ELECTRONIC LETTER

Phenylketonuria screening registry as a resource for population genetic studies

U Hannelius, C M Lindgren, E Melén, A Malmberg, U von Dobeln, J Kere

J Med Genet 2005;**42**:e60 (http://www.jmedgenet.com/cgi/content/full/42/10/e60). doi: 10.1136/jmg.2005.032987

Background: Neonatal screening for metabolic diseases, involving samples stored on filter paper (Guthrie spots), provides a potential resource for genetic epidemiological studies.
Objective: To develop a method to make these dried blood spots available for large scale genetic epidemiology.
Methods: DNA from untraceable Guthrie spots was extracted using a saponin and chelex-100 based method and preamplified by improved primer preamplification. Analyses were done on 38 samples each of fresh, 10, and 25 year old Guthrie spots and the success rate determined for PCR amplification for five amplicon lengths.
Results: The method was applicable even on 25 year old samples. The success rate was 100% for 100 bp amplicons and 80% for 396 bp amplicons. Ninety four Guthrie samples were genotyped, including carriers of two different PKU mutations; all carriers were found (six R158Q, four R252W), with no false positives. Finally, 2132 anonymous samples from the Swedish PKU registry were extracted and pre-amplified and the allele frequencies of *APOε4*, *PPARγ* Pro12Ala, and the *CCR5* 32 bp deletion determined. Local variations in allele frequencies suggested subpopulation structuring. There was a significant difference ($p<0.01$) in regional allele frequencies for the CCR5 32 bp deletion in the Swedish population.
Conclusion: Whole genome amplification makes it feasible to conduct large genetic epidemiological studies using PKU screening registries.

Today's medical research is faced with the challenge of understanding genetic factors that give susceptibility to a wide variety of genetic disorders. To screen the genome effectively for the effect and presence of variants in genetically predisposed diseases, it is of great help to have a picture of what the allelic architecture, or the spectrum of different alleles influencing the disease, looks like.[1 2] With today's genotyping technologies and the efforts of the International HapMap Project,[3] it is now possible for smaller laboratories to undertake large genetic studies. The bottleneck is not the genotyping capacity but the sample collection and the availability of samples, as well as the financial and quality issues related to these.[4–6]

In 1963 Guthrie published data that showed the feasibility of using blood collected on filter paper for screening newborn infants for phenylketonuria (PKU). Since then, dried blood spots have been used routinely for determining the levels of myriad disease markers,[7 8] and retrospective studies on metabolite levels using these kinds of samples have been possible for some time.[9] The Swedish PKU registry encompasses all newborn infants born since 1975—a collection of approximately three million individual samples. Such repositories, in combination with the possibility of linkage with population based health registries, could provide invaluable information for molecular epidemiological studies exploring the effects of combined genetic and environmental risk factors in the causation of disease.

With the availability of several protocols designed to amplify the whole genome, small samples, even single cells, are no longer an obstacle to undertaking large scale genotyping.[10–13] However, though dried blood spots have already been used in some studies for genotyping purposes,[14] no one has clearly addressed the obvious limitations posed by the small sample size and long term storage of DNA derived from such samples if screening registries are opened to the scientific community.

The focus of our study was to develop a method that would make the dried blood spots available for undertaking genetic epidemiology on a large scale. By combining and optimising proven techniques we came up with a streamlined protocol that is robust, cost-effective, and easy to perform. We used the technique to determine the allele frequencies of three important disease variants—*APOε4*, *PPARγ* Pro12Ala, and the 32 bp deletion in *CCR5*[15–21]—in 2132 samples derived from the Swedish PKU registry. Our data show a significant geographical difference in Sweden for the *CCR5* 32 bp deletion, highlighting the importance of understanding the genetic structure even in a non-founder population.[22]

METHODS

The patient sample distribution for the PKU genotyping was unknown to us in order to keep the study unbiased. All samples, including two negative controls, were run as duplicates.

We collected 2058 samples, representing a cross section of Sweden from all children who were born in Sweden during one week in 2003. The samples were punched from irrevocably anonymous Guthrie cards, but preserving a record of the birth hospital, thus giving the geographical origin of each sample. Initial power calculations showed that approximately 100 samples from the northern part of Sweden needed to be included in order to make comparisons of region specific allele frequencies with reasonable power (see Statistical analysis). For practical reasons we decided to include 89 samples—that is, one 96 well plate excluding positive and negative controls. These samples originated from the counties of Norrbotten and Vasterbotten. Four negative controls and three positive controls were included on each 96 well plate. The study was approved by the ethics committee of Karolinska Institutet, Stockholm.

Abbreviations: I-PEP-L, improved primer preamplification; NCBI, National Center for Biotechnology Information; SNP, single nucleotide polymorphism; WGA, whole genome amplified

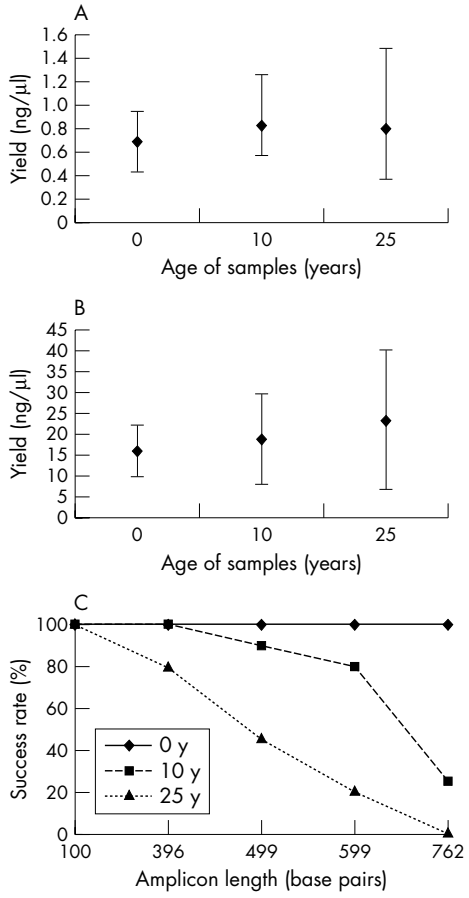

Figure 1 Yields and success rates of extraction and whole genome amplification measured on 0, 10, and 25 year old samples. In all, 38 sample assays in duplicate from each age group were measured in (A) and (B) and 20 samples in (C). Error bars in (A) and (B) represent the standard deviations. (A) The yield of single stranded DNA following DNA extraction, and of double stranded DNA following whole genome amplification (B), was constant between the sample groups, but the standard deviation was larger for the older samples. (C) The 25 year old samples were 100% successful for the shortest amplicons, sufficient in size for single nucleotide polymorphisms genotyping. The fresh samples were 100% successful even for the longest fragments.

Sample handling and DNA extraction

Using an automated puncher, 3 mm diameter punches where excised from the Guthrie cards. The filter paper used was S&S2992 for the 10 year old and 25 year old samples, and S&S903 for the fresh samples. Blood-free spots were used as negative controls. The DNA extraction was carried out in 200 µl polymerase chain reaction (PCR) certified DNAse and human DNA-free strips (Robbins Scientific Corporation, Sunnyvale, California, USA). A 150 µl aliquot of a phosphate buffered saline (PBS) solution containing 0.5% (wt/vol) saponin (VWR International AB, Stockholm, Sweden) was added, and the strips were centrifuged at 1800×*g* in a plate centrifuge for one minute or longer, to ensure that the punches were immersed in liquid. After overnight incubation at 4℃, the brown solution was removed and three washes of 200 µl PBS—one quick and two 30 minute washes at 4℃ each—were carried out. The PBS was removed and 150 µl of 5% chelex-100 in PBS was added to each well. The samples were vortexed vigorously for 30 seconds and incubated in a PCR machine (Biometra T100). During the first 30 minute incubation step at 60℃ the samples were vortexed twice more, and during the second 25 minute incubation at 95℃ they were vortexed three times. After the incubation the samples were cooled to 4℃ and centrifuged for 10 minutes at 1800×*g* in a plate centrifuge.

DNA quantification

The extracted DNA and the whole genome amplified samples were quantified in duplicate in 96-well format untreated black microtitre well plates (Nunc A/S, Roskilde, Denmark) using both the OliGreen and PicoGreen reagents (Molecular Probes, Eugene, Oregon, USA) according to the manufacturer's protocol.

Whole genome amplification

Improved primer preamplification (I-PEP-L; Dietmaier W, personal communication, 2002) was carried out in a 96-well format using a Biometra T100 thermocycler. A 50 µl sample of a mix containing 5% dimethylsulphoxide (DMSO), 8 mM totally degenerate 15-mer primer (Proligo, Boulder, Colorado, USA), 0.1 mM dNTP (Roche Diagnostics, Basel, Switzerland), 3.6 U Expand High Fidelity polymerase (Roche Diagnostics), and 2.5 mM $MgCl_2$ in 1×PCR buffer (No 3) was added to 10 µl of template (approximately 8–10 ng). The temperature profile for the reaction was as follows: an initial denaturation step of four minutes at 94℃ was followed by 20 cycles of a ramping step of 1℃/s to 94℃, 94℃ for 30 seconds, a ramping step of 1℃/s to 28℃, 28℃ for one minute, a ramping step of 0.1℃/s to 55℃, and 55℃ for 45 seconds; this was followed by 30 cycles of a ramping step of 1℃/s to 94℃, 94℃ for 30 seconds, a ramping step of 1℃/s to 60℃, 60℃ for 45 seconds, a ramping step of 1℃/s to 72℃, and 72 ℃ for one minute; a final elongation step of 72℃ for eight minutes finished the programme. Reagents and templates were kept on ice for the duration of the set up of the reaction.

PCR protocol: SNP genotyping

Amplification and detection primers for the polymorphisms in question were designed using the SpectroDesigner software (Sequenom, San Diego, California, USA). The specific PCR measuring 5 µl was undertaken in a 384-well format using Biometra T100 and Applied Biosystems 9700 thermocyclers. A 4 µl sample of PCR mix containing 0.2 µM PCR primer 1 or 2 (MetaBion, Martinsried, Germany), 0.2 mM dNTP (Roche Diagnostics), 2.5 mM $MgCl_2$ (Qiagen, Hilden, Germany), and 0.2 U Qiagen Hotstart Taq in 1×PCR buffer was added to 1 µl template (I-PEP reaction). All PCR reactions used the same temperature profile, with 15 minutes of denaturation at 95℃ followed by 45 cycles of 94℃ for 30 seconds, 60℃ for 15 seconds, and 72℃ for 15 seconds. A final elongation step of 72℃ for five minutes ended the programme. The PCR was set up using a Hamilton mph96 (Hamilton Co, Reno, Nevada, USA) pipetting station for pipetting mix and primers (3 and 1 µl, respectively) and a Beckman Multimek pipetting robot for dispensing template. The PCR products were analysed on a 2% agarose gel (NuSieve, Cambrex Corp, New Jersey, USA). All primer sequences are available in supplementary table 4 (available on the journal website: www.jmedgenet.com/supplemental).

PCR product clean up: SNP genotyping

Unincorporated deoxynucleotides where removed by adding 2 µl of SAP (shrimp alkaline phosphatase) mix containing 0.3 U SAP (Amersham Biosciences, Uppsala, Sweden) in 1×hME buffer (Sequenom) and incubating the samples at 37℃ for 20 minutes. The SAP enzyme was heat inactivated for five minutes at 85℃.

Primer extension protocol: SNP genotyping

Primer extension was undertaken in a total volume of 9 µl. The reaction contained 0.58 units of Thermosequenace enzyme (Amersham Biosciences), 1× termination mix (for

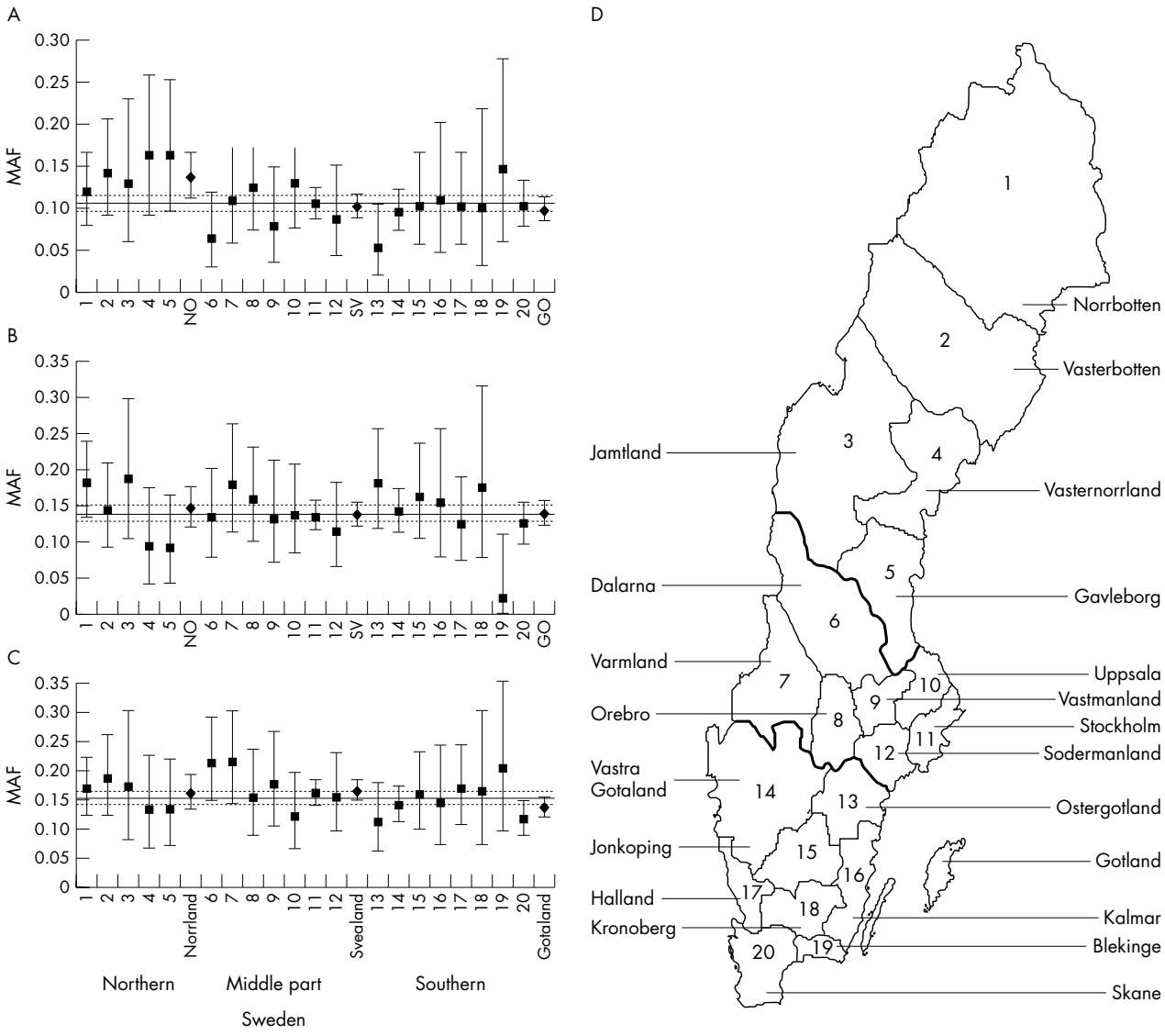

Figure 2 Geographically stratified minor allele frequencies. The minor allele frequencies for (A) CCR5, (B) PPARγ, and (C) APOε4 stratified according to the counties and regions in Sweden. The vertical continuous lines in (A)–(C) represent the allele frequencies for the whole of Sweden, and the dotted lines the corresponding 95% confidence intervals (CI). The error bars represent the 95% CI for each separate county and region. Counties are represented as numbers on the x axis in a north to south order. The corresponding names for the counties are depicted in (D), which represents a map of Sweden with the county borders marked with thin lines and regional borders marked with bold lines. The northernmost region of Norrland (NO) covers the greatest area but is at the same time the most scarcely populated region. Svealand (SV, the middle part of Sweden) and Gotaland (GO, the southern part of Sweden) have roughly the same size and population. Eighteen per cent of the approximately nine million Swedes are foreign born or have at least one non-native parent. Most of the immigrants are from other Nordic countries. There are two official minorities comprising Finnish speaking people living in the northeast along the Finnish border, and the Sami people living in the north (information derived from www.sweden.se).

example, dTTP, ddATP, ddCTP, ddGTP) and 5 μM of test specific MASSEXTEND primer (Metabion, Martinsried, Germany). Salt was removed by using an ion exchange resin (Sequenom, San Diego, California, USA) after which approximately 10 nl of the samples were spotted onto Maldimatrix-containing SpectroCHIPS (Sequenom).

Detection and analysis: SNP genotyping

The SpectroCHIPS were analysed by an Autoflex MassARRAY mass spectrometer (Bruker Daltonics, Billerica, Massachusetts, USA). Data were analysed independently by two persons using the SpectroTyper software (Sequenom Inc).

Statistical analysis

Power calculations were carried out using two-sample comparisons of proportions. After oversampling the northern part of Sweden we had, when comparing the regions of, for example, Norrland and Gotaland, at least 80% power for detecting differences in allele frequencies for *CCR5* 32 bp ins/del of 4.5% or higher ($p = 0.05$). This difference or greater has previously been reported for *CCR5* 32 bp ins/del when comparing northern and southern Europe. Based on the success rate (75%) for the *APOε4* (rs429358) assay on one plate (89 samples), we decided to exclude those data from the final results. Allele frequencies and success rates were calculated according to standard procedures, and the observed genotype frequencies were fitted to the Hardy–Weinberg equilibrium and evaluated using the χ^2 test (one sided p value). Calculations of 95% confidence intervals (CI) of the allele frequencies were based on the binomial distribution. Probability (p) values for differences in allele frequencies between different regions and counties were

obtained using the χ^2 test. Multiple testing was not corrected for, so many of the variations reported as significant may in fact result from sampling variation and represent false positive results when comparing counties (about 200 single tests performed). A Bonferroni correction would imply that only p values of less than 0.0002 would be significant when comparing allele frequencies on the county level. p Values for differences in DNA yields following DNA extraction and whole genome amplification were calculated using the Kruskal–Wallis rank test. All analyses were done in Stata.

RESULTS

DNA yield and quality

First, to determine the amount and quality of DNA extracted and amplified from dried blood spots, we took 38 punches each of fresh, 10 year old, and 25 year old Guthrie spots. Following DNA extraction, no significant difference in yields ($p>0.2$) between the fresh (mean (SD), 97 (36) ng), 10 year old (116 (62) ng), and 25 year old (112 (95) ng) blood spots were detected (fig 1A). The theoretical DNA yield from a 3 mm diameter punch is approximately 230 ng, assuming 6 pg of DNA per cell, 25 µl of blood on one 13 diameter spot, and seven million white blood cells per ml blood.

Next, we subjected the samples to whole genome amplification and measured the double stranded DNA yield. A 10 µl sample, approximately 8 ng of the extracted DNA, was used as template for one whole genome amplification reaction. This amount was empirically derived from pilot experiments (data not shown) and from previously reported guidelines.[13] Difference in amplification was non-significant ($p>0.5$) when comparing the fresh (mean (SD), 16 (6) ng/µl), 10 year old (19 (11) ng/µl), and 25 year old (23 (17) ng/µl) sample groups. Based on these yields (fig 1B) we obtained a 20-fold amplification of the genome, and about 13 individual whole genome amplified (WGA) reactions can be run on the DNA extracted from a single punch. The negative controls, being punches from bloodless filter paper that had been subjected to the DNA extraction phase, yielded high molecular weight material as well, but in the subsequent specific PCR assays that use human specific primers they turned out to be negative. Three different CEPH DNAs were used as positive controls.

Next we determined the success of the templates in supporting specific PCR assays giving rise to amplicons ranging from 100 to 760 bp. The PCR products were visualised on an agarose gel and the success rates were plotted against amplicon length (fig 1C). All fresh samples amplified for all amplicon lengths. The older samples had a success rate of 100% for the 100 bp amplicon, while 80% of the 25 year old and 100% of the 10 year old samples amplified for the 396 bp amplicon. The success rate for the oldest samples fell drastically when products of 499 base pairs and more were amplified: 90% the 10 year old samples were successful for the 499 base pair amplicon and 80% for the 599 base pair amplicon.

Validation for genotyping purposes

To validate our method for genotyping purposes we next determined whether we could detect any allele dropouts following the I-PEP-L amplification of DNA extracted from dried blood spots. We genotyped 90 WGA samples twice for three SNP variants where the genotypes were known for the unamplified samples. These results showed that there were no discrepancies between non-amplified and whole genome amplified DNA as template (data not shown).

Further validation was conducted by blindly characterising two different PKU variants, rs5030843 and rs5030847, in 94 samples including an unknown number of PKU patients, representing mutations in one or both of the two variants. All patients representing the different mutations were correctly identified. We have also conducted experiments showing that single stranded DNA that is too degraded to work as template for microsatellite genotyping can be remedied by preamplification (data not shown).

Allele frequency determination

We applied these methods on a large sample set representing a cross section of the Swedish population in order to study geographical differences for three disease associated genes.[23] The genes with the polymorphism and reported frequencies of risk alleles were *APOE* (ε2/3/4, 0.16–0.24), *CCR5* (32 bp Ins/Del, 0.05–0.07), and *PPARγ* (Pro12Ala, 0.85), respectively.[23] All the characterised polymorphisms were within Hardy–Weinberg equilibrium, and the allele frequencies were in the range of previously reported data (table 1). We analysed the data by geographical location in Sweden (county and region, fig 2 A–D). The total number of successfully genotyped alleles and the allele frequencies for each geographical region were calculated (supplementary tables 1–3). While most areas in Sweden showed homogeneous allele distributions and no

Table 1 Minor allele frequencies for the three variants as reported in different populations

Variant	Population sample set	Minor allele frequency	Hardy–Weinberg goodness of fit*
APOε4 (rs429358)	Our samples	0.154	0.035
	NCBI†	0.148	N/A
	Sweden‡	0.220	N/A
CCR5 32Bp ins/del	Our samples	0.106	0.388
	Sweden§	0.142	0.281
	Finland§	0.158	0.062
	Norway§	0.105	0.24
	Denmark§	0.110	0.216
PPARγ (rs1801282)	Our samples	0.138	0.727
	Scandinavian population¶	0.150	N/A
	NCBI	0.056	0.439

*The concordance of the observed allele frequencies in our sample set with the Hardy–Weinberg equilibrium are represented as a one sided p value of a χ^2 test. Corresponding values for the other populations are reported where available.
†The NCBI data are taken from the relevant home page (www.ncbi.nlm.nih.gov).
‡The allele frequency in the Swedish population for *APOε4* is listed as reported by Liljedahl *et al*.[24]
§The allele frequencies in the Scandinavian populations for the *CCR5* 32Bp ins/del variant are listed as reported by Libert *et al*.[25]
¶The allele frequency in a Scandinavian population for the *PPARγ* variant as reported by Altshuler *et al*.[26]
N/A, not assessed.

overall heterogeneity was present on the county level ($p>0.3$ for all three variants, 19 df), there were some counties that differed significantly when pairwise comparisons were made. These comparisons were not corrected for multiple testing and therefore the differences should be interpreted with caution. Most notably, the southern county of Blekinge had a significantly lower frequency of the protective Ala12 allele in the *PPARγ* gene than the counties of Stockholm ($p<0.05$), Vastra Gotaland ($p<0.02$), Varmland ($p<0.01$), Ostergotland ($p<0.01$), and Norrbotten ($p<0.01$) (fig 2B). We also observed a difference in the ϵ4 allele distribution between the counties of Dalarna and Skane ($p<0.01$). Here, the susceptibility allele was present at a higher frequency in Dalarna (fig 2C). A significant difference ($p<0.01$) for the *CCR5* 32 bp deletion was observed between the counties of Norrbotten and Ostergotland, the protective deletion being more common in the northern county of Norrbotten (fig 2A). On the region level we could not detect any significant differences in *PPARγ* gene for the Pro12Ala distribution. For the ϵ4 allele there was a moderate difference ($p<0.05$) between Svealand and Gotaland, the rare ϵ4 allele being present at a slightly lower frequency in the southern region of Gotaland (fig 2C). On studying possible large scale differences between three regions in Sweden, the most interesting finding was that the northernmost region of Norrland showed a higher prevalence of the *CCR5* 32 bp deletion variant than the two southern regions of Svealand ($p<0.02$) and Gotaland ($p<0.01$) (fig 2A).

DISCUSSION

Large population based sample repositories, such as the Swedish PKU registry, provide a potentially valuable source for large genetic studies, as considerable savings can be made in sample collection.[14] Even considering the multiplexing capabilities of today's genotyping platforms, the problem is to make the samples last beyond the relatively few PCR reactions that can normally be run on the small amounts of DNA extracted from 3 mm diameter punches. Without a way of ensuring a considerable increase in lifetime and usage possibilities for individual samples it is neither rational nor ethical to make these kinds of registries available to the research community. We have therefore developed a method that uses a DNA extraction step which produces inhibitor-free DNA that works well for a subsequent whole genome amplification step. As the DNA extraction and whole genome amplification works even on 25 year old samples for the 100 bp long amplicons (fig 1A–C) it makes all samples stored in similar kinds of registries available for SNP genotyping on most platforms. Based on the amplification success for the 396 and 499 bp amplicons, the genotyping can be extended to microsatellites when using 10 year old samples as source material (fig 1C). As 13 60 μl WGA reactions can be run from a single 3 mm diameter punch, from which in turn 1 μl is used for a subsequent genotyping step, the total quantity of available singleplex PCRs amounts to about 800.

The sample repository of 2132 extracted and preamplified Guthrie spots provides us with an opportunity to investigate the genetic structure in the Swedish population. Even though all the samples are anonymous, the place of birth is known and hence this makes it possible for us to stratify the genotype data on several levels of geographical resolution. The suggested differences in allele frequencies for the *CCR5* 32 bp deletion at a regional level were very interesting (fig 2A). Significant heterogeneity between European populations for the *CCR5* mutant allele ($p<0.001$, 17 European populations sampled) and a similar north to south downhill gradient has previously been reported in Europe as a whole.[25] Although the heterogeneity in our sample set is only moderate when comparing all three regions for the *CCR5* variant ($p<0.02$, 2 df), the data clearly showed that the protective allele is more prevalent in the northern part of Sweden. Notable allele frequency differences for many genes have been observed in the neighbouring country Finland.[27] These observations support the notion that sample stratification may pose a problem in case–control studies, even when a non-founder population is sampled, and attention should be paid to hidden differences between cases and controls, especially in certain regions of Sweden.[22] We will address this subject in an upcoming project where several anonymous markers will be typed and analysed by clustering algorithms in order to study the overall geographical differences more thoroughly.

There are potential problems in using a PCR based whole genome amplification method. Preferential allele amplification has been observed when using too little template, five cells or less, for the PEP reaction,[28] and allelic dropouts may occur when too little template is used for the specific PCR following whole genome amplification.[11] Significant bias of the order of 10^2 to 10^4 for the amount of amplification between genetic loci using PEP has also been reported. An alternative technique to ours, called MDA (multiple displacement amplification), is based on the strand displacing enzyme from phage ϕ24 that facilitates isothermal amplification of the whole genome. MDA excels in the areas mentioned above, as well as in amplification efficiency.[10] However, the protocol for PEP we are using has on the other hand successfully been used to investigate loss of heterozygosity (LOH) in hyperplasic and concomitant cancerous tissue.[29] Also, the experiments we have conducted here show no ambiguities in genotypes when comparing the use of unamplified and whole genome amplified templates, or when comparing our results to results confirmed previously in the PKU screening laboratory. Other techniques, such as MDA—which are otherwise superior in many aspects to PEP—encounter problems when degraded or short stranded DNA is used as template. As older dried blood spots generally have not been stored optimally[30] they might not be suitable as a template source for MDA. We have also compared the genotyping results when using either I-PEP-L or MDA (GenomiPhi kit, Amersham Biosciences) amplified DNA from 20 buccal swabs. In this case, success rate for genotyping was only 61% (of 160 genotypes) for the MDA method, while it was 90% (of 160 genotypes) for the I-PEP-L method. There were also nine contradicting genotyping results between the MDA and I-PEP-L amplified groups. Of these, four were determined to be wrong in the MDA group owing to Mendelian inconsistencies. Five other contradicting genotypes could not be determined as all were correctly inherited (Peyrard-Janvid M, personal communication, 2004). The weak results for MDA might reflect the fact that the method used for extracting DNA from the swabs includes a heating step which probably fragments the DNA to some extent.

Although the PKU registries are constantly growing because of routine screening, they constitute a finite number of samples and should be treated with the utmost care. The method we have presented here has proven to be very efficient for processing these kinds of samples, and the basic idea of using whole genome amplification should be adopted by anyone thinking of accessing PKU registry samples.

ACKNOWLEDGEMENTS

We wish to dedicate this publication to the memory of the late professor Lars Hagenfeldt who was the head of Centre for Inherited Metabolic Diseases and the Swedish Neonatal Screening Laboratory and took part in the planning of this investigation. We thank Ville Makela for his technical assistance in the microsatellite genotyping, Marianne Soderqvist for providing the sample plates, and Henrik Ahlman for his assistance in the DNA quantification. This study was

supported by the Swedish Research Council. J Kere is also affiliated to the Chemical Research Centre, Karolinska University Hospital and the Department of Medical Genetics, University of Helsinki, and is a member of Biocentrum Helsinki, C M Lindgren is also affiliated to the Department of Biosciences at Novum, Karolinska Institutet.

Supplementary tables can be viewed on the journal website (www.jmedgenet.com/supplemental).

Authors' affiliations

U Hannelius, J Kere, Department of Biosciences at Novum, Karolinska Institutet, Stockholm
C M Lindgren, Clinical Research Centre, Karolinska University Hospital, Stockholm Sweden
E Melén, Institute of Environmental Medicine, Karolinska Institutet
A Malmberg, U von Dobeln, Centre for Inherited Metabolic Diseases, Karolinska University Hospital

Competing interests: none declared

Correspondence to: Ulf Hannelius, Karolinska Institutet, Department of Biosciences, Hälsovägen 7-9, 14157 Huddinge, Sweden; Ulf.Hannelius@biosci.ki.se

Received 18 March 2005
Revised version received 11 May 2005
Accepted for publication 17 May 2005

REFERENCES

1 **Smith DJ**, Lusis AJ. The allelic structure of common disease. *Hum Mol Genet* 2002;**11**:2455–61.
2 **Pritchard JK**, Cox NJ. The allelic architecture of human disease genes: common disease-common variant.or not? *Hum Mol Genet* 2002;**11**:2417–23.
3 **Anonymous**. The International HapMap Project. *Nature* 2003;**426**:789–96.
4 **Gray IC**, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 2000;**9**:2403–8.
5 **Syvanen A**. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001;**2**.
6 **Holland NT**, Smith MT, Eskenazi B, Bastaki M. Biological sample collection and processing for molecular epidemiological studies. *Mutat Res* 2003;**543**:217–34.
7 **Guthrie R**. Screening for phenylketonuria. *Triangle* 1969;**9**:104–9.
8 **Guthrie R**. The origin of newborn screening. *Screening* 1992;**1**:5–15.
9 **Alm J**, Hagenfeldt L, Larsson A, Lundberg K. Incidence of congenital hypothyroidism: retrospective study of neonatal laboratory screening versus clinical symptoms as indicators leading to diagnosis. *BMJ (Clin Res Ed)* 1984;**289**:1171–5.
10 **Dean FB**, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du, Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002;**99**:5261–6.
11 **Zhang L**, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci USA* 1992;**89**:5847–51.
12 **Zheng S**, Ma X, Buffler PA, Smith MT, Wiencke JK. Whole genome amplification increases the efficiency and validity of buccal cell genotyping in pediatric populations. *Cancer Epidemiol Biomarkers Prev* 2001;**10**:697–700.
13 **Lovmar L**, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucl Acids Res* 2003;**31**:e129.
14 **Hamvas A**, Trusgnich M, Brice H, Baumgartner J, Hong Y, Nogee LM, Cole FS. Population-based screening for rare mutations: high-throughput DNA extraction and molecular amplification from Guthrie cards. *Pediatr Res* 2001;**50**:666–8.
15 **Corder EH**, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;**261**:921–3.
16 **Strittmatter WJ**, Saunders AM, Schmechel D, *et al*. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci USA* 1993;**90**:1977–81.
17 **Tang MX**, Maestre G, Tsai WY, Liu XH, Feng L, Chung WY, Chun M, Schofield P, Stern Y, Tycho B, Mayeux R. Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. *Am J Hum Genet* 1996;**58**:574–84.
18 **Dragic T**, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 1996;**381**:667–73.
19 **Deng H**, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 1996;**381**:661–6.
20 **Samson M**, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 1996;**382**:722–5.
21 **Liu R**, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996;**86**:367–77.
22 **Freedman ML**, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschorn JN, Altshuler D. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;**36**:388–93.
23 **Hirschhorn JN**, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;**4**:45–61.
24 **Liljedahl U**, Karlsson J, Melhus H, Kurland L, Lindersson M, Kahan T, Nyvstrom F, Lind L, Syvanen AC. A microarray minisequencing system for pharmacogenetic profiling of antihypertensive drug response. *Pharmacogenetics* 2003;**13**:7–17.
25 **Libert F**, Cochaux P, Beckman G, Samson M, Aksenova M, Cao A, Czeizel A, Claustres M, de la Rua C, Ferrari M, Ferrec C, Glover G, Grindle B, Guran S, Kucinskas V, Lavinha J, Mercier B, Ogur G, Peltonen L, Rosatelli C, Schwartz M, Spitsyn V, Timar L, Beckman L, Parmentier M, Vassart G. The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet* 1998;**7**:399–406.
26 **Altshuler D**, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;**26**:76–80.
27 **Kere J**. Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet* 2001;**2**:103–28.
28 **Dietmaier W**, Hartmann A, Wallinger S, Heinmoller E, Kerner T, Endl E, Jauch KW, Hofstadter F, Ruschoff J. Multiple mutation analyses in single tumor cells with improved whole genome amplification. *Am J Pathol* 1999;**154**:83–95.
29 **Obermann EC**, Junker K, Stoehr R, Dietmaier W, Zaak D, Schubert J, Hofstaedter F, Knuechel R, Hartmann A. Frequent genetic alterations in flat urothelial hyperplasias and concomitant papillary bladder cancer as detected by CGH, LOH, and FISH analyses. *J Pathol* 2003;**199**:50–7.
30 **Karlsson H**, Guthenberg C, von Dobeln U, Kristensson K. Extraction of RNA from dried blood on filter papers after long-term storage. *Clin Chem* 2003;**49**:979–81.