Analysis of Genetic Variation in the GenomEUtwin Project

Kaisa Silander[1], Tomas Axelsson[2], Elisabeth Widén[3], Andreas Dahlgren[2], Aarno Palotie[3,4], and Ann-Christine Syvänen[2]

[1] *Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland*
[2] *Molecular Medicine, Department of Medical Sciences, Uppsala University, University Hospital, Uppsala, Sweden*
[3] *The Finnish Genome Centre, University of Helsinki, Finland*
[4] *Department of Clinical Chemistry, University of Helsinki, Finland*

Multiallelic short tandem repeat polymorphisms, or microsatellites, are useful markers in genome wide scans to identify chromosomal regions containing genes underlying disease loci. The biallelic single nucleotide polymorphism (SNP) can be used to fine map previously identified large candidate regions or to test functional candidate genes by association analysis. In the GenomEUtwin project the population based impact of susceptibility genes for six multifactorial traits will be studied. A genome wide panel of informative human microsatellite markers will be analyzed by fluorescent capillary electrophoresis in well characterized twin and population samples. Contrary to microsatellites, selection of the most informative panels of SNPs is hampered by imperfect data on the allele frequencies and population distribution of SNPs markers in the databases. Therefore, selection of SNPs requires a substantial amount of bioinformatics, and, the SNPs need to be validated experimentally in the relevant populations prior to genotyping large sample sets. In the GenomEUtwin project, large scale genotyping of SNPs will be performed using the SNPstreamUHT and MassARRAY genotyping systems that are based on the primer extension reaction principle combined with fluorescent and mass spectrometric detection, respectively. Production of the genotyping data will be a joint effort by GenomEUtwin partners at the University of Helsinki, the National Public Health Institute in Helsinki, Finland and Uppsala University, Sweden. All genotyping data will be stored in a common database established specifically for the GenomEUtwin project, from where it can be accessed by the twin research centres that provided the samples for genotyping.

Comparison of the genomic DNA sequence of different individuals reveals that our genome displays both variation in length and variation in the nucleotide sequence. Length variation polymorphisms are frequently due to variation in the number of tandem repeated units, and the single nucleotide polymorphisms (SNPs) are nucleotide positions at which two (or more) bases can occur, or a base is inserted or deleted. The polymorphisms are utilised as markers for identification of genomic trait loci by linkage analysis and linkage disequilibrium mapping.

Short tandem repeat (STR) polymorphisms or microsatellites are most useful in linkage mapping studies where the objective is to map a chromosomal region co-segregating with a trait of interest. These markers usually have many alleles and are highly informative. They are mainly found in non-coding regions throughout the genome at an average density of 3 kilobase pairs (Subramanian et al., 2003). Well defined genome wide panels of STR markers have been used in linkage analysis with great success in hundreds of studies that have discovered genes causing monogenic disorders and in studies that have assigned disease loci in multifactorial, complex traits. The identification of the actual genes underlying the complex disorders, however, has been less successful; for a review, see (Botstein & Risch, 2003). Therefore, an alternative strategy, based on whole genome linkage disequilibrium mapping using genome wide panels of SNP markers, has been proposed for identification of complex disease genes (Risch & Merikangas, 1996).

It has been estimated that SNPs occur on average at one out of a thousand of the 3×10^9 nucleotides contained in the human genome (Sachidanandam et al., 2001). Functional SNPs in coding regions of genes are the direct causes of most of the monogenic disorders. Many SNPs in coding or regulatory regions of genes represent allelic variants that may predispose individuals to complex, multifactorial disorders. SNPs are also useful as markers in linkage disequilibrium (LD) mapping or association studies.

The rationale in genome wide linkage LD studies would be to genotype a collection of SNPs that occur at regular intervals and that cover the whole genome to detect genomic regions in which the frequencies of the SNP alleles differ between cases and controls, or that differ according to quantitative measures of a trait. The required number of markers is determined by the range of LD in the genome, of which there is only preliminary information available (Carlson et al., 2003; Dawson et al., 2002). The goal of the recently initiated haplotype mapping (HapMap) project is to characterise in detail the structure of the genetic variation and range of linkage disequilibrium in the human genome (Couzin, 2002; Gabriel et al., 2002). This project will, if successful, clarify how many SNP markers would be required for whole genome mapping studies and provide

Address for correspondence: Ann-Christine Syvänen, Molecular Medicine, Department of Medical Sciences, Uppsala University, Entrance 70, Research Department 2, University Hospital, 75185 Uppsala, Sweden.Email: Ann-Christine.Syvanen@medsci.uu.se

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

guidelines for selecting the markers. In the largest published SNP genotyping study so far, a genome wide panel of 65,000 gene-based SNPs were analysed for their association with susceptibility to myocardial infarction in 1000 case-control samples (Ozaki et al., 2002). This study, which generated over 60 million genotypes, identified a SNP haplotype in the lymphotoxin-α gene that was associated to myocardial infraction. No other SNP study comparable in size to this one has been published.

