

## ARTICLE

# Dissecting a population genome for targeted screening of disease mutations

Tomi Pastinen<sup>1</sup>, Markus Perola<sup>1,2</sup>, Jaakko Ignatius<sup>3,4</sup>, Chiara Sabatti<sup>2</sup>, Päivi Tainola<sup>1</sup>, Minna Levander<sup>1</sup>, Ann-Christine Syvänen<sup>1,5</sup> and Leena Peltonen<sup>1,2,4,\*</sup>

<sup>1</sup>Department of Molecular Medicine, National Public Health Institute, Biomedicum, 00250 Helsinki, Finland,

<sup>2</sup>Department of Human Genetics, University of California, Los Angeles, CA 90095-7088, USA, <sup>3</sup>Department of Clinical Neurophysiology, Helsinki University Hospital (Jorvi Hospital), 02740 Espoo, Finland, <sup>4</sup>Department of Medical Genetics, University of Helsinki, 00250 Helsinki, Finland and <sup>5</sup>Molecular Medicine, Department of Medical Sciences, S-751 85 Uppsala, Sweden

Received August 6, 2001; Revised and Accepted October 10, 2001

Compared to mixed populations, population isolates such as Finland show distinct differences in the prevalence of disease mutations. However, little information exists of the differences on the prevalence of different disease alleles in regional populations with different history of multiple bottlenecks. We constructed a DNA-array and monitored the prevalence of 31 rare and common disease mutations underlying 27 clinical phenotypes in a large population-based study sample. Over 64 000 genotypes were assigned in 2151 samples from four geographical areas representing early and late settlement regions of Finland. Each sample was analyzed in duplicate and a total of 142 000 array-derived genotyping calls were made. On average one in three individuals was found to be a carrier of one of the 31 monitored mutations. This should remove fears of the stigmatizing effect of a carrier-screening program monitoring multiple diseases. Regional differences were found in the prevalence of mutations, providing molecular evidence for the deviating population histories of regional subisolates. The mutations introduced early into the population revealed relatively even distribution in different subregions. More recently introduced rare mutations showed local clustering of disease alleles, indicating the persistence of population subisolates and the effect of multiple bottlenecks in molding the population gene pool. Regional differences were observed also for common disease alleles. Such precise information of the carrier frequencies could form the basis for targeted genetic screens in this population. Our approach describes a general paradigm for large-scale carrier-screening programs also in other populations.

## INTRODUCTION

DNA variants in over 1000, mostly monogenic, traits have been identified (1). Identification of mutations has not resulted in immediate DNA diagnosis due to the high heterogeneity of disease alleles. Isolated populations provide special advantages for DNA diagnostics and carrier-screening programs due to a limited spectrum of disease mutations; one test having high diagnostic specificity and sensitivity (2,3). Targeted screening of ethnically restricted disease mutations in the appropriate population subgroups has demonstrated its efficiency in disease prevention (4). However, for most populations, rational design of genetic-screening programs requires large population-based pilot studies to monitor for the diversity and prevalence of specific disease mutations.

DNA-array technology is a promising approach to monitor large numbers of sequence variants in one assay (5–8). If a moderate number of variants in a large number of samples are to be analyzed, a custom-made, spotted oligonucleotide arrays used for enzymatic allele discrimination provides a high-throughput system (8). Such arrays are flexible and can be tailored for population-specific carrier screening of several disease variants. Information on the frequency and regional distribution of individual disease mutations as well as the validation of array-based assays must exist before such tests can be implemented population-wide as a standard component of the health care system.

We constructed a DNA-array for the detection of 31 mutations in a study sample of 2400 individuals from different geographical

\*To whom correspondence should be addressed at: UCLA Department of Human Genetics, Gonda Neuroscience and Genetics Research Center, Room 6506, 695 Charles E. Young Drive South, Box 708822, Los Angeles, CA 90095-7088, USA. Tel: +1 310 794 5631; Fax: +1 310 794 5446; Email: lpeltonen@mednet.ucla.edu

Present address:

Tomi Pastinen, Montreal Genome Centre, Montreal, H3G 1A4 Quebec, Canada

regions of Finland to monitor for the regional differences in the prevalence of disease alleles. The majority of the mutations included on the array, 'the Finnish Chip', are recessive disease mutations enriched in the Finnish population, belonging to the 'Finnish disease heritage'. In addition, we incorporated to the array some disease mutations common in most Caucasian populations, like mutations of  $\alpha$ -1-antitrypsin, factor V and the hereditary hemochromatosis gene.

We did not include on the array dominant disease mutations like those of hereditary breast and ovarian cancer (BRCA1 and 2), familial non-polypotic colon cancer or familial hypercholesterolemia due to their rarity. Recent studies in cancer patients (9,10) or population-based surveys (11) demonstrate the low prevalence of these mutations in the Finnish population (<1:400).

The microarray-based primer extension assay proved to be highly robust and provided reliable genotyping results at a relatively low cost. Interestingly, we found distinct regional differences in the carrier frequency of both rare and common disease mutations within this genetically homogeneous, isolated population, the findings providing support for multiple historical population bottle necks. The population-wide carrier frequency of disease alleles detected by this panel of 31 mutations was 1:3, illustrating the prediction that if a wider panel of disease mutations were to be included on the array, every Finn would be a carrier of at least one of the tested disease mutations. This should alleviate fears of the stigmatizing effect of carrier-screening programs monitoring numerous disease mutations.

## RESULTS

### Mutations screened with the DNA array

The mutation-screening panel on the DNA-array comprised the known major mutations of Finnish diseases (12), a total of 19 mutations in 16 different genes. Most mutations show a strong founder effect and the coverage of the mutation detection varied between 74 and 99% for different diseases. The common Caucasian mutations consisted of 10 mutations in nine genes including two common polymorphisms of factor V (13) and the prothrombin gene (14). A list of the mutations included on the array as well as the corresponding diseases and OMIM symbols are provided in Table 1. A pair of allele-specific detection primers for each of the mutations was spotted on derivatized microscopic glass-slides (8). Eighty replicate arrays were spotted onto each slide and custom-made reaction chambers were used to analyze up to 80 separate samples per slide. This facilitated the monitoring of 2480 mutations (31 genotypes for 80 individuals) on a single microscopic glass slide (Fig. 1). The allele-specific primer extension reactions provided reliable discrimination between the obtained genotypes (8).

We determined the carrier frequencies of disease mutations by genotyping 2151 anonymous DNA samples from four geographical regions (Fig. 2). One rural sample was collected from southern Botnia, representing the 'early settlement' region, which was inhabited some 2000 years ago. Another rural sample was collected from North Karelia, representing the 'late settlement' northeastern region, permanently inhabited after the 16th century. The two urban community

samples were from Helsinki and from the city of Oulu on the northwestern coastline. Both of these cities have been targets for internal population movement during the past eight to nine decades. However, according to Y chromosomal haplotype analysis, the influence from early immigration from Western Europe is much less evident in Oulu region than in southern Finland (15). Further, immigration to Helsinki, the capital, has been much more excessive than to Oulu. As positive controls, 212 samples from carriers of the disease mutations were allocated blindly among the study samples.

### Quality of the data from the arrays

Each sample was amplified by PCR and assayed in duplicate on separate microarrays. The genotyping results for each sample were read and the genotypes were independently assigned from the two arrays. Moreover, the genotypes were called on coded samples, without knowledge of the sample status to serve as quality control of the array-based screening procedure. The reference methods were systematically used to verify the genotyping result if a sample indicated a carrier status, if the results from the duplicate assays were discrepant or if the signal intensities on a microarray were too low to assign the genotype. For samples with low signal intensity on the array, the reference method was performed only once, and the sample was omitted from further analysis if this reference assay failed. In the cases of conflicting duplicate genotype calls and for every carrier genotype sample the reference method was used to confirm the genotype.

