Genotyping core workshop, 8.-9.3.2004, Helsinki

Minutes

Session 1. DNA extraction unit at KTL

Chair: Janna Saarela

Mervi Alanne was introducing the sample handling procedure at KTL. Samples are recorded to the database and provided to genotyping labs either in adjusted concentration in liquid format or as ready-to-use PCR plates. The gender of each sample is checked by test (multiplex) PCR as well as the quality of each DNA. For this the information about the gender is needed from twin centers.

In addition to Finnish twin cohort (n=1443, additional family members included) the following samples have arrived at KTL:

Cohort	**LSDAT-97**	**Geminikar**	**Geminikar-Hvidovre**	**UK**	**Total**
Samples total @ KTL	335	338	257	231	1161
Processed for genotyping	325	338	0	212	875
On plates	Eutwin_P1 - P3, P7	Eutwin_P4 - P7		Eutwin_P8 - P10	
Send to Uppsala	325	338	0	0	663
Not on plate	10	0			
comment	MZ				

In most cases the DNA samples have been of high quality and correct concentration, but there have been some issues with sample IDs and the amount of DNA with certain cohorts.

Discussion

It was agreed that for microsatellite genotyping one water-control and one duplicate sample should be included in each plate. The amount of duplicates and water-controls for SNP genotyping was to be determined after discussion of QC issues.

There have been some challenges in the communication between KTL/FGC and the twin centers: 1) it has not been clear for people outside Helsinki that KTL DNA unit and FGC are separate units and they have assumed that the information they have given to either unit is readily available for both. 2) It has been difficult for DNA and genotyping units to find out whom to contact in each twin center about sample issues.

- It would be preferable to have a contact person and his/hers contact information in each center. As a general rule it was agreed that if one is uncertain whom to contact, the PI of the group should be contacted. Before sending DNA to any center KTL DNA unit should have a phone number of the person at the receiving end to be able to check and agree on the details.

Status of genotyping in each of the genotyping centers

Helsinki, ms

Data (chrs 1-11) ready for 6 Finnish and 7 Danish DNA plates, total of 119000+135000 = 254 000 genotypes. Also genotyped chromosomes 12-22 in 7 Finnish plates (= 214 000

genotypes, 3 more plates to be genotyped) and in 2 UK-twin plates (=70 000 genotypes, 40 more samples to be genotyped).
- 750 samples from Australia coming.

Uppsala, ms

Data ready for twin plates 1-3 (chrs 11-22), plates 4-7 in the process, estimated to be ready by summer.

Helsinki, SNP genotyping

No twin samples genotyped yet, instead the candidate genes in the Morgam cohort
- 242 SNP selected and validated
- 71 SNPs genotyped on ~1250 samples (Finrisk92 & trios & controls)

Uppsala, SNP genotyping

No twin samples genotyped yet, instead X-chromosomal SOX gene genotyped in 30 trios for haplotypes as a model for variable area of the genome

Session 2. Streamlining SNP selection in candidate genes

Chair: Ann-Christine Syvänen

Based on presentations by Per Lundmark (Uppsala) and Markus Perola (Helsinki) the principles for SNP selection adapted by the two centers are highly similar. Therefore issuing a recommendation for GenomEUtwin was straightforward:

1) SNPs with minor allele frequencies > 5% will be selected for genotyping
2) SNPs are selected in candidate gene region to have a density of one SNP per 2,5 – 5 kb. Exceptions will be made for very large genes, where the density may be lower, not to exceed 10-20 SNPs per gene.

It was recognized that in some cases the selection of SNPs is determined by the purpose of the study, for example when reported functional SNPs of allele frequencies < 5% are to be verified in twin samples.

It was decided to share information on working and failed SNPs between the two centers by placing this information at a site accessible by both center. Kaisa Silander will give instructions on what information should be included and Juha Muilu will set up a site at the Biomedicum SITRIX server. Per Lundmark and Tomas Axelsson will place the SNP information from Uppsala at this site.