Both the capacity and cost of the available technology for SNP genotyping are obstacles for executing studies using tens of thousands of SNP markers. Thus, a more feasible alternative to whole genome SNP-mapping is to analyse SNP markers in or close to candidate genes, or in candidate genomic regions previously identified by genetic linkage studies. This is also the approach that will be applied in the GenomEUtwin project which, with its unascertained population sample aims to study the population based genetic impact of susceptibility genes and gene loci for a number of multifactorial traits (see Peltonen, in this issue of Twin Research). Additionally this project aims, in collaboration with other projects of the participating groups to identify new predisposing loci and genes by combining data from whole genome scans, dense mapping and genotyping of candidate genes both from family based and twin studies.

Selection of Markers

Microsatellite Markers

Comprehensive human genetic maps have incorporated more than 8000 STR polymorphisms, primarily from Généthon, the Cooperative Human Linkage Centre, the Utah Marker Development Group, and the Marshfield Clinic Research Foundation[1]. The most widely used STR or microsatellite panels are the "Marshfield" and "Applied Biosystems (ABI)" marker collections. The Marshfield panel is a public domain resource consisting mostly of tetranucleotide markers (Broman et al., 1998). The ABI panel is provided by Applied Biosystems (Forster City, CA, USA), and consists of dinucleotide repeat markers. The standardization of the marker panels for genome wide scans is of major importance in large, continuous multicenter studies such as GenomEUtwin. In principle, tetranucleotide repeat markers provide better allele separation than dinucleotide markers, but the modern capillary electrophoresis instruments have sufficient resolution for convenient and reliable separation of both types of markers. Thus, the choice between marker sets is determined by project or site specific factors.

One of the aims of the GenomEUtwin study is to provide a genotype resource that can be used for years in subsequent linkage and association studies, and that can be combined with genome wide scans from other projects. Since one of the partners in GenomEUtwin had already performed a large genome wide scan using the ABI panel, it was decided that the subsequent scans were to be run with the ABI marker set.

The density of microsatellite markers applied in a genome wide scan is a compromise between cost and statistical power. For statistical analyses, a dense as possible marker set is preferable, but since the financial resources are limiting, the best cost-benefit ratio has to be estimated. Most classical STR panels contain about 400 markers providing a 10 centimorgan (cM) resolution. This is typically sufficient for linkage based family studies. Especially in sib pair studies, genome wide scans at higher resolution would be beneficial. However, it has been calculated that it is more cost-beneficial to increase the number of sib pairs for genotyping than the number of markers (Hauser et al., 1996). Therefore, in the GenomEUtwin study, the genome wide scans will be performed at a 10 cM resolution.

SNP Markers

It has been estimated that the human genome contains 5–6 million bi-allelic SNP markers with minor allele frequencies over 10%, but only half of these high-frequency SNPs are likely to be included in the current SNP databases (Reich et al., 2003). Most of the approximately 4 million candidate SNPs, that so far have been included in the databases, have not yet been validated. Owing to ongoing studies that are validating SNPs and determining their population distributions, the situation may improve radically in the near future. The GenomEUtwin study can contribute to these efforts with information on SNPs in European populations. Currently, there is not a single database from which all SNP markers can be retrieved, and the same SNPs may be identified by different names in different databases. The most commonly used SNP databases are those of Celera[2], the SNP Consortium[3] and dbSNP[4].

For the initial fine-mapping aimed to identify a disease-associated marker within a candidate region a few megabases in size, we will select a set of informative SNPs based on available information on LD and haplotype diversity, for example, the one produced by the HapMap project. Such a set of SNPs may be smaller and more effective than simply choosing all haplotype-tagging SNPs within a region (Cardon & Abecasis, 2003). If such data is not yet available for the region of interest, we will primarily select candidate SNPs that have been confirmed to be true polymorphisms, or SNP that have been identified by two independent discovery efforts. This strategy is supported by (Reich et al., 2003), who found that in a total of 173 kb sequence comprising 17 loci, 97% of the SNPs that have been independently identified by both Celera and the public SNP discovery effort had a minor allele frequency over10%. For SNP markers lacking population specific validation data, DNA pools may be used for estimating their allele frequencies in the relevant populations (Syvanen et al., 1993), followed by typing the polymorphic SNPs in a small number of individuals to identify SNPs that are in strong LD with each other. For a cost-effective approach, we will also consider using pooled case and control DNA samples for the initial fine-mapping step (Mohlke et al., 2002).

