

Eurexpress II

Overview of image and data production and transmission

Eurexpress II

Overview of data production and transmission

1. Table of contents

1.	Table of contents	2
2.	Preface.....	3
3.	E-mail Addresses of Eurexpress Informatics Contact Persons	4
4.	Data transmission structure	5
5.	Data production structure.....	6
6.	Hard- and software for data production	7
7.	Production procedures.....	8
7.1	i1 – Microscope.....	8
7.2	i2 - Image processing	8
7.3	i3 - Image upload	9
7.4	i4 - Network.....	10
7.5	d1 - Using metadata tables.....	10
7.6	d2 - Entering metadata	11

2. Preface

All Eurexpress II ISH production units will initially upload images and accompanying metadata to the GenePaint.org database at the GWDG in Goettingen by means of the same tools and procedures that are currently used at the Max Planck Institute Hannover. In later phases of the project, this will be complemented by a tracking database for data entry and a Eurexpress II portal for data retrieval by public users.

The purpose of this document is to provide an overview of the *status quo* of image production and metadata handling in Hannover. This will be the basis for data production and transfer for all ISH production units within the Eurexpress II project.

The section “Production Procedures” briefly describes the main steps of the process and provides a few examples for metadata entry. More detailed instructions that describe all steps will be provided in the SOPs for image processing and metadata handling.

3. E-mail Addresses of Eurexpress Informatics Contact Persons

Please report updates and additions to Axel Visel (axel.visel@mpihan.mpg.de)

