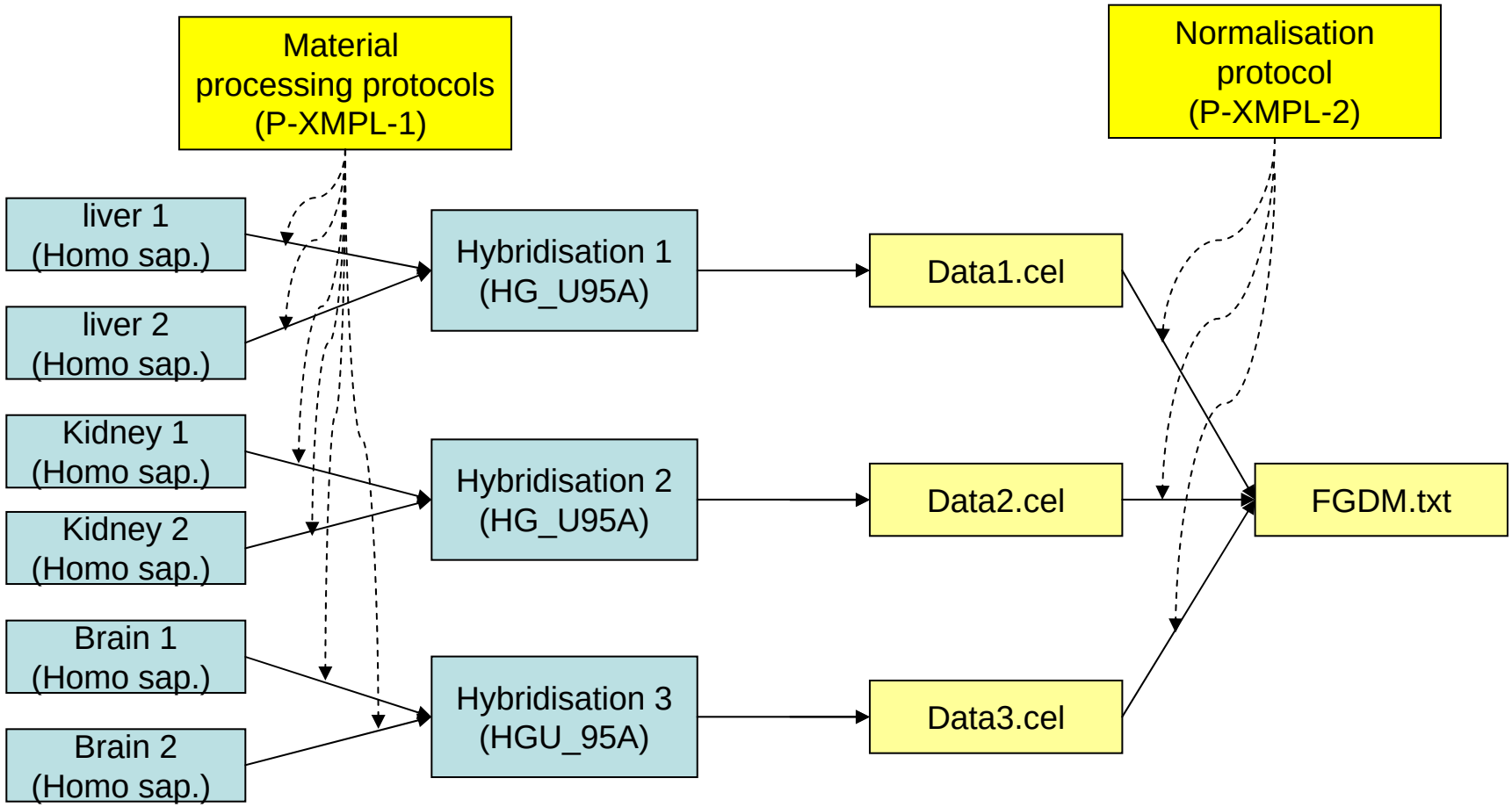
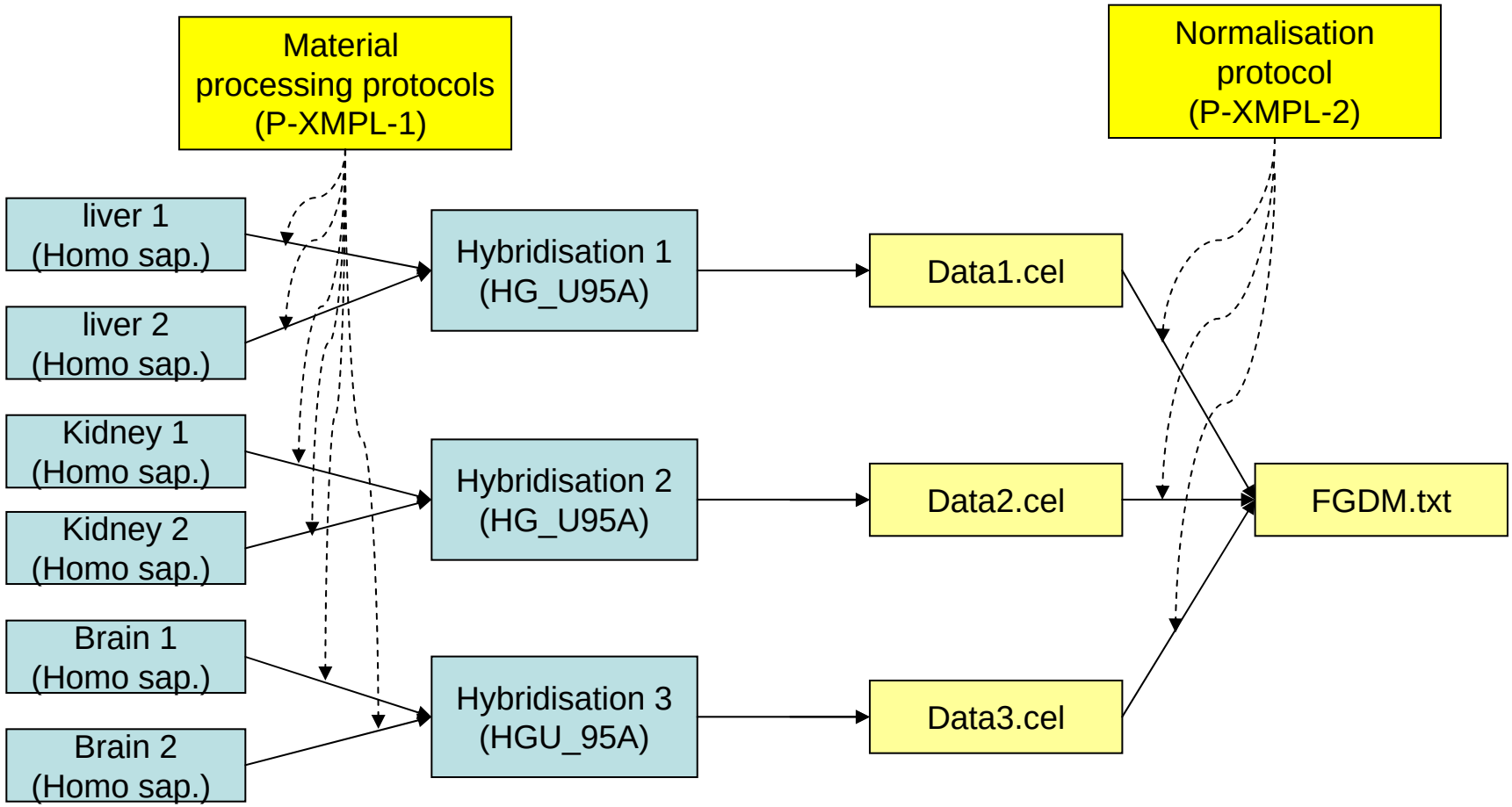
MAGE-simple or MAGE-TAB

-a platform specific implementation of
MAGE-OM v2 (simple layer)

MAGE workshop, NCI

May 8, 2006


Material processing protocols (P-XMPL-1)
Normalisation protocol (P-XMPL-2)
liver 1 (Homo sap.)
liver 2 (Homo sap.)
Kidney 1 (Homo sap.)
Kidney 2 (Homo sap.)
Brain 1 (Homo sap.)
Brain 2 (Homo sap.)
Hybridisation 1 (HG_U95A)
Hybridisation 2 (HG_U95A)
Hybridisation 3 (HGU_95A)
Data1.cel
Data2.cel
Data3.cel
FGDM.txt



Sample ID	Characteristics[Organism]	Characteristics[OrganismPart]	Protocol REF	Hybridization ID	ArrayDesign REF	ArrayData URI	Protocol REF	DerivedArrayData URI
liver 1	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
liver 2	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
kidney 1	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
kidney 2	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
brain 1	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt
brain 2	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt

MIAME

Minimum **I**nformation **A**bout a **M**icroarray **E**xperiment that is needed to interpret its results unambiguously and potentially to reproduce the experiment

MIAME checklist

- Array design description
  - What is where on the array and what biological objects (e.g., genes) we want to measure
- Experiment
  - Experiment design – what is the relationship between samples, arrays, datafiles
  - Each sample – e.g., organism, body part, developmental stage
  - Hybridisation procedures, protocols, parameters
  - Data, data transformation protocols (algorithms), parameters

nature

26 September 2002 Volume 419 Issue no 6905

Microarray standards at last

Not a moment too soon, the microarray community has issued guidelines that will make their data much more useful and accessible. *Nature* and the Nature research journals will respond accordingly.

You read a paper with a fascinating conclusion about the expression of several genes. You decide to use some of the same experiments on your system of choice. But when you wade through hundreds of pages of supplementary information, you find that crucial details needed for replication are missing.

Welcome to the exciting but frustrating world of DNA microarray research. Microarrays are plastic or glass chips spotted with tiny amounts of thousands of probes, used to query the activity levels of that many genes in any tissue or organism at one time. Variables in every step of the experiment often make cross-paper comparison virtually impossible. Microarray papers also pose a considerable strain on the refereeing process; the vast amounts of data mean that critical review is a monumental task.

Yet referees sometimes feel they are not given enough details, leading cautious reviewers to think that they must reanalyse the primary data set. In other cases, the primary data provided are in proprietary software and so are impossible to comment on. Many journals allowed authors to put the huge data files on their own websites for the review process, until it became clear that unscrupulous authors compromised the anonymity of referees by tracking who had visited the website.