Haploblock analysis of European populations using the migrane locus on Chr 4

Session 3. Whole genome amplification

Chair: Tomas Axelsson

Whole genome amplification (WGA) is a promising method to make the source of a DNA sample unlimited. Currently there is a concern that low amount of DNA in certain samples in a cohort and in some cases all samples in a cohort, included in the GenomEUTwin, restricts genotyping in these samples.

Ann-Christine Syvänen, Elisabeth Widén and Kaisa Silander presented data showing that SNP- and STR genotyping results obtained from using genomic DNA is very similar or equal to results obtained from WGA products amplified from genomic DNA from the same sample using Phi 29 polymerase in a reaction called Multiple displacement amplification (MDA) given that the DNA samples is of high quality and sufficient amounts of DNA is used in the

WGA reaction. Discrepancies in SNP genotyping results were seen in DNA samples with poor quality and low yield.

Recommendations/ Decisions

- It was decided to use Dutch twin samples in a pilot project to further evaluate the WGA method using MDA/Phi 29 polymerase.
- Apart from the Dutch twin pilot it was decided not to use WGA products from genomic DNA as a source for genotyping
- In case WGA will be used in large-scale a special facility has to be set up for this purpose

Session 4. QC for genotyping procedures

Chair: Kaisa Silander

Microsatellite markers: Elisabeth Widén & Inger Jonasson
Both genotyping centers are using the same DNA plates, prepared by the DNA core at NPHI. Each 96-well plate includes 2 CEPH control samples in shifting positions, 1 blank and 1 blind duplicate sample. Runs are read independently by two individuals, and CEPH reference samples and blanks are checked.
Further checks are performed currently only at FGC, which include genotypes in (technically) bad quality runs, identity check for MZ twins, bin checks, GRR, and Sibmed. HWE will be applied in future.

Action items:

1. Should implement QC checks similar to those done at FGC also at Uppsala lab.
2. Check whether sex of individuals is consistent with X-chromosome genotype data from Uppsala lab.
3. Check genotype consistency for blind duplicates for available genotype data from both FGC and Uppsala lab.

SNP markers: Kaisa Silander & Tomas Axelsson
QC procedures differ between the two SNP genotyping centers due to different platform considerations. See powerpoint presentations for more details on specific QC procedures.
In both labs only 50-85% of SNPs are accepted following QC procedures, the rest are excluded due to various reasons (monomorphic, bad quality assays, HWE inconsistency, etc.).

Action item:

- For actual Genomeutwin sample SNP genotyping, attempt to use same DNA plate format to save work for DNA core. Tomas and Kaisa to discuss what format would be acceptable for both labs.

Joint QC pilot between genotyping cores: Aarno Palotie
A joint QC pilot for SNP genotyping is planned, that would include several interested laboratories in Europe, that use various SNP genotyping methods. DNA from 46 individuals of Finnish origin will be distributed to labs. The DNA samples were obtained from NPHI's paternity lab, and the samples are under work in the DNA core at NPHI.

Action items and decisions:

1. Standard DNA concentration measurement will be done for the samples, but no further QC (i.e. no sex-PCR check). DNA samples at ≥50 ng/μl.

2. The SNP genotyping result comparisons will be handled by Labquality (www.labquality.fi/english/).
3. 18 SNPs will be selected from public databases by an independent partner (Juha Muilu). Deadline for SNP selection: March 26.
4. The selected SNPs should be validated, common, not in repetitive sequence and not having other SNPs nearby.
5. Provide to participating genotyping labs only rs# for selected SNPs.
6. SNPs will be genotyped using standard lab procedures (e.g. in multiplex if this is the common way to genotype at lab).
7. Chrisse and Aarno to contact labs to see if interested to participate.
8. Project should be completed by summer.

Discussion on a joint genotyping manual for GenomEUtwin: Kaisa Silander
Due to lack of time this topic was discussed only in a small group. Decided that as a first step would summarize in the manual joint QC procedures and SNP selection procedures.

Session 5. Database-related issues

Chair: Elisabeth Widen

Juha Muilu presented the Gt-DB scheme

The GenomEUtwin phenotype data will be contained in a federated database. All twin-centers maintain their own database and share the information via the federation. The federated database is flexible and gives the centers a better opportunity to control their own data.

The genotype data will be contained in one single database, Gt-DB. The data is firstloaded into a submission database, in which all necessary harmonization and QC procedures will take place. Harmonized data is transferred to a separate data-warehouse database.

The twin cohorts own their data, both phenotypes and genotypes. If a twin-center wishes to withdraw from the federation, the genotypes have to be withdrawn too.

The genotypes and necessary data about markers will be stored in Gt-DB. Also pedigree information will be stored here. According to the Database core’s plans, no pedigree information is going to be stored in the phenotype database.

The meeting accepted the presented database structure. However, all details concerning naming conventions and required information about reference sequence could not be solved during the meeting.

1. Marker names: The meeting agreed on using either DS or rs-numbers whenever they are available. If not, Juha Muilu will coordinate the naming procedure. Each laboratory will be given a unique prefix and Juha’s responsibility is to give each individual marker a unique number. A more detailed description of the naming convention will be sent out to all genotyping laboratories to comment upon. Agreement on the procedure should be reached fairly soon.

2. Reference sequence. The details concerning the reference sequence used to locate the markers was discussed. Juha Muilu will send out a detailed suggestion about how the

reference sequence will be used in Gt-DB for the genotyping laboratories to comment upon.

3. The Helsinki laboratory and the Uppsala SNP-center all have databases and can provide each genotype with a unique identifier. The Rudbäck-laboratory doesn't have a genotyping database at the moment. Juha Muilu will work out how to solve the genotyping identifier issue together with the Rudbäck laboratory separately.

Mats Jonsson presented the Chiasma SNP database in Uppsala

Handling of samples, handling of results and workflow tracking have been implemented in the database

Currently only the results handling part is in use. The next step will be to start using the sample handling part.

The raw genotyping data is primarily stored in genotyping platform specific databases. The final genotyping results are stored in the Chiasma database

Chiasma has links to both the internal (platform specific) databases and to external databases

The graphical user interfaces seemed very nice. There are GUIs for results entry and analysis tools

Juha and Mats will together work out whether Mats's GUIs could be implemented in part also in Gt-DB

Discussion about the required QC-procedures required in Gt-DB

Each genotyping laboratory is responsible for doing as thorough QC as possible on their own data, including checking for Hardy Weinberg equilibrium. In addition, the genotyping laboratories also clean the data from genotypes violating Mendelian segregation.

Data-loading: Juha Muilu and Anne Leinonen are responsible for carrying out all the QC:s related to data-loading, and harmonization

The people performing the statistical analyses are responsible for checking for Hardy-Weinberg equilibrium in the complete data-set they are about to analyze. Also mutation detection analysis, the results of which in part is dependent on the map, will be performed by the person carrying out the statistical analysis.

Combining pedigree-data. In case additional pedigree-members are genotyped, the genotyping laboratory adding data on new pedigree-members will be responsible for carrying out segregation checks on the complete pedigree data.

Required feedback

Each genotyping laboratory should provide feedback about the zygosity status on the twins to the submitting twin center. However, the zygosity information obtained through genotyping in this project does not exclude the possibility of sample-mixup, therefore the twin centers should not communicate the results back to the twin-pairs without confirming the diagnosis by genotyping a separate sample.

The genotyping core wanted feedback about the statistical analysis. It was decided that information about the analysis process should be incorporated in the database. The genotyping core wanted feedback both on analysis events when no problems with the genotypes have been encountered as well as feedback about possible problems.

Genomeutwin genotyping core and the database and statistical cores

Communication between cores

The genotyping core found it useful to have representatives present at the db-core meeting and the statistical core meeting. Also, the genotyping core wishes that the db-core and statistical core participate in gt-core meetings. Next db-core meeting will be in Helsinki March 15-16, and the Helsinki laboratories will have their representatives attending. Elisabeth Widén will go to the next statistical core meeting in Leiden in May.