Once a marker or region displaying association has been identified, the region will be saturated with additional SNPs. These SNPs will be typed in a well-selected subset of cases and controls and in samples from family trios to determine the haplotype structure and the extent of LD in the region. We will then choose a set of informative SNPs

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

representing different haplotypes for typing a larger set of cases and controls. Sequencing of the region of interest will be required to identify a comprehensive set of SNPs.

Genotyping Methods

Within the GenomEUtwin consortium there are three high throughput genotyping facilities. The SNP genotyping "platform" located at Uppsala University Hospital has a long experience in developing and applying primer extension based techniques for SNP genotyping (for a review, see Syvanen, 1999). This partner serves as the coordinator for the genotyping activities, and will provide about half of the SNP genotypes for the project. The genotyping unit at the Rudbeck laboratory in Uppsala specializes in high throughput genotyping of multiallelic markers and will provide about half of the microsatellite genotypes. The Finnish genome center will perform genotyping of both microsatellite and SNP markers and will, in a joint effort with the National Public Health Institute in Helsinki, provide about half of the multiallelic and SNP genotypes for the GenomEUtwin project. Figure 1 shows a flow sheet of the genotyping process.

Genotyping of Microsatellites

Microsatellite genotyping is a three step process. First, the DNA samples are amplified by the polymerase chain reaction (PCR) using a defined set or primer pairs. One primer of each pair is fluorescently labelled with one out of three possible fluorophores. Then three fluorescent PCR prod-

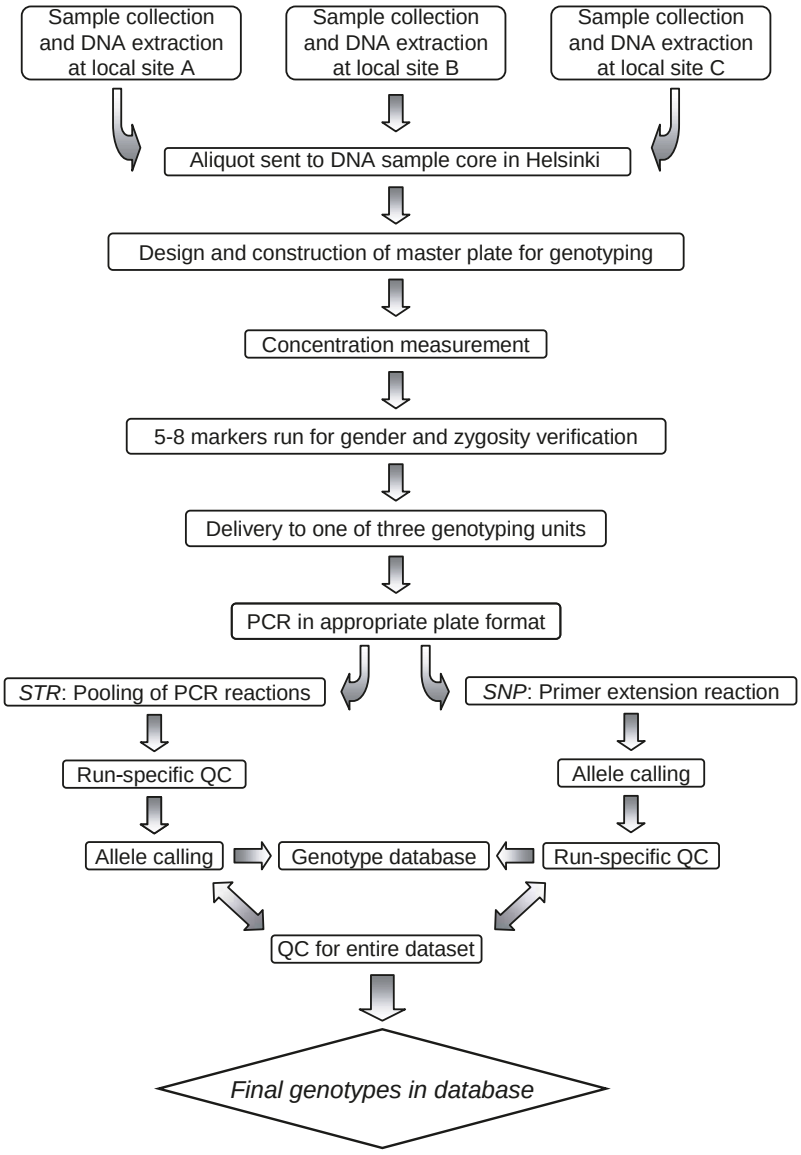

Figure 1

The stages of the genotyping process in GenomEUtwin. In the top part of the flow chart, three local sites (A, B, C) are included to symbolically illustrate all the participating twin research centres. Each centre can access the final genotyping data of their own samples from the database.