Since each sample was analyzed in duplicate, a total of 142 000 array derived genotyping calls were made. None of the heterozygous controls was assigned as normal homozygotes. Of the 71 000 genotypes generated, 95% were called successfully. In 5% of the samples the signal intensities were too low to make a reliable genotype call, and the reference methods were used to assign the genotypes. Less than 0.1% of the samples gave conflicting results in the duplicate array-based assay, due to inconsistencies resulting from the spotting of the primer-arrays. The final assignment of genotype in the study samples was made in 99.15% of cases (64 076 out of 64 625 genotypes were assigned). The genotyping results are summarized in Tables 2 and 3.

### Carrier frequencies of the 'Finnish disease heritage' mutations

A particularly informative approach to the analysis of the regional prevalence of the Finnish mutations is to divide them based on the time of their assumed introduction into the population. Recent introduction of a disease allele is characterized by the geographically restricted occurrence of patients, a genealogical history revealing ancestors in the same communities and a demonstration of linkage disequilibrium over a wide genetic interval flanking the disease mutation (12,16). Representative examples of such mutations in our screen include vLINCL (17), EPMR (18), Salla disease (19), CCD (20) and two X-linked retinoschisis (RS) (21) mutations (Table 1). Figure 3 demonstrates the carrier frequencies of five young mutations in the four regional samples. Notably, all these mutations show evidence of local clustering of disease alleles, indicating the persistence of the subisolates. A striking example of such clustering is the high frequency (1:44) of the vLINCL mutation

**Table 1.** A list of the mutations included on the array as well as the corresponding OMIM symbols and diseases

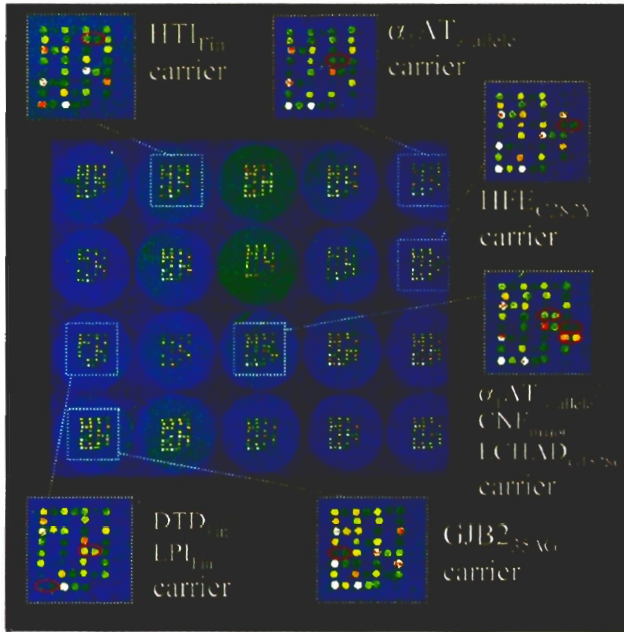
Mutation	Gene name (SYMBOL)	OMIM symbol <sup>a</sup>	Common disease name
<b>Finnish mutations</b>			
AGU	<i>Aspartylglucosaminidase (AGA)</i>	208400	Aspartylglycosaminuria
APECED	<i>Autoimmune regulator (AIRE)</i>	240300	Autoimmune polyendocrine syndrome, type I
CCD	<i>Down-regulated in adenoma (DRA)</i>	214700	Familial chloride diarrhea
CNF <sub>Major</sub> , CNF <sub>Minor</sub>	<i>Nephrin (NPHN)</i>	256300	Congenital nephrosis 1, Finnish type
DTD	<i>Human sulfate transporter (DTDST)</i>	222600	Diastrophic dysplasia
EPMR	<i>Ceroid lipofuscinosis, neuronal 8 (CLN8)</i>	600143	Progressive epilepsy with mental retardation
GA1 <sub>OAT181</sub> , GA2 <sub>OAT402</sub>	<i>Ornithine-Δ-aminotransferase (OAT)</i>	258870	Hyperornithinemia with gyrate atrophy of Choroid and retina
HFI <sub>A149P</sub>	<i>Aldolase B (ALDOB)</i>	229600	Hereditary fructose intolerance
HTI	<i>Fumarylacetoacetate hydrolase (FAH)</i>	276700	Tyrosinemia, type I
INCL	<i>Palmitoyl-protein thioesterase (PPT)</i>	256730	Infantile neuronal ceroid lipofuscinosis
LPI	<i>Solute carrier family 7, member 7 (SLC7A7)</i>	222700	Lysinuric protein intolerance
NKH	<i>Glycine cleavage system P protein (GCSP)</i>	238300	Nonketotic hyperglycinemia 1
ODG	<i>Follicle-stimulating hormone receptor (FSHR)</i>	233300	Hereditary hypergonadotropic ovarian failure
RS <sub>Pori</sub> , RS <sub>Oulu</sub>	<i>Retinoschisis 1 (XLR1)</i>	312700	X-linked, juvenile RS 1
Salla	<i>Solute carrier family 17, member 5 (SLC17A5)</i>	604369	Sialuria, Finnish type
VLINCL	<i>Ceroid lipofuscinosis, neuronal 5 (CLN5)</i>	256731	Variant late infantile neuronal ceroid lipofuscinosis
<b>Common Caucasian mutations</b>			
α1AT <sub>z-allele</sub>	<i>Protease inhibitor 1 (PI)</i>	107400	α-1-Antitrypsin deficiency
CFΔTT394, CFΔF508	<i>Cystic fibrosis transmembrane conductance regulator (CFTR)</i>	219700	Cystic fibrosis
FV <sub>Leiden</sub>	<i>Coagulation factor V (F5)</i>	227400	Thrombophilia due to deficiency of cofactor for activated protein C
GJB2Δ35G	<i>Gap junction protein, β-2 (GJB2)</i>	220290	Neurosensory deafness, autosomal recessive 1
HFE <sub>C282Y</sub>	<i>Hemochromatosis (HFE)</i>	235200	Hereditary hemochromatosis
PKU <sub>R408W</sub>	<i>Phenylalanine hydroxylase (PAH)</i>	261600	Phenylketonuria
PT <sub>G20210A</sub>	<i>Coagulation factor II (F2)</i>	176930	Hyperprothrombinemia
<b>Others</b>			
BATTEN	<i>Ceroid lipofuscinosis, neuronal 3</i>	204200	Juvenile neuronal ceroid lipofuscinosis, Batten disease
LCHAD	<i>Hydroxyacyl-CoA dehydrogenase (HADHA)</i>	600890	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency
MCAD	<i>Acyl-CoA dehydrogenase, medium-chain (ACADM)</i>	201450	Deficiency of medium-chain acyl-CoA dehydrogenase
MPS-I	<i>α-1-Iduronidase (IDUA)</i>	252800	Mucopolysaccharidosis type I

<sup>a</sup>The reference in which the common Finnish or Caucasian allele included in the mutation-screening panel was described can be found at <http://www.ncbi.nlm.nih.gov/Omim/> [online Mendelian inheritance in man (OMIM). Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD); 1999] by using this number.

in the southern Botnia sample whereas the carrier frequency in the 'mixed' Helsinki population is only 1:1000. Clustering is also evident for CCD and Salla disease in which the prevalence of disease mutation was highest in North Karelia and lowest in the Helsinki sample. However, the distribution for the Salla mutation was relatively even across all geographical regions when compared with vLINCL, which would provide evidence for a relatively early introduction of Salla mutation to the Finnish population. Only one of the two major mutations causing X-linked RS could be detected as one RS<sub>Oulu</sub> carrier was identified in the Oulu sample. No carriers for the other mutation, RS<sub>Pori</sub>, were seen in any of our samples, probably reflecting the fact that the mutation is highly concentrated in a

single community, Noormarkku (21), located in southwestern Finland and outside the regions sampled for this study. Similar findings were obtained for EPMP, a rare recessive disease known to occur in a very restricted region in Kainuu in north-eastern Finland (18) which was also not covered in the sampling for the present study.

