An integrated system for high throughput TaqMan™ based SNP genotyping

Jochen Hampe¹,*, Andreas Wollstein¹, Timothy Lu¹, Hans-Jürgen Frevel¹, Marcus Will¹, Carl Manaster² and Stefan Schreiber¹

¹1st Medical Department, Christian-Albrechts-University, D-24105 Kiel, Germany and ²Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA

Received on October 23, 2000; revised on January 30, 2001; accepted on February 19, 2001

ABSTRACT

Summary: We have developed an integrated laboratory information system that allows the flexible handling of pedigree, phenotype and genotype information. Specifically, it includes client applications for an integrated data import from TaqMan typing files, Mendel checking, data export, handling of pedigree and phenotype information and analysis features.

Availability: The SQL source code, sources and binaries of the client applications (NT and Windows95/98 platforms) and additional documentation are available at http://www.mucosa.de/.

Contact: J.Hampe@mucosa.de

High throughput single nucleotide polymorphism mapping is an enabling technology for many areas of genetic research: the identification of susceptibility genes for complex phenotypes using disequilibrium mapping or candidate gene studies requires the genotyping of large cohorts (>2000 individuals) for large numbers of genetic markers (Hampe et al., 2000; Kruglyak and Lander, 1996). It is estimated that for the mapping of a typical candidate region of 30 cM genetic size, 500–1000 SNPs will be needed to identify the disease gene.

TaqMan typing is based on using the 5′ exonuclease activity of Taq-polymerase and employs a combination of PCR and competitive hybridization (Livak et al., 1995). The process is integrated into a single PCR reaction. First pass success rates for assay design of 92% are reported (Morin et al., 1999).

To use this capability of the TaqMan technology, an integrated database system is needed to achieve the required throughput and to ensure the validity of the data. A number of proprietary systems are available, that mostly target corporate customers (e.g. http://www.genomica.com, http://www.appliedbiosystems.com, http://www.bio.licor.com). No public domain database solution for integrated SNP genotyping exists.

We have therefore designed a database system that integrates the following information:

- Pedigree and phenotype information. The pedigree relationships are stored in the patient table in a similar format to the linkage files. The trait information is completely normalized through the traits and patient_trait tables. Pedigrees can be grouped into populations for more structured access and for later analysis.

- Sample and plate information. The relation between patients and samples is enforced with a trigger in order to allow the users to enter samples into the database in the absence of patient information. The genotyping is performed in a microtiter plate format, of which any rectangular format is supported by the database.

- Marker information. Genetic markers are stored by chromosome and genetic location. Assay conditions and primers are kept in a separate table, thus allowing the definition of multiple assays per genetic variant.

- Genotypes. TaqMan data are imported in a wizard-guided process into a primary data table that is organized by experiment. All primary experiments are stored and consolidated into a final genotype table. Each genotype can be traced to the original experiment through the taq_data_id number. The genotypes are checked for Mendelian inheritance errors (‘Mendel checking’) with an integrated SQL script. The user can graphically review the primary data and edit the conflicting genotypes (see Figure 1).

- Data export. The genotype data can be exported into linkage file format (i.e. the ‘.pre’ and ‘.dat’ file format used by the LINKAGE programs; Terwilliger and Ott, 1994).

The database has an open design that will easily allow the incorporation of other primary genotyping technology through the modification of the client applications. At

*To whom correspondence should be addressed.
Fig. 1. The figure shows the interface used in the correction of Mendel errors in an SNP genotyping study. The left panel shows the problematic pedigree as flagged by the automatic Mendel checking procedure. The code ‘MomX’ indicates a misinheritance originating from the maternal genotype. The genotypes can be edited in this mode. The right panel shows the genotypes from the whole plate, where the marker was typed. The genotype calls are marked with green circles, red triangles and blue diamonds for the 1/1, 1/2 and 2/2 genotypes respectively. All genotypes from the pedigree in question are marked in black. A tool-tip displays the individual identification. The axes correspond to the fluorescent readings from the TaqMan (ABI 7700) machine (x-axis: FAM, y-axis: TET). Data from the genotyping of the TNF-alpha promoter polymorphism ‘-308’ are shown in from a German family sample (Hampe et al., 1999).

our site, the database runs on an MS-SQL server 7 platform. The client applications were implemented on the NT/Windows platform in order to make the database clients available on the most prevalent laboratory device controllers. The applications are written in Visual Basic.

The system has evolved in our laboratory since 1998 and has been used in several high throughput genotyping studies (Hampe et al., 1999, 2001; Olavesen et al., 2000). We anticipate that the presented database system will facilitate the use of high throughput typing technology in laboratories that do not want to develop their own integrated database system. The implementation uses widely available technology and should therefore easily be transferrable to other settings.

ACKNOWLEDGEMENTS

Grant support: this work was supported by the Deutsche Forschungsgemeinschaft (Schr 512/5-1, SFB 415), a Training and Mobility of Research (TMR) Network grant of the European Union (ERB-4061-PL-97-0389), by MFG, the German Human Genome Project (DHGP) and a ‘Kompetenznetz chronisch-entzündliche Darmerkrankungen’ of the German Federal Department for Research and Education (BmBF).

REFERENCES