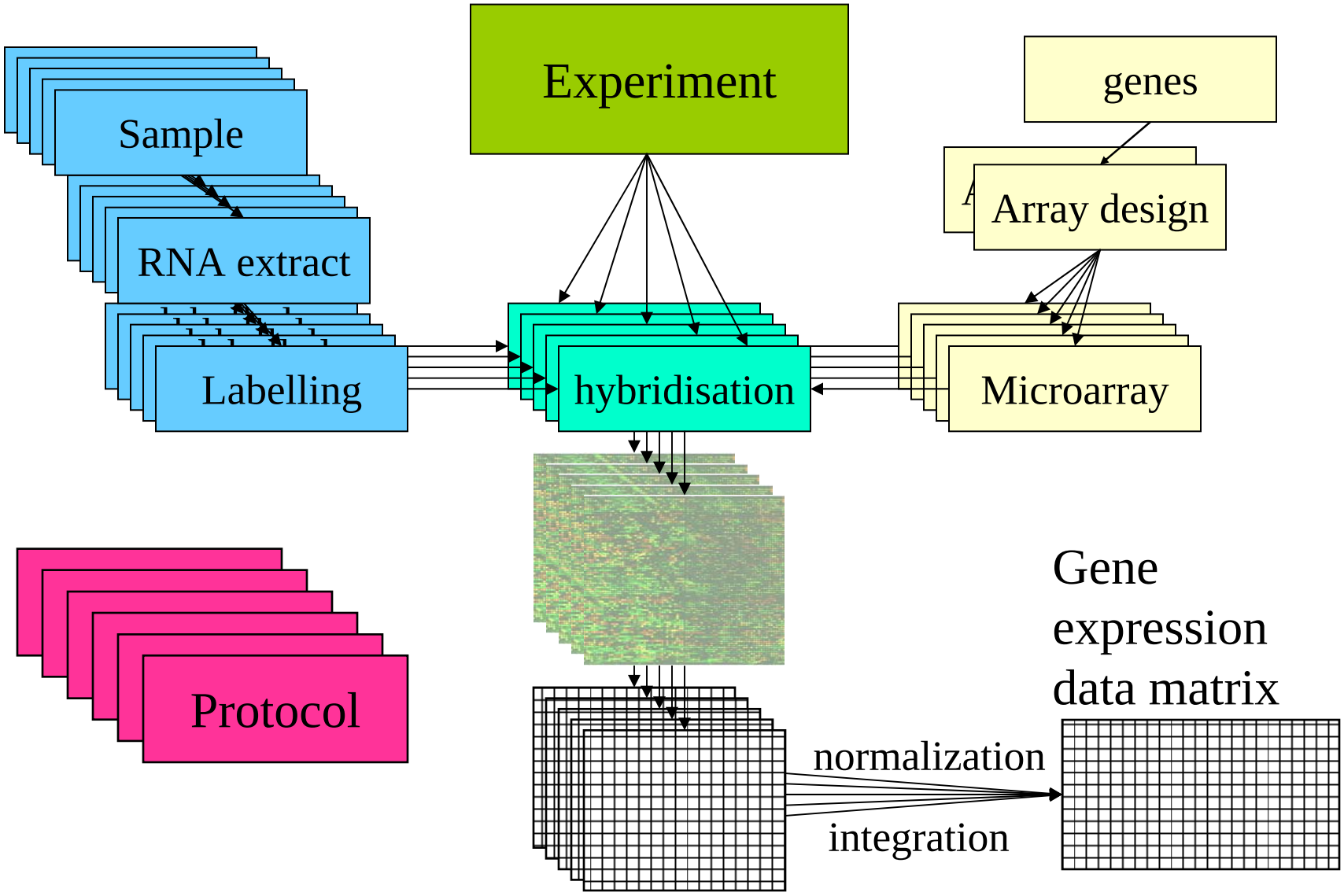
In a move to remedy these problems, the international Microarray Gene Expression Data (MGED) group has written an open letter to scientific journals proposing standards for publication. Other members of the microarray community welcomed these steps, designed to clarify the Minimal Information About a Microarray Experiment (MIAME) guidelines (*Nature Genetics* **29**, 365–371; 2001).

For authors, the proposal provides a checklist of variables that should be included in every microarray publication, at http://www.mged.org/Workgroups/MIAME/miame_checklist.html. This checklist, with all variables completed, would be supplied as supplementary information at the time of submission. The MGED group suggests that journals require submission of microarray data to either of two databases emerging as the main public repositories: GEO (www.ncbi.nlm.nih.gov/geo/) or ArrayExpress (www.ebi.ac.uk/arrayexpress).

Harried editors can rejoice that, at last, the community is taming the unruly beast that is microarray information. Therefore, all submissions to *Nature* and the Nature family of journals received on or after 1 December containing new microarray experiments must include the mailing of five compact disks to the editor. These disks should include necessary information compliant with the MIAME standard. The information must be supplied in a format that could be read by widely available software packages. Data integral to the paper's conclusions should be submitted to the ArrayExpress or GEO databases, with accession numbers where available, supplied at or before acceptance for publication.

How much data should authors provide to the community? Specifically, do other researchers really need to recreate the exact microarray just to test the expression level of a few key genes, which could presumably be done through other methods? Perhaps with further evolution and standardization of microarray technology, the need to specify so many variables will decrease, but the MGED standards are surely appropriate for the current state of the field. ■

MIAME is not a format, but…



Some important MIAME concepts

- **Experimental design** – a graph or a table representing relationships between biomaterials, e.g., which sample goes on which array

Microarray experiment

Samples

Extracts

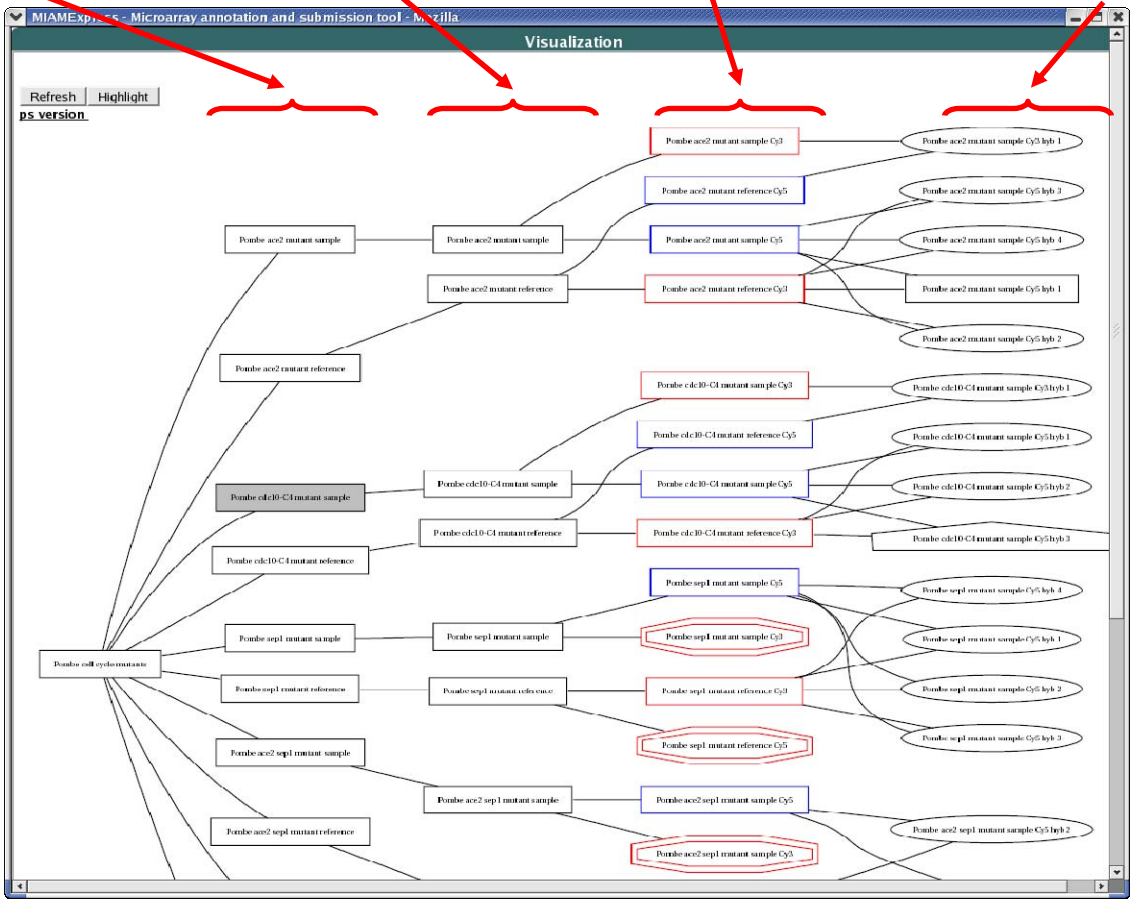
Labelled Extracts

Colours related to labels

Hybridizations

Shapes related to array designs

Experiment



Rustici et al., *S. pombe* cell-cycle mutant data (2004)

Some important MIAME concepts

- **Experimental design** – a graph or a table representing relationships between biomaterials, e.g., which sample goes on which array
- **Experimental factors** – the most important variables studied in the experiment, e.g., time in the time course, does, compound, in a dose response element, etc

MIAME is not a format, but

- MIAME defines an experiment structure
- **A precisely defined format for** communicating data between LIMS, databases, data analysis tools **is needed** to cope with the growing amounts of data
- **MAGE-ML** provides such means and is largely successful, but too heavy for small users without dedicated software support, particularly until good tools are develped