Somewhat differently from young mutations, the old Finnish mutations show relatively even distribution of patients and their carrier frequencies (Fig. 4). The most common of the Finnish diseases, congenital nephrosis (22), is caused by two major mutations, CNF<sub>major</sub> and CNF<sub>minor</sub>, accounting for 78 and 16% of disease alleles, respectively. We found a relatively even distribution of the CNF<sub>major</sub> allele across all the geographical



**Figure 1.** Examples of screening for Finnish mutations using allele-specific primer extension on microarrays. A section showing the results from 20 samples is illustrated. Eighty samples are analyzed per microscope slide with the aid of rubber grids forming 80 separate reaction wells. On a single slide 2480 genotypes were scored. Differences in signal patterns indicate mutations, and the six carriers identified are enlarged with circles around probe pairs with signals indicative of heterozygosity for the given site. Each array contains a pair of allele-specific probes for each mutation to be detected, one probe corresponding to the wild-type allele (odd vertical columns) and another probe corresponding to the mutated allele (even columns). Carriers are identified by calculating the ratio between the signal intensities obtained from the wild-type probe sites and the mutant probe sites for each mutation. The average signal ratios that discriminate between the genotypes are >10-fold. They vary between the mutations, but are relatively constant for each mutation. For example, the discrimination against the z-allele of the  $\alpha$ -1-antitrypsin gene is over 20-fold due to the low background signal for the mutant probe in row four, column six of the array in non-carriers and high signal intensities at this site in carriers of the mutation. The non-carriers of the common mutation underlying lysinuric protein intolerance show relatively high background signals at the mutant probe sites in row eight, column two, but the very high signal intensity from the wild-type probe in non-carriers discriminates the carriers from non-carriers by 5-fold differences in signal ratios. The accuracy of genotype discrimination by our microarray system for this set of mutations was validated previously in a sample of >400 individuals including known carriers in Pastinen *et al.* (8).

regions whereas we saw a significant clustering of the CNF minor allele in the sample from the early settlement southern Botnia region (Fig. 4). In this region, the combined carrier frequency of the two mutations was as high as 1:27 (suggesting birth incidence of 1:2500), whereas in the other study samples the combined carrier frequency varied from 1:48 to 1:61. The second most common Finnish disease is aspartylglucosaminuria, in which one major mutation ( $AGU_{Fin}$ ) accounts for 98% of disease alleles (23). Again the carrier frequency varied, the prevalence being lowest in the old settlement, southern Botnia region (1:132) and highest in the late settlement North Karelia (1:63). Diastrophic dysplasia (24), the third most common representative of Finnish diseases, has a major mutation identified in 90% of disease alleles. Interestingly we found a high frequency of the DD major mutation in the regional samples from North Karelia and Oulu, a finding not expected



**Figure 2.** Map of Finland. The geographical areas where the study samples were obtained are shaded in gray.

based on the results of earlier epidemiological and genealogical studies (25).

The combined carrier frequency of all tested Finnish mutations varied between ~1:11 and ~1:6 in regional study populations, being highest in the rural eastern and western study samples.

### Mutations common in Caucasian populations

The carrier frequencies for common Caucasian mutations are shown in Figure 5. The Z-allele of the  $\alpha$ -1-antitrypsin gene revealed a higher prevalence in Helsinki and early settlement southern Botnia, when compared to the late settlement northeastern sample (1:26 versus 1:46). The frequency of the common mutation of hereditary hemochromatosis,  $HFE_{C282Y}$ , varied across Finland, showing high prevalence in the early settlement western sample (1:10), with decreasing frequency in the late settlement northeastern Finland (1:33). An opposite regional distribution was evident for  $FV_{Leiden}$  (13) and prothrombin $G20210A$  (14), SNPs predisposing to venous thromboembolism. Both were more commonly encountered in the late settlement sample. Another common disease mutation which was monitored was a common frameshift mutation in the connexin 26 gene referred to as  $GJB2_{\Delta35G}$ , resulting in congenital non-syndromic deafness in several populations (26,27). The  $GJB2_{\Delta35G}$  mutation had a relatively uniform geographical distribution across Finland with a carrier frequency of 1:43–1:63. The combined carrier frequency of the common Caucasian mutations in Finland varied between ~1:7 and ~1:5, being highest in the rural, early settlement region.

### Mutations with exceptionally low prevalence in Finland

We monitored two mutations,  $\Delta F508$  and  $\Delta TT394$  in the CFTR gene, reported to account for 45 and 30% of Finnish CF chromosomes, respectively (28). The carrier rate for the

**Table 2.** The genotyping results and extrapolated carrier frequencies

	Rural areas						Urban areas					
	Early settlement Botnia/western			Late settlement North Karelia/eastern			Helsinki			Oulu		
	No. of scored samples	No. of carriers	Carrier frequency	No. of scored samples	No. of carriers	Carrier frequency	No. of scored samples	No. of carriers	Carrier frequency	No. of scored samples	No. of carriers	Carrier frequency
Finnish mutations												
AGU	396	3	1:132	385	7	1:55	1005	16	1:63	365	4	1:91
APECED	396	5	1:79	385	4	1:96	1005	8	1:126	365	1	1:365
CCD	396	2	1:198	385	4	1:96	1002	3	1:334	363	0	–
CNF <sub>Major</sub>	395	10	1:40	385	8	1:48	1005	18	1:56	365	5	1:73
CNF <sub>Minor</sub>	396	5	1:79	385	0	–	1005	3	1:335	365	1	1:365
DTD	395	4	1:99	385	9	1:43	1001	14	1:72	365	8	1:46
EPMR <sup>a</sup>	149	0	–	150	0	–	150	0	–	148	0	–
GA <sup>1</sup> <sub>OAT181</sub>	396	0	–	385	0	–	1005	0	–	365	1	1:365
GA <sup>2</sup> <sub>OAT402</sub>	394	3	1:131	384	3	1:128	1002	15	1:67	365	1	1:365
HFI <sub>A149P</sub>	396	3	1:132	385	1	1:385	1005	7	1:143	365	2	1:183
HTI	396	4	1:99	385	0	–	1004	1	1:1004	365	0	–
INCL	396	13	1:30	385	9	1:43	1004	12	1:84	365	4	1:91
LPI	391	3	1:130	385	4	1:96	972	5	1:194	365	4	1:91
NKH	391	2	1:196	385	3	1:128	1001	1	1:1001	365	4	1:91
ODG	395	0	–	385	2	1:193	996	2	1:498	365	0	–
RS <sub>Pori</sub>	395	0	–	385	0	–	1005	0	–	365	0	–
RS <sub>Oulu</sub>	396	0	–	385	0	–	1004	0	–	365	1	1:365
Salla	343	4	1:86	385	6	1:64	941	6	1:157	365	3	1:122
VLINCL	396	9	1:44	385	1	1:385	1000	1	1:1000	365	2	1:183
Common mutations												
$\alpha$ 1AT <sub>z</sub> -allele	396	14	1:28	349	14	1:39	1005	38	1:26	365	8	1:46
CF $\Delta$ TT394	396	0	–	384	0	–	1003	4	1:251	365	2	1:183
CF $\Delta$ F508	396	4	1:99	385	2	1:193	1002	6	1:167	365	1	1:365
FV <sub>Leiden</sub>	394	6	1:66	385	16	1:24	1005	39 <sup>c</sup>	1:26	365	11	1:33
GJB2 $\Delta$ 35G	395	8	1:49	385	9	1:43	1002	16	1:63	365	8	1:46
HFE <sub>C282Y</sub>	394	38 <sup>d</sup>	1:10	384	19 <sup>d</sup>	1:20	1001	56	1:18	365	11	1:33
PKUR408W	396	1	1:396	385	0	–	1004	3	1:334	365	2	1:183
PT <sub>G20210A</sub>	396	3	1:132	385	4	1:96	1004	6	1:167	365	4	1:91
Others												
BATTEN	312	2	1:156	385	5	1:77	827	5	1:165	365	1	1:365
LCHAD <sup>b</sup>	395	3	1:132	385	2	1:193	492	3	1:164	365	1	1:365
MCAD	396	2	1:198	385	0	–	1005	2	1:503	365	1	1:365
MPS-I	395	1	1:395	385	1	1:385	988	5	1:198	365	1	1:365
Combined carrier frequencies	39.1%			33.5%			29.4%			25.2%		

Due to the rarity of some mutations, regional differences in carrier frequencies may be affected by random fluctuation. In Table 3 confidence intervals for binomial proportions on the base of likelihood evaluations (66,67) are provided for more extensive comparisons of allele frequencies.