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

Table 1

Approaches Used for Error Detection in High Throughput Genotyping in the GenomEUtwin Project

Method	Type of error detected	Error checking software	Reference
Plate controls	Plate mix-ups and inversions		
Segregation of alleles in pedigrees	Sample mix-ups, non-paternities Genotyping errors Null-alleles Marker mutations	PedCheck	O'Connell & Weeks, (1998)
Allele sharing in sibling-pairs	Sample mix-ups	GRR	Abecasis et al., (2001)
Hardy-Weinberg equilibrium	Common null alleles		
Mutation and error detection through multipoint mapping	Genotyping errors Null alleles	SIBMED	Douglas et al., (2000)
Blinded duplicate samples	Estimate of overall error rate		

ucts are combined, and the amplified alleles are separated based on their difference in length in a capillary electrophoresis instrument (MegaBACE 1000 at the Finnish genome center and ABI 3700 at the Rudbeck laboratory in Uppsala). Finally, the different fragments representing the microsatellite alleles are identified by fragment analysis using a reference sample with a fourth fluorophore included in each sample. The allele calling process is semi-automated and facilitated by genotyping platform specific software (Genetic Profiler, Amersham Biosciences and Genemapper, Applied Biosystems). All genotypes will be manually reviewed by two independent specialists, and the error checking methods detailed in Table 1 will be applied. To ensure the consistency of allele binning across the whole project, the results for every new genotyping run will be adjusted to fit the prior genotyping data. Alleles will be tethered according to the reference samples.

Genotyping of SNP Markers

Most of the methods used for genotyping SNPs depend on PCR amplification of the genomic regions, which span the SNPs prior to the actual genotyping reaction so that the required sensitivity and specificity are achieved (for a review, see Syvanen, 2001). In primer extension methods, also known as minisequencing, the distinction between SNP genotypes is based on the high accuracy of nucleotide incorporation by the DNA polymerases (Syvanen, 1999). A detection primer that anneals adjacent to the site of an SNP is extended by one (or a few) nucleotides, the identity of which define the genotype of the SNP. The primer extension reaction is robust, allowing specific genotyping of most SNPs at similar reaction conditions. This feature is advantageous for high-throughput applications because the efforts required for assay design and optimization are minimized. Consequently, primer extension is the reaction principle of choice for the large-scale SNP genotyping that will be performed in the GenomEUtwin project (see Figures 2 and 3).

The Uppsala SNP genotyping "platform" uses the SNPstream UHT ("Ultra High Throughput") system (Beckman Coulter, CA, USA), in which multiplex (12-plex) PCR followed by multiplex (12-plex) primer extension reactions with fluorescently labelled dideoxy nucleotides have been streamlined in a 384-well microarray

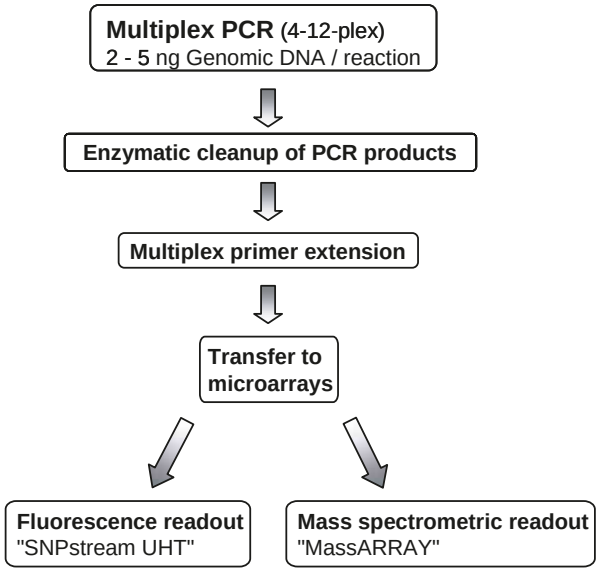

Figure 2

The steps of the SNP genotyping procedure. The whole procedure is carried out in a 384-well microtiter plate format.

format (Bell et al., 2002). The PCR and minisequencing primers are designed after down loading the sequences that flank all SNPs of a panel from an SNP database into a web-based software for primer design[5] (Beckman Coulter). The program automatically selects optimal PCR and minisequencing primers for each SNP, and combines the SNPs with the same of the six possible nucleotide variations (A/C, A/G, A/T, C/G, C/T or G/T) into multiplex panels of up to12 SNPs. Multiplex PCR amplification is performed in 5 µl volumes in a 384 microtiter-plate well format using 2–5 ng of genomic DNA. The amplified samples are subjected to multiplex, thermally cycled minisequencing reactions in solution. For each panel the extension mix contains two fluorescently labelled (Bodipy-Fluorescein and TAMRA) nucleotide terminators specific for the SNP type (see Figure 3). Each minisequencing primer has a 5'-"tag" sequence, and this "tag" is used to capture the extension products to immobilised complementary probes