GWDG Göttingen

Ulrich Degenhardt: ulrich.degenhardt@gwdg.de

Cc: Dagmar Ullrich: dullric@gwdg.de

MRC Edinburgh

Richard Baldock: Richard.Baldock@hgu.mrc.ac.uk

Cc: Duncan Davidson: d.davidson@hgu.mrc.ac.uk

ORGARAT Essen

Stefan Kruse: stefan.kruse@orgarat.de

UNIGE Geneva

Nathalie Lin-Marq: Nathalie.Lin-Marq@medecine.unige.ch

Cc: Alexandre Reymond: Alexandre.Reymond@medecine.unige.ch

IGBMC Strasbourg

Laurent Vasseur: vasseur@igbmc.u-strasbg.fr

Cc: Pascal Dolle: dolle@titus.u-strasbg.fr

TIGEM Naples

Elia Stupka: stupka@tigem.it

Graciana Diez Roux: diezroux@tigem.it

MPI Berlin

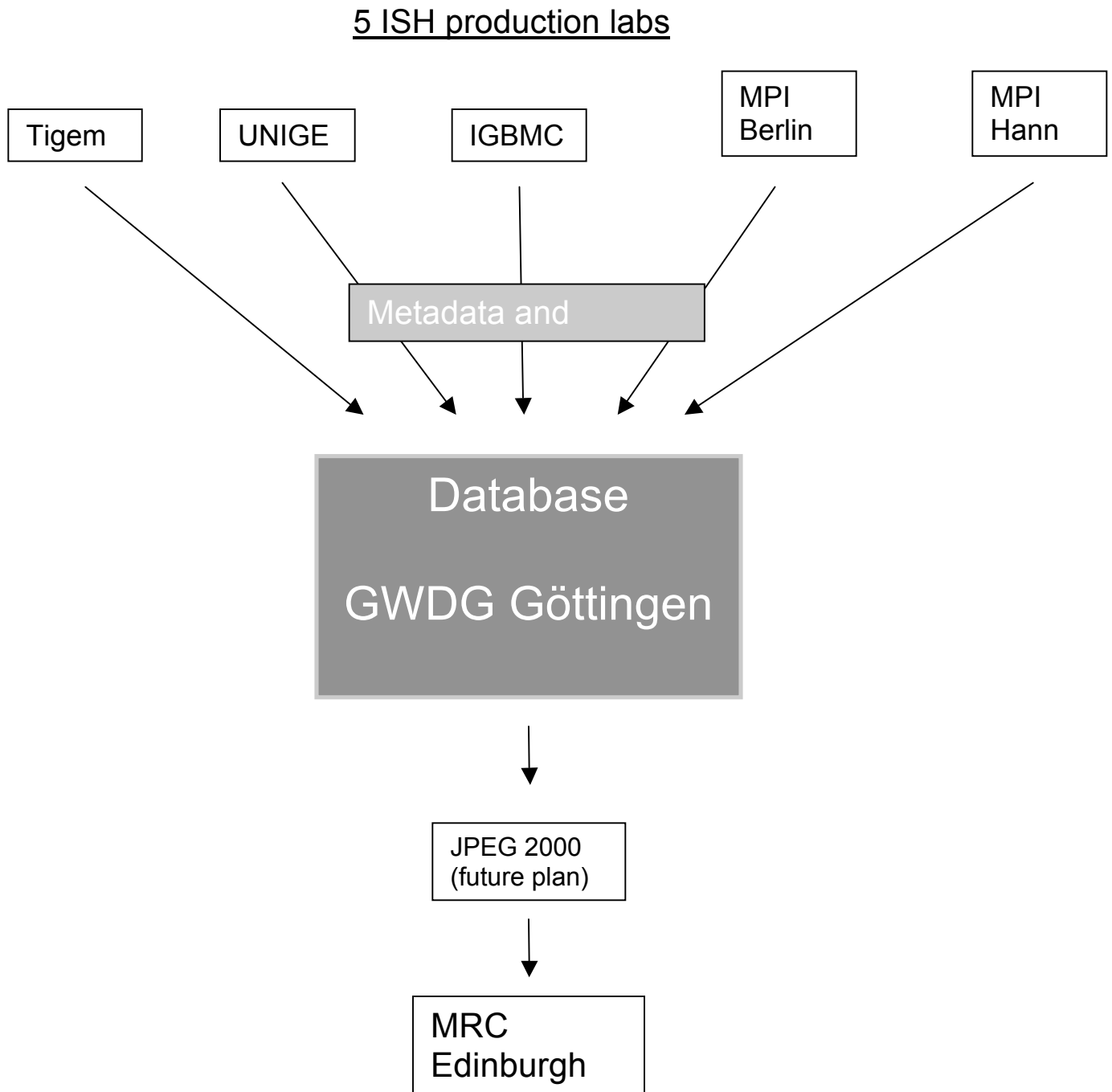
Marc Sultan: sultan@molgen.mpg.de

MPI Hannover

Markus Uhr: markus.uhr@mpihan.mpg.de

Cc: Axel Visel: axel.visel@mpihan.mpg.de

4. Data transmission structure



5. Data production structure

Overview of image and data flow

Images



Bmp mosaics taken from Microscope assembled to TIF images



TIF images and image control files

External USB2 hard drives



Image processing:

- cropping
- manual control
- turning
- conversion to FPX
- conversion to JPG

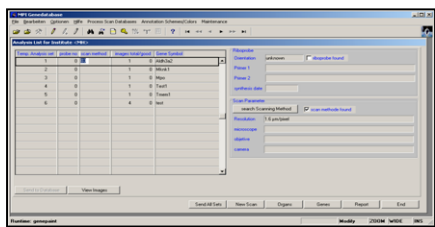
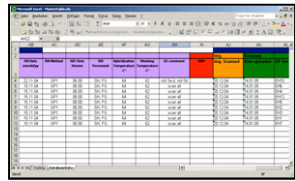


Image uploads by Magic application

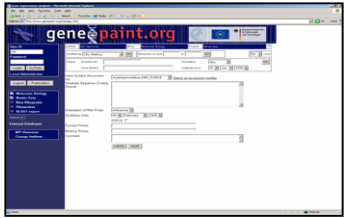


File store and Database at GWDG

Data



Metadata from histology, molecular biology and ISH



Entering metadata into web browser interfaces of GenePaint / Eurexpress II database

6. Hard- and software for data production

(The information given here is based on the hard- and software we currently use in Hannover.)

i1:

- Microscope and computer from Leica. Genepaint microscope software is included.

i2:

- 5 external USB2 hard drives from Maxtor, each with 300 GB to transport the images and image control files from the microscope to the image processing computer.
- Dell Precision 670 with 2 x P4, 3.2 GHz, 2 GB RAM, 2 x 250 GB SATA hard drive with Windows XP Professional for image processing.
- Adobe Photoshop 7.0 for cropping, manual control, recropping, turning and mirroring.
- A batch script for Photoshop for automatic cropping.
- Mystic Media Autoimager to generate FPX images.
- Cerious Software Thumbsplus 2002 to generate JPG images.

