J Mol Med (1999) 77:754–760
© Springer-Verlag 1999
Digital Object Identifier (DOI)
10.1007/s001099900059

Bioinformatics:
Bits and Bytes

Shamil Sunyaev · Jens Hanke · Atakan Aydin · Ute Wirkner
Inga Zastrow · Jens Reich · Peer Bork

Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes

Published online: 4 November 1999

Abstract Analysis of human genetic variation can shed light on the problem of the genetic basis of complex disorders. Nonsynonymous single nucleotide polymorphisms (SNPs), which affect the amino acid sequence of proteins, are believed to be the most frequent type of variation associated with the respective disease phenotype. Complete enumeration of nonsynonymous SNPs in the candidate genes will enable further association studies on panels of affected and unaffected individuals. Experimental detection of SNPs requires implementation of expensive technologies and is still far from being routine. Alternatively, SNPs can be identified by computational analysis of a publicly available expressed sequence tag (EST) database following experimental verification. We performed *in silico* analysis of amino acid variation for 471 of proteins with a documented history of experimental variation studies and with confirmed association with human diseases. This allowed us to evaluate the level of completeness of the current knowledge of nonsynonymous SNPs in well studied, medically relevant genes and to estimate the proportion of new variants which can be added with the help of computer-aided mining in EST databases. Our results suggest that approx. 50% of frequent nonsynonymous variants are already stored in public databases. Computational methods based on the scan of an EST database can add significantly to the current knowledge, but they are greatly limited by the size of EST databases and the nonuniform coverage of genes by ESTs. Nevertheless, a considerable number of new candidate nonsynonymous SNPs in genes of medical interest were found by EST screening procedure.

Key words Single nucleotide polymorphism · Disease-associated genes · Expressed sequence tag database

Abbreviations *EST:* expressed sequence tag · *SNP:* single nucleotide polymorphism

S. Sunyaev · U. Wirkner · Peer Bork (✉)
EMBL, Meyerhofstrasse 1,
D-69012 Heidelberg, Germany
e-mail: bork@embl-heidelberg.de
Fax: +49-6221-387517

S. Sunyaev · J. Hanke · I. Zastrow
J. Reich · P. Bork
Max Delbrück Center
for Molecular Medicine,
Berlin-Buch, Germany

A. Aydin
Franz-Volhard Clinic,
Max Delbrück Center
for Molecular Medicine,
Virchow Klinikum, Berlin-Buch,
Germany

Please send articles to:
Peer Bork
Max-Dehlbrück-Center
for Molecular Medicine (MDC)
Robert-Rössle-Strasse 10
D-13122 Berlin, Germany
and:
EMBL
Meyerhofstrasse 1
D-69117 Heidelberg, Germany
E-mail: bork@embl-heidelberg.de
http://www.embl-heidelberg.de/~bork/

Introduction

The genetic diversity of human individuals is dominated by single nucleotide polymorphisms (SNPs); nonsynonymous SNPs, which affect the amino acid sequence of proteins, are the main source of phenotypic variation in humans. Large-scale identification of SNPs is one of the major goals of the human genome project in the years [1] to come, with various implications for the understanding of genetic diversity. Completion of a human high-density SNP map is under way [2, 3], and SNPs also present an entry point to study multifactorial diseases and variation in drug response [4]. Ideally, if all nonsynonymous SNPs in all human genes were known, one could search for associations of these variants with disease phenotypes in a pool of unrelated individuals and select variants which increase the risk of human complex disorders. This approach is known as “direct association” studies [5]. There is no hope, however, that direct association studies will be performed on the whole genome level in the near future. An alternative approach is to perform association studies for a much smaller set of SNPs, with the aim of identifying neutral variants which are likely to be linked with variants responsible for a disease phenotype. The latter approach is called “indirect association studies” [5]. Recent investigations [6], however, point out that linkage disequilibrium of variants is unlikely to extend to significantly long distances. This means that large-scale indirect studies would entail analysis of approx. 500,000 SNPs, and therefore even indirect studies on the whole genome do not seem realistic in the short term.