MAGE-TAB goals

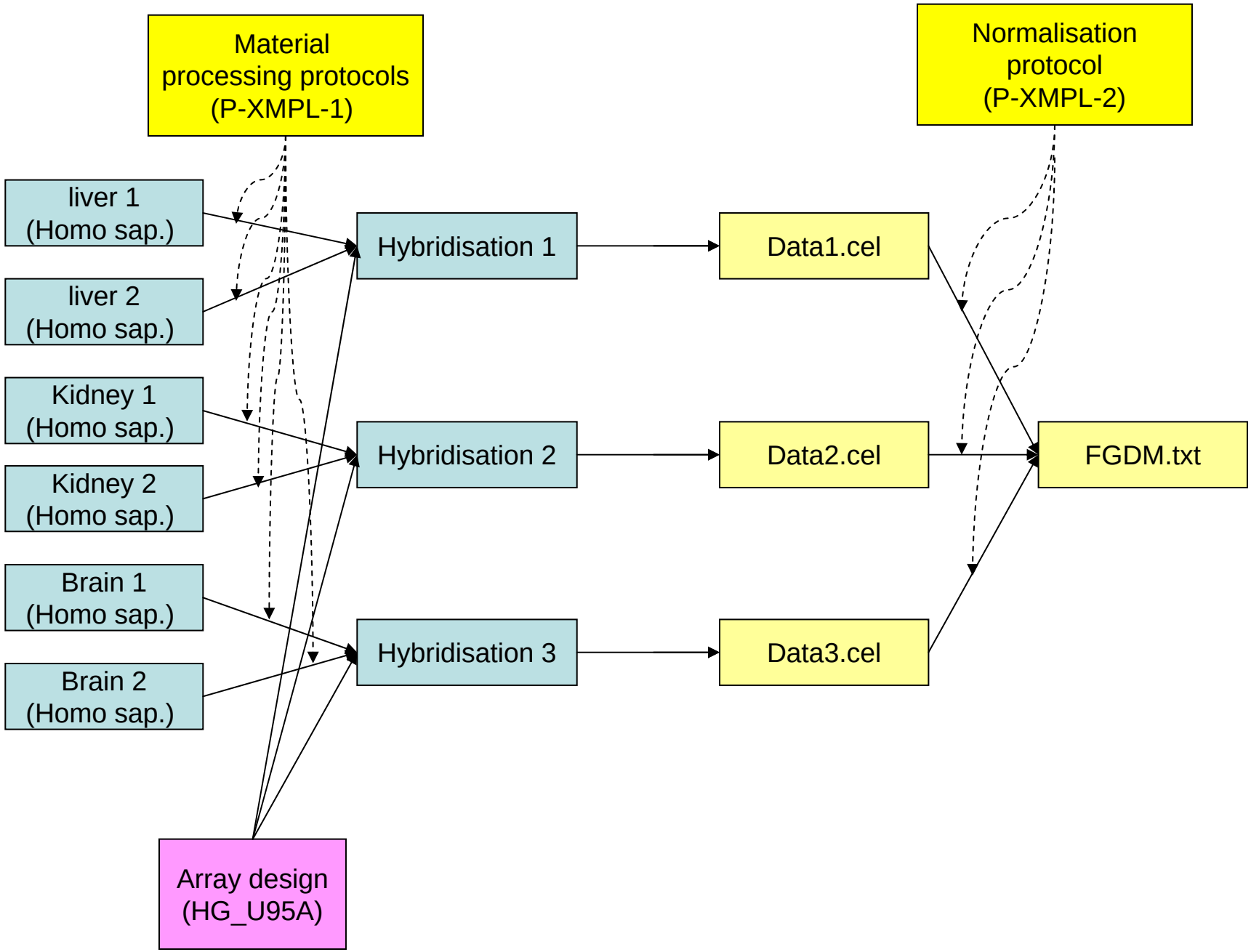
- A simple MIAME supportive format (it should be possible to describe all MIAME requirements in an explicit way)
- It should be possible to describe any microarray experiment
- It should be easy and natural to describe the most popular experimental designs
- It should be possible to create, edit and read files in this format without any specialised software

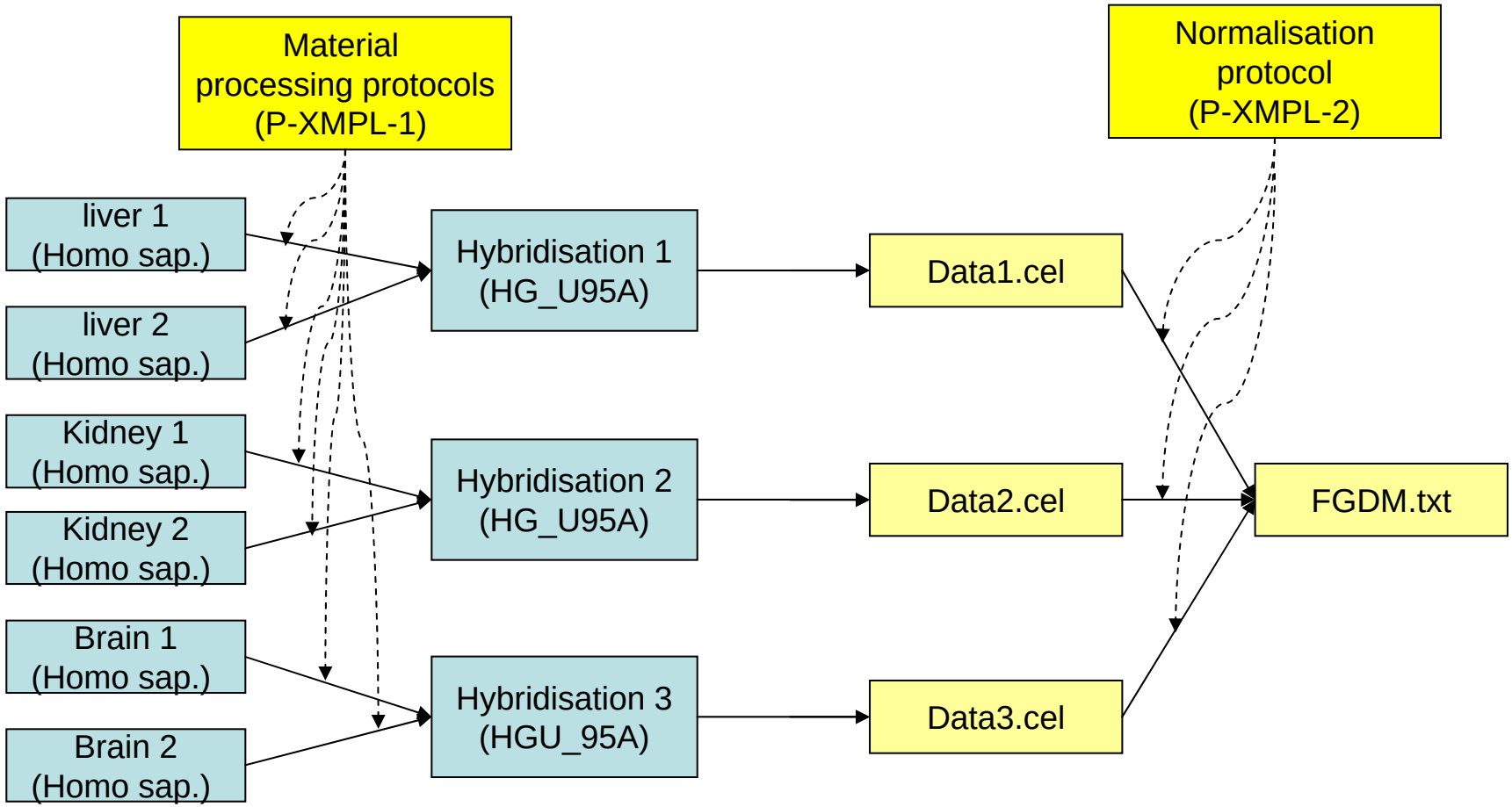
MAGE-TAB in a nutshel

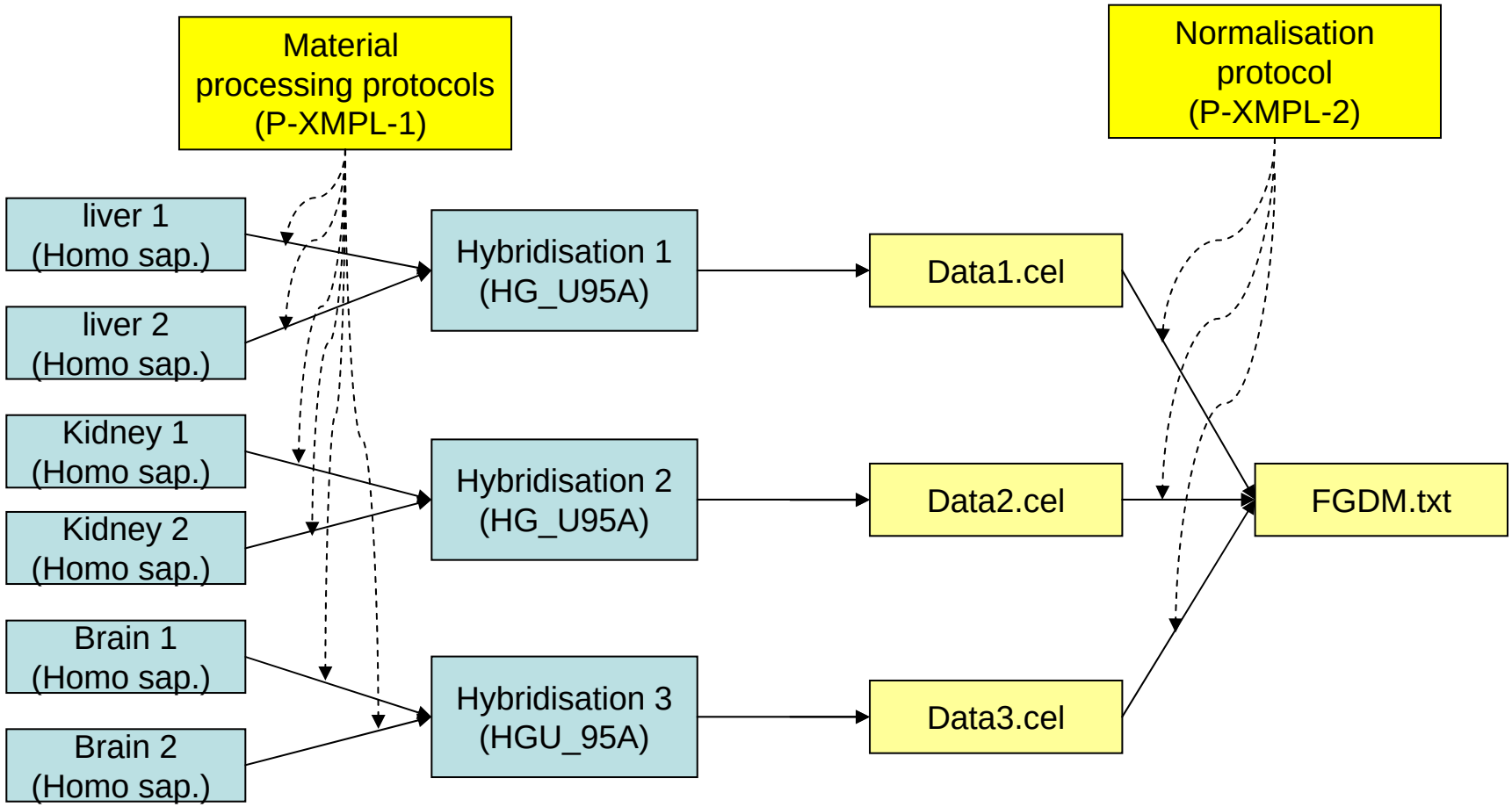
- A ‘spread-sheet’ based format for describing microarray experiments
- Description consists of four types of documents
  - General information about the experiment
  - Coding the experiment (sample) design
  - Coding the array design (ADF)
  - Data – raw in the native formats, processed – a spreadsheet where rows represent ‘genes’ (or any objects measured on the array), and each column refer to one or more ‘samples’