<sup>a</sup>EPMR mutation was included on the array screening at a later phase and only 150 samples from each geographic region were included in the analysis.

<sup>b</sup>The first design of the reference method produced false negative results due to amplification of a pseudogene with the selected PCR primers, thus the 500 first samples analyzed in Helsinki were omitted from the data.

<sup>c</sup>One homozygote for FVLeiden mutation was found in the Helsinki population.

<sup>d</sup>Two homozygotes for HFE C282Y mutation were found in the Southern Botnia population.

CF mutations was low in all study samples, the combined carrier frequency varying from 1:99 in the western to 1:193 in

the eastern study sample. The common PKU mutation, PAH R408W, accounts for 50% of Finnish PKU alleles (29). The

**Table 3.** Confidence intervals (based on binomial likelihoods) for the allele frequency estimation

Disease	Botnia		North Karelia		Helsinki		Oulu	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
AGU	0.00076	0.01104	0.00369	0.01867	0.00457	0.0129	0.00152	0.01397
APECED	0.00202	0.0147	0.00139	0.01324	0.00174	0.00784	0.00006	0.00762
CCD	0.00029	0.0091	0.00139	0.01324	0.00027	0.00438	0	0.00509
CNFMajor	0.00612	0.02318	0.00449	0.02037	0.00534	0.01409	0.00221	0.01591
CNFMinor	0.00202	0.0147	0	0.00476	0.00027	0.00437	0.00006	0.00762
DTD	0.00135	0.01292	0.00537	0.02207	0.00381	0.01169	0.00471	0.02148
EPMR*	0	0.01227	0	0.01219	0	0.01219	0	0.01235
GA1OAT181	0	0.00462	0	0.00476	0	0.00184	0.00006	0.00762
GA2OAT402	0.00076	0.0111	0.00078	0.01136	0.00417	0.0123	0.00006	0.00762
HFIA149P	0.00076	0.01104	0.00006	0.0072	0.00139	0.00716	0.00032	0.00984
HTI	0.00135	0.01289	0	0.00476	0	0.00278	0	0.00501
INCL	0.00877	0.02792	0.00537	0.02207	0.0031	0.0104	0.00152	0.01397
LPI	0.00077	0.01119	0.00139	0.01324	0.00083	0.00599	0.00152	0.01397
NKH	0.00029	0.00922	0.00077	0.01133	0	0.00279	0.00152	0.01397
ODG	0	0.00463	0.0003	0.00935	0.00012	0.00359	0	0.00501
RSPori	0	0.00463	0	0.00476	0	0.00184	0	0.00501
RSOulu	0	0.00462	0	0.00476	0	0.00184	0.00006	0.00762
Salla	0.00156	0.01483	0.00286	0.01689	0.0012	0.00692	0.00086	0.01196
VLINCL	0.00522	0.02146	0.00006	0.0072	0	0.00279	0.00032	0.00984
a1ATz-allele	0.0097	0.02945	0.01099	0.03344	0.0134	0.02586	0.00471	0.02148
CFDTT394	0	0.00462	0	0.00477	0.00052	0.00508	0.00032	0.00984
CFDF508	0.00135	0.01289	0.0003	0.00935	0.00112	0.0065	0.00006	0.00762
FVLeiden	0.00279	0.01651	0.01192	0.03354	0.01384	0.02643	0.00752	0.02683
GJB2D35G	0.00436	0.01986	0.00537	0.02207	0.00459	0.01294	0.00471	0.02148
HFEC282Y	0.03434	0.0656	0.01494	0.03836	0.02122	0.03617	0.00752	0.02683
PKUR408W	0.00006	0.00698	0	0.00476	0.00027	0.00438	0.00032	0.00984
PTG20210A	0.00076	0.01104	0.00139	0.01324	0.00111	0.00649	0.00152	0.01397
BATTEN	0.00038	0.01153	0.00209	0.01512	0.001	0.00704	0.00006	0.00762
LCHAD**	0.00076	0.01107	0.0003	0.00935	0.00063	0.00888	0.00006	0.00762
MCAD	0.00029	0.0091	0	0.00476	0.00012	0.00358	0.00006	0.00762
MPS-I	0.00006	0.00706	0.00006	0.0072	0.00083	0.00589	0.00006	0.00762

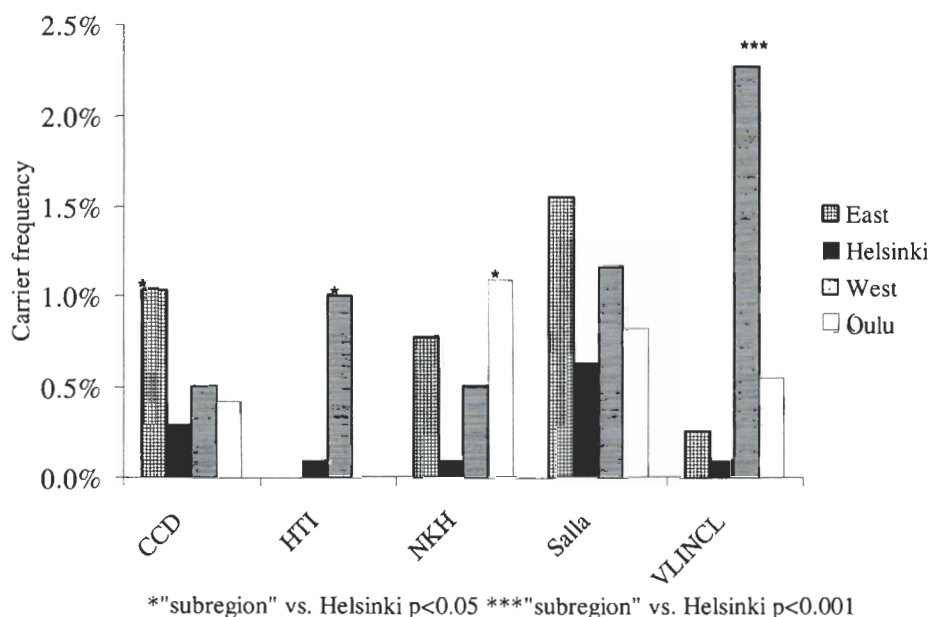
calculated carrier frequency of PAH mutations was <1:200, consistent with estimated PKU incidence of less than 1:100 000 in Finland (30). The common deletion in the CLN3 gene, causing Batten disease (31) and accounting for 90% of Finnish Batten chromosomes, was surprisingly rare. The carrier frequency for this mutation was 1:165–1:365 in all regional samples, except the eastern, late settlement study sample, which had a frequency of 1:77, being comparable to that reported from other Caucasian populations (32). The prevalence of the medium-chain acyl CoA dehydrogenase (MCAD) mutation was very low, the carrier frequency varying between 1:503 and 1:198. This would suggest that the MCAD deficiency represents still another example of a 'negative founder effect' in the Finnish population.

#### Linkage disequilibrium in common disease alleles

SNPs flanking the common mutations in the HFE (33,34) and FV (35) genes were selected (Materials and Methods; Table 4)

to monitor for the linkage disequilibrium of common disease alleles and the haplotypes of disease alleles in regional samples. After correction for multiple testing, the only significant departure from linkage equilibrium between the SNP variants of the HFE allele was observed in the late settlement North Karelia sample (Table 4). In the case of the factor V gene SNPs, no regional differences in linkage disequilibrium could be observed.

The predicted haplotypes of HFE and FV alleles suggested that the mutation occurred in the same haplotype background in all geographic regions. This provided evidence that many of these are old founder mutations. The old age of mutations is supported by the fact that for both loci these haplotypes were in agreement with the previously described haplotypes observed in other Caucasian populations (33,34,36,37). Allelic information of these common variants provided no evidence of genetic distance between regional subpopulations based on the Nei's identity estimation (38).



