal transduction
3.8	9.1	7.4	9.9	4.0	<i>G-protein mediated signaling</i>
3.2	5.7	9.0	6.4	7.5	Intracellular signaling cascade
0.5	1.4	3.7	1.3	1.8	<i>Calcium mediated signaling</i>
7.8	16.4	19.6	11.5	13.6	Developmental processes
2.6	6.2	10.1	3.8	7.0	Ectoderm development
2.3	5.9	9.0	3.5	6.9	<i>Neurogenesis</i>
3.3	4.2	9.5	3.8	3.8	Cell proliferation and differentiation
11.1	10.8	10.1	11.1	15.2	Nucleoside, nucleotide and nucleic acid metabolism
6.1	9.1	7.9	6.7	11.2	mRNA transcription
4.7	7.4	6.9	6.4	9.3	<i>mRNA transcription regulation</i>
0.2	0.6	0.0	2.2	0.5	Metabolism of cyclic nucleotides

Different signalling mechanisms are used, according to known physiology

microarray.no



What more can you do?

- Validation by independent method
 - e.g. QPCR or *in situ* hybridisation
 - preferably on independent samples
 - including both high- and low-intensity-genes, as well as genes with high and poorer significance.
- Are your genes translated?
 - Check by Western blotting or immunohistochemistry



What more can you do?

- Find as much info about your genes/processes/pathways as possible
 - Are they normally present in the tissue you are studying?
 - Does it make sense that they are differentially expressed?
 - How do your genes behave in other available data sets?
 - What factors have previously been shown to affect their expression?
 - Do your genes have any known interaction partners and how are the expression patterns of these?
- i.e. try to make a story of what may have happened in your system, supported by your original data, validations, follow-up studies and the literature.



Public repositories for microarray data

NCBI **GEO** Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO Not logged in | Login

Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

QUERY

DataSets

Gene profiles

GEO accession

GEO BLAST


BROWSE

DataSets

GEO accessions

Site contents

Public data

Platforms	8,555
Samples	542,058
Series 	22,013
DataSets	2,721

Documentation



The **ArrayExpress Archive** is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. **Gene Expression Atlas** contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.

Experiments Archive

20982 experiments, 596762 assays

Experiment, citation, sample and factor annotations

 [Browse experiments](#) | [platform designs](#) 

 [Submitter/reviewer login](#)

 [ArrayExpress Query Help](#)

Gene Expression Atlas

5667 experiments, 138864 assays, 18398 conditions

Genes

up/down in

Conditions

Any species

[Gene Expression Atlas Home](#)

microarray

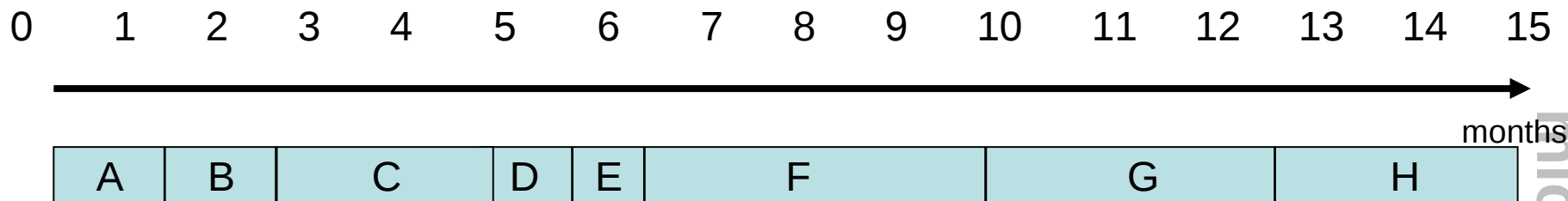
BASE

- Organise all your related files (images, quantitation numbers, protocols, etc) into an Experiment
- BASE helps you document all the details needed about each step (MIAME)
- Share data between collaborators
- The server have large disks with very good backup routines (nightly backup)
- It's easy to export data from BASE to analyse in other software tools such as J-Express, SciCraft, MEV etc.
- Export your data in standard format (MAGE-ML) required by public repositories like Array Express at EBI (UK) (Publicly available data is commonly a requirement for publication in more and more journals).





Timeline



A: Start planning the experiment, design meeting

B: Getting REK approval etc

C: Waiting for patients to be included, animals, optimizing experiment protocol....

D: Perform experiment, sampling and extraction of RNA

E: Microarray experiment- generation of data

F: Interpretation of data, dig into the biology, validation, additional experiments...

G: Make figures, write paper

H: Revise paper, prepare data for export to f.ex Array Express. Documentation!





Good luck analysing your data!

Online help with analysis:

<http://www.microarray.no/magma>

<http://www.molmine.com/forum>

Workshop mailinglist: nmc-analysis-ws@bioinfo.no

Contact information:

Bergen: kjell@microarray.no

Oslo: solveig@microarray.no

microarray.no