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

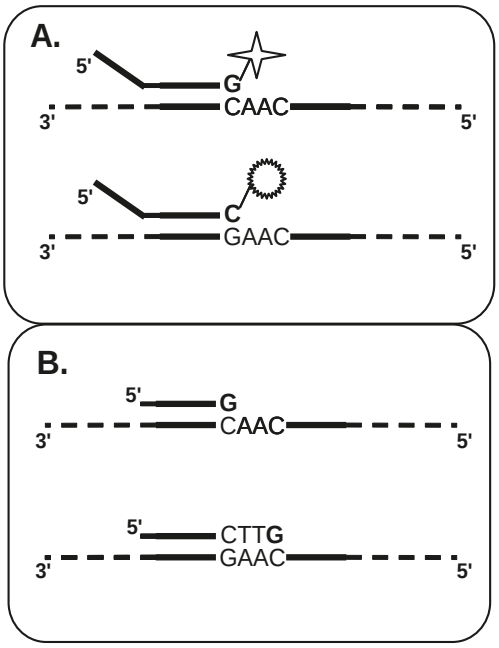

Figure 3

Schematic illustration of the primer extension principles used for SNP genotyping in the SNPstream UHT and MassARRAY systems. (A) In the SNPstream UHT system a primer that anneals immediately adjacent to the site of the SNP is extended by the DNA polymerase with two fluorescently labelled terminating nucleotide analogues that are complementary to nucleotide at the site of the SNP. In homozygous samples the primers will become extended with one of the fluorescent nucleotides, and in heterozygote samples both fluorescent nucleotides will be incorporated. Each primer has a specific "tag" sequence in its 5'-end for capture of the extension products in the wells of the 348-well microarray plate. (B) In the MassARRAY system, a primer that anneals adjacent to the SNP site is subjected to the primer extension reaction using combinations of regular nucleotides and terminating nucleotide analogues. The nucleotide combinations are designed specifically for each SNP according to the flanking sequence to generate primer extension products with different masses. For the SNP in the example the terminating nucleotide G and the regular nucleotides C and T have been used.

spotted in the wells of a 384-well microarray plate. Finally, the fluorescent signals on the microarray plate are read by a two-colour CCD-camera, and the genotypes of the SNPs are assigned by cluster analysis based on the fluorescence intensities measured from each spot on the array (Figure 4). The data analysis is fully automated and generates 4608 individual genotypes per 384 plate in approximately 15min. Spots with low intensity or ambiguous genotype calls are automatically failed, however, manual quality checks of the scatter plots can also performed.

At the Finnish genome center, SNP genotyping is performed using the MassARRAY system (Sequenom, San Diego, CA, USA), in which the genotype assignment is based on differences in mass between the primer extension products corresponding to the two SNP alleles (see Figure 3). Approximately 100 base pair fragments of genomic sequence flanking the SNP are PCR-amplified. A detection primer is extended into the diagnostic SNP site by a DNA polymerase in the presence of a mixture of regular nucleotides and terminating nucleotides, which are designed specifically for each SNP by a software provided with the genotyping system. A small aliquot of the primer extension product is spotted onto a chip, which is then analyzed by the mass spectrometer. The protocols used are as recommended by the system manufacturer, using approximately 5 ng of genomic DNA in the PCR amplification. The genotypes are automatically called based on the mass of the primer extension products (Figure 5). The allele calling algorithm gives a quality score to each genotype, and genotypes are accepted only if they have the two highest quality scores, or if they have been manually reviewed and marked as accepted by the user. In addition, 10–20 % of the data from every plate is manually reviewed to ensure that the data is of good quality.

Error Detection in High-throughput Genotyping

High-throughput genotyping imposes multiple challenges on the work-flow in the laboratory. In the beginning of the genotyping process (see Figure 1), the samples are diluted and dispensed into microtiter plates. Thereafter, all steps are carried out in this plate format. Since thousands of genotyping reactions may be set up during a working day, specific check-points, to identify possible errors in the handling of samples and reagents are required. Errors that may occur during the genotyping process are sample and plate mix-ups, technical problems related to PCR and the visualization of the alleles, errors in allele calling, and marker mutations. Since all errors will interfere with the statistical analysis, and may obscure the linkage or association result, they should be detected and controlled as carefully as possible throughput the genotyping process. The different methods and software used for controlling the genotyping work-flow are listed in Table 1. Plate controls with water, reference samples with previously known genotypes and duplicate samples in asymmetric, shifting positions will identify plate inversion or mix-up. Usually the errors discussed above will lead to genotyping results that are inconsistent with Mendelian inheritance, especially when multiallelic markers are genotyped. Detection of genotyping errors is more challenging when only sibling pairs without parents are genotyped, as in GenomeEUtwin, and even more so in case-control studies. When sibling pairs are genotyped with markers in a whole genome scan, the deviations from the expected allele sharing will reveal possible sample mix-ups, but individual genotyping errors and marker mutations will remain undetected. If the markers genotyped are closely spaced (i.e., at 1–3 cM intervals) some of the errors will appear as double recombinants when haplotypes are constructed based on the genotypes. Subjecting the determined genotype frequencies within the studied sample set to a test for Hardy-Weinberg equilibrium will reveal markers for which one of the alleles is missing in some of the genotyped individuals (null alleles). Since all errors cannot be detected, we will include 1% blinded duplicate samples in the analyses, which will provide important information of the residual error rate in the data-set. These estimated genotyping error rates can then be used in the analysis of the data (Gordon et al., 2002; Rice & Holmans, 2003).

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

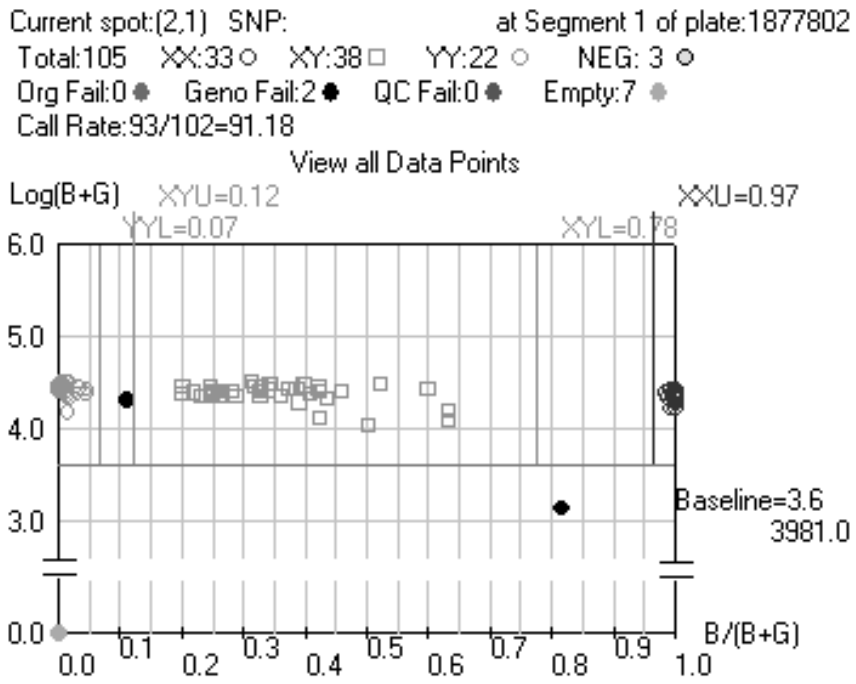

Figure 4

SNPstream UHT scatter plot showing genotyping results from one SNP typed in about one hundred samples. The scatter plot was generated by the GetGenos software (Beckman Coulter). The logarithms of the sums of the fluorescent signals corresponding to the two alleles in each sample are plotted on the vertical axis, and the ratios between one of the signals divided by the sum of both signals are plotted on the horizontal axis. Three distinct, non-overlapping clusters of signal ratios are formed, with clusters corresponding to samples from homozygous individuals close to the vertical axes (signal ratios < 0,07 and > 0,97) and clusters representing heterozygotes in the middle part of the plot (signal ratios 0,12 – 0,78). For this SNP, the cut-off value for acceptance of the results is 3981 fluorescence units (3,6 on the log-scale). Black dots represent samples that have been failed, either due to low signal or data points outside the clusters.

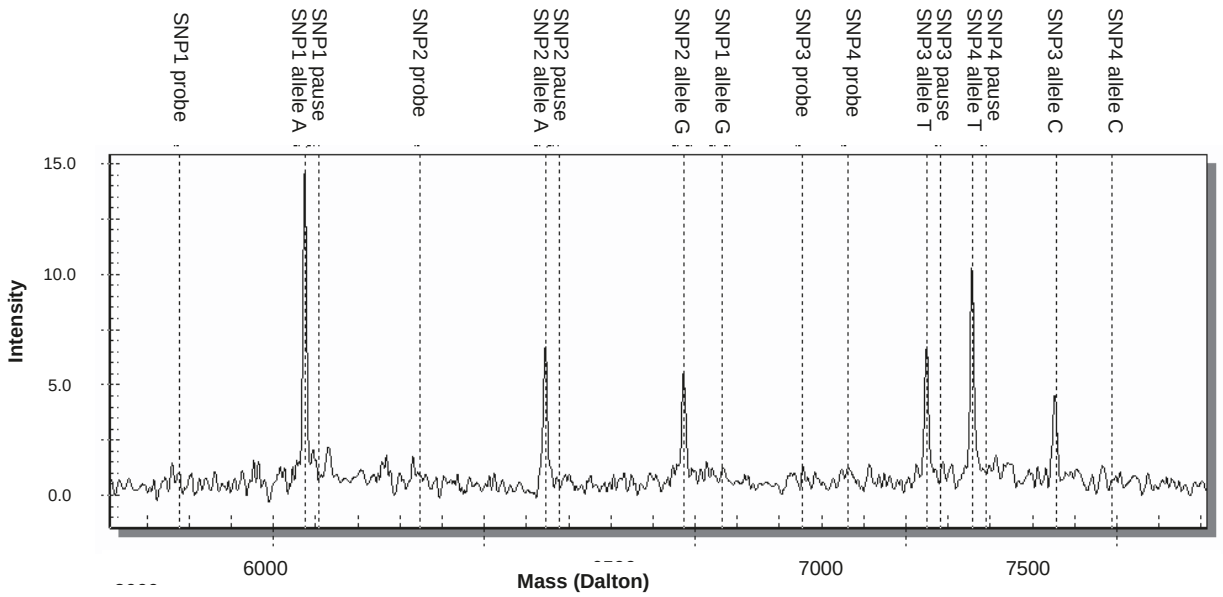

Figure 5

MassARRAY genotype profile obtained for a sample, in which 4 SNPs have been genotyped in a multiplex reaction. The peak heights correspond to the amount of the detected products and the mass of the measured or expected extension products are given on the horizontal axis. The alleles were called with SpectroREADER software (Sequenom). The SNPs are denoted with 1-4, and their genotypes are SNP1: AA, SNP2: AG, SNP3: TC, and SNP4: TT. Probe = unextended assay-specific primer; Pause = paused extension reaction.

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

Table 2

Preliminary Validation of SNP Genotyping Results

Genotyping system	Validation method	Number of genotypes	Number of discrepant s genotype	Discrepancies per 1000 genotypes
MassARRAY individual SNPs	MassARRAY multiplex	1000	1	1
MassARRAY multiplex	Solid-phase minisequencing [1]	780	0	< 1,2
MassARRAY individual SNPs	Primer extension on microarrays [2]	270	0	< 3,7
SNPstream UHT	SNPstream UHT	4000	18	4,5
SNPstream UHT	Homogeneous minisequencing [3]	620	4	6,4
SNPstream UHT	Tag-array minisequencing [4]	900	7	7,7
MassARRAY	Mendelian inheritance check	2700	1	0,3
SNPstream UHT	Mendelian inheritance check	19000	25	1,3

Note: [1]Syvänen et al., (1993), [2] Pastinen et al., (2000), [3] Chen et al., (1999), [4] Lindroos et al., (2002)

Validation of Genotyping Results

The error rate for microsatellite genotyping at the Finnish Genome Center has been estimated both by comparing the genotyping results from 52 monozygotic twin pairs (~39,500 genotypes) and based on violation of Mendelian inheritance in extended twin pedigrees (~28,500 genotypes). In both cases, the whole genome was scanned using the ABI marker set. The observed error rate in fragment analysis was 0.5 errors per 1000 genotypes for monozygotic twin pairs. A similar error rate, 0.4 errors per 1000 genotypes was seen in the extended twin pedigrees. In addition to the allele scoring errors, 0.8 unresolved errors per 1000 genotypes were revealed by the inheritance check in the extended twin pedigrees. These errors were due to marker mutations, non-amplified "null" alleles or technical problems. This evaluation indicates that the overall success rate for genotyping at the Finnish Genome Center is very high. In a sample set consisting of twin-pairs without parents, the error rate is estimated to 1.2–1.5 errors per 1000 genotypes. If pedigree structures are available, the residual error rate after inheritance checks is estimated to be as low as one per 10,000 genotypes.

The accuracy of SNP genotyping was assessed during the establishment of the two genotyping systems. The data were validated by repeating the genotyping assays performed on the MassARRAY and the SNPstream UHT systems (~5000 genotypes), respectively, by comparison of the genotyping results with those obtained by well established "in house" reference methods (~3000 genotypes) (Lindroos et al., 2002; Pastinen et al., 2000; Syvanen et al., 1993) and based on Mendelian inheritance checking (~22,000 genotypes). Table 2 present a summary of these preliminary results. As can be seen in Table 2, the error rates are low, varying between 1 and 8 errors per 1000 genotypes. A quality assessment program is being planned, in which the genotyping results will be validated between the two laboratories at regular intervals by analyzing common reference samples and SNPs.

Concluding Remarks

The GenomeEUtwin project will provide a unique genotype database for genetic epidemiological studies (see separate article on databases elsewhere in this issue of Twin Research). The first phase of the project will produce about 5 million multiallelic and 10 million SNP genotypes. This first phase will also provide an important infrastructure and harmonization for subsequent genotyping and twin studies, including quality control and genotype data management formats. A long-term goal is to expand the genotyping effort to as many as possible of the 600,000 twin pairs and their family members and thus provide one of the largest genotyped population cohorts globally. The extensive follow up and life style information collected in these twin cohorts combined with their genotype information provides truly exciting prospects for studies targeted on gene–environment interactions.

Endnotes

1. Available from http://research.marshfieldclinic.org/genetics
2. Celera data base is available from http://www.celera.com
3. The SNP Consortium is available from http://snp.cshl.org
4. dbSNP is available from http://www.ncbi.nlm.nih.gov/SNP/
5. The software is available from www.autoprimer.com.

References

Abecasis, G. R., Cherny, S. S., Cookson, W. O., Cardon, L. R. (2001). GRR: Graphical representation of relationship errors. *Bioinformatics, 17*, 742–743.

Bell, P. A., Chaturvedi, S., Gelfand, C. A., Huang, C. Y., Kochersperger, M., Kopla, R., et al. (2002). SNPstream UHT: Ultra-high throughout SNP genotyping for pharmocogenomics and drug discovery. *Biotechniques, Suppl.,* 70–72, 74, 76–77.

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian diseas, future approaches for complex disease. *Nature Genetics, 33* (Suppl.), 228–237.

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L., & Weber, J. L. (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *American Journal of Human Genetics, 63*, 861–869.

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press

Cardon, L. R., & Abecasis, G. R. (2003). Using haplotype blocks to map human complex trait loci. *Trends in Genetics, 19*, 135–140.

Carlson, C. S., Eberle, M. A., Rieder, M. J., Smith, J. D., Kruglyak, L., & Nickerson, D. A. (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nature Genetics, 33*, 518–521.

Couzin, J. (2002). Human genome. HapMap launched with pledges of $100 million. *Science, 298,* 941–942.

Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., et al. (2002) A first-generation linkage disequilibrium map of human chromosome 22. *Nature, 418,* 544–548.

Douglas, J. A., Boehnke, M., & Lange, K. (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *American Journal of Human Genetics, 66*, 1287–1297.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science, 296*, 2225–2229.

Gordon, D., Finch, S. J., Nothnagel, M., & Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: Application to single nucleotide polymorphisms. *Human Heredity, 54,* 22–33.

Hauser, E. R., Boehnke, M., Guo, S. W., & Risch, N. (1996). Affected-sib-pair interval mapping and exclusion for complex genetic traits: Sampling considerations. *Genetic Epidemiology, 13*, 117–137.

Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L., & Syvanen, A. C. (2002). Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Research, 30,* e70.

Mohlke, K. L., Erdos, M. R., Scott, L. J., Fingerlin, T. E., Jackson, A. U., Silander, K., Hollstein, P., Boehnke, M., & Collins, F. S. (2002). High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proceedings of the National Academy of Science, USA, 99*, 16928–16933.

O'Connell, J. R., & Weeks, D. E. (1998). PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *American Journal of Human Genetics, 63*, 259–266.

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Hori, M., Nakamura, Y., & Tanaka, T. (2002). Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature Genetics, 32*, 650–654.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. & Syvanen, A. C. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Research, 10*, 1031–1042.

Reich, D. E., Gabriel, S. B., & Altshuler, D. (2003). Quality and completeness of SNP databases. *Nature Genetics, 33*, 457–458.

Rice, K. M., & Holmans, P. (2003). Allowing for genotyping error in analysis of unmatched case-control studies. *Annual Human Genetics, 67*, 165–174.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*, 1516–1517.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*, 928–933.

Subramanian, S., Mishra, R. K., & Singh, L. (2003). Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. *Genome Biology, 4*, R13.

Syvanen, A. C. (1999). From gels to chips: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Human Mutations, 13,* 1–10.

Syvanen, A. C. (2001). Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nature Review Genetics, 2*, 930–942.

Syvanen, A. C., Sajantila, A., Lukka, M. (1993). Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *American Journal of Human Genetics, 52*, 46–59.

https://doi.org/10.1375/twin.6.5.391 Published online by Cambridge University Press