However, sets of candidate genes are known for many human disorders. This simplifies the task of analyzing all nonsynonymous SNPs in a relatively small set of genes. Even direct association studies may be realistic on such a reduced set of genes. As a necessary first step of this analysis all potential nonsynonymous SNPs in these genes must be enumerated. For many medically relevant genes variation data are already available. Two groups [7, 8] have recently analyzed two sets of disease-related genes using novel chip technology. Their analysis shows that nonsynonymous SNPs are likely to have a low frequency of a minor allele, which makes their detection more difficult. It is clear that in spite of the rapid accumulation of SNP data, currently available list of nonsynonymous SNPs is far from complete even for the most studied genes.

Independent computer-aided SNP detection for well studied genes make it possible to estimate the sufficiency of accumulated data and current SNP detection techniques for SNP-based association studies. Searching publicly available databases of expressed sequence tags (ESTs) [9] is becoming a common method for the large-scale identification of candidate SNPs [10, 11, 12]. Two teams [11, 12] have performed experiments on mining SNPs in EST databases and have reported a high accuracy of the method, confirming a significant proportion of identified variants in an independent sample of individuals.

Here we report the results of *in silico* detection of nonsynonymous SNPs for a large set of human genes having verified associations with diseases and with a documented history of variation studies. The results were compared with the information currently available in public databases (Swiss-Prot, OMIM, dbSNP). This allowed us to estimate the proportion of SNPs detectable by scanning EST database and to evaluate the current state of completeness of databases containing the information about human genetic variations. Newly discovered nonsynonymous SNPs in disease-associated genes will be submitted to HGBASE (www address: hgbase.interactiva.de) to make them available to the scientific community. The dataset was derived by identifying all disease-associated proteins stored in the Swiss-Prot database [13] that have known variations cross-linked to the OMIM (http://www.ncbi.nlm.nih.gov/Omim/) database. In addition, a number of sequences extracted from the recent literature were added.

Methods

Nonsynonymous SNPs are most easily found by aligning the respective protein sequences with six frame translations of a nonredundant set of 1.08×10^6 publicly available human ESTs stored in dbEST [8]. Of these, 80% could be classified into 627 distinct libraries (based on library identifiers), most of which are from distinct individuals. The large number of positions and individuals should yield a high rate of SNPs, although a considerable number of ambiguities, sequencing errors, and other artifacts hamper such analysis.

As a high rate of sequencing errors leading to amino acid replacements can be expected [5], all variations were subjected to a filtering procedure to account for known problems in sequencing [14, 15, 16]. From the BLAST output we extracted EST-protein similarity regions with at least 95% sequence identity (counted in amino acids) and at least five identical flanking residues. Similarity regions shorter than 30 amino acids were omitted. All positions in three amino acids proximity of sequence ambiguities were ignored. Long, nearly identical repeats such as those from collagens were discarded as no unambiguous alignment is possible. Many sequencing errors were found in regions around “double frameshifts” or due to consecutive base-calling problems. To eliminate such regions all variations requiring two nucleotide changes per codon were ignored, as were correlated mutations in a window of five residues. For candidates that passed all of these filters, the nucleotide sequence around suspected variation candidates was also analyzed to detect synonymous mismatches and ambiguities and to identify motifs known to cause sequencing problems.

Each matching EST was cross-checked with 11,827 distinct human proteins extracted from current protein sequence databases to avoid misassignments due to matches with highly similar paralogous sequences. Despite this consistency check we cannot rule out the possibility that some of the ESTs classified in our study as representing different alleles are in fact from human paralogues that have not been sequenced yet. As only 8% of all the ESTs matched distinct paralogues, we consider this proportion as minor.

For approximately 55% of EST sequences representing variation sites, fluorescent traces were available. Automated analysis of all available traces was performed using Phred software [17]. Phred is a base-calling routine known to minimize the frequency of miscalls. Phred also accompanies base calls with quality values related to miscall probabilities.