MAGE-TAB in a nutshel

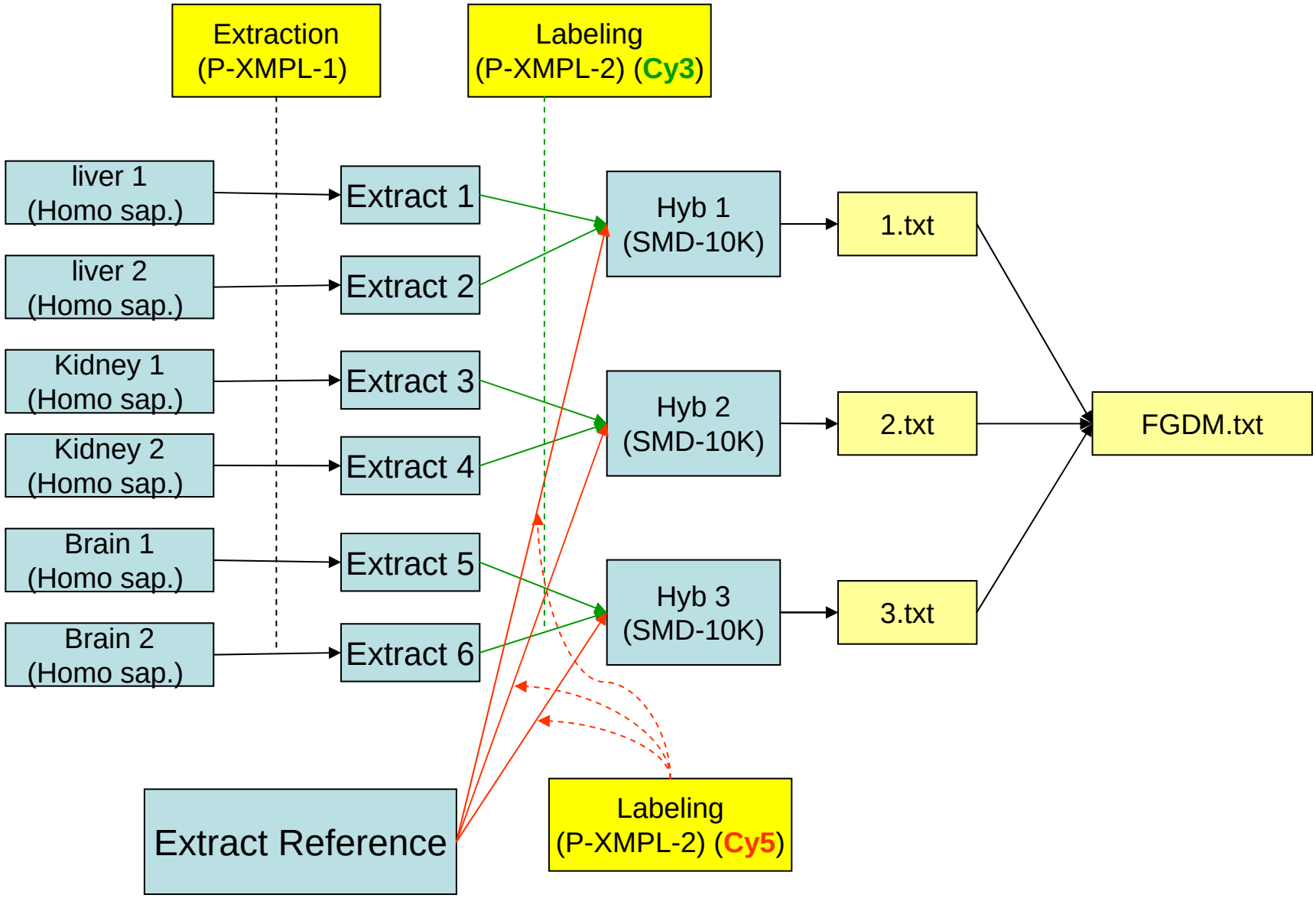
- A ‘spread-sheet’ based format for describing microarray experiments
- Description consists of four types of documents
  - General information about the experiment
  - **Coding the experiment (sample) design - EDF**
  - Coding the array design (ADF)
  - Data – raw in the native formats, processed – a spreadsheet where rows represent ‘genes’ (or any objects measured on the array), and each column refer to one or more ‘samples’


Material processing protocols (P-XMPL-1)
Normalisation protocol (P-XMPL-2)
liver 1 (Homo sap.)
liver 2 (Homo sap.)
Kidney 1 (Homo sap.)
Kidney 2 (Homo sap.)
Brain 1 (Homo sap.)
Brain 2 (Homo sap.)
Hybridisation 1
Hybridisation 2
Hybridisation 3
Data1.cel
Data2.cel
Data3.cel
FGDM.txt
Array design (HG_U95A)


Material processing protocols (P-XMPL-1)
Normalisation protocol (P-XMPL-2)
liver 1 (Homo sap.)
liver 2 (Homo sap.)
Kidney 1 (Homo sap.)
Kidney 2 (Homo sap.)
Brain 1 (Homo sap.)
Brain 2 (Homo sap.)
Hybridisation 1 (HG_U95A)
Hybridisation 2 (HG_U95A)
Hybridisation 3 (HGU_95A)
Data1.cel
Data2.cel
Data3.cel
FGDM.txt



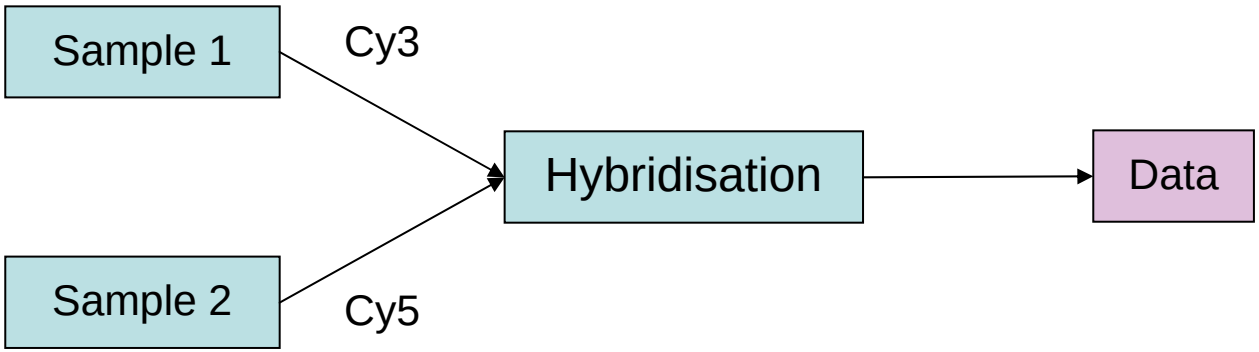
Sample ID	Characteristics[Organism]	Characteristics[OrganismPart]	Protocol REF	Hybridization ID	ArrayDesign REF	ArrayData URI	Protocol REF	DerivedArrayData URI
liver 1	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
liver 2	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
kidney 1	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
kidney 2	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
brain 1	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt
brain 2	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt


Extraction
(P-XMPL-1)
Labeling
(P-XMPL-2) (Cy3)
liver 1
(Homo sap.)
liver 2
(Homo sap.)
Kidney 1
(Homo sap.)
Kidney 2
(Homo sap.)
Brain 1
(Homo sap.)
Brain 2
(Homo sap.)
Extract 1
Extract 2
Extract 3
Extract 4
Extract 5
Extract 6
Hyb 1
(SMD-10K)
Hyb 2
(SMD-10K)
Hyb 3
(SMD-10K)
1.txt
2.txt
3.txt
FGDM.txt
Extract Reference
Labeling
(P-XMPL-2) (Cy5)

