ACTA MEDICA LITUANICA. 2005. VOLUME 12 No. 2. P. 1–7

Validation and selection of an informative set of single-nucleotide-polymorphisms in the *TGFB3* gene

Auðra Morkûnienë,

Vaidutis Kuèinskas

Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

Single nucleotide polymorphisms (SNPs) have become favourable markers in the association studies of the genetic basis of complex traits. The adequacy of such studies critically depends on the validity of the selected SNPs as well as on the existence and extent of linkage disequilibrium (LD) between functional alleles related to the susceptibility to a multifactorial pathology and surrounding SNP markers. The aim of the present study was to validate a set of SNPs in the *TGFB3* locus and to select the maximally informative ones for the LD-based investigation of *TGFB3* gene allele association with cleft lip and/or palate phenotype. Fifty unaffected individuals from the general Lithuanian population were genotyped for nine SNPs in the non-coding parts of the *TGFB3* gene. Five SNPs (rs2359992, rs3917194, rs3917177, rs3917168, rs3917147) were shown to be true polymorphisms with a minor allelic frequency (MAF) >1%. Out of them, rs3917177 and rs3917168 were defined as common variation (MAF >10%) while rs3917211, rs7158338 were shown to be monomorphic. Pairwise LD was observed between rs2359992, rs3917177 and rs3917168.

Key words: single nucleotide polymorphisms (SNPs), *TGFB3*, linkage disequilibrium (LD)

INTRODUCTION

Investigation of molecular genetic basis of human pathology at present is increasingly being shifted from monogenic to complex diseases. Mapping genes and identifying alleles that contribute to the susceptibility for such pathology will be a major challenge of the post-genome era (1). The majority of complex disease-associated nucleotide sequence changes are expected to be single nucleotide polymorphisms (SNPs). The prevailing hypothesis proposes that the genetic factors underlying common complex diseases are a combination of a number of alleles that are relatively common in the population, each having a modest impact on the function of a gene. It is based on the assumption that an allele of ancient origin with the functional impact in the development of a complex pathology insufficient to be eliminated by natural selection might become relatively prevalent in human populations. Alleles having a frequency of <20% are likely to be of recent origin. Investigation of the association between relatively prevalent alleles and the clinical phenotype of complex genetic origin in large cohorts of affected individuals is one of the most efficient strategies in revealing the genetic basis of the susceptibility for such pathology. Association analysis on a genome-wide scale requires immense resources and particularly high throughput technologies, thus currently it is focused on candidate genes. To date, most candidate gene studies have directly analysed the association between the disease status and a small number of candidate SNPs in the coding or regulatory sequences of candidate genes (2, 3). Such approach enables determining whether an allele confers a considerable risk of the disease. An essential precondition is the already known or predicted functional importance of the nucleotide sequence variant under investigation. This implies a profound knowledge of the allelic diversity of the gene and the role of its product in biological processes, which is far from being the case for the majority of human genes.

An alternative to the direct analysis of functional SNPs in the candidate genes is indirect analysis using a dense set of assayed DNA markers in the locus of interest, which does not require an a priori identification of functional SNPs (3). It is based on the assumption that DNA mutations which generate SNPs

Correspondence to: Prof. Vaidutis Kuèinskas, Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariðkiø 2, LT-08661 Vilnius, Lithuania. Fax: +370 5 2365199. E-mail: vaidutis.kucinskas@santa.lt

are occasional and generally are not recurrent, leading to an association among alleles along the chromosomes, such that the presence of one nucleotide sequence variant provides information about the presence of other nonpareil variants (including unassayed ones which confer susceptibility to a complex pathology), *i.e.* characterises a definite haplotype which is of common origin in the population. Such association is called linkage disequilibrium (LD). Most LD-based methodologies pinpoint only a single genomic region or gene. Recent studies have shown that the distribution of meiotic recombination sites in human genome is relatively non-random (4), leading to the existence of a series of high LD regions ("blocks") separated by short (< 20 kb) discrete segments of very low LD ascribed to "recombination hotspots" (5). Regions of high LD exhibit a limited haplotype diversity: a small number of distinct haplotypes (termed haplotype blocks) account for most of the chromosomes in the population. Such haplotype blocks can be characterized by a small number of selected DNA markers (haplotype tagging) (6). Knowledge of local LD patterns may help to identify common haplotypes carrying the unassayed functional SNP alleles which confer susceptibility to a complex disease. The forces affecting the pattern of LD (mutation, gene conversion, recombination, natural selection, demographic structure of populations) vary along the genome and are generated by highly stochastic processes, thus the extent of LD at any particular region of the genome is not predictable and must be assessed empirically.