Multiple occurrence of a variant can be a good indicator of correct base calling. We assigned *P* values to suspected variation sites assuming that sequencing errors are independent, binomially distributed events, and that the probability of a sequencing error does not depend on a sequence context or on the type of the nucleotide. *P* values depend on the number of different ESTs that match a position, on the number of consistently deviating ESTs in that position, and on the type of amino acid substitution. *P* values do not take into account hidden dependencies of various clones such as regions that are difficult to sequence. Given *N* ESTs covering the position and *k* of them representing a substitution from amino acid *a* to amino acid *b*, the *P* value can be defined as:

$$P=\sum_{i=k}^{N} C_N^i p_{ab}^i (1-p_{ab})^{N-i}$$

where p_{ab} is an estimate of the probability that an *a* to *b* substitution is a result of a sequencing error.

To evaluate the accuracy of the method, 100 clones were directly resequenced, and fluorescent traces were visually checked for 100 additional clones. The direct resequencing experiment demonstrated that for our purpose SNPs which pass all filters and have Phred quality values [17] higher than 20 are very reliable candidates [18]. In the cases in which fluorescent data are not available *P* values can be used to estimate the likelihood of a mismatch resulting from sequencing error. Approximately 75% of suspected variation sites with *P* values less than 0.001 have Phred quality values higher than 20. The use of lower *P* value thresholds cannot, however, significantly improve the accuracy.

The clones with candidate cSNPs were obtained from the Resource Center/Primary Database of the German Human Genome Project or directly from the Image consortium [19]. The respective EST plasmids were sequenced and analyzed using standard equipment and procedures provided by Applied Biosystems (ABI).