i3:

- Magic Client (customized by Stefan Kruse, Orgarat) for controlled image upload.
- Neuhaus13 EasyRename to rename images in the case of a special manual upload situation (freeware, download from <http://www.neuhaus13.com/software/tools/easyrename.html>).

i4:

- A 34 Mbps Network connection to the internet for image upload and data entering over the internet.
- At present, we store the images on an additional server at the ISH production site (Hannover) and the images are pulled by an overnight batch job to the GWDG. This procedure has to be discussed based on the existing network connections in the labs and on optimizations suggested by the GWDG.

d1:

- Any PC with Microsoft Excel to use the metadata tables.

d2:

- Any PC with Microsoft Internet Explorer connected to the internet.

Remark:

We also use a high-performance PC (1 x P4, 3.0 GHz, 1 GB Ram, 1 x 120 GB hard drive) at points d1 and d2. It is equipped with the same software described for i2, which makes it possible to dynamically distribute image processing jobs or, if necessary, let both systems run parallel.

Production procedures

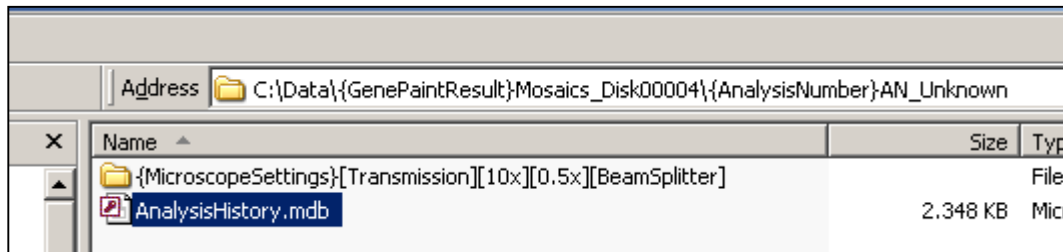
7.1 i1 – Microscope

After the microscope software is set up, ~190 bitmap images are automatically taken for each embryo. The total number of images depends on the settings made on the slide plate editor. After the scanning process is completed, the mosaic process is started. The microscope software generates the TIF images in folders on the external hard drive connected to the microscope PC.

Structure of directories and file names:

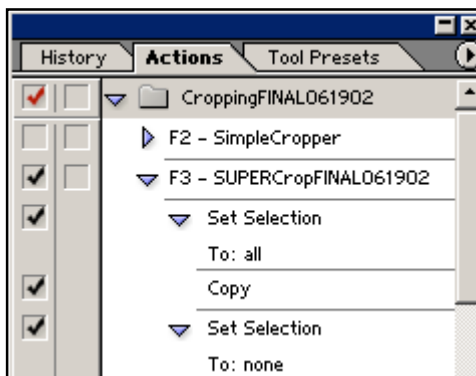


Please notice the file AnalysisHistory.mdb in the folder structure. It is a Microsoft Access database file that contains the image control information from the microscope used later by the Magic Client software to check proper image upload.



7.2 i2 - Image processing

The external hard drive is taken to the PC for image processing. After connecting, the first process is automatic cropping with Photoshop, using a batch job included in the Photoshop action file “CroppingFINAL061902.atn”.



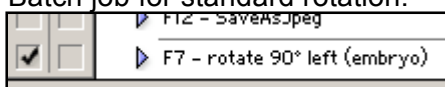
The source is the external hard drive with the raw, uncropped TIF files, the target is the internal drive on the image processing PC.

The cropping process takes a long time and should be run over night.

Dust and tissue debris can interfere with automated tissue edge detection, therefore all auto-cropped files have to be manually reviewed and, where necessary, recropped with Photoshop.

Next, the images have to be rotated (and mirrored, where necessary) into the standard orientation: head on top of the image, face looking to the left. Usually this means a 90° counter-clockwise rotation of the image. However, this may be different if the tissue on the original slide was not in the default orientation. Rotating the images, will again take several hours and should also run over night.