**Figure 3.** Frequency of carriers of 'young' Finnish disease heritage mutations in different geographical regions of Finland. The carrier frequencies in the 'subregions' were compared to those in the Helsinki region.

## DISCUSSION

Our study demonstrates the power of array-based mutation detection techniques in parallel screening for numerous disease mutations. Miniaturization and multiplexing resulted in the low cost for genotyping, the price per genotype being \$0.30USD/mutation/sample (8). Several approaches for parallel genotyping by DNA-microarrays have been presented [reviewed by Hacia and Collins (39)], but there is a notable lack of their application beyond the proof-of-principle level. ASO-hybridization-based approaches suffer from complicated array designs and imperfect allelic discrimination (40,41) rendering them inflexible and costly for custom genotyping applications. Primer extension-based approaches (7,8,42) with low complexity array designs provide a better accessibility of the microarray technology for high-throughput genotyping. However, dissemination of high-throughput primer extension-based microarray systems to clinical diagnostics and carrier-screening programs are currently hindered by the considerable set-up optimization required for multiplex PCR (8,43). Generic arrays of 'tag' (44) or 'zip code' (45) sequences can be used to capture and visualize primer extension or ligation products carrying complementary tag- or sequence-sequences at their 5' ends in a highly parallel manner. These generic arrays will obviously be more accessible to non-specialized laboratories.

Despite the relatively limited sample size and statistical power of our pilot study, the obtained carrier frequencies of 'Finnish diseases' were generally in very good agreement with the reported carrier frequencies of the individual diseases, evidencing for the accuracy of the data (12,25). Furthermore, for most of the diseases studied, the observed geographic distribution of the carrier frequencies agreed well with earlier epidemiological and genealogical data (16,46). It should be noted that our study samples reflect the effect of at least two features molding the gene pools of the analyzed regions. First,

the western sample represents the typical early settlement region with the population history of some 100 generations, whereas the eastern sample represents the late settlement communities, permanently inhabited only some 15 generations ago. Secondly, both western and eastern samples were collected from rural, isolated communities, whereas Helsinki and Oulu samples represent more mixed, urban populations. However, some distinct differences in the population history as well as in the distinction of Y chromosomal haplotypes between these urban populations have been shown to exist (15) and this genetic diversity could explain some of the variation seen in the carrier frequencies between these two urban areas. Very low carrier frequencies were seen for many 'Finnish mutations' in rural population samples. An enrichment of these disease alleles is not obvious at the whole population level but only in highly restricted communities created by a small number of founders and by a long-standing isolation.

The regional distribution varied considerably for the rare mutations that had been recently introduced into the population. These disease alleles are known to reveal linkage disequilibrium over a wide genetic interval (12). With regard to the older, common Finnish mutations, such as CNF, DD and LPI (16), no dramatic differences could be observed in carrier frequency between the regional study samples. Interestingly, the INCL<sub>Fin</sub> mutation (47) as well as the most common APECED mutation (48) suspected to represent old mutations based on the short LD interval in disease chromosomes showed enrichment in both east and west rural populations as compared to the more mixed Helsinki and Oulu study samples (Tables 2 and 3). Further, the AGU mutation, otherwise revealing an even prevalence across the regional samples showed a lower frequency of carriers in the early settlement region. These findings would be in agreement with multiple regional bottlenecks in the population history and suggest that local events mold the gene pool

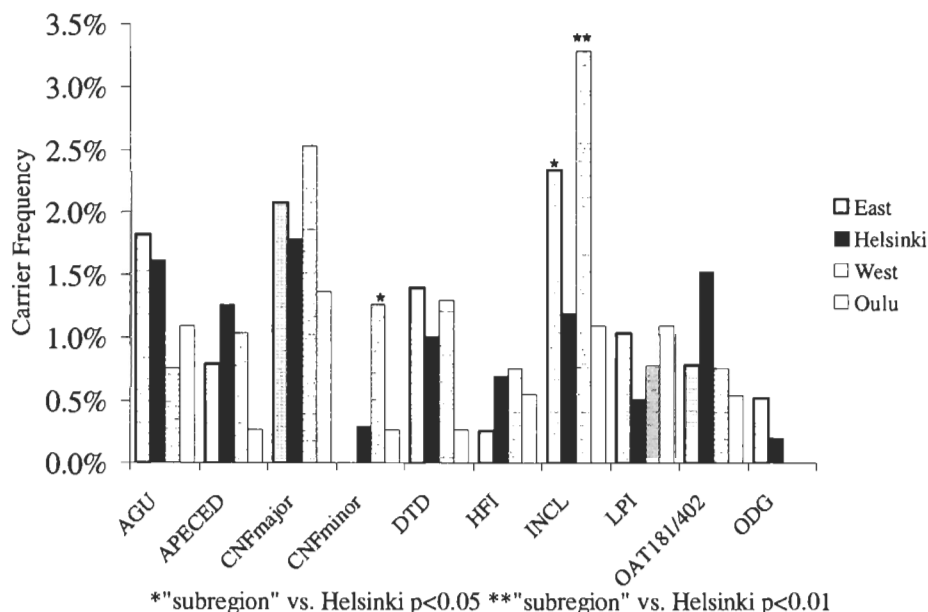


Figure 4. Frequency of carriers of 'old' Finnish disease heritage mutations in different geographical regions of Finland.

Table 4. Uncorrected P-values for the linkage disequilibrium analyses between different pairs of SNPs

Pair of SNPs	Botnia	North Karelia	Helsinki	Oulu
<b>HFE</b>				
HLA-HIVS2 C282Y	0.0283	0.0010	0.1132	0.4510
C282Y HLA-H5569	0.6480	1.0000	0.1026	0.6994
HLA-HIVS2 HLA-H5569	0.0838	0.0043	0.1131	0.4510
<b>Factor V</b>				
FVLeiden 11091A_C	1.0000	1.0000	0.3562	1.0000
FVLeiden 11063G_A	0.6053	0.7379	0.8721	0.2841
11091A_C 11063G_A	0.7432	0.3255	0.8597	0.6414
FVLeiden 11064G_A	0.6049	0.6357	0.8662	0.5152
11091A_C 11064G_A	0.7431	0.3093	0.8602	0.6447
11063G_A 11064G_A	<0.0001	<0.0001	<0.0001	<0.0001
FVLeiden 10964A_G	1.0000	0.1717	0.1922	1.0000
11091A_C 10964A_G	1.0000	0.3236	1.0000	1.0000
11063G_A 10964A_G	0.7594	0.1595	1.0000	0.3985
11064G_A 10964A_G	0.7628	0.2001	1.0000	0.4574

continuously and can have a significant effect on the frequency of disease alleles.

It should be noted that although the major mutations of old Finnish diseases showed a relatively even distribution across the country, minor alleles of these diseases revealed dramatic differences in their regional distribution supporting their more recent introduction into the population, as exemplified by the prevalence of the CNF minor mutation. This mutation was found to be highly enriched in the western coastline communities

while being very rare in other parts of Finland. This is in agreement with the recent study of 2000 Finns from the eastern area (49). Most probably, the CNF minor mutation has been introduced to this early settlement region quite recently, possibly after the major internal migration movement in the 16th century (16). This illustrates the fact that even in isolated populations with relatively little immigration several different mutations for a rare disease may at times be found simultaneously. Further, not all of these mutations are 'original' founder mutations nor are they of similar age. Finally, CNF also provides a demonstrative example of how the introduction of a second mutation into the population can result in patients who are compound heterozygotes, a feature that may markedly increase the incidence of the disease.

The data on the common Caucasian mutations revealed several interesting features. A geographic gradient has been reported in the prevalence of the deletion mutation of connexin 26 (50). In Europe the carrier frequency is highest (1:31) in some Mediterranean populations and gradually decreases towards central and northern Europe (1:79). Despite this, the Finnish frequency (1:45–1:60) represents an average European population frequency. There were no distinct differences in the regional distribution of this mutation providing evidence for its early introduction and even spreading during the inhabitation of the country.