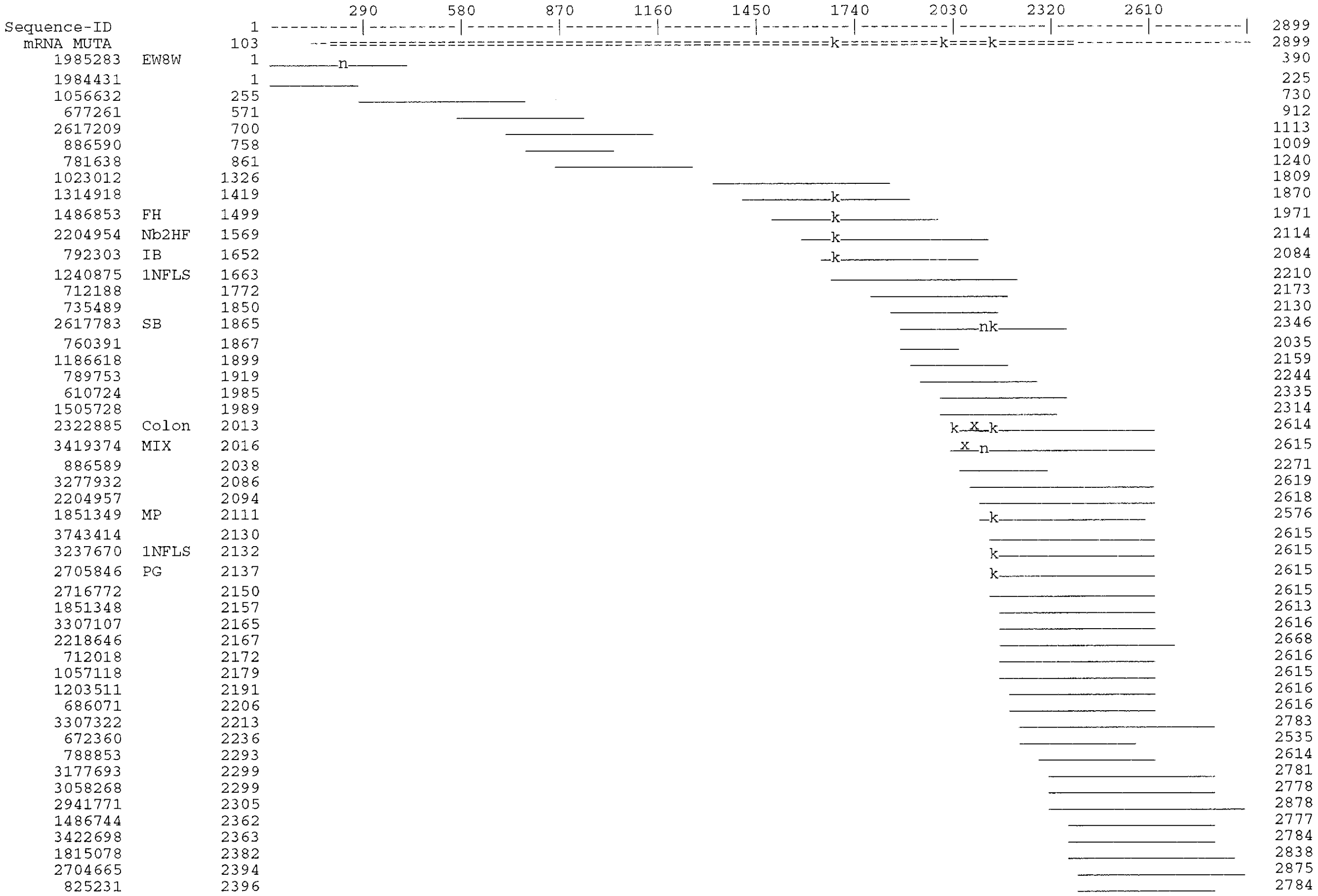

Fig. 1 EST coverage and detection of nonsynonymous SNPs in the methylmalonyl-CoA mutase precursor (*MUTA*). ESTs are aligned to the appropriate mRNA whose coding region is shown by *double dashed lines*. Accession numbers and positions of the aligned regions within the ESTs are given. The MUTA mRNA is covered by a total of 49 ESTs. The three known polymorphic sites [23, 24] are marked on the mRNA by *k*; amino acid variant in the same site detected in ESTs are also denoted by *k*; *n* confirmed novel variants (Table 2); *x* false-positive predictions. None of the rare mutations with a disease phenotype (see Swiss-Prot and OMIM) were present in the 37 ESTs. The library abbreviations of the ESTs with confirmed nonsynonymous SNPs are as follows: *EM8W* embryo 8 weeks; *FH* fetal heart; *Nb2HF8* Soares total fetus Nb2HF8 9w; *IB* Soares infant brain 1NIB; *1NFLS* Soares fetal liver and spleen; *SB* Stratagene schizo brain S11; *Colon*, bulk colon villous adenoma; *MIX* pooled human melanocyte, fetal heart, and pregnant uterus; *MP* malignant prostate; *PG* Soares pineal gland N3HPG

Table 1 Density of predicted nonsynonymous SNPs

	Proteins with EST match	Proteins with candidate SNPs
No. of proteins	424	347
No. of libraries	424	326
No. of total positions[a]	293,869	259,782
EST-covered positions[a]	147,324	139,430
Percentage EST coverage	50	54
No. of raw nucleotide mismatches	2994	2994
Percentage available fluorescent data	55	55
No. of candidates with Phred values higher than 20	311	311
No. of candidates with *P* value less than 0.001	147	147

[a] Amino acid positions

Results

In the set of 471 disease-associated genes we identified 2994 raw candidates (see Table 1). About 50% of these passed all filters. In 147 candidates the *P* value was less than 0.001. Fluorescent data were available for 55% of ESTs. Of the predicted SNPs 311 have Phred peak quality values higher than 20. Selected SNPs that were confirmed by direct resequencing are shown in Fig. 1 and Table 2.