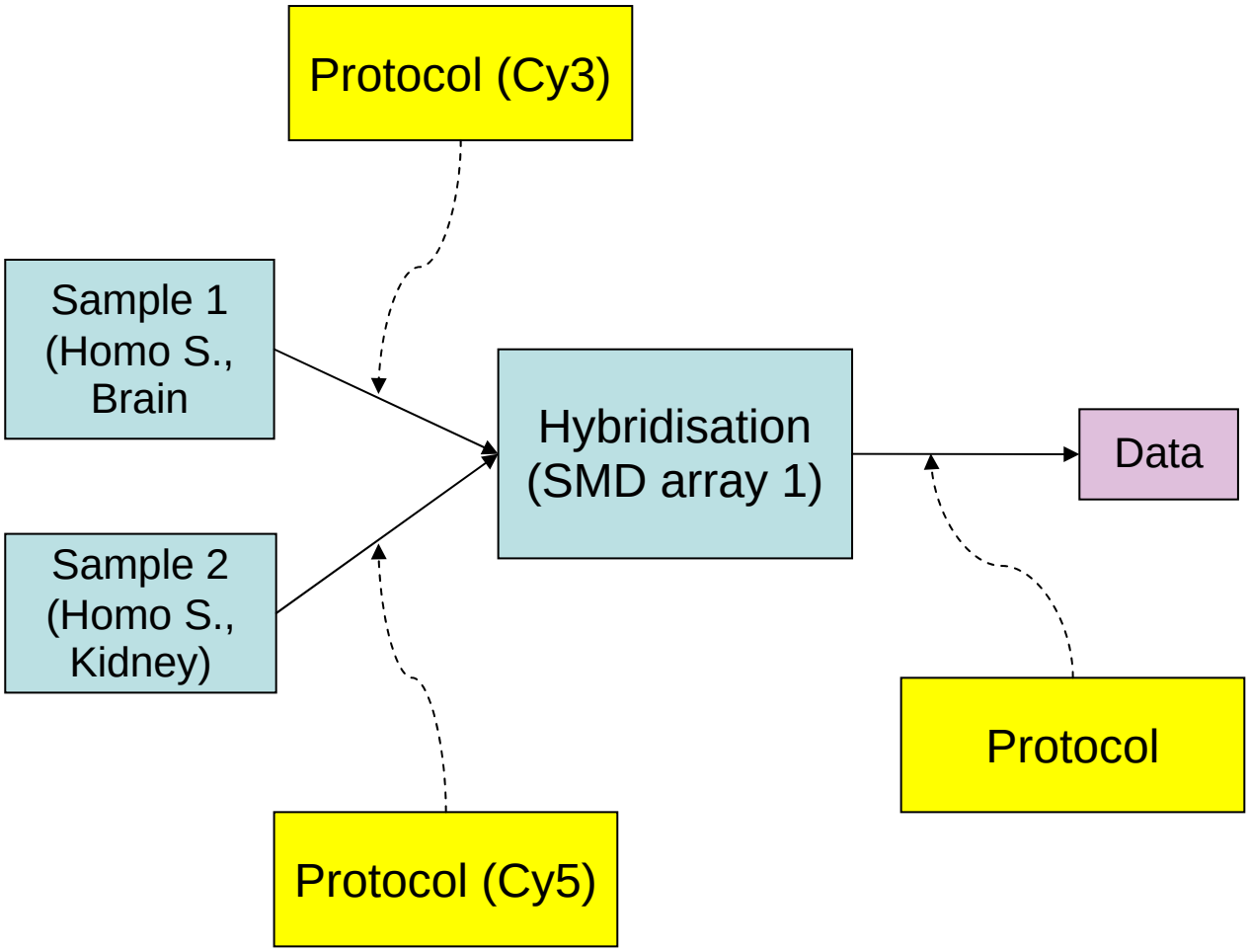
MAGE-TAB – dual channel iterated design

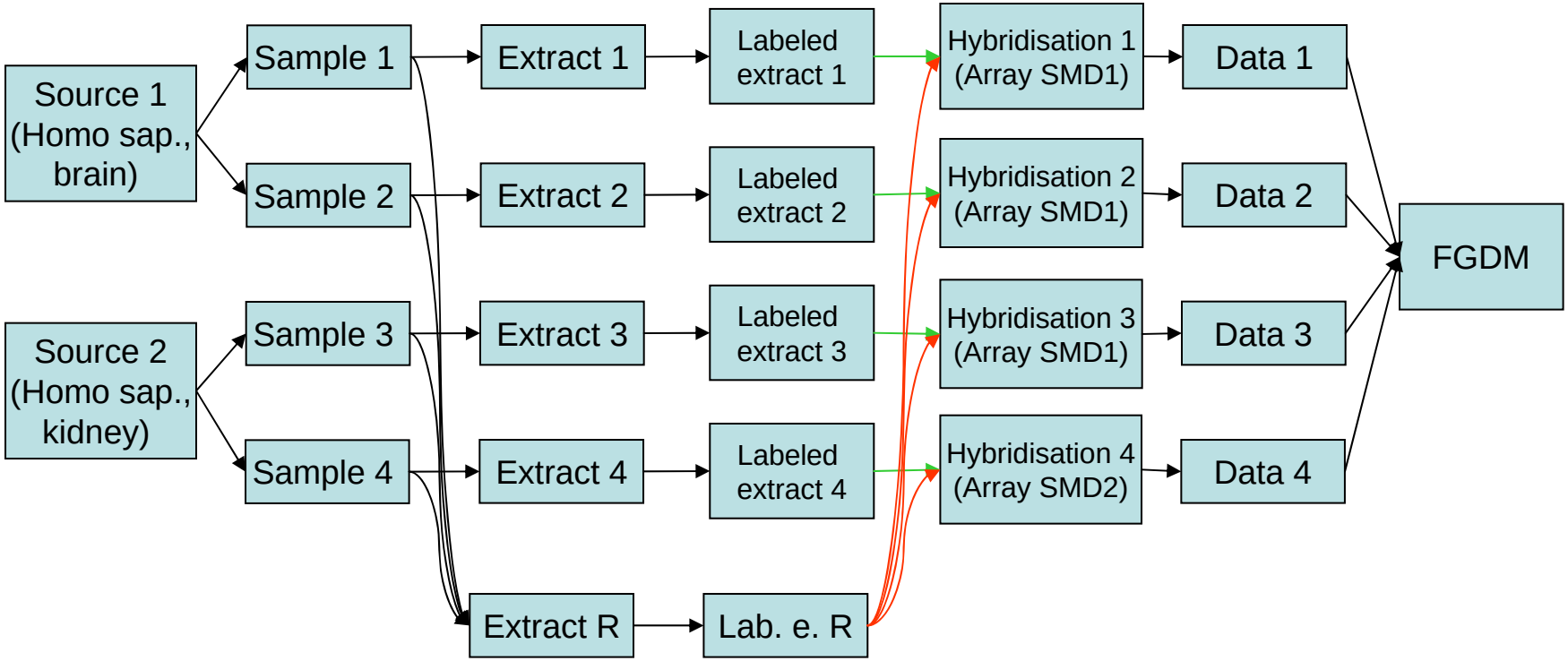
Source ID	Characteristics[Organism]	Characteristics[OrganismPart]	Protocol REF	Extract ID	Label	Protocol REF	Hybridization ID	ArrayDesign REF	ArrayData URI	DerivedArrayData URI
liver 1	Homo sapiens	liver	P-XMPL-1	extract 1	Cy3	P-XMPL-2	hyb 1	SMD_10k_Print3	1.txt	FGEM.txt
liver 2	Homo sapiens	liver	P-XMPL-1	extract 2	Cy3	P-XMPL-2	hyb 1	SMD_10k_Print3	1.txt	FGEM.txt
kidney 1	Homo sapiens	kidney	P-XMPL-1	extract 3	Cy3	P-XMPL-2	hyb 2	SMD_10k_Print3	2.txt	FGEM.txt
kidney 2	Homo sapiens	kidney	P-XMPL-1	extract 4	Cy3	P-XMPL-2	hyb 2	SMD_10k_Print3	2.txt	FGEM.txt
brain 1	Homo sapiens	brain	P-XMPL-1	extract 5	Cy3	P-XMPL-2	hyb 3	SMD_10k_Print3	3.txt	FGEM.txt
brain 2	Homo sapiens	brain	P-XMPL-1	extract 6	Cy3	P-XMPL-2	hyb 3	SMD_10k_Print3	3.txt	FGEM.txt
				reference	Cy5	P-XMPL-2	ALL			

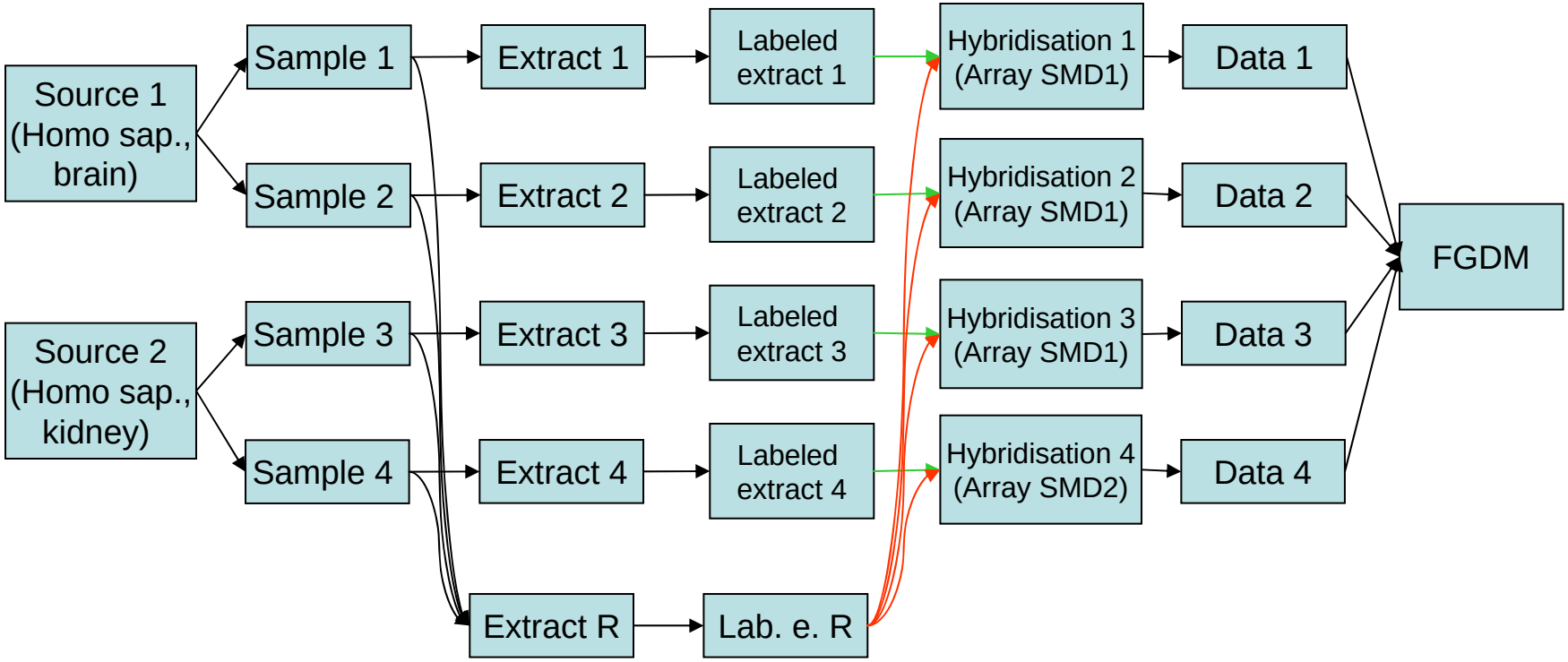
Experiment design graph



Experiment design graph

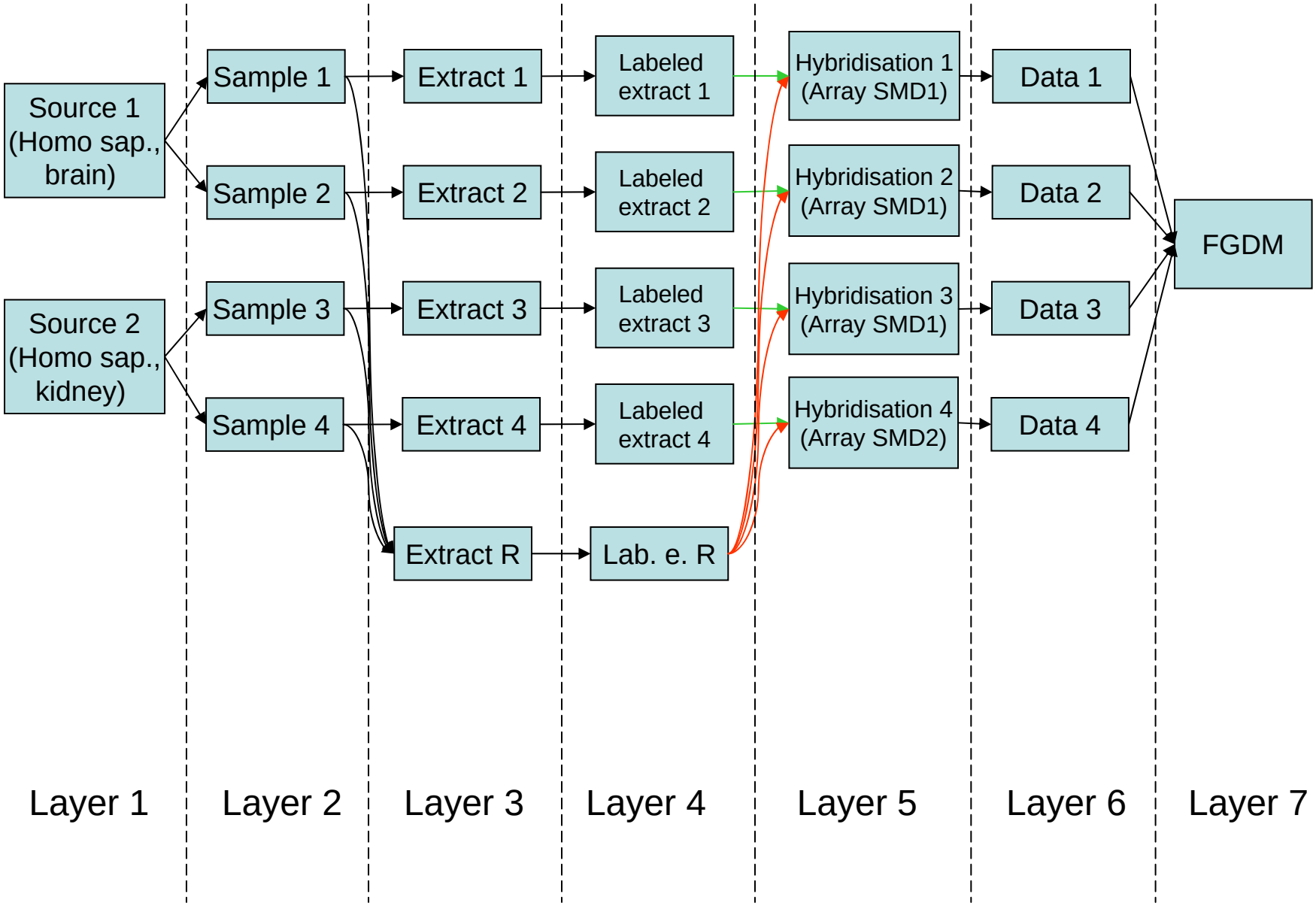



Source 1 (Homo sap., brain)
Source 2 (Homo sap., kidney)
Sample 1
Sample 2
Sample 3
Sample 4
Extract 1
Extract 2
Extract 3
Extract 4
Extract R
Labeled extract 1
Labeled extract 2
Labeled extract 3
Labeled extract 4
Lab. e. R
Hybridisation 1 (Array SMD1)
Hybridisation 2 (Array SMD1)
Hybridisation 3 (Array SMD1)
Hybridisation 4 (Array SMD2)
Data 1
Data 2
Data 3
Data 4
FGDM