α-1-Antitrypsin deficient individuals are most commonly homozygous for a recurrent point mutation referred to as the Z-allele (51). Nearly 30 years ago a study was conducted among Finnish students to determine the prevalence of the Z-allele by protein electrophoresis in different parts of Finland (52). The carrier frequency of the Z-allele in Helsinki was reported to be 2.7% (1:38) and a higher frequency was seen in the western early settlement region when compared to north-eastern Finland. Our data are in good agreement with these

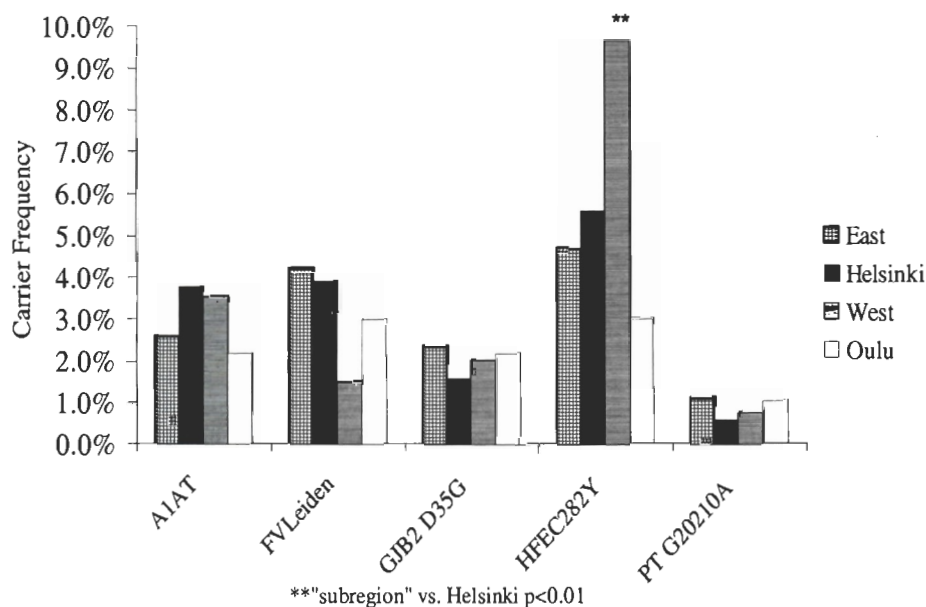


Figure 5. Frequency of carriers of Finnish mutations for common Caucasian traits.

findings exposing a similar decreasing gradient (1:26–1:46) of the Z allele from west to east.

The prevalence of another common disease mutation,  $HFE_{C282Y}$ , causing hereditary hemochromatosis had been previously studied in two Finnish population-based samples. The first study reported a carrier frequency of 9.9% (1:10) in 141 Finns (53), whereas a larger study in eastern Finland reported a carrier frequency of 6.7% (1:15) (54). Our data exposes differences in tested geographical regions, the prevalence being highest in the early settlement area of Finland. The high global prevalence of both of these mutations is reflected by the high prevalence among founders of the Finnish population. Further, the mutation has become relatively evenly distributed, even during the internal migration. However, later bottlenecks, especially those affecting the eastern, late settlement region have molded the population prevalence of the mutations. An opposite regional distribution was observed in the prevalence of two other common Caucasian mutations:  $FV_{Leiden}$  and prothrombin $_{G20210A}$ , both predisposing to venous thrombosis. These were most prevalent in the eastern samples. It remains to be seen if this opposite direction of increasing gradients in the frequency of common mutations simply reflects a random drift in regional subsamples or rather mirrors the dual theory concerning the settlement history of Finland (15,55). The combined carrier frequency of these common Caucasian mutations in Finland is slightly higher (1:5) than or similar to the pooled frequency of recessive 'Finnish' mutations (1:6). As expected, population bottlenecks that have had a distinct effect on rare alleles, have less dramatically influenced the common Caucasian alleles. Even the recent bottlenecks have resulted in only minor regional differences in the prevalence of common mutations. However, when analyzing the SNPs flanking a common mutation (HFE), a more significant linkage disequilibrium between the SNPs was detected in the eastern rural sample than in the other

populations. This reflects a younger age and a less mixed population in the late settlement region when compared to the rest of Finland. Our finding also emphasizes the importance of population selection in linkage disequilibrium-based disease gene mapping studies. The fact that even closely spaced SNPs were in linkage equilibrium for a clinically important factor V Leiden mutation, underlines the potential difficulties in performing and interpretations of possible SNP-based studies relying on linkage disequilibrium.

The negative founder effect reflecting genetic drift in small founder populations of Finland has resulted in almost complete exclusion of some relatively common Caucasian disease alleles. We found a very low prevalence of CF and PKU alleles, approximately one-third of the prevalence reported for other Caucasian populations (56,57). Also, in agreement with an earlier study the carrier frequency of the common mutation for MCAD was very low adding another mutation to the list of disease alleles showing exceptionally low prevalence in Finland (58). It appears that by chance these mutations had not been introduced into the Finnish population reflecting the small number of founders.

Our mutation screen of Finns represents the largest direct determination of the carrier frequencies of numerous disease mutations in a representative population sample. Furthermore, it is thus far the only study considering regional differences in the mutational spectrum in a population isolate, extensively studied both genealogically and genetically. Most rare mutations of Finnish disease heritage showed relatively constant population frequency indicating the old age of these mutations; however, some mutations had remarkable high local incidences, reflecting the distinct effect of recent population bottlenecks. Although being less affected than rare recessive mutations, the prevalence of common Caucasian mutations also reflected the stochastic effect of population bottlenecks, founder effects and genetic drift: some common mutations showing distinct

regional enrichment such as HFE<sub>C282Y</sub> in Botnia or FV<sub>Leiden</sub> in North Karelia whereas others became significantly under-represented in Finland (i.e. CFAF508).

## MATERIALS AND METHODS

### Samples

Four sets of population samples representing different geographical areas in Finland were analyzed (Fig. 5). A sample from the city of Helsinki, the capital of Finland and considered to represent the most genetically heterogeneous subpopulation of the country consisted of 510 samples of the large epidemiological population-based FINRISKI study (59) and 495 samples from the Finnish Twin Cohort Study (<http://kate.pc.helsinki.fi/twin/twinhome.html>) (60). The sample from the county of southern Botnia consisted of 396 random blood donors from the rural counties of Kurikka, Alajärvi and Lapua. The North Karelia (eastern Finland) sample consisted of 385 random blood donors from the rural counties of Eno, Ilomantsi, Juuka, Lieksa and Polvijärvi. The Oulu sample was composed of 365 random inhabitants of the city of Oulu in northern Finland born in the year of 1966. All samples were analyzed anonymously. In addition to the study subjects, 212 samples from verified carriers of the disease mutations studied were included to serve as blinded positive controls. Following the initial screening, the remaining positive samples (some DNA-aliquots were exhausted after the array-based assay and confirmatory genotyping) for either the FV<sub>Leiden</sub> and HFE<sub>C282Y</sub> mutations from each geographic region were further analyzed for SNPs in the FV and HFE genes. Along with these samples approximately 20 negative samples for both mutations were picked from each geographic region and similarly analyzed for the SNPs. A total of 218 samples were included for this part of the study.

### Genotyping

A system based on allele-specific primer extension in a microarray format was used to screen for carriers of the 31 disease mutations as well as the FV and HFE SNPs as described in detail previously for these panels (8). Standard solid-phase minisequencing in a microtiter plate format served as the reference method (61). For the NKH and LPI mutations PCR-RFLP digestion was used as the reference method (62,63), and for the Salla disease mutation an allele-specific PCR reaction (18) with an internal control amplicon was used. The genotypes of the carriers identified by the array-based screening were confirmed using one of the above mentioned reference methods. Samples yielding signal ratios not falling within distinct clusters or having low signal intensities were reanalyzed using a reference method.

### Statistical analysis

Statistically significant differences between carrier frequencies were assessed using Fisher's exact test. In cases where, based on published literature, carriers were expected to cluster in a particular region one-tailed tests were used, otherwise two-tailed tests were employed. Hardy-Weinberg proportion and linkage disequilibrium were assessed using the Genepop web version 3.1c program (<http://wbiomed.curtin.edu.au/genepop/>) (64). Nei's identity estimation (38) was carried out to compare

genetic distances between different study samples using the program PopDist (65). In the case of the Factor V gene SNPs (WIAF-10964 A/G, WIAF-11529 T/G, WIAF-11065 G/A, WIAF-11064 A/G, WIAF-11063 A/G, WIAF-11091 A/C, WIAF-11062 C/T) no regional differences in linkage disequilibrium values could be observed. For the HFE gene, tests of disequilibrium based on multilocus genotypes were conducted within each population to avoid spurious results due to admixture. Fisher exact tests could not reject the hypothesis of equilibrium between SNPs HH<sub>C282Y</sub> and HFE<sub>5569</sub> in any of the populations, despite the close positions of the two markers. Equilibrium could not be rejected between any marker for the Helsinki and Oulu populations. To correct for multiple comparisons across populations (4) and marker pairs (three for each population) using the Bonferroni method, one should inflate the *P*-values of each single test by a number between 4 and 12 (corresponding, respectively, to the extreme of total dependence and independence between marker pairs). This leaves only one significant departure from equilibrium in the HFE gene in North Karelian population (the pair 282Y and HLA-IVS2 has an initial *P*-value of 0.01 and the pair HLA-H5569 HLA-IVS2 has an initial *P*-value of 0.004—these *P*-values are before the Bonferroni correction).

## ACKNOWLEDGEMENTS

We owe our deepest gratitude to several researchers and clinicians for making the control samples available to us. Vlad Kustanovitch is appreciated for his advice in grammar and style. The following grants made this study possible: the Technology Development Centre of Finland, EC Biomed2 Contract no. BMH4-972013, The Hjelt Fond of the Pediatric Research Foundation, The Finnish Cultural Foundation, The Academy of Finland and The Maud Kuistila Foundation.

## REFERENCES

1. Antonarakis, S.E. and McKusick, V.A. (2000) OMIM passes the 1000-disease-gene mark. *Nat. Genet.*, **25**, 11–11.
2. Hietala, M., Aula, P., Syvänen, A.-C., Isoniemi, A., Peltonen, L. and Palotie, A. (1996) DNA-based carrier screening in primary healthcare: screening for aspartylglucosaminuria mutations in maternity health offices. *Clin. Chem.*, **42**, 1398–1404.
3. Eng, C.M., Schechter, C., Robinowitz, J., Fulop, G., Burgert, T., Levy, B., Zinberg, R. and Desnick, R.J. (1997) Prenatal genetic carrier testing using triple disease screening. *J. Am. Med. Assoc.*, **278**, 1268–1272.
4. Mitchell, J.J., Capua, A., Clow, C. and Scriver, C.R. (1996) Twenty-year outcome analysis of genetic screening programs for Tay-Sachs and  $\beta$ -thalassaemia disease carriers in high schools. *Am. J. Hum. Genet.*, **59**, 793–798.
5. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P. and Collins, F.S. (1996) Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.*, **14**, 441–447.
6. Cronin, M.T., Fucini, R.V., Kim, S.M., Masino, R.S., Wespi, R.M. and Miyada, C.G. (1996) Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum. Mutat.*, **7**, 244–255.
7. Pastinen, T., Perola, M., Niini, P., Terwilliger, J., Salomaa, V., Vartiainen, E., Peltonen, L. and Syvänen, A. (1998) Array-based multiplex analysis of candidate genes reveals two independent and additive genetic risk factors for myocardial infarction in the Finnish population. *Hum. Mol. Genet.*, **7**, 1453–1462.
8. Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. and Syvänen, A.C. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.*, **10**, 1031–1042.