We redetected and confirmed by resequencing ten risk mutations. An example is EST AA601655 (matching argininosuccinate synthase) containing the (confirmed) R272C mutation that has previously been found to cause citrullinemia in Japanese patients [20]. Furthermore, some of the confirmed variants suggest severe effects on the phenotype of the corresponding pro-

Table 2 Selected examples of confirmed amino acid variants in human disease genes

EST accession no.	Library	Position	Substitution	*P* value
Glucosylceramidase[a]				
AA442114	Nb2HF8_9w	241	G→R	2.00×10^{-2}02
AA442114	Nb2HF8_9w	252	F→I	1.40×10^{-2}02
R60051	1NIB	448	D→H	1.70×10^{-5}05
R40200	1NIB	483	L→P	9.30×10^{-19}19
R56138	3×1NIB	495	A→P	1.40×10^{-16}16
Plasminogen[b]				
H73620	1NFLS	31	A→P	5.10×10^{-5}
AA382677	Testis I	31	A→P	5.10×10^{-5}
AA343648	Gall bladder	31	A→P	5.10×10^{-5}
AA382677	Testis I	46	I→R	1.10×10^{-9}
AA343648	Gall bladder	46	I→R	1.10×10^{-9}
H73620	1NFLS	46	I→R	1.10×10^{-9}
H73620	1NFLS	57	E→K	1.70×10^{-8}
AA382677	Testis I	57	E→K	1.70×10–
AA343648	Gall bladder	57	E→K	1.70×10–
Methylmalonyl-CoA mutase[c]				
AA333008	Embryo 8w	34	H→Q	1.30×10^{-2}
AA022744	Fetal heart	532	H→R	1.10×10^{-8}
AA476743	Total fetus	532	H→R	1.10×10^{-8}
R35402	Infant brain	532	H→R	1.10×10^{-8}
N78174	1NFLS	532	H→R	1.10×10^{-8}
AA552631	Colon	648	G→V	6.00×10^{-2}
AA663792	Schizo brain	651	F→S	1.70×10^{-3}
AI082582	Mix	651	F→S	1.70×10^{-3}
AA552631	Colon	671	V→I	4.00×10^{-8}
AA663792	Schizo brain	671	V→I	4.00×10^{-8}
AA228688	M. prostate	671	V→I	4.00×10^{-8}
AI022429	1NFLS	671	V→I	4.00×10^{-8}
AA702733	Pinal gland	671	V→I	4.00×10^{-8}
α_1-Antitrypsin[d]				
AA291386	Ovarian tumor	125	R→H	4.90×10^{-2}
AA633935	Lung	237	V→A	2.20×10^{-19}
W94701	Fetal heart	302	E→[f]	3.70×10^{-1}
AA961570	NCI-KID6	379	A→V	5.30×10^{-1}
AA928105	NFL-TGBC-S1	400	E→D	3.90×10^{-5}
H53635	Ovary tumor	400	E→D	3.90×10^{-5}
N93368	fetal lung	400	E→D	3.90×10^{-5}
W59995	Pancreatic islet	400	E→D	3.90×10^{-5}
AA716727	Fetal heart	400	E→D	3.90×10^{-5}
Isovaleryl-CoA dehydrogenase[e]				
T69370	Liver	66	Q→R	1.30×10^{-3}
R51288	1NIB	66	Q→R	1.30×10^{-3}
AA578147	2× NCI_PR4	126	Y→H	3.40×10^{-3}
AA071259	Neuroepithelium	254	I→V	2.70×10^{-2}

Only one EST is given if several clones from one library contain the same cSNP

[a] Associated with Gaucher disease; Swiss-Prot ID: GLCM-human

b Associated with plasminogen deficiency; Swiss-Prot ID: PLMN-human

[c] Associated with methylmalonicaciduria; Swiss-Prot ID: MUTA-human

[d] Associated with pulmonary emphysema; Swiss-Prot ID: A1AT-human

[e] Associated with isovalericacidemia; Swiss-Prot ID: IVD-human

[f] Note that five of the seven cSNPs in this gene are found in a single cDNA library, prepared from a brain of a 2-month-old girl who died of muscular atrophy [22]

teins (see, e.g., confirmed mutation in α_1-antitrypsin that lead to stop codons; Table 2), although such mutations may be recessive.