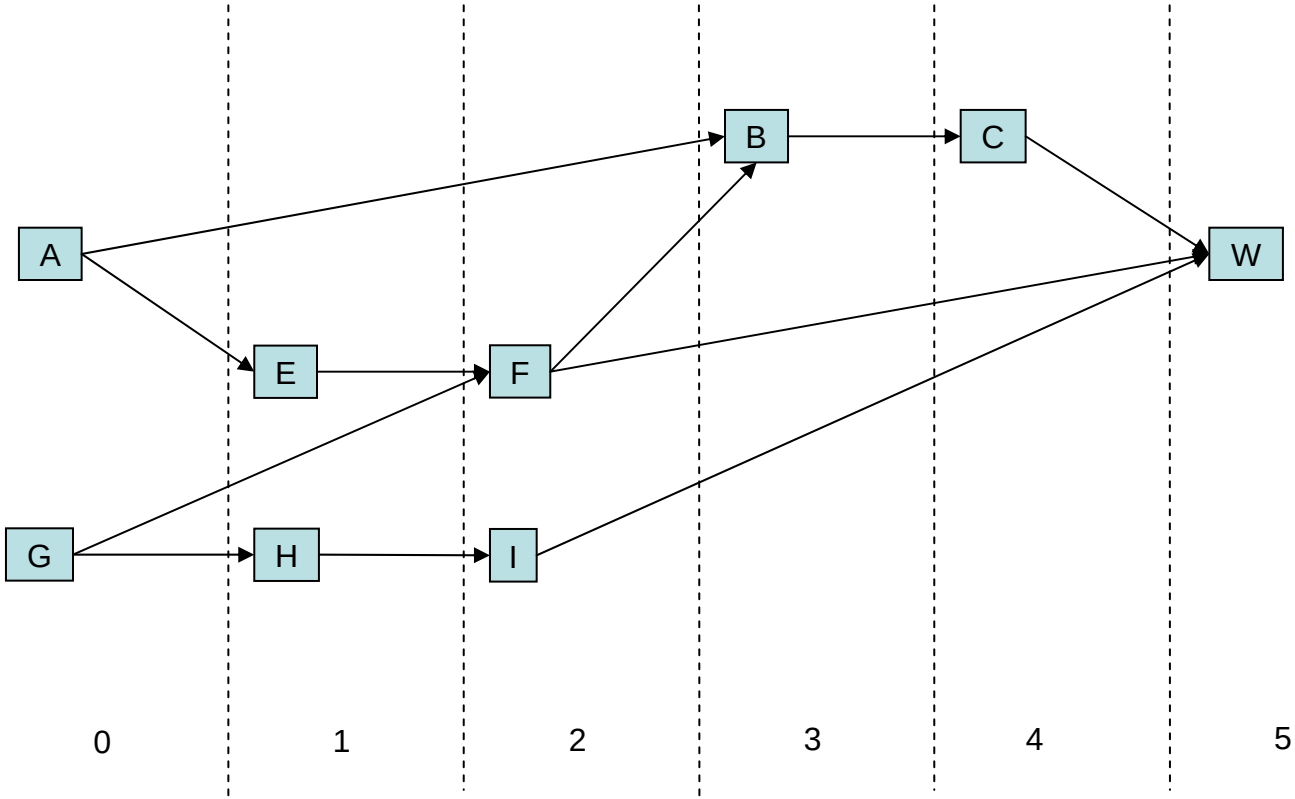


Array design graphs representing high throughput experiments are

1. **Regular** (the same structure is repeated)
2. **Small in and out-degree**, except 'reference' nodes
3. **Layered** – it is possible to group nodes in consecutive layers of each containing nodes of the same type


Source 1 (Homo sap., brain)
Source 2 (Homo sap., kidney)
Sample 1
Sample 2
Sample 3
Sample 4
Extract 1
Extract 2
Extract 3
Extract 4
Extract R
Labeled extract 1
Labeled extract 2
Labeled extract 3
Labeled extract 4
Lab. e. R
Hybridisation 1 (Array SMD1)
Hybridisation 2 (Array SMD1)
Hybridisation 3 (Array SMD1)
Hybridisation 4 (Array SMD2)
Data 1
Data 2
Data 3
Data 4
FGDM
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7

Layers in a more complex DAG



It is natural to represent such ‘regular layered’ DAGs by spreadsheets

- Each column in the spreadsheet represents a layers
- Each raw in the spreadsheet represents a column

The representation of experiment design graphs that we have in practice is rather simple

MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

Investigation Title	University of Heidelberg H sapiens TK6	
Experimental Designs	genetic_modification_design	
Experimental Factors	genetic_modification	
Person Last Name	Maier	
Person First Name	Patrick	
Person Email	patrick.maier@radonk.ma.uni-heidelberg.de	
Person Phone	+496213833773	
Person Address	Theodor-Kutzer-Ufer 1-3	
Person Affiliation	Department of Radiation Oncology, University of Heidelberg	
Person Roles	submitter	investigator
Quality Control Types	biological_replicate	
Replicate Types	biological_replicate	
Date of Experiment	2005-02-28	
Public Release Date	2006-01-03	
PubMed ID	12345678	
Publication Author List	Patrick Maier; Katharina Fleckenstein; Li Li; Stephanie Laufs; Jens Zeller; Stefan Fruehauf; Carsten Herskind; Frederik Wenz	
Publication Status	submitted	
Experiment Description	Gene expression of TK6 cells transduced with an oncoretrovirus expressing MDR1 (TK6MDR1) was compared to untransduced TK6 cells and to TK6 cell transduced with an oncoretrovirus expressing the Neomycin resistance gene (TK6neo). Two biological replicates of each were generated and the expression profiles were determined using Affymetrix Human Genome U133 Plus2.0 GeneChip microarrays. Comparisons between the sample groups allow the identification of genes with expression dependent on the MDR1 overexpression.	
Protocol Name	GROWTHPRTCL10653	
Protocol Type	grow	
Protocol Description	TK6 cells were grown in suspension cultures in RPMI 1640 medium supplemented with 10% horse serum (Invitrogen, Karlsruhe, Germany). The cells were routinely maintained at 37 C and 5% CO2.	
Protocol Parameters	media	
Protocol Name	EXTPRTCL10654	
Protocol Type	nucleic_acid_extraction	
Protocol Description	Approximately 10^6 cells were lysed in RLT buffer (Qiagen).Total RNA was extracted from the cell lysate using an RNeasy kit (Qiagen).	
Protocol Parameters	Extracted Product	Amplification
Protocol Name	TRANPRTCL10656	
Protocol Type	bioassay_data_transformation	
Protocol Description	Mixed Model Normalization with SAS Micro Array Solutions (version 1.3).	
EDF Files	e-mexp-428_tab.txt	

MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

Block Column	Block Row	Column	Row	Reporter ID	Reporter Sequence	Role	Control Type	Composite Element ID	CompositeElement DatabaseEntry [refseq]
1	1	1	1	R1	ATGGTTGGTTACGTGT	experimental		PTEN	NM_000314
1	1	1	2	R2	CCGCGTTGCCCCGCC	experimental		PAX2	NM_003989
1	1	1	3	R3	CGTAGCTGATCGATGA	experimental		WWOX	NM_016373
1	1	1	4	R4	GGTTGGCTGAGATCGT	experimental		MAPK8	NM_139047
1	1	2	1	R1	ATGGTTGGTTACGTGT	experimental		PTEN	NM_000314
1	1	2	2	R2	CCGCGTTGCCCCGCC	experimental		PAX2	NM_003989
1	1	2	3	R3	CGTAGCTGATCGATGA	experimental		WWOX	NM_016373
1	1	2	4	R4	GGTTGGCTGAGATCGT	experimental		MAPK8	NM_139047
4	6	20	20	462020	TCCCTTCCGTTGTCCT	control	**control_spike_calibration**		

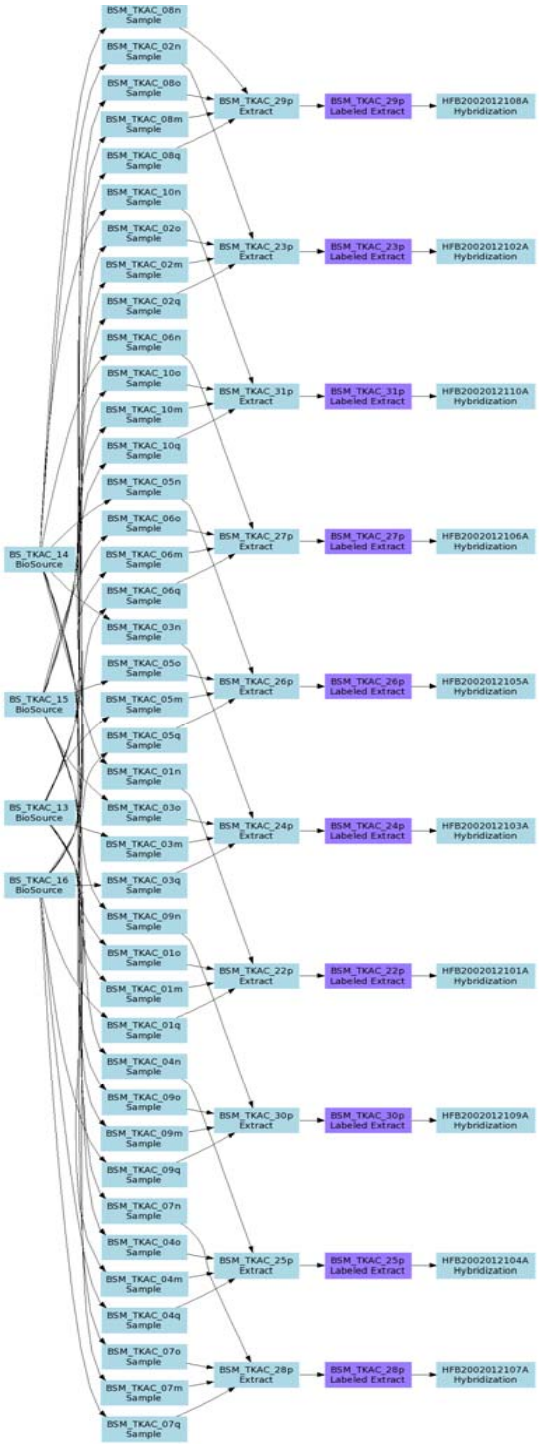
MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

Loop design with dye swap

Loop design with dye swap

Source ID	Sample ID	Extract ID	LabeledExtract ID	Label	Hybridization ID
Source 1	Sample 1	Extract 1	LabeledExtract 1a	Cy3	1-Cy3 vs 2-Cy5
Source 2	Sample 2	Extract 2	LabeledExtract 2b	Cy5	1-Cy3 vs 2-Cy5
Source 1	Sample 1	Extract 1	LabeledExtract 1c	Cy5	1-Cy5 vs 2-Cy3
Source 2	Sample 2	Extract 2	LabeledExtract 2d	Cy3	1-Cy5 vs 2-Cy3
Source 2	Sample 2	Extract 2	LabeledExtract 2a	Cy3	2-Cy3 vs 3-Cy5
Source 3	Sample 3	Extract 3	LabeledExtract 3b	Cy5	2-Cy3 vs 3-Cy5
Source 2	Sample 2	Extract 2	LabeledExtract 2c	Cy5	2-Cy5 vs 3-Cy3
Source 3	Sample 3	Extract 3	LabeledExtract 3d	Cy3	2-Cy5 vs 3-Cy3
Source 1	Sample 1	Extract 1	LabeledExtract 1b	Cy5	3-Cy3 vs 1-Cy5
Source 3	Sample 3	Extract 3	LabeledExtract 3a	Cy3	3-Cy3 vs 1-Cy5
Source 1	Sample 1	Extract 1	LabeledExtract 1d	Cy3	3-Cy5 vs 1-Cy3
Source 3	Sample 3	Extract 3	LabeledExtract 3c	Cy5	3-Cy5 vs 1-Cy3


BS_TKAC_14
BioSource
BS_TKAC_15
BioSource
BS_TKAC_13
BioSource
BS_TKAC_16
BioSource
BSM_TKAC_08n
Sample
BSM_TKAC_02n
Sample
BSM_TKAC_08o
Sample
BSM_TKAC_08m
Sample
BSM_TKAC_08q
Sample
BSM_TKAC_10n
Sample
BSM_TKAC_02o
Sample
BSM_TKAC_02m
Sample
BSM_TKAC_02q
Sample
BSM_TKAC_06n
Sample
BSM_TKAC_10o
Sample
BSM_TKAC_10m
Sample
BSM_TKAC_10q
Sample
BSM_TKAC_05n
Sample
BSM_TKAC_06o
Sample
BSM_TKAC_06m
Sample
BSM_TKAC_06q
Sample
BSM_TKAC_03n
Sample
BSM_TKAC_05o
Sample
BSM_TKAC_05m
Sample
BSM_TKAC_05q
Sample
BSM_TKAC_01n
Sample
BSM_TKAC_03o
Sample
BSM_TKAC_03m
Sample
BSM_TKAC_03q
Sample
BSM_TKAC_09n
Sample
BSM_TKAC_01o
Sample
BSM_TKAC_01m
Sample
BSM_TKAC_01q
Sample
BSM_TKAC_04n
Sample
BSM_TKAC_09o
Sample
BSM_TKAC_09m
Sample
BSM_TKAC_09q
Sample
BSM_TKAC_07n
Sample
BSM_TKAC_04o
Sample
BSM_TKAC_04m
Sample
BSM_TKAC_04q
Sample
BSM_TKAC_07o
Sample
BSM_TKAC_07m
Sample
BSM_TKAC_07q
Sample
BSM_TKAC_29p
Extract
BSM_TKAC_23p
Extract
BSM_TKAC_31p
Extract
BSM_TKAC_27p
Extract
BSM_TKAC_26p
Extract
BSM_TKAC_24p
Extract
BSM_TKAC_22p
Extract
BSM_TKAC_30p
Extract
BSM_TKAC_25p
Extract
BSM_TKAC_28p
Extract
BSM_TKAC_29p
Labeled Extract
BSM_TKAC_23p
Labeled Extract
BSM_TKAC_31p
Labeled Extract
BSM_TKAC_27p
Labeled Extract
BSM_TKAC_26p
Labeled Extract
BSM_TKAC_24p
Labeled Extract
BSM_TKAC_22p
Labeled Extract
BSM_TKAC_30p
Labeled Extract
BSM_TKAC_25p
Labeled Extract
BSM_TKAC_28p
Labeled Extract
HFB2002012108A
Hybridization
HFB2002012102A
Hybridization
HFB2002012110A
Hybridization
HFB2002012106A
Hybridization
HFB2002012105A
Hybridization
HFB2002012103A
Hybridization
HFB2002012101A
Hybridization
HFB2002012109A
Hybridization
HFB2002012104A
Hybridization
HFB2002012107A
Hybridization

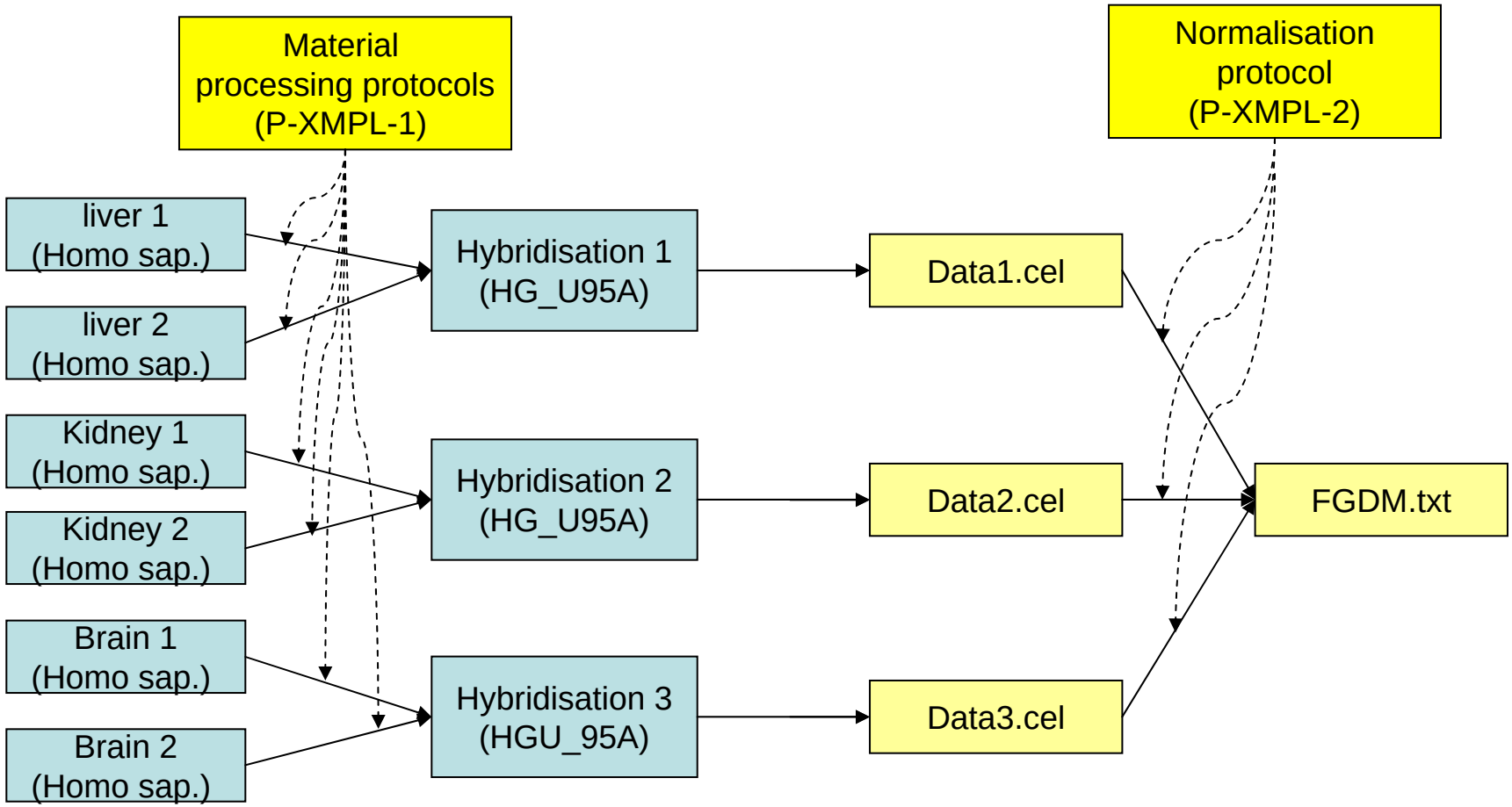
Source ID	Sample ID	Extract ID	LabeledExtract ID	Label	Hybridization ID
BS_TKAC_13	BSM_TKAC_01m	BSM_TKAC_22p	BSM_TKAC_22p	biotin	HFB2002012101A
BS_TKAC_13	BSM_TKAC_02m	BSM_TKAC_23p	BSM_TKAC_23p	biotin	HFB2002012102A
BS_TKAC_13	BSM_TKAC_03m	BSM_TKAC_24p	BSM_TKAC_24p	biotin	HFB2002012103A
BS_TKAC_13	BSM_TKAC_04m	BSM_TKAC_25p	BSM_TKAC_25p	biotin	HFB2002012104A
BS_TKAC_13	BSM_TKAC_05m	BSM_TKAC_26p	BSM_TKAC_26p	biotin	HFB2002012105A
BS_TKAC_13	BSM_TKAC_06m	BSM_TKAC_27p	BSM_TKAC_27p	biotin	HFB2002012106A
BS_TKAC_13	BSM_TKAC_07m	BSM_TKAC_28p	BSM_TKAC_28p	biotin	HFB2002012107A
BS_TKAC_13	BSM_TKAC_08m	BSM_TKAC_29p	BSM_TKAC_29p	biotin	HFB2002012108A
BS_TKAC_13	BSM_TKAC_09m	BSM_TKAC_30p	BSM_TKAC_30p	biotin	HFB2002012109A
BS_TKAC_13	BSM_TKAC_10m	BSM_TKAC_31p	BSM_TKAC_31p	biotin	HFB2002012110A
BS_TKAC_14	BSM_TKAC_01n	BSM_TKAC_22p	BSM_TKAC_22p	biotin	HFB2002012101A
BS_TKAC_14	BSM_TKAC_02n	BSM_TKAC_23p	BSM_TKAC_23p	biotin	HFB2002012102A
BS_TKAC_14	BSM_TKAC_03n	BSM_TKAC_24p	BSM_TKAC_24p	biotin	HFB2002012103A
BS_TKAC_14	BSM_TKAC_04n	BSM_TKAC_25p	BSM_TKAC_25p	biotin	HFB2002012104A
BS_TKAC_14	BSM_TKAC_05n	BSM_TKAC_26p	BSM_TKAC_26p	biotin	HFB2002012105A
BS_TKAC_14	BSM_TKAC_06n	BSM_TKAC_27p	BSM_TKAC_27p	biotin	HFB2002012106A
BS_TKAC_14	BSM_TKAC_07n	BSM_TKAC_28p	BSM_TKAC_28p	biotin	HFB2002012107A
BS_TKAC_14	BSM_TKAC_08n	BSM_TKAC_29p	BSM_TKAC_29p	biotin	HFB2002012108A
BS_TKAC_14	BSM_TKAC_09n	BSM_TKAC_30p	BSM_TKAC_30p	biotin	HFB2002012109A
BS_TKAC_14	BSM_TKAC_10n	BSM_TKAC_31p	BSM_TKAC_31p	biotin	HFB2002012110A
BS_TKAC_15	BSM_TKAC_01o	BSM_TKAC_22p	BSM_TKAC_22p	biotin	HFB2002012101A
BS_TKAC_15	BSM_TKAC_02o	BSM_TKAC_23p	BSM_TKAC_23p	biotin	HFB2002012102A
BS_TKAC_15	BSM_TKAC_03o	BSM_TKAC_24p	BSM_TKAC_24p	biotin	HFB2002012103A
BS_TKAC_15	BSM_TKAC_04o	BSM_TKAC_25p	BSM_TKAC_25p	biotin	HFB2002012104A
BS_TKAC_15	BSM_TKAC_05o	BSM_TKAC_26p	BSM_TKAC_26p	biotin	HFB2002012105A
BS_TKAC_15	BSM_TKAC_06o	BSM_TKAC_27p	BSM_TKAC_27p	biotin	HFB2002012106A
BS_TKAC_15	BSM_TKAC_07o	BSM_TKAC_28p	BSM_TKAC_28p	biotin	HFB2002012107A
BS_TKAC_15	BSM_TKAC_08o	BSM_TKAC_29p	BSM_TKAC_29p	biotin	HFB2002012108A
BS_TKAC_15	BSM_TKAC_09o	BSM_TKAC_30p	BSM_TKAC_30p	biotin	HFB2002012109A
BS_TKAC_15	BSM_TKAC_10o	BSM_TKAC_31p	BSM_TKAC_31p	biotin	HFB2002012110A
BS_TKAC_16	BSM_TKAC_01q	BSM_TKAC_22p	BSM_TKAC_22p	biotin	HFB2002012101A
BS_TKAC_16	BSM_TKAC_02q	BSM_TKAC_23p	BSM_TKAC_23p	biotin	HFB2002012102A
BS_TKAC_16	BSM_TKAC_03q	BSM_TKAC_24p	BSM_TKAC_24p	biotin	HFB2002012103A
BS_TKAC_16	BSM_TKAC_04q	BSM_TKAC_25p	BSM_TKAC_25p	biotin	HFB2002012104A
BS_TKAC_16	BSM_TKAC_05q	BSM_TKAC_26p	BSM_TKAC_26p	biotin	HFB2002012105A
BS_TKAC_16	BSM_TKAC_06q	BSM_TKAC_27p	BSM_TKAC_27p	biotin	HFB2002012106A
BS_TKAC_16	BSM_TKAC_07q	BSM_TKAC_28p	BSM_TKAC_28p	biotin	HFB2002012107A
BS_TKAC_16	BSM_TKAC_08q	BSM_TKAC_29p	BSM_TKAC_29p	biotin	HFB2002012108A
BS_TKAC_16	BSM_TKAC_09q	BSM_TKAC_30p	BSM_TKAC_30p	biotin	HFB2002012109A
BS_TKAC_16	BSM_TKAC_10q	BSM_TKAC_31p	BSM_TKAC_31p	biotin	HFB2002012110A

MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

Data files

- Raw data – native formats (cel, genpix)
- Normalised, summarised data – columns may be individual references



Sample ID	Characteristics[Organism]	Characteristics[OrganismPart]	Protocol REF	Hybridization ID	ArrayDesign REF	ArrayData URI	Protocol REF	DerivedArrayData URI
liver 1	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
liver 2	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
kidney 1	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
kidney 2	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
brain 1	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt
brain 2	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt

Sample ID	Characteristics[Organism]	Characteristics[OrganismPart]	Protocol REF	Hybridization ID	ArrayDesign REF	ArrayData URI	Protocol REF	DerivedArrayData URI
liver 1	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
liver 2	Homo sapiens	liver	P-XMPL-1	hyb 1	HG_U95A	1.CEL	P-XMPL-2	FGEM.txt
kidney 1	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
kidney 2	Homo sapiens	kidney	P-XMPL-1	hyb 2	HG_U95A	2.CEL	P-XMPL-2	FGEM.txt
brain 1	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt
brain 2	Homo sapiens	brain	P-XMPL-1	hyb 3	HG_U95A	3.CEL	P-XMPL-2	FGEM.txt

FGEM.txt:

Genes	Hyb 1 (liver)	Hyb 2 (kidney)	Hyb 3 (brain)
Gene 1			
Gene 2			
Gene 3			
….			
Gene n			

Experimental Factors

- Most important experimental variables (e.g., organs – liver, kidney, brain – in the examples above)
- Any column in the EDF can be marked as an experimental factor
- These can be propagated down the edges of the graph to columns in FGEM
- In FGEM they serve as concise annotation

Sample Properties - Microsoft Internet Explorer

File Edit View Favorites Tools Help

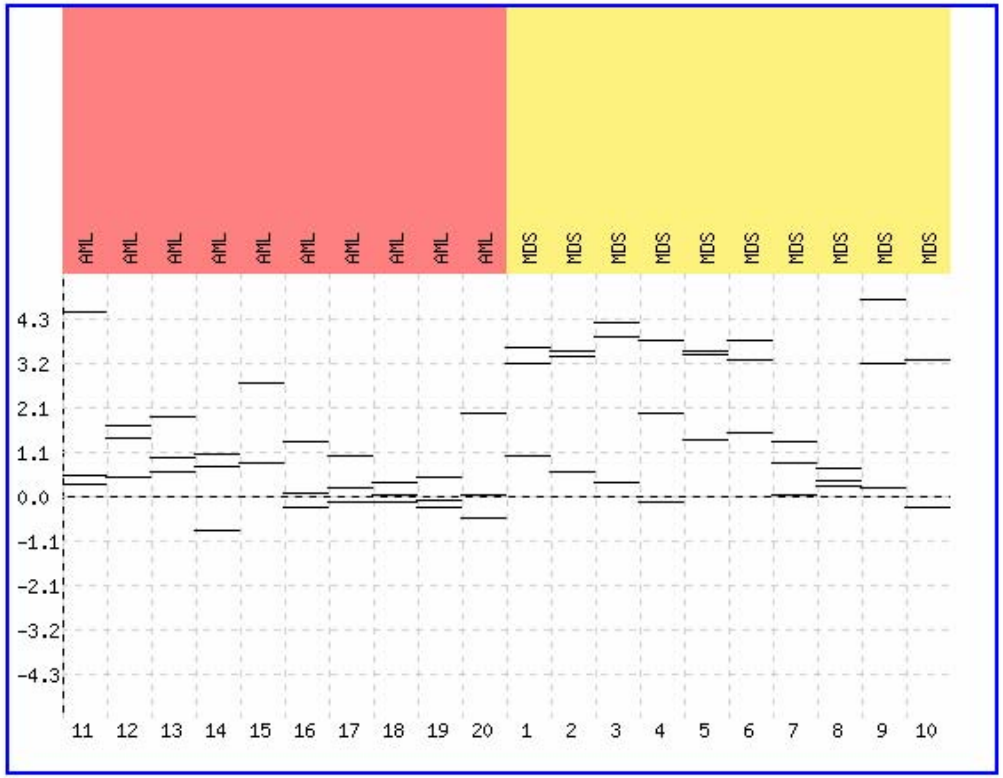
Back Search Favorites

Address http://www.ebi.ac.uk/aedw/DW?queryType=13BioSample&11experiment_id_list.query=10944204&nextPage=BioSample_List.vm&session=3767AF21B9BADAF8ABD7379D6FB80154&expId=109 Go

BRCA1 ENSG00000012048 cell cycle checkpoint , molecular function unknown , cellular component unknown

E-MEXP-25 experimental factors and gene plots

diseasestate



E-MEXP-25 sample properties

Disease state

Internet

MAGE-TAB

- Investigation design file – IDF
- Array design file ADF
- Experiment (sample) design file EDF (SDF)
- Data files

Standard design templates

- simple iterated design;
- iterated design with technical replicates;
- iterated design with pooling;
- iterated designs for dual channel experiments;
- dual channel iterated designs with dye swap;
- dual channel iterated designs with a reference sample;
- dual channel iterated design with a reference and dye swap;
- dual channel iterated design with a pooled reference;
- loop design;
- loop design with die swap;
- time series experiments;

http://www.mged.org/Workgroups/MAGE/mage.html#mage-tab

MAGE-TAB

- Any experiment can be represented in MAGE-TAB in a structured way to MIAME granularity
- Large experiments with a regular design can be represented in a natural way
- It should be possible to create MAGE-TAB files using generic spread-sheet software

It is proposed that MAGE-TAB becomes a platform specific implementation of MAGE

For discussion

- EDF or SDF – Experiment design graph – asks for EDF?
- What to do with IDF then? – Header File? (HF)
- Biomaterial and data objects (source, sample, extract, … - subtypes)?
- All microarray experimental designs can be described by MAGE-TAB! Which not?
- Standard QTs for processed data

Whenever a major organization develops a new system as an official standard for X, the primary result is the widespread adoption of some simpler system as a de facto standard for X

The Law of Standards by John F. Sowa (http://www.jfsowa.com/computer/standard.htm) –

Acknowledgements

- Tim Rayner
- Cathy Ball (IDF)
- Philippe Rocca-Sera (ADF)
- Michael Miller, Helen Parkinson, Ugis Sarkans, Paul Spellman, Anna Farne
- MAGE working group

Funding - NHGRI/NIBIB, MGED sponsors