Eurexpress II

Overview of image and data production and transmission

Max Planck Institute of Experimental Endocrinology, Hannover
Axel Visel, Markus Uhr, Feb. 2005
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 ISH production labs

- Tigem
- UNIGE
- IGBMC
- MPI Berlin
- MPI Hann

Metadata and images

Database

GWDG Göttingen

JPEG 2000 (future plan)

MRC Edinburgh
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

Data

- Metadata from histology, molecular biology and ISH

External USB2 hard drives

Image processing:
- cropping
- manual control
- turning
- conversion to FPX
- conversion to JPG

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

Image uploads by Magic application

File store and Database at GWDG
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.


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 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.

<table>
<thead>
<tr>
<th>Acc. #</th>
<th>Gene Symbol</th>
<th>Template No.</th>
<th>Synthesis Date</th>
<th>Specimen</th>
<th>ISH Date mm/dd/yy</th>
<th>Scan date mm/dd/yy</th>
<th>Date uploaded</th>
<th>GenePaint Set ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>XM_136135</td>
<td>2900002GC4Rik</td>
<td>Han2</td>
<td>10.11.2004</td>
<td>C1747_4</td>
<td>22.11.04</td>
<td>20.12.04</td>
<td>14.01.05</td>
<td>EH10</td>
</tr>
<tr>
<td>NM_025357</td>
<td>Smp3</td>
<td>Han1</td>
<td>10.11.2004</td>
<td>C1764_5</td>
<td>15.11.04</td>
<td>20.12.04</td>
<td>14.01.05</td>
<td>EH6</td>
</tr>
<tr>
<td>NM_000031</td>
<td>Fbbp7</td>
<td>Han3</td>
<td>10.11.2004</td>
<td>C1764_6</td>
<td>15.11.04</td>
<td>20.12.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_011098</td>
<td>Fitx2</td>
<td>Han4</td>
<td>10.11.2004</td>
<td>C1773_1</td>
<td>15.11.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_009700</td>
<td>Sfr</td>
<td>Han5</td>
<td>10.11.2004</td>
<td>C1773_2</td>
<td>15.11.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_001730</td>
<td>Gli3</td>
<td>Han6</td>
<td>10.11.2004</td>
<td>C1773_3</td>
<td>15.11.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_144844</td>
<td>Otx2</td>
<td>Han7</td>
<td>10.11.2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_007770</td>
<td>Crx</td>
<td>Han8</td>
<td>10.11.2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_011912</td>
<td>Vax2</td>
<td>Han9</td>
<td>10.11.2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

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):

<table>
<thead>
<tr>
<th>Step</th>
<th>Mask name</th>
<th>Path to database mask in menu</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accession no.</td>
<td>Molecular Biology / Master Data / Accession Number</td>
<td>We recommend to copy and paste field entries from “Entrez Gene”</td>
</tr>
<tr>
<td>2</td>
<td>Dissection and Preparation</td>
<td>Histology / Dissection</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Riboprobes</td>
<td>Molecular Biology / New Riboprobe</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ISH</td>
<td>ISH / ISH</td>
<td></td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Slide per Set</th>
<th>Probe #</th>
<th>ISH Data</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1367</td>
<td>2900002604Rik</td>
<td></td>
</tr>
</tbody>
</table>

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.

Clicking on “images” …

...shows the uploaded images as thumbnails.

Confirming the images leads to the last step:

analysis
Accession Number 1649
Gene RIKEN cDNA 5730509C05 gene

Process State Transfer
If all images (JPG) are visible confirm images

Click