Batch job for standard rotation:



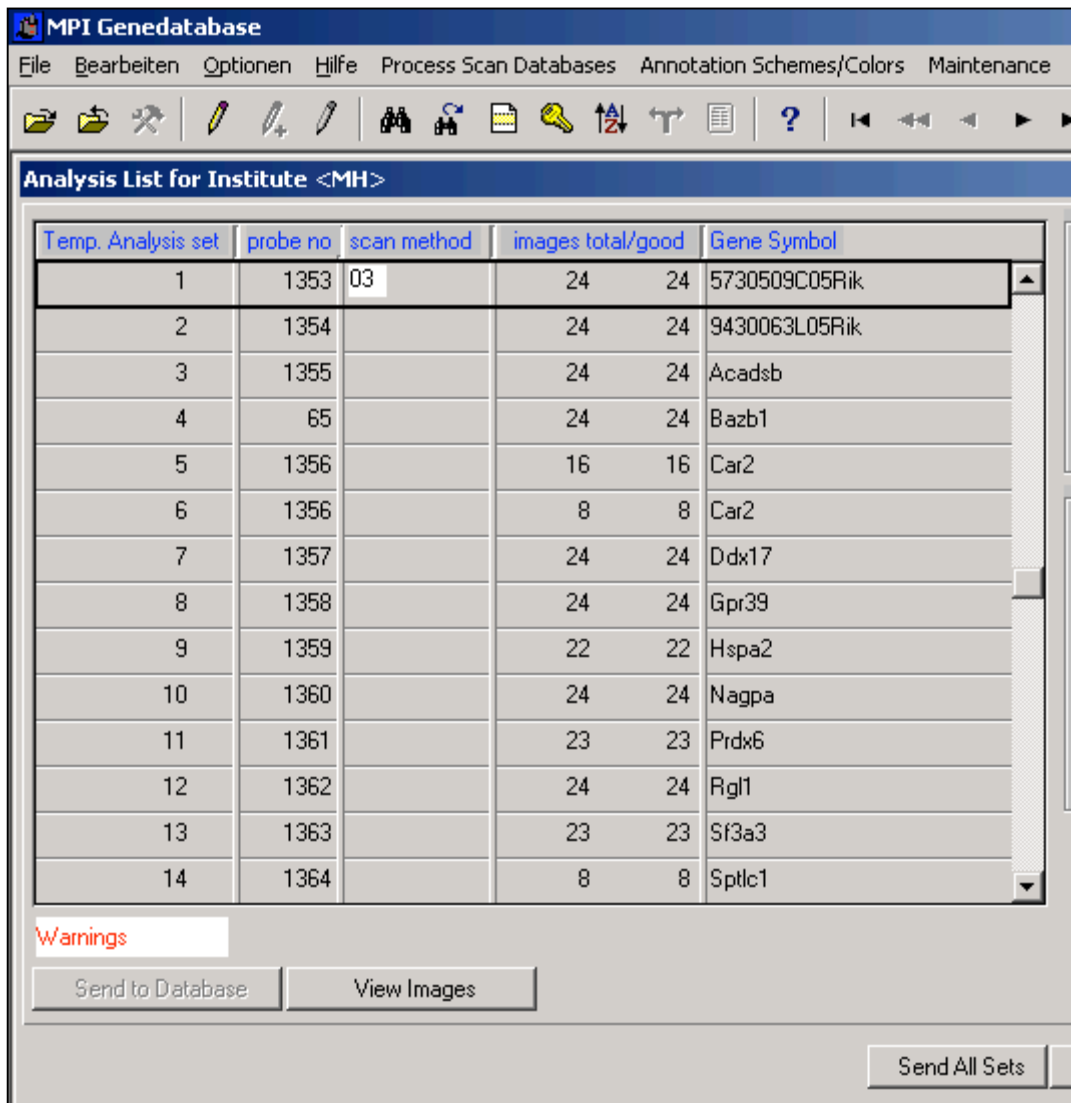
Next, FPX and JPG images are generated from the TIF files.

FPX is the Kodak flashpix format used by the ZoomImageServer on the server side to zoom into the images on the web site. FPX files are generated by a Software from Mystic Media called AutoImager (www.autoimager.com).

For JPG generation, we use Cerious Software Thumbsplus 2002. We use a batch file "genepaint_tif2jpg.TPB" which includes the conversion parameters (33% image size, JPG quality 40%).

7.3 i3 - Image upload

Magic Client is the customized software (by Stefan Kruse, Orgarat) for image upload to the file server. In the user interface, you can select the set of images you want to upload to the according data set on the GenePaint/Eurexpress II Database.



The screenshot shows the MPI Genedatabase software interface. The title bar reads "MPI Genedatabase". The menu bar includes "File", "Bearbeiten", "Optionen", "Hilfe", "Process Scan Databases", "Annotation Schemes/Colors", and "Maintenance". The toolbar contains various icons for file operations and navigation. The main window displays "Analysis List for Institute <MH>". Below this is a table with the following data:

Temp. Analysis set	probe no	scan method	images total/good	Gene Symbol
1	1353	03	24 24	5730509C05Rik
2	1354		24 24	9430063L05Rik
3	1355		24 24	Acadsb
4	65		24 24	Bazb1
5	1356		16 16	Car2
6	1356		8 8	Car2
7	1357		24 24	Ddx17
8	1358		24 24	Gpr39
9	1359		22 22	Hspa2
10	1360		24 24	Nagpa
11	1361		23 23	Prdx6
12	1362		24 24	Rgl1
13	1363		23 23	Sf3a3
14	1364		8 8	Sptlc1

Below the table, there is a "Warnings" section and three buttons: "Send to Database", "View Images", and "Send All Sets".

The software does three things:

1. It checks, with the help of the above-mentioned AnalysisHistory.mdb, if a suitable data set has already been generated on the GenePaint/Eurexpress II Database. For this purpose, it opens several database communication channels over the network.
2. It checks if all images found in AnalysisHistory.mdb are in fact located on the hard disk.
3. It uploads the images and sets a flag in the data set on the GenePaint/Eurexpress II database that the images are in the state of transportation and can be reviewed and published.

Important:

Image upload can only be done after a suitable data set has been entered in the GenePaint/Eurexpress II database! So, step “d2 – Entering metadata” has to be finished before step i3 is started.

7.4 i4 - Network

We use a 34 Mbps Network connection to the internet for image upload and entering data over the internet. For example: Transportation of images for 15 genes means that approximately 31 GB have to be moved over the network.

As mentioned above: This procedure has to be discussed based on the existing network connections in the labs and on optimizations suggested by the GWDG.

7.5 d1 - Using metadata tables

Until a Eurexpress II tracking database has been established (expected to be usable in late 2006), a set of Excel spreadsheets will be used for sample tracking and metadata recording at the ISH production sites.

The metadata tables consist of 4 Excel files:

1. **Probe Data.xls** – all information about templates, riboprobes, sequences etc.
2. **Histology.xls** – all information about embedded specimens and cryosections
3. **ISH DB Data.xls** – all information about ISH runs and results
4. **Mastertable.xls** – selected information for workflow tracking and database data entry (two worksheets)

As an example, the “Tracking” worksheet of mastertable.xls is shown below.

First, the lab unit responsible for **molecular biology** enters its data both to Probe Data.xls and Mastertable.xls. Next, the **ISH** unit will perform ISH and then fill the according fields in the table. After scanning, the **scanning** personnel will fill in the scan date. Finally, the **database** personnel will upload the data and fill in the last two columns.

Acc. #	Gene Symbol	Template No.	Synthesis Date	Specimen	ISH Date mm/dd/yy	Scan date mm/dd/yy	Date uploaded	GenePaint Set ID
XM_136135	2900002G04Rik	Han2	10.11.2004	C1747_4	22.11.04	20.12.04	14.01.05	EH10
NM_025357	Smpx	Han1	10.11.2004	C1764_5	15.11.04	20.12.04	14.01.05	EH6
NM_009031	Rbbp7	Han3	10.11.2004	C1764_6	15.11.04	20.12.04		
NM_011098	Pitx2	Han4	10.11.2004	C1773_1	15.11.04			
NM_009170	Shh	Han5	10.11.2004	C1773_2	15.11.04			
NM_008130	Gli3	Han6	10.11.2004	C1773_3	15.11.04			
NM_144841	Otx2	Han7	10.11.2004					
NM_007770	Crx	Han8	10.11.2004					
NM_011912	Vax2	Han9	10.11.2004					

Acc. #	Gene Symbol	Template No.	Synthesis Date	Specimen	ISH Date mm/dd/yy	Scan date mm/dd/yy	Date uploaded	GenePaint Set ID
XM_136135	2900002G04Rik	Han2	10.11.2004	C1747_4	22.11.04	20.12.04	14.01.05	EH10
NM_025357	Smpx	Han1	10.11.2004	C1764_5	15.11.04	20.12.04	14.01.05	EH6
NM_009031	Rbbp7	Han3	10.11.2004	C1764_6	15.11.04	20.12.04	14.01.05	EH4
NM_011098	Pitx2	Han4	10.11.2004	C1773_1	15.11.04	20.12.04	14.01.05	
NM_009170	Shh	Han5	10.11.2004	C1773_2	15.11.04	20.12.04		
NM_008130	Gli3	Han6	10.11.2004	C1773_3	15.11.04	20.12.04		
NM_144841	Otx2	Han7	10.11.2004	C1773_4	15.11.04			
NM_007770	Crx	Han8	10.11.2004	C1773_5	15.11.04			
NM_011912	Vax2	Han9	10.11.2004	C1773_6	15.11.04			

= new entries

Such consecutive entry of data by the different lab units results in a cascade appearance of the mastertable, which provides a To-Do list for each lab unit and at the same time allows to follow the progress in all units of the lab.

More detailed experimental and sample-tracking information can be found in tables 1-3. They are filled with data by the lab personnel, either by pasting from other tables (e.g. template information provided by Eurexpress template production facilities) or after completion of experimental procedures.

Prior to upload of metadata to the database, the person responsible for data upload copies all relevant parts from tables 1-3 into mastertable.xls to the more detailed "Database Data" sheet.

For instance:

1. Probe Data.xls

2. Histology.xls

3. ISH DB Data.xls

Acc. #	Gene Symbol	Template	Synthesis Date	Probe Direction (sense/antisense)	Sequence	Forward Primer	Reverse Primer	Spec. Prefix	Spec. ID	Strain	Tissue	Age	Dissector	Dissect. Date	Dissect. Time	Preparation Method	Total No. of Slides	Section Type	µm	Section Person	Section Date	Slide per Set	Gene Name	Hybr. slides	Specimen	ISH Date mm/dd/yy	ISH Method	ISH Time h:mm	ISH Personnel	Hybridization Temperature C	Washing Temperature C	QC comment
XM_136135	2900002G04Rik	Han2	10.11.2004	AS	CGAGAGATTC	not available	not available	C	1747_4	NMRI	Embryo	E14.5	HK	25.08.04	9:00	OCT	6	sagittal	25	Polina	21.09.04	6	2900002G04Rik	16	C1747_4	22.11.04	GP1	06:00	SH, FG	64	62	not 3a-5, not 5d
NM_007770	Crx	Han2	10.11.2004	AS	TCGAGAGATTC	not available	not available	C	1773_5	NMRI	Embryo	E14.5	PS	09.08.04	11:05	OCT	6	sagittal	25	Doreen	02.11.04	6	Crx	16	C1773_5	15.11.04	GP1	06:00	SH, FG	64	62	scan all
NM_009031	Rbbp7	Han2	10.11.2004	AS	TCGAGAGATTC	not available	not available	C	1764_5	NMRI	Embryo	E14.5	HK	07.07.04	9:00	OCT	6	sagittal	25	Doreen	28.10.04	6	Gli3	16	C1773_3	15.11.04	GP1	06:00	SH, FG	64	62	scan all
NM_011098	Pitx2	Han2	10.11.2004	AS	TCGAGAGATTC	not available	not available	C	1764_6	NMRI	Embryo	E14.5	HK	07.07.04	9:00	OCT	6	sagittal	25	Doreen	28.10.04	6	Otx2	16	C1773_4	15.11.04	GP1	06:00	SH, FG	64	62	scan all
NM_009170	Shh	Han2	10.11.2004	AS	TCGAGAGATTC	not available	not available	C	1773_1	NMRI	Embryo	E14.5	PS	09.08.04	11:05	OCT	6	sagittal	25	Doreen	02.11.04	6	Pitx2	16	C1773_1	15.11.04	GP1	06:00	SH, FG	64	62	scan all
NM_008130	Gli3	Han2	10.11.2004	AS	TCGAGAGATTC	not available	not available	C	1773_2	NMRI	Embryo	E14.5	PS	09.08.04	11:05	OCT	6	sagittal	25	Doreen	02.11.04	6	Pitx2	16	C1773_1	15.11.04	GP1	06:00	SH, FG	64	62	scan all

Mastertable.xls, worksheet "Database Data".

Thereby, all data necessary to fill out the browser entry interfaces for GenePaint/Eurexpress II data sets is pre-compiled in a single table. The coloring of the columns helps to locate the positions where the data has to be inserted.

7.6 d2 - Entering metadata

GenePaint/Eurexpress II contains several data entering masks that have to be filled for each dataset. This information is copied/pasted from the "database data" worksheet in mastertable.xls.

As an example, the accession number-related data entry is shown (detailed instructions for all entry forms will be provided later in separate SOPs):

Probe Data							
Acc. #	Gene Symbol	Template	Synthesis Date	Probe Direction (sense or antisense)	Sequence	Forward Primer	Reverse Primer
XM_136135	2900002G04Rik	Han2	10.11.2004	AS	GGAAGAGT	not available	not available

The screenshot shows the genePaint.org web interface. The top navigation bar includes 'Institute MPI Hannover', 'Menu Molecular Biology', and 'Program Riboprobes'. A search bar is present with 'modified by Uhl, Markus' and 'find probe no. from' to 'to' fields. The main form area has the following fields:

- Gene Symbol (Accession no): RIKEN cDNA 2900002G04 gene (XM_136135) [Select an accession number]
- Template Sequence (Coding Strand): CTGCTTCTCTCCGTGCTTTTGTGTGTGTATGGGGTGAGGGAAATCTAGCTACTTTCTCCCCGTCTCAGGATTCTGTTATTTTATTATTAAGCTCCATCATCTGCAGAGCA TTA CTCTCCTTACTGTTTAGGGTTGAAGAAAGGTTTGGTTGAATGAATAAGTAA ATAGTGGCAATTATTTCTTTCCCTTATGGGAGCTGAATCTAGCAAGGACCCTTTC ATGTCACCCCTTAAAAATCGCCAAATTGTTGCTCCAAACCATTGAATGGCTTTTCA GCTGAATGAATCAAGTAACATACAGGTGGCTA
- Orientation of RNA Probe: Antisense
- Synthesis Date: 10 November 2004
- RTPCR:
- Forward Primer: [empty field]
- Reverse Primer: [empty field]
- Comment: [empty text area]

Buttons for 'submit' and 'reset' are at the bottom of the form.

The database personnel has to fill out the following masks to get a complete set:

Step	Mask name	Path to database mask in menu	Remark
1	Accession no.	Molecular Biology / Master Data / Accession Number	We recommend to copy and paste field entries from "Entrez Gene"
2	Dissection and Preparation	Histology / Dissection	
3	Riboprobes	Molecular Biology / New Riboprobe	
4	ISH	ISH / ISH	

In addition, the "Database Data" sheet in mastertable.xls is used to store information which is necessary to go through steps in the table above.

For example: In step 3, a probe number is generated that is required for data entry in step 4:

ISH Data		
Slide per Set	Probe #	Gene Name
6	1367	29000002G04Rik

Hybridization

Probe Number select probe no.

Method

Operator

from Mastertable.xls / Database Data

from ISH entering mask

After image upload, the database personnel has to confirm the successful image upload by using "Scan and Annotate / View and annotate" in the database menu.

Set	show	State	Accession No.	Probe No.	view
1649	images	Transfer	NM_197940	1321	

clicking on "images" ...

analysis 1649

Accession Number NM_197940

Gene RIKEN cDNA 5730509C05 gene

Process State Transfer

If all images (JPG) are visible

Image no	section	image name	image
1	Embryo_E14.50_C1122_1_1a	MH00001649_00001B	
2	Embryo_E14.50_C1122_1_1b	MH00001649_00002B	

...shows the uploaded images as thumbnails.

Confirming the images leads to the last step:

Set	show	S
1649	images	A

Click

analysis no. **1649**

accession no. **NM_197940**

analysis started

analysis issued by

analysis status

storing location

Last modified by User: Uhr date: 10-Feb-2005