SNPs are increasingly becoming the markers of choice in the analysis of allelic association as well as in genetic analysis in general. Such polymorphisms are highly abundant, occurring approximately in every 1000th bp of human genome. The haplotypic patterns, ease of genotyping, allele frequency drift and low mutation rate also make SNPs excellent markers for the investigation of complex genetic traits and a tool for the understanding of genome evolution. SNPs are far more prevalent than microsatellite DNA markers and, therefore, may provide a high density of markers near a locus of interest. Current approaches for carrying out genetic association studies focus in particular on the strategies that seek to use patterns of statistical association among SNPs to identify a subset of SNPs that efficiently represent the other SNPs in a genomic region (tagging SNPs) on the basis of LD patterns. The ultimate value of SNPs for association mapping studies depends in part on the distribution of SNP allele frequencies and intermarker LD.

Currently, the public SNP databases comprise > 5×10^6 raw SNPs spread throughout the human genome. The major public repositories are the National Center for Biotechnology Information (NCBI) database of genetic variation (dbSNP) (7). The dbSNP database contains ~4×10^6 non-redundant human SNPs derived from ~?300 sources. Estimates of the number of SNPs in the genome range up to ~?15×10^6, for polymorphisms with a rare allele frequency > 1%, but a considerable fraction of the public SNPs are just experimental artefacts. However, only a small fraction of these SNPs is well characterized and validated (8), thus validation of selected SNPs before using them in linkage or association analysis is still an important problem.

One of the relatively prevalent complex traits in humans is syndromic and non-syndromic oral-facial clefts (OFC). Despite extensive studies their etiology remains largely unknown, with only a few cases of rare OFC syndromes being associated with identified gene mutations or being shown as secondary to a recognized teratogen exposure (9). In recent years, a number of independent groups have investigated approximately 30 candidate genes for the involvement in nonsyndromic OFC, but these studies produced conflicting results. *MSX1* (4p16.1), *TGFA* (2p13), *TGFB1* (19q13.1), *TGFB2* (1q41), *TGFB3* (14q24), *RARA* (17q12), *MTHFR* (1p36.3) genes are among the strongest candidates for the involvement in nonsyndromic OFC, thus the presence of functional SNPs within them is likely.

The present study focuses on the transforming growth factor beta three (*TGFB3*) gene, which has been suggested to be involved in the development of cleft lip and/or palate phenotype both by animal models (10, 11) and positive results of association studies (12–14). The aim of the study was to validate a set of SNPs in the locus and to select efficient ones by LD-based methodology for the subsequent use in the investigation of association of the *TGFB3* gene alleles with the cleft lip and/or palate phenotype. In the present paper, we estimate the allele frequency and intermarker LD distributions for SNPs in the *TGFB3* gene in a sample of 50 unaffected individuals from the general Lithuanian population.

MATERIALS AND METHODS

SNPs in the *TGFB3* gene (14q24) to be analysed were drawn from the published data on cleft lip and/or palate studies as well as from public SNP databases by selecting SNPs with a known allelic frequency determined by methods other than *in silico*.

DNA samples from 50 unrelated unaffected Lithuanian individuals collected from all regions of Lithuania in 1996 were used in the present study. The total DNA was isolated from peripheral blood by phenol–chloroform extraction.

The genotyping procedure for all selected SNPs was similar, involving polymerase chain reaction (PCR) amplification of a DNA fragment covering a relevant SNP with a subsequent restriction fragment length polymorphism (RFLP) assay. Primer pairs have

Table 1. **Information about selected SNPs in the *TGFB3* gene**

SNP			Primers for PCR amplification* (5'→3')	PCR products length (bp)	Restriction enzyme	Restriction fragment length (bp)	
Reference in dbSNP	Chromosome position	1/2**				Genotype** [1][1]	Genotype** [2][2]
rs4903361	74402015	A/G	F – CTCCCAAAGTGCTGGGATTA				
			R – CCTTCTCAAATGGCTTTCCA	161	*Bbv*I	29, 132	161
rs2359992	74406964	C/G	F – CCAGGAAACAAATGTGAAGAGA				
			R – TTACAGGTGTGAGCCAATGC	170	*Bse*GI	170	50, 120
rs3917211	74416764	A/G	F – AACATTGCCTTGAGCCATTT				
			R – GAGAGTCCACAACCCCTGAA	156	*Hin*4I	156	27, 42, 87
rs3917194	74421574	A/T	F – CTCAGTTGTCTGGGCACTCA				
			R – AGCCTCTGGAAGACAACACG	235	*Xap*I	39, 196	235
rs3917177	74426671	C/G	F – CCCCAGCTCTGGTTATAGCA				
			R – CTGCCACACCTGGCTAATTT	246	*Pae*I	105, 141	246
rs3917168	74431605	A/T	F – TGGCAAAACCCCATCTCTAC				
			R – TTTGATCTCCCAGGCAATTC	234	*Xap*I	234	32, 202
rs3917147	74436690	C/T	F – GGTGGGGAAGGAGATAGAGC				
			R – CTTTATGGCACCCAGATGCT	201	*Sat*I	72, 129	201
rs7158338	74440083	G/T	F – AAAAAGTGCTGAGCCAGGTG				
			R – CCGGGCTAATTTTTCCATTT	154	*Mnl*I	13, 47, 94	154
rs2075765	74445450	C/T	F – TTAAGAGTTAGTGATTCCTTAATGACC				
			R – AAATGCCAAGGAAGGGAAT	201	*Mbo*II	72, 129	201

** F, forward primer; R, reverse primer.*
*** 1, allele 1; 2, allele 2.*

been designed for each of the selected SNP markers by the *Primer3* software.

PCRs were performed separately for each SNP in 25 μl of reaction mixture (100 ng DNA, 0.2 mM each dNTP, 2.5 mM $MgCl_2$, 10 pmol each primer (Table 1) and Taq DNA polymerase). Standard cycling conditions were applied for all PCRs with a variation in annealing temperature dependent on the PCR primers. Detailed PCR conditions are available from authors. 5 μl aliquots of each PCR product were run on agarose gel to verify successful amplification.

For RFLP analysis, each PCR product was digested with a corresponding restriction endonuclease (RE) (Table 1). Reaction mixture: 4–7 μl of each PCR product, 10 U RE, 1.5 μl 10 × RE buffer, distilled water to a final volume of 15 μl. The mixture was incubated overnight at RE optimal temperature. RE-digested PCR products were separated electrophoretically on agarose gel. DNA samples of known SNP genotypes and a sample of an undigested PCR product served as controls. Sites for a corresponding RE were either present or absent in the alleles of a biallelic SNP, thus different SNP genotypes yielded different patterns of RE-digested DNA fragments (Table 1).

The test of pairwise LD and the permutation test were performed using the EM algorithm. Calculations were carried out with the aid of the ARLEQUIN Ver 2.000 and SeattleSNPs vg2 program packages.

RESULTS AND DISCUSSION

At present, knowledge of common genetic variation is incomplete for the majority of genes, including *TGFB3*. Therefore, indirect association analysis was considered to be more appropriate for the investigation of the involvement of *TGFB3* gene alleles in the susceptibility to oral–facial clefting. The strategy of the study was to test a set of densely spaced SNPs in the *TGFB3* locus and to evaluate LD between the genotyped SNPs. To design the study, it was necessary to define true SNPs in a region and to select the most informative ones according to the patterns of LD between these polymorphic sequence variants. A two-stage LD-based methodology was applied. In this paper, we present the results of the stage when a low-density SNP map was used to identify the potential LD in the *TGFB3* gene (for the second stage the high-density SNP map should be used).

The first step of the current study was to select a set of SNPs in the nonconding parts (introns, 5' and 3' untranslated regions) of the *TGFB3* gene. Out of all publicly available (*i.e.* registered in the dbSNP) SNPs a set of nine candidate SNPs was selected (Table 1) with an average density of 1 SNP every 5 kb expecting their minor allelic frequency (MAF) to be >1%. The 5 kb length between SNPs was chosen for the initial (first stage) investigation. It should be noted that all DNA markers used for association analysis should meet the main criteria: they must be validated; they must be unambiguously identifiable using the experimental approach applied in the study; they must be heterozygous in the population to be investigated.

As the SNPs stored in dbSNP are mostly "candidate" SNPs found by computer data-mining procedures and have not been well characterized yet or even are validated by allele frequencies in some populations, the SNPs in any genetic analysis must be

Table 2. **Results of the genotyping of 50 unrelated individuals from general Lithuanian population according to selected SNPs in the *TGFB3* gene**

SNP		Identified genotype (number of cases)			Allele frequency		
Reference SNP (dbSNP)	1 / 2*	[1][1]	[1][2]	[2][2]	1	2	Heterozygosity
rs2359992	C/G	0	10	40	0.1	0.9	0.2
rs3917211	A/G	50	0	0	1	0	–
rs3917194	A/T	0	1	49	0.1	0.9	0.02
rs3917177	C/G	0	14	36	0.15	0.85	0.28
rs3917168	A/T	0	13	37	0.15	0.85	0.26
rs3917147	C/T	0	3	47	0.025	0.975	0.06
rs7158338	G/T	50	0	0	1	0	–

** 1, allele 1; 2, allele 2.*

Table 3. **Comparison of allele frequency and heterozygosity of verified SNPs in dbSNP and our data**

SNP		SNP data in dbSNP			Data of the present study		
Reference (dbSNP)	1/2*	Frequency of allele 1	Frequency of allele 2	Heterozygosity	Frequency of allele 1	Frequency of allele 2	Heterozygosity
rs2359992	C/G	0.246	0.754	0.371	0.1	0.9	0.2
rs3917211	A/G	0.761	0.239	0.324	1	0	–
rs3917194	A/T	0.844	0.156	0.263	0.1	0.9	0.02
rs3917177	C/G	0.747	0.253	0.378	0.15	0.85	0.28
rs3917168	A/T	0.618	0.382	0.472	0.15	0.85	0.26
rs3917147	C/T	0.131	0.869	0.228	0.025	0.975	0.06
rs7158338	G/T	?	?	?	1	0	–

** 1, allele 1; 2, allele 2,*
?, not validated SNP.

verified before use in the population on the point of testing. Thus, the second step of the study was validation of the selected SNPs using an RFLP assay of PCR-amplified DNA fragments covering each SNP. The results of genotyping 50 unrelated individuals from the general Lithuanian population revealed that all selected SNPs were presented at a low allele frequency (Table 2). Five SNPs (rs2359992, rs3917194, rs3917177, rs3917168, rs3917147) were shown to be true polymorphisms with a minor allelic frequency (MAF) >1%. Out of them, two (rs3917177 and rs3917168) were defined as a common variation with MAF >10%. Two candidate SNPs (rs3917211, rs7158338) were shown to be monomorphic. For each validated SNP allele and genotype frequencies, observed heterozygosity are displayed in Table 2. We failed to genotype SNPs rs4903361 and rs2075765 using the RFLP approach applied in this study due to the unspecific PCR product (rs4903361) and the star activity of *Mbo*II (rs2075765). Not only ungenotyped SNPs were excluded from the further study, but also genotyped monomorhic SNPs were excluded from LD studies between loci.

Allele frequencies estimated in the present study appear to be inconsistent with those from the dbSNP (Table 3). Such differences might be a consequence of different populations tested to validate the SNP. Another possible reason is a small number of chromosomes tested in previous studies in other populations, which was insufficient for statistically significant results. Besides that, polymorphisms in dbSNP come from a large number of sources and cohorts and have been discovered using a variety of techniques leading to numerous inconsistencies.

The third step of the study was selecting maximally informative SNPs out of the set of validated ones. In LD-based indirect association analysis, a limited number of informative SNPs may represent the whole haploblock covering a large set of marker loci. There are two possibilities to select the most informative set of SNPs: 1) using a haplotype-based algorithm (htSNPs) when haplotypes can be deduced from family data; 2) using an LD-based algorithm (15).

A wide variety of statistics have been proposed to measure the amount of LD, and these have different strengths depending on the context (16). Most of the measures of LD that are in wide use quantify the degree of association between pairs of markers. In part, they differ according to the way in which they

Table 4. **Results of pairwise linkage disequilibrium test among five SNPs**

Reference SNP (dbSNP)	rs2359992	rs3917194	rs3917177	rs3917168	rs3917147
rs2359992	♦	– 0.190722	+ 0	+ 0	– 0.109332
rs3917194	– 0.190722	♦	– 0.282544	– 0.260582	– 0.0595176
rs3917177	+ 0	– 0.282544	♦	+ 0	– 0.414717
rs3917168	+ 0	– 0.260582	+ 0	♦	– 0.175189
rs3917147	– 0.109332	– 0.0595176	– 0.414717	– 0.175189	♦

0 (or +), complete LD (white cells).

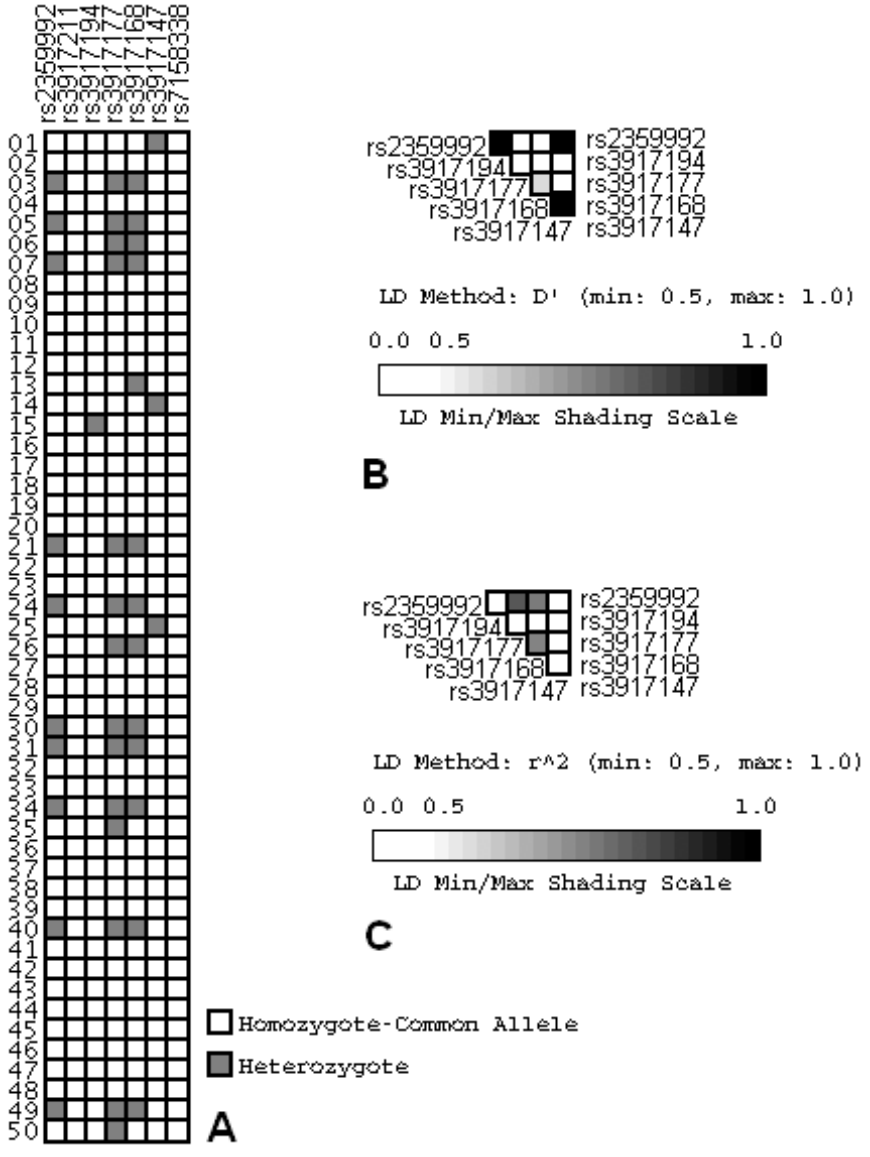

Figure. Visualised common variation and LD (using SeattleSNPs vg2) among 5 SNPs in *TGFB3* gene. Patterns of genotype at each SNP are shown as a visual genotype plot, in which each column represents a site and each row represents a sample (50 individuals, A). The LD statistics D' (B) and r^2 (C) describe the similarity of pattern among pairs of polymorphic sites (triangles)

depend on the allele frequencies. The two most popular measures of LD between pairs of biallelic markers are D' and r^2 (16). Besides, estimating the amount of disequilibrium between the pairs of markers, it is also natural to test the null hypothesis of independence between marker pairs (i.e. linkage equilibrium).

Different measures can give vastly different estimates of LD (16). D' is best if plotting the association between disease and markers for the purpose of "simple fine mapping". D' is directly related to recombination fraction, and its generalization to more than two loci is the only measure of LD not sensitive to allele frequencies. D' = 1 if 2 or 3 haplotypes are present and r^2 = 1 at 2 haplotypes present (for pairwise LD). Intermediate values of D' hard to interpret r^2 are best for association mapping, because there is a simple linear relationship with sample size. r^2 depends on marker allele frequencies and can be difficult to interpret when comparing multiple markers in a region.

In the present study, genotyping of unrelated individuals does not allow to determine the gametic phase (*i.e.* haplotypes) of the double heterozygotes, when considering two loci at a time. The problem of unavailable haplotypes in association studies can be overcome by association analyses that make use of the LD-selected minimal site set allowing

detecting either haplotype-specific or clade-specific effects within each nonrecombinant region, without prior inference of haplotypes (3). It is based on the fact that some SNPs will be specifically associated with a single haplotype, whereas other SNPs will be associated with clades of related haplotypes. LD-selected tagged SNPs describe both haplