European Journal of Human Genetics (2000) 8, 653–660
© 2000 Macmillan Publishers Ltd All rights reserved 1018–4813/00 $15.00
www.nature.com/ejhg

ARTICLE

A new human genetic resource: a DNA bank established as part of the Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC)

Richard W Jones[1], Susan Ring[2], Linda Tyfield[2], Renata Hamvas[1], Hugh Simmons[3], Marcus Pembrey[1], Jean Golding[3] and ALSPAC Study Team[3]

[1]Institute of Child Health, University College London; [2]Molecular Genetics Unit, The Lewis Laboratories, Southmead Hospital, Westbury-on-Trym, Bristol; [3]ALSPAC, Unit of Paediatric and Perinatal Epidemiology, Division of Child Health, University of Bristol, Bristol, UK

We describe a unique human DNA resource forming part of the Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC), a longitudinal cohort study involving 14 000 children and their families living in a geographically defined area of England. The DNA bank will underpin the search for associations between genetic polymorphisms and common health outcomes. The opportunities to collect blood samples suitable for DNA extraction are necessarily limited, and the samples themselves have often been treated in different ways and have varied storage histories. With the need to maximise yields, the choice of DNA extraction method is critical to the success of the bank and we have investigated the suitability of various commercial and in-house methods of DNA extraction. Various steps have been taken to minimise errors in sample address and identification, including the use of a pipetting robot for dilution and transfer of samples between 96-well arrays to provide aliquots suitable for PCR. The robot has been programmed to cope with concentrated viscous DNA solutions. *European Journal of Human Genetics* (2000) **8**, 653–660.

Keywords: DNA extraction; DNA bank; robot

Introduction

A strategy which incorporates genetic variation and other direct biological information into population-based epidemiological studies of common disorders has gained wide support over the last 5 years[1] and it is now widely accepted that gene/environment interactions can be best studied in total population samples. As candidate genes and their common sequence variations emerge from the human genome project, such studies will play a central role in relating this wealth of data to health outcomes. The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) is a study of 14 000 children and their families born in 1991–1992 in an area centered on Bristol, formerly the Avon district of the UK. Full details of the study are in the ALSPAC protocol which may be found on the World Wide Web (http://www.ich.bristol.ac.uk). The children are being followed by means of regular questionnaires and periodic clinical examinations designed to measure and record growth, physical and behavioural development, educational attainment, the prevalence of common childhood diseases and a wide range of environmental variables. Biological samples are being collected which will allow biochemical and genetic analyses. The interests of the children and their families are protected by an Ethics and Law Committee[2] within ALSPAC and by local medical ethics committees.

ALSPAC is designed to illuminate the ways in which physical and social environments interact with genotype to affect health, behaviour and development. Identifying the genetic components of common outcomes in childhood and adult life is therefore an integral part of the study. An association between the insulin gene *VNTR* and infant size at birth for a subset of the cohort has been published.[3,4] Studies on associations between various polymorphisms and childhood infection,[5] atopy, asthma, and with intermediate

Correspondence: Richard W Jones, Institute of Child Health, University College London, 30 Guilford Street, London WC1N 1EH, UK.
Tel: +44 20 7242 9789 ext. 0732; Fax: +44 20 7404 6191;
E-mail: r.jones@ich.ucl.ac.uk
Received 22 December 1999; revised 21 March 2000; accepted 4 April 2000

phenotypes related to diabetes and heart disease, are under way or planned.

Maternal samples for DNA extraction were obtained during the course of routine antenatal care and cord blood samples were collected at birth. Blood samples have also been requested, with maternal consent, from cohort children as they attend an ALSPAC clinic at 7 years of age. About 70% of children agree to have a sample taken. The source, storage and processing of each sample are logged in a computer stock control database. DNA is extracted and processed to provide 96-well arrays containing DNA aliquots at a uniform concentration. These are distributed for analysis by external groups collaborating with ALSPAC through its Genetics Advisory Committee. Given that the opportunities to collect blood samples from cohort members are limited, ensuring that the arrays will support the maximum possible number of analyses is a prime concern. Consequently the DNA extraction methods used in the study must maximise DNA yield from a variety of sample types. The extracted DNA must then be processed to provide aliquots suitable for PCR. Whilst robotic systems have been developed to streamline many aspects of genetic analysis,[6] the wide range of concentrations and the viscosity of DNA extracted from blood samples presents particular problems for liquid handling and has required development of a robotic protocol that allows such solutions to be pipetted with adequate precision.

Materials and methods

Sample collection

Samples have been obtained from several sources. Cord blood was collected in heparin at birth and the white cells removed and stored at –70°C for 5 to 8 years before DNA was extracted. Further blood samples were taken from a sub-cohort of children (children in focus) who attended ALSPAC clinics at approximately 43 and 61 months of age. These samples were anticoagulated with EDTA and the white cells separated and stored at –70°C for 1 month to 2 years before extraction. The majority of children, now aged 7 years, are attending the 'Focus@7' ALSPAC clinic where approximately 70% of children provide a 4 ml EDTA blood sample specifically for DNA extraction. These samples are immediately frozen unseparated and stored at –20°C for up to 3 weeks before DNA extraction. Approximately 8000 of these samples will have been collected from children by the end of this clinic. Coverage of the cohort will be extended by extracting DNA from mouth wash samples from some of the children who did not wish to give blood. Finally, extraction of DNA from cord blood will add a further 4000 samples to give an overall coverage of approximately 85% of the cohort children.

Blood samples were collected from approximately 10 000 mothers during their normal antenatal care and either anticoagulated with EDTA and stored unseparated at –20°C or collected in heparin and the white cells removed and stored at –70°C. DNA has been extracted from these samples approximately 5–7 years after they were collected.

Sample identification

Each sample is allocated a number linking it to the individual's confidential record in a database with restricted access (Figure 1). Samples for DNA extraction are given a laboratory accession number which identifies them through the extraction process. This number is logged into a DNA database (Microsoft Access) which assigns each sample to a position in a bar-code numbered 96-well array. After extraction, DNA samples are transferred to their allocated wells. An identical control DNA sample is at a different position in each plate so that a plate's identity can be checked by the position of a unique genotype at the correct control position; 3–7% of samples, depending on the set, are from the same individuals sampled at different times. These duplicates act as controls for sample processing and genotyping.

DNA extraction

Commercial methods explored as candidate methods for DNA extraction were: GenoMag genomic DNA Kit (Advanced Biotechnologies, Epsom, Surrey, UK); Wizzard genomic DNA purification kit (Promega, Southampton, Hants, UK); Puregene Genomic DNA purification kit (Flowgen Instruments Ltd, Lichfield, Staffs, UK); QIAamp blood Midi Kit and QIAamp blood Maxi Kit (Qiagen, Crawley, West Sussex, UK).

All samples from long-term storage are now extracted using a standard phenol–chloroform extraction method.[7] Blood samples stored at –20°C for less than one month are extracted

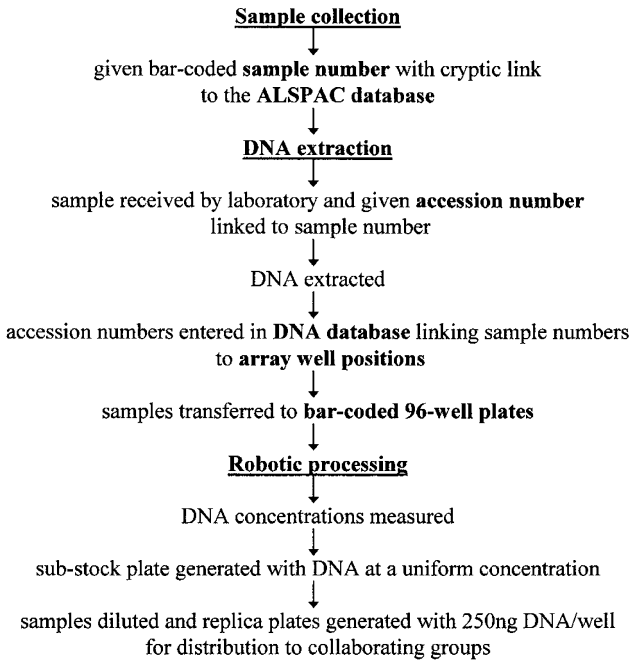

Figure 1 Extracting and processing DNA for genetic analysis.

using a salting-out method.[8] All batches of DNA extractions include a control sample to monitor recovery. Extracted DNA is transferred to deep 96-well plates (Advanced Biotechnologies) which are heat sealed with aluminium foil and stored at –50°C.

To minimise transposition errors during manual processing, two laboratory workers are involved in all transfers from tube to tube or tube to 96-well array. In transfers to 96-well arrays, a system of masking is employed to prevent samples being placed in the wrong well. Once placed in an array, a robot performs all subsequent liquid transfers.

Robotic processing of DNA samples

A Biomek 2000 (BeckmanCoulter Ltd, High Wycombe, Bucks, UK) is used to process DNA stocks and generate replica plates containing DNA aliquots at a uniform concentration suitable for PCR (Figure 2). The robot uses a combination of single and eight-gang air-displacement pipetting tools with disposable tips. Samples from DNA stocks are first serially diluted in 10 mM TrisCl pH 8.0 to allow DNA concentrations to be measured. Standards (λ-DNA) are added manually to the plate and DNA concentrations measured by fluorimetry. The volumes of sample and diluent (water) required to generate a plate of samples at uniform concentration are automatically calculated and inserted in a program (PlateTek Ltd, High Wycombe, Bucks, UK), written in Bioscript (BeckmanCoulter Ltd) and Tcl,[9] which directs the robot to produce a further array of samples at a uniform concentration. This plate is then used to generate replica plates which are distributed to collaborating groups for genotyping.

Each robotic process has a source and one or more destination 96-well plates. To access a programme the source and destination plate bar-codes are read into the computer. The prefix letters determine which programme is selected. Access is prevented if the numbers on the two plates are different or not presented in the correct order.

All DNA solutions are mixed before quantitative transfers. Mixing consists of 10 or more aspirate and dispense cycles in which solutions pass through a 0.5 mm orifice polypropylene tip (P200, BeckmanCoulter Ltd) at 70 μl/s. For quantitative transfers, sample volumes under and over 20 μl are aspirated at pipetting speeds of 1.5 μl/s and 12 μl/s and with biases of 2 μl and 4 μl, respectively. Tips are withdrawn slowly from solutions at 1 mm/s after mixing, aspirating or dispensing.

DNA used to test robotic processing was either high molecular weight calf thymus DNA (Roche Diagnostics Ltd, Lewes, Sussex, UK) or human DNA extracted from duplicated ALSPAC samples.

Measurement of DNA concentration

In tests of the different DNA extraction methods, DNA concentration was calculated from the A^{260}. As part of sample processing, DNA concentrations were measured in a Fluoroskan Ascent (Labsystems, ThermoQuest, Basingstoke, Hants, UK) using PicoGreen[10] (Molecular Probes, Lieden, Netherlands) with λ-DNA (Life Technologies Ltd, Paisley, UK) as standards. The callibration curve is forced through zero to reduce inaccuracy at low DNA concentrations. The factor used to convert fluorescence to DNA concentrations in the original stocks was calculated by taking λ-DNA of known concentration through the dilution and measurement process, thus avoiding assumptions about the exact volumes pipetted by the robot.

PCR

To detect errors in sample handling, PCR was carried out on presumed duplicate samples using primers at the HLA–DBR locus.[8] When the PCR products are melted and allowed to re-anneal they generate a complex pattern of heteroduplexes which is unlikely to be duplicated in any two individuals in the ALSPAC cohort.

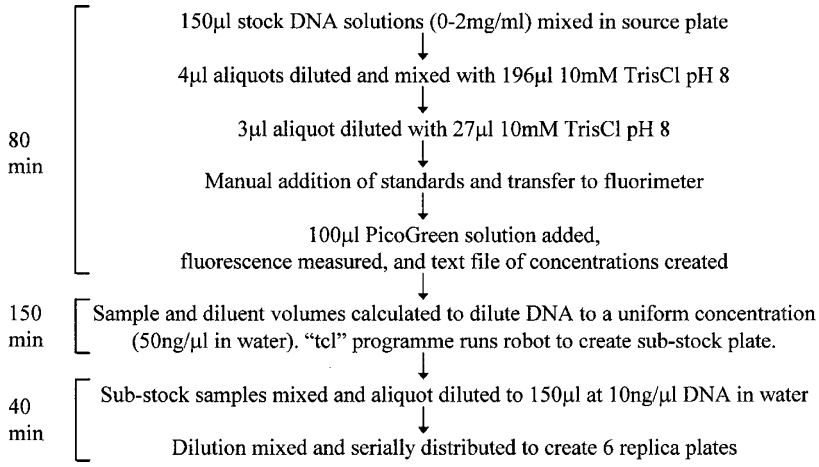

Figure 2 Robotic processing of DNA sample arrays. The time required for each step in the process is shown on the left of the figure.

Results

DNA extraction

A number of commercial DNA extraction methods were compared with an in-house salting-out method. All the methods gave adequate DNA yields from small fresh blood samples and there were no PCR failures. The commercial systems, because they required fewer tube transfers and in some cases used microfuges, were faster. However some of this advantage was lost when the commercial methods were scaled up to extract DNA from larger sample volumes because of the need to use full-size centrifuges. Furthermore, the consumables cost of the in-house method was about half that of the commercial systems.

In the Qiagen system, DNA is adsorbed on to a membrane and the extraction process can be automated using a vacuum manifold system. Consequently, further comparisons were made between this method, the in-house salting-out method, and phenol–chloroform extraction. With whole blood (Figure 3A), phenol–chloroform extraction gave the highest DNA yields (mean = 42 μg/ml blood, n = 186) with salting-out (mean = 37 μg/ml blood, n = 999) giving only slightly less. The membrane adsorption method gave half the yield of either of these methods (mean = 15 μg/ml blood, n = 16).

Much of the material available in the ALSPAC study has been stored for a number of years as whole blood or as white cell pellets. These samples often contain clot, despite anti-coagulation, and DNA prepared from them by the in-house salting-out method or commercial methods was often coloured by retained haemoglobin indicating incomplete removal of protein. DNA extraction from stored white cell pellets by the membrane adsorption method and phenol–chloroform extraction were compared (Figure 3B). As expected, both methods gave approximately half the yields obtained when they were applied to whole blood but the membrane method (mean = 9 μg/ml blood, n = 98) was still less efficient than phenol–chloroform extraction (mean = 23 μg/ml blood, n = 102). For all three methods, whether used to extract white cell pellets or whole blood, the distribution of yields about their respective means is very similar and the salting-out and phenol–chloroform methods are therefore expected to have the lowest rate of absolute failures.

Phenol-chloroform extraction was chosen to prepare DNA from whole blood samples and white-cell pellets that have been in long-term frozen storage. DNA is extracted from whole blood samples that are fresh, or frozen for less than a month, using the salting-out method since the advantage of using non-toxic reagents outweighs the small reduction in yield compared with phenol–chloroform extraction.

Robotic processing of DNA samples

Figure 2 summarises the steps involved in robotic processing. Solutions of high molecular weight DNA are viscous and therefore difficult to pipette accurately or precisely. The small orifices in 20 μl capacity tips obstruct DNA solution flow and only 200 μl tips are used to handle DNA solutions, even at small volumes. Tips cannot be re-used for quantitative transfers because the surface tension of residual DNA solution at the tip orifice obstructs aspiration, especially of small volumes. For quantitative transfers, slow pipetting speeds encourage non-turbulent flow of viscous solution through the tip orifice and prevent bubbles forming during aspiration,

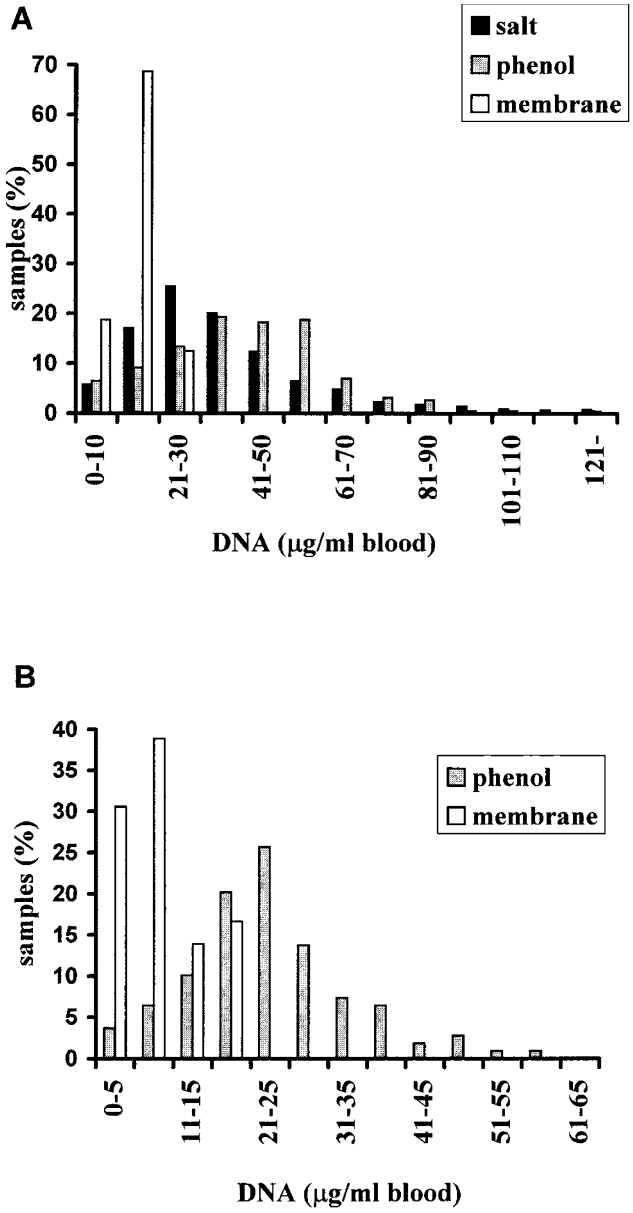

Figure 3 Distributions of DNA yields from blood **A** and buffy coat **B** using a salting-out method (salt), phenol/chloroform extraction (phenol) and a commercial (Quiagen) membrane adsorption method (membrane). DNA concentration was estimated by measuring absorbence at 260 nm.

or droplets being left on the inside of the tip during dispensing. Generous bias volumes and delays at the end of all aspirating or dispensing moves assist equilibration against the resistance of solution flow through the tip orifice. Touching the tip on the side of wells after aspirating or dispensing fails to remove drops adhering to the outside of the tip and instead the robot is set to withdraw tips slowly from samples so that the viscous DNA solution drains down the outside of the tips and a clean break is made as the tip leaves the meniscus.

Incorporating these provisions into the robot protocols ensures accurate transfers which are independent of sample viscosity. Diluent is always added to destination wells first. Aliquots are drawn up slowly into the pipette tip from the source well and the tip withdrawn slowly from the solution. At the destination well the aliquot is dispensed directly into diluent followed by three wash cycles to ensure complete emptying without any blow-out action and the attendant risk of generating aerosols. The tip is then slowly withdrawn. Dilute DNA solutions are dispensed into empty wells with the tip in contact with the well wall. The tip is then wiped on the well wall as it is slowly withdrawn. DNA aliquots disperse slowly when diluted so that subsequent transfers are very imprecise unless dilutions are adequately mixed. Acceptable precision was only achieved if all dilution steps were followed by a mixing step involving 10 aspirate and dispense cycles performed at 75% of maximum pipetting speed.

The performance of the robotic process was tested using commercially prepared high molecular weight calf thymus DNA at concentrations up to 2 mg/ml. Figure 4 shows DNA concentrations measured by PicoGreen fluorescence for samples prepared by dilution of calf thymus DNA. The relationship is linear but with a positive bias for PicoGreen measurements. Table 1 shows the intra-assay (A) and inter-assay (B) precision for DNA concentration measurements. Good precision was achieved across the whole range of concentrations showing that the pipetting protocols adopted to handle viscous DNA solutions were successful in ensuring consistent performance.

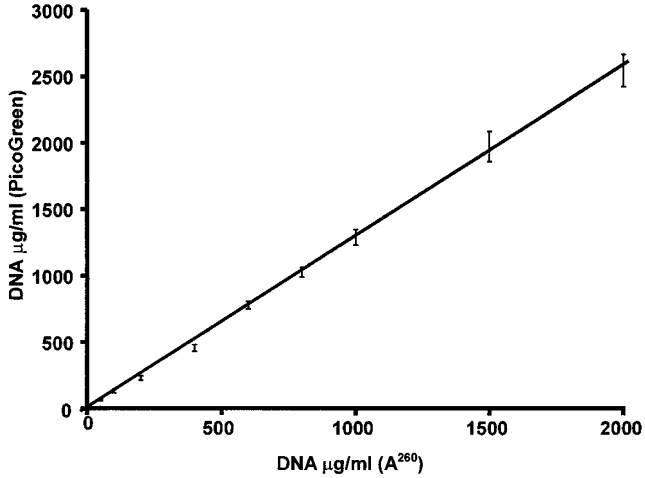

Figure 4 The graph shows the regression of high molecular weight calf thymus DNA concentrations measured by PicoGreen fluorescence on concentrations measured by absorbance at 260 nm (A^{260}). Each point is the mean +/– standard error for eight estimations.

The precision of the overall robotic process was measured as the variation in DNA concentrations in the final replica plate in which all samples, other than those prepared from stocks containing less than 50 ng/µl, should contain DNA at 10 ng/µl. Using calf thymus DNA samples at starting concentrations ranging from 0.1 to 2 mg/ml. (Table 2, A) a CV of 16.3% was obtained. Human DNA samples from the ALSPAC cohort are likely to differ from highly purified commercial DNA preparations since they are extracted in large numbers and will inevitably vary in quality according to the state of the original blood samples and the efficiency of the extraction process. We therefore tested the ability of the process to generate replica plates containing DNA samples at a uniform concentration from arrays of these samples. Starting DNA concentrations ranged from zero to 2.6 mg/ml (Table 2, B) and zero to 0.9 mg/ml (Table 2, C). Initially, poorer precision was obtained than with commercial calf thymus DNA, but acceptable precision was achieved by increasing the number

Table 1 Intra-(**A**) and inter-(**B**) assay precision at a range of calf thymus DNA concentrations measured using PicoGreen. Intra-assay precision was calculated from a single plate with eight samples at each concentration and inter-assay precision was calculated from a single set of DNA concentrations in each of nine plates. The mean, standard deviation (SD) and coefficient of variation (CV) are shown for each DNA concentration

	DNA µg/ml (A260)										
	0	*50*	*100*	*200*	*400*	*600*	*800*	*1000*	*1500*	*2000*	
A *n*=8											
DNA µg/ml	–0.5	51	106.6	200.2	427.2	797.6	1055.3	1328.6	2139.5	2602.5	mean
(PicoGreen)	0.1	1.1	1.3	3.9	9.2	34.8	22.5	45.4	59.4	80.1	SD
		2.2	1.2	2.0	2.2	4.4	2.1	3.4	2.8	3.1	CV (%)
B *n*=9											
DNA µg/ml	–0.5	66.4	128.0	228.0	450.8	775.8	1004.1	1248.6	1911.2	2420.0	mean
(PicoGreen)	0.4	7.5	10.1	15.6	23.0	28.9	36.3	43.8	127.2	179.5	SD
		11.2	7.9	6.8	5.1	3.7	3.6	3.5	6.7	7.4	CV (%)

Table 2 Precision of the robotic process measured as variation in DNA concentration in the final (uniform DNA concentration) replica plates for PCR. Each plate contains a maximum of 88 samples. Samples of DNA in plates B and C at original concentrations less than 50μg/ml give replicates at less than 10ng/μl and are excluded. Standard deviations (SD) and coefficients of variation (CV) are shown

	Plates		
	A	*B*	*C*
DNA	calf thymus	human	human
n	88/88	75/88	83/88
mean (ng/μl)	8.1	9.0	9.9
SD	1.3	1.6	1.8
CV (%)	16.3	17.8	18.0

and speed of cycles in the mixing steps. This may reflect the presence of more contaminating protein in these samples compared with the calf thymus DNA.

Sample probity

Seventy-one pairs of DNA samples, each of each pair putatively taken from the same individual on different occasions, extracted from cord blood or from samples taken at clinics attended by some children at age 43 or 61 months, were tested for identification errors using PCR at the HLA–DBR locus.[7] Four pairs did not match. Assuming that only one of each of the unmatched pairs was wrongly identified this gives an error rate of 2.8%. No errors have been found in 35 pairs of presumed duplicate samples among 2333 DNA samples from children at the 7-year clinic.

PCR success rates

Table 3 shows the success rates for PCRs performed on arrays of DNA aliquots from heparinised or EDTA anti-coagulated blood. Success rates probably vary in part depending on the persistence of individual laboratories in attempting to get refractory DNA samples to PCR. Heparin anti-coagulation was thought to adversely affect PCR for the detection of the insulin gene *VNTR* polymorphisms and a preliminary whole genome amplification step was introduced to overcome this. Pre-amplification was also used to detect the insulin receptor substrate-1 gene polymorphism. Heparin anti-coagulation did not appear to have an adverse effect on other PCRs.

Table 3 Success rates for PCRs performed by collaborating groups using arrays containing DNA aliquots prepared by phenol–chloroform extraction from blood anticoagulated with EDTA or heparin. PCRs were for (1) the insulin VNTR,[3] (2) the insulin receptor substrate-1 gene 972 polymorphism,[11] (3) the angiotensin converting enzyme gene intron 16 insertion/deletion polymorphism,[12] (4) the interleukin-4 gene R-576 polymorphism,[13] and (5) the mannose-binding lectin exon-1 polymorphisms[5]

DNA	*polymorphism*	*pre-amplification*	*anti-coagulant*	*total*	*failed %*
mother	InsVNTR (1)	yes	heparin	534	3.5
			EDTA	273	3.3
child	InsVNTR (1)	yes	heparin	768	5.4
			EDTA	349	1.7
child	IRS-1,972 (2)	yes	heparin	812	3.1
			EDTA	355	0.6
child	ACE (3)	no	heparin	755	2.3
			EDTA	338	9.5
child	IL4,R576 (4)	no	heparin	354	4
			EDTA	128	7.8
child	MBL, codons 52, 54 and 57 (5)	no	heparin	680	0.4
			EDTA	303	0

Discussion

Selecting a DNA extraction method

The choice of DNA extraction method is critical in the creation of a DNA resource for a cohort study of this size. Any method must provide the maximum yield so that as many genetic analyses as possible can be supported, and the DNA must be of sufficient quality to minimise the number of unsuccessful PCRs. Our initial expectation was that an automatable commercial system would fulfil these requirements. Some commercial DNA extraction methods involve adsorption of DNA on to membranes or magnetic particles, removing the need for centrifugation and allowing automation. However, these methods have limited DNA capacity and are therefore unsuited to the needs of the ALSPAC bank. This leaves the smaller number of tube transfers in commercial as opposed to in-house methods as the main source of time-saving, but that advantage diminishes when considered as a proportion of the total preparation time including, as it does, time-consuming pre- and post-extraction steps.

Of all the methods tested, only phenol–chloroform extraction coped with the high DNA content of cord blood, with the fibrin clot in maternal pregnancy samples, or with samples which had been stored frozen for several years. It is salutary that phenol–chloroform extraction is still the most robust of the DNA extraction methods and provides the highest yields. Otherwise, the in-house salting-out method was as effective as commercial methods in extracting DNA from fresh blood samples but, like phenol–chloroform extraction, was cheaper because of lower consumable costs. Using either of the in-house methods, two laboratory workers comfortably extract 140 samples per week as well as performing the various administrative tasks required to log the passage of samples through the laboratory and their transfer to 96-well arrays. Allowing for the varied history of the blood samples and the presence of some small samples from which very little DNA can be extracted, the overall PCR success rates for DNA prepared by the in-house methods appear acceptable.

Robotic processing

Simple replication of an array of samples into daughter plates can be performed manually using various hand tools. However, this is tedious and risks inaccuracy and cross-contamination of samples. Transferring different volumes of

solution from individual well positions to create arrays at a uniform DNA concentration is particularly prone to error if performed manually.

Robots provide an obvious solution to the problem of maintaining consistency and accuracy in processing thousands of samples, but their application is not straightforward because of the need to find strategies which overcome the problem of handling viscous DNA solutions. However, the indefatigability of a robot has made it easier to resolve this problem than would have been the case if pipetting were performed manually. Thus aspirating and dispensing rates, and the rate at which tips leave solution can be accurately controlled, while manual strategies commonly adopted to manage DNA solutions, such as wiping tips on the side of the well, can be reproduced. DNA solutions disperse only slowly when diluted, even when solutions are already quite dilute, and dilutions must be thoroughly mixed to generate a homogenous solution and avoid inaccurate transfers at subsequent pipetting steps. A robot can be set to perform this task in a completely consistent manner with the operator able to control the vigour and number of mixing cycles. The precision with which a wide range of genomic DNA concentrations can be measured attests to the effectiveness of the robotic protocol described here. It is unlikely that the same precision could be achieved manually.

Sample probity

Duplicated samples are included among samples taken for DNA extraction as a means of measuring misidentification rates. Among a limited number of samples collected and extracted at an early stage in the study, there was an identification error rate of 2.8%. Such errors will include errors in labelling the original blood samples, in recording or transcribing variables identifying the individual or numbers identifying the sample, in separating blood constituents, during transfers in the course of manual DNA extraction, and in transfers to multi-well plates. Samples collected from children at the 7-year clinic, which is being attended by the majority of cohort members and is a principle source of DNA from the cohort, are taken specifically for DNA extraction, are bar-code labelled, and are not separated before DNA extraction. All transfers in the course of DNA extraction are double-checked. So far there is no evidence of mis-identification, though it is noteworthy that even an identification error rate of 2.8% will have only a limited effect on the strengths of associations between genotypes and outcomes in the cohort.

Future maintenance and use of the DNA arrays

Because of the wide range of yields (Figure 3) the proportion of useful samples in arrays will decline as the resource is used. The first 60 replicates will on average be missing 3 out of 88 samples. After 500 replicates, about half the samples will be exhausted. Whole genome amplification prior to analysis, as already practised by one of the collaborating groups, will greatly extend the lifetime of the bank, as will improvements in the efficiency of PCR, including multiplexing, which should allow smaller quantities of DNA to be consumed per polymorphism. However, a bank depending on direct DNA extraction from blood samples is a finite resource and the foreseeable expansion in the number of candidate polymorphisms over the next decade could run down the existing samples quite quickly. Strategies for patching arrays as individual samples are exhausted are likely to be organisationally complex. Although ALSPAC holds alternative samples for DNA extraction for many cohort members, the eventual need to request further blood samples will pose ethical and practical problems.

Lymphoblastoid cell lines are expensive to set up (approximately £100 per sample), but once established they are able to provide, repeatedly, relatively consistent amounts of DNA across samples. Arrays would have a longer life with a sharper end-point and their replacement would be relatively straightforward. Thus, long-term organisation and maintenance would be simpler than for a collection based on direct extraction from blood. A bank set up on this basis would be likely to repay much of its initial cost by improving the viability of the resource and eliminating the need to reorganise arrays, or return to cohort members for more blood samples.

As the number of candidate genes and polymorphisms expands as a result of the human genome project, ALSPAC will become an increasingly important resource helping to determine the significance of newly discovered genetic variation. Direct measurement of continuous variables across the whole cohort avoids assumptions about disease definitions and increases the likelihood of discovering significant associations. Furthermore, the accumulation of both genetic and outcome data from the cohort will allow hypotheses to be tested with increasing efficiency since a single infrastructure and data set will serve multiple projects.

Acknowledgements

DNA is extracted and banked using funds from the Medical Research Council. We are extremely grateful to the children and families of the ALSPAC cohort for their co-operation in providing samples and to the midwives and others who helped to make these samples available. The ALSPAC study team comprises interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers and managers. ALSPAC is part of the WHO initiated European Longitudinal Study of Pregnancy and Childhood and is supported by the Wellcome Trust, the Medical Research Council, the Department of Health, the Department of Environment and many others.

References

1 NIGMS: The Genetic Architecture of Complex Traits. Report from National Institute of General Medical Sciences, USA, Jan. 1998, http://www.sfbr.org/nigms/.
2 Mumford SE: Children of the 90s: ethical guidance for a longitudinal study. *Arch Dis Child* 1999; **81**: F146–F151.
3 Dunger DN, Ong KKL, Huxtable SJ *et al*: Association of the *INS* VNTR with size at birth. *Nat Genet* 1998; **19**: 98–100.

4 Ong KKL, Phillips DI, Fall C *et al*: The insulin gene VNTR, type 2 diabetes and birth weight. *Nat Genet* 1999; **21**: 262–263.
5 Mead R, Jack D, Pembrey M *et al*: Analysis of mannose-binding lectin alleles in a prospectively recruited Bristol population. *Lancet* 1997; **349**: 1669–1670.
6 Nizetic D, Zehetner G, Monaco AP *et al*: Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc Natl Acad Sci USA* 1991; **88**: 3233–3237.
7 Sambrook J, Fritsch EF, Maniatis F: *Molecular Cloning, A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York, USA: 1989.
8 Bidwell J, Wood N, Clay TM *et al*: DNA heteroduplex technology. In: Chrambach A, Dunn MJ, Radola BJ (eds), *Advances in Electrophoresis*. VCH Press: Weinheim, 1994; vol 7: pp 311–351.
9 Ousterhout JK: *Tcl and the Tk Toolkit*. Addison-Wesley, Mass., USA: 1994.
10 Ahn SJ, Costa J, Emanuel JR: Picogreen quantitation of DNA: effective evaluation of samples pre- and post-PCR. *Nucleic Acids Res* 1996; **24**: 2623–2625.
11 Zhang Y, Wat N, Stratton IM, Warren-Perry MG *et al*: UKPDS 19: Heterogeneity in NIDDM: Separate contributions of IRS-1 and beta 3-adrenergic-receptor mutations to insulin resistance and obesity respectively with no evidence for glycogen synthase gene mutations. *Diabetologia* 1996; **39**: 1505–1511.
12 O'Dell SD, Humphries SE, Day INM: Rapid methods for population-scale analysis for gene polymorphisms: The ACE gene as an example. *Br Heart J* 1995; **73**: 368–371.
13 Hershey GK, Friedrich MF, Esswein LA *et al*: The association of atopy with a gain of function mutation in the alpha subunit of the interleukin-4 receptor. *New Engl J Med* 1997; **337**: 1720–1725.