9. Syrjäkoski, K., Vahteristo, P., Eerola, H., Tamminen, A., Kivinummi, K., Sarantaus, L., Holli, K., Blomqvist, C., Kallioniemi, O.P., Kainu, T. *et al.* (2000) Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. *J. Natl. Cancer. Inst.*, **92**, 1529–1531.
10. Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomäki, P., Chadwick, R.B., Kaariainen, H., Eskelinen, M., Jarvinen, H. *et al.* (1998) Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *New Engl. J. Med.*, **338**, 1481–1487.
11. Vuorio, A.F., Turtola, H., Piilähti, K.M., Repo, P., Kanninen, T. and Kontula, K. (1997) Familial hypercholesterolemia in the Finnish north Karelia. A molecular, clinical, and genealogical study. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 3127–3138.
12. Peltonen, L., Jalanko, A. and Varilo, T. (1999) Molecular genetics of the Finnish disease heritage. *Hum. Mol. Genet.*, **8**, 1913–1923.
13. Bertina, R.M., Koeleman, B.P., Koster, T., Rosendaal, F.R., Dirven, R.J., de Ronde, H., van der Velden, P.A. and Reitsma, P.H. (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*, **369**, 64–67.
14. Poort, S.R., Rosendaal, F.R., Reitsma, P.H. and Bertina, R.M. (1996) A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*, **88**, 3698–3703.
15. Kittles, R.A., Perola, M., Peltonen, L., Bergen, A.W., Aragon, R.A., Virkkunen, M., Linnoila, M., Goldman, D. and Long, J.C. (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.*, **62**, 1171–1179.
16. Varilo, T. (1999) The age of the mutations in the Finnish heritage: a genealogical and linkage disequilibrium study. Academic Dissertation., University of Helsinki, Finland.
17. Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L. (1998) CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. *Nat. Genet.*, **19**, 286–288.
18. Ranta, S., Zhang, Y., Ross, B., Lonka, L., Takkunen, E., Messer, A., Sharp, J., Wheeler, R., Kusumi, K., Mole, S. *et al.* (1999) The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nat. Genet.*, **23**, 233–236.
19. Verheijen, F.W., Verbeek, E., Aula, N., Beerens, C.E., Havelaar, A.C., Joosse, M., Peltonen, L., Aula, P., Galjaard, H., van der Spek, P.J. *et al.* (1999) A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nat. Genet.*, **23**, 462–465.
20. Hoglund, P., Haila, S., Socha, J., Tomaszewski, L., Saarialho-Kere, U., Karjalainen-Lindsberg, M.L., Airola, K., Holmberg, C., de la Chapelle, A. and Kere, J. (1996) Mutations of the down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. *Nat. Genet.*, **14**, 316–319.
21. Huopaniemi, L., Rantala, A., Forsius, H., Somer, M., de la Chapelle, A. and Alitalo, T. (1999) Three widespread founder mutations contribute to high incidence of X-linked juvenile retinoschisis in Finland. *Eur. J. Hum. Genet.*, **7**, 368–376.
22. Kestilä, M., Lenkkeri, U., Mannikko, M., Lamerdin, J., McCready, P., Putaala, H., Ruotsalainen, V., Morita, T., Nissinen, M., Herva, R. *et al.* (1998) Positionally cloned gene for a novel glomerular protein—nephrin—is mutated in congenital nephrotic syndrome. *Mol. Cell*, **1**, 575–582.
23. Ikonen, E., Baumann, M., Gron, K., Syvänen, A.C., Enomaa, N., Halila, R., Aula, P. and Peltonen, L. (1991) Aspartylglucosaminuria: cDNA encoding human aspartylglucosaminidase and the missense mutation causing the disease. *EMBO J.*, **10**, 51–58.
24. Hastbacka, J., de la Chapelle, A., Mahtani, M.M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton, B.A., Kusumi, K., Trivedi, B., Weaver, A. *et al.* (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell*, **78**, 1073–1087.
25. de la Chapelle, A. and Wright, F. (1998) Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl. Acad. Sci. USA*, **95**, 12416–12423.
26. Kessel, D.P., Dunlop, J., Stevens, H.P., Lench, N.J., Liang, J.N., Parry, G., Mueller, R.F. and Leigh, I.M. (1997) Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature*, **387**, 80–83.
27. Zelante, L., Gasparini, P., Estivill, X., Melchionda, S., D'Agruma, L., Govea, N., Mila, M., Monica, M.D., Lutfi, J., Shohat, M. *et al.* (1997) Connexin 26 mutations associated with the most common form of non-syndromic neurosensory autosomal recessive deafness (DFNB1) in Mediterraneans. *Hum. Mol. Genet.*, **6**, 1605–1609.
28. Kere, J., Estivill, X., Chillon, M., Morral, N., Nunes, V., Norio, R., Savilahti, E. and de la Chapelle, A. (1994) Cystic fibrosis in a low-incidence population: two major mutations in Finland. *Hum. Genet.*, **93**, 162–166.
29. Guldberg, P., Henriksen, K.F., Sipila, I., Guttler, F. and de la Chapelle, A. (1995) Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. *J. Med. Genet.*, **32**, 976–978.
30. Visakorpi, J.K., Palo, J. and Renkonen, O.V. (1971) The incidence of Pku in Finland. *Acta Paediatr. Scand.*, **60**, 666–668.
31. TIBD Consortium (1995) Isolation of a novel gene underlying Batten disease, CLN3. *Cell*, **82**, 949–957.
32. Järvelä, I., Mitchison, H.M., Munroe, P.B., O'Rawe, A.M., Mole, S.E. and Syvänen, A.C. (1996) Rapid diagnostic test for the major mutation underlying Batten disease. *J. Med. Genet.*, **33**, 1041–1042.
33. Beutler, E., West, C. and Gelbart, T. (1997) HLA-H and associated proteins in patients with hemochromatosis. *Mol. Med.*, **3**, 397–402.
34. Jeffrey, G.P., Chakrabarti, S., Hegele, R.A. and Adams, P.C. (1999) Polymorphism in intron 4 of HFE may cause overestimation of C282Y homozygote prevalence in haemochromatosis. *Nat. Genet.*, **22**, 325–326.
35. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, **22**, 231–238.
36. Bernardi, F., Faioni, E.M., Castoldi, E., Lunghi, B., Castaman, G., Sacchi, E. and Mannucci, P.M. (1997) A factor V genetic component differing from factor V R506Q contributes to the activated protein C resistance phenotype. *Blood*, **90**, 1552–1557.
37. Faioni, E.M., Franchi, F., Bucciarelli, P., Margaglione, M., De Stefano, V., Castaman, G., Finazzi, G. and Mannucci, P.M. (1999) Coinheritance of the HR2 haplotype in the factor V gene confers an increased risk of venous thromboembolism to carriers of factor V R506Q (factor V Leiden). *Blood*, **94**, 3062–3066.
38. Nei, M. (1972) Genetic distance between populations. *Am. Nat.*, **106**, 283–292.
39. Hacia, J.G. and Collins, F.S. (1999) Mutational analysis using oligonucleotide microarrays. *J. Med. Genet.*, **36**, 730–736.
40. Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J. *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.
41. Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N. *et al.* (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.*, **23**, 203–207.
42. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. and Syvänen, A.C. (1997) Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *PCR Methods Appl.*, **7**, 606–614.
43. Hacia, J.G., Sun, B., Hunt, N., Edgemon, K., Mosbrook, D., Robbins, C., Fodor, S.P., Tagle, D.A. and Collins, F.S. (1998) Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. *Genome Res.*, **8**, 1245–1258.
44. Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J. and Chakravarti, A. (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.*, **10**, 853–860.
45. Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., Barany, G. and Barany, F. (1999) Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.*, **292**, 251–262.
46. Norio, R., Nevanlinna, H.R. and Perheentupa, J. (1973) Hereditary diseases in Finland; rare flora in rare soul. *Ann. Clin. Res.*, **5**, 109–141.
47. Vesa, J., Hellsten, E., Verkruyse, L.A., Camp, L.A., Rapola, J., Santavuori, P., Hofmann, S.L. and Peltonen, L. (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature*, **376**, 584–587.
48. The Finnish-German APECED Consortium (1997) An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat. Genet.*, **17**, 399–403.
49. Romppanen, E.L. and Mononen, I. (2000) Detection of the Finnish-type congenital nephrotic syndrome by restriction fragment length polymorphism and dual-color oligonucleotide ligation assays. *Clin. Chem.*, **46**, 811–816.
50. Gasparini, P., Rabionet, R., Barbujani, G., Melchionda, S., Petersen, M., Brondum-Nielsen, K., Metspalu, A., Oitmaa, E., Pisano, M., Fortina, P. *et al.* (2000) High carrier frequency of the 35delG deafness mutation in

- European populations. Genetic Analysis Consortium of GJB2 35delG. *Eur. J. Hum. Genet.*, **8**, 19–23.
51. Crystal, R.G. (1989) The  $\alpha$  1-antitrypsin gene and its deficiency states. *Trends Genet.*, **5**, 411–417.
  52. Varpela, E., Saris, N.E. and Lokki, O. (1973) The prevalence of serum  $\alpha$  1-antitrypsin deficiency in Finland. *Duodecim*, **89**, 818–824.
  53. Heliö, T., Färkkilä, M., Halme, L., Karlsson, M., Palotie, M. and Kontula, K. (1998) Genetic background of hereditary hemochromatosis and DNA-diagnostics [Finnish]. *Duodecim*, **114**, 1404–1409.
  54. Tuomainen, T.P., Kontula, K., Nyysönen, K., Lakka, T.A., Helio, T. and Salonen, J.T. (1999) Increased risk of acute myocardial infarction in carriers of the hemochromatosis gene Cys282Tyr mutation: a prospective cohort study in men in eastern Finland. *Circulation*, **100**, 1274–1279.
  55. Lahermo, P., Savontaus, M.L., Sistonen, P., Beres, J., de Knijff, P., Aula, P. and Sajantila, A. (1999) Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.*, **7**, 447–458.
  56. Beaudet, A.L. (1990) Carrier screening for cystic fibrosis. *Am. J. Hum. Genet.*, **47**, 603–605.
  57. Guthrie, A. and Susi, A. (1963) A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics*, **32**, 338–343.
  58. Romppanen, E.L., Mononen, T. and Mononen, I. (1998) Molecular diagnosis of medium-chain acyl-CoA dehydrogenase deficiency by oligonucleotide ligation assay. *Clin. Chem.*, **44**, 68–71.
  59. Vartiainen, E., Puska, P., Jousilahti, P., Korhonen, H.J., Tuomilehto, J. and Nissinen, A. (1994) Twenty-year trends in coronary risk factors in north Karelia and in other areas of Finland. *Int. J. Epidemiol.*, **23**, 495–504.
  60. Kaprio, J. (1994) Lessons from twin studies in Finland. *Ann. Med.*, **26**, 135–139.
  61. Syvänen, A.C., Ikonen, E., Manninen, T., Bengtstrom, M., Soderlund, H., Aula, P. and Peltonen, L. (1992) Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: application to aspartylglucosaminuria in Finland. *Genomics*, **12**, 590–595.
  62. Kure, S., Takayanagi, M., Narisawa, K., Tada, K. and Leisti, J. (1992) Identification of a common mutation in Finnish patients with nonketotic hyperglycinemia. *J. Clin. Invest.*, **90**, 160–164.
  63. Torrents, D., Mykkanen, J., Pineda, M., Feliubadalo, L., Estevez, R., de Cid, R., Sanjurjo, P., Zorzano, A., Nunes, V., Huoponen, K. *et al.* (1999) Identification of SLC7A7, encoding y+LAT-1, as the lysinuric protein intolerance gene. *Nat. Genet.*, **21**, 293–296.
  64. Raymond, M., and Rousset, F. (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.*, **86**, 248–249.
  65. Guldbrandtsen, B., Tomiuk, J. and Loeschcke, V. (2000) POPDIST, version 1.1.1: a program to calculate population genetic distance and identity measures. *J. Hered.*, **91**, 178–179.
  66. Conover, W. (1971) *Practical Nonparametric Statistics*. John Wiley & Sons, New York.
  67. Hollander, M. and Wolfe, D. (1973) *Nonparametric Statistical Inference*. John Wiley & Sons, New York, pp. 15–22.