In most confirmed cases both frequent and rare mutations are observed. Only few of these have already been described (e.g., isovaleric acid CoA dehydrogenease, methylmanoyl CoA mutase, α_1-antitrypsin in Table 2). Often the frequency of distinct libraries in which an amino acid variant has been predicted is correlated with the allele frequency. For example, in α_1-antitrypsin five distinct libraries contain the well known [21] and frequent E400D variant (Table 2). In a number of cases only polymorphisms that appear to

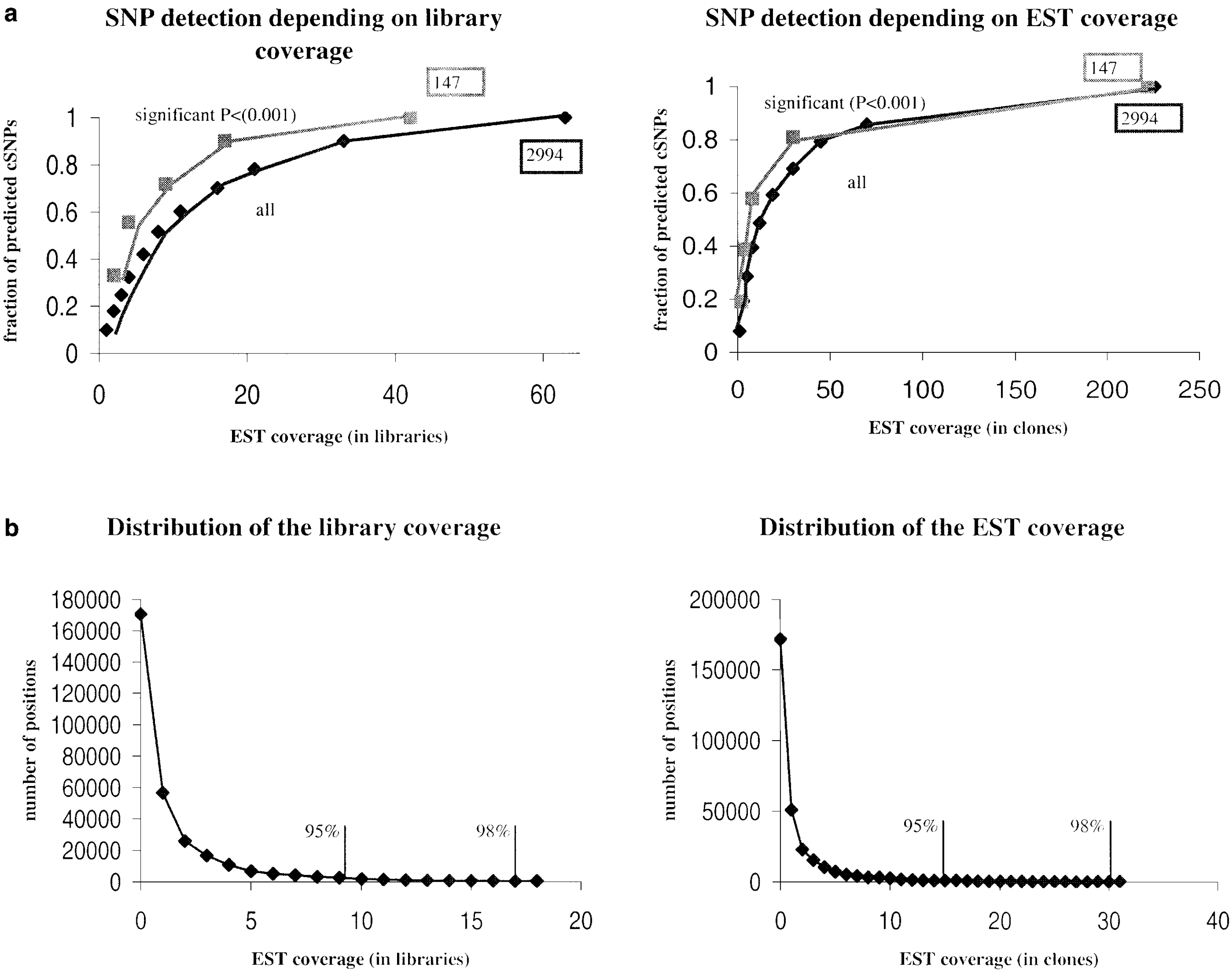

Fig. 2a,b Dependence of the SNP detection rate on EST coverage. **a** The cumulative distribution of raw mismatches (total 2994) and significant candidates (total 147). Significant candidates are defined by *P* values of 0.001. The distribution is shown with respect to library (*upper left*) and clone coverage (*upper right*). The data were pooled in approximately equally populated bins, i.e., each bin contains the same number of positions. As expected, the curve shows an asymptotic behavior, but indicates that with an increase in EST and library coverage considerably more candidate SNPs can be detected. **b** Histogram of cDNA library (*lower left*) and EST coverage (*lower right*) per position. The far tails of the distributions are not shown. The 95% and 98% points of the position coverage are given, for example, 95% of all positions are covered by nine or fewer libraries. The average of the library coverage is about two and of the EST coverage about three. For positions with high coverage the number of libraries is about half of the number of clones

have a high allele frequency were identified, as judged from multiple occurrences in ESTs from distinct libraries (e.g., plasminogen, Table 2).

The fraction of SNPs detectable by *in silico* analysis can be estimated by the redetection of known neutral polymorphisms. We have reidentified 26% (10 from 38) of neutral variants described in OMIM database as “allele” or “polymorphism.” These variants are very frequent and have no association with any disease phenotype. To include also rare variants we analyzed all substitutions included in the Swiss-Prot database but not contained in OMIM. We excluded all OMIM variants because they are likely to be associated with severe diseases and thus have negligible frequencies in normal population. Of these putatively neutral variants 6% (87 of 1553) were redetected. The result suggests that, although SNP mining in EST databases can add a significant amount of new variants to our current knowledge, it is unable to replace direct experimental studies since only about one-quarter of the variants can be seen in ESTs even for very frequent alleles. However, as the main cause of missing SNPs is low coverage (Fig. 2), the performance of EST-based SNP detection can be significantly improved with the growth of EST databases.

To estimate what proportion of our data is already included in public databases we restricted our analysis to high-quality candidates of potentially frequent amino acid variants. We selected 48 candidate SNPs which they have high Phred value and also have a significant *P* value, which implies that the variant is represented by at least two ESTs. These variations are likely to represent true SNPs with a high frequency of a minor allele. Only 21 of those were unknown, 17 were classified in the Swiss-Prot database as “variant” and 10 as “conflict” (which can represent a true variant or a sequencing error in protein sequence). OMIM and dbSNP databases contained no additional variants from this set. This suggests that a significant proportion of frequent nonsynonymous SNPs (more than 50% in our small sample) in well studied genes are already incorporated into the public databases. Although about 60% of all nonsynonymous SNPs have a minor allele frequency of less than 5% [8], given current progress in the field the enumeration of all nonsynonymous SNPs in the key candidate genes for disease association studies seems feasible, and electronically mined candidates can add significantly to our current knowledge of human variation.

To analyze the overlap of EST-based methods and novel experimental SNP detection techniques we compared our results with those of Halushka et al. [7]. Ten reliable candidates (three with high Phred values and seven with significant *P* values but without support by fluorescent data) were detected in 5 genes from the set of 75 genes analyzed by Halushka et al. [7]. Among these ten candidates two were previously included in the Swiss-Prot database and were reported by Halushka et al. [7], one additional variant is listed only in the Swiss-Prot database, and other seven candidates are likely to represent newly identified amino acid variants.

Discussion

Identification of amino acid variation in proteins with known or suspected disease association is a key point in association studies. Since we do not know a priori how many variants exist in these genes, it is unclear how many variants are still to be found. We have shown that particularly frequent SNPs which are identifiable in ESTs significantly overlap with the set of previously identified variants. We conclude that a considerable proportion of amino acid variants needed for association studies are already known in well studied genes.

Computer-aided SNP mining in EST databases has a limited capacity. The main difficulty here is low EST coverage. Only about 50% of protein positions are covered by at least one EST; the proportion of positions covered by many different EST libraries is negligible (Fig. 2). We know nothing about the ethnic diversity of EST-based samples. Numerous sequencing errors force us to accept only candidates with very reliable fluorescent peaks, which leads to the loss of many true SNPs. A similar effect may arise due to the very high sequence identity threshold used to avoid inclusion of ESTs of paralogous genes. Some false positives probably appear as a result of cloning artifacts and somatic mutations. However, in spite of all problems described above, computer-aided SNP mining is able to identify many SNPs which have not yet been included in publicly available databases. For example, we detected 311 putative amino acid variants in a set of 471 well studied disease-associated genes; most of these are likely to represent new polymorphisms potentially useful for association studies. This will provide a considerable increase in known alleles.

References

1. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L (1998) New goals for the U.S. Human Genome Project: 1998–2003. Science 282:682–689
2. Wang DG, et al (1998) Large-scale identification, mapping, genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082
3. Lai E, Riley J, Roses A (1999) A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics 54:31–38
4. Kleyn PW, Vesell ES (1998) Genetic variation as a guide to drug development. Science 281:1820–1821
5. Collins FS, Guyer MS, Chakravarti A (1997) Variation on a theme: cataloging human DNA sequence variation. Science 278:1580–1581
6. Kruglyak L, (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144
7. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247
8. Cargill M, Altschuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238
9. Boguski MS, Lowe TMJ, Tolstoshev CM (1993) dbEST – data base for expressed sequence tags. Nat Genet 4:332–333
10. Taillon-Miller P, Gu Y, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. Genome Res 8:748–784
11. Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate SNPs from public EST data. Nat Genet 21:323–325
12. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167–174
13. Bairoch A, Apweiler R (1999) The SWISS-PROT protein sequence data bank, its supplement TrEMBL in 1999. Nucleic Acids Res 27:49–54
14. Lipshutz RJ, Taverner F, Hennessy K, Hartzell G, Davis R (1994) DNA sequence confidence estimation. Genomics 19:417–424
15. Lawrence CB, Solovyev VV, (1994) Assignment of position-specific error probability to primary DNA sequence data. Nucleic Acids Res 22:1272–1280
16. Yamakawa H, Osamu O (1997) A DNA cycle sequencing reaction that minimizes compressions on automated fluorescent sequencing. Nucleic Acids Res 25:1311–1312
17. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185
18. Sunyaev S, Hanke J, Brett D, Aydin A, Zastrow I, Lathe W, Bork P, Reich J (1999) Individual variation in protein-coding sequences of the human genome. Adv Protein Chem (in press)

19. Lennon G, Auffray C, Polymeropoulos M, Soares MB, Lennon G (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33:151–152
20. Kobayashi K, Shaheen N, Terazono H, Saheki T (1994) Mutations in argininosuccinate synthetase mRNA of Japanese patients, causing classical citrullinemia. Am J Hum Genet 55:1103
21. Graham A, Hayes K, Weidinger S, Newton CR, (1990) Markham, A.F., Kalsheker, N.A. Characterisation of the alpha-1-antitrypsin M3 gene, a normal variant. Hum Genet 85:381
22. Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91:9228–9232
23. Crane AM, Martin LS, Valle D, Ledley FD (1992) Phenotype of disease in three patients with identical mutations in methylmalonyl CoA mutase. Hum Genet 89:259–264
24. Ledley FD, Rosenblatt DS (1997) Mutations in mut methylmalonic acidemia: clinical and enzymatic correlations. Hum Mutat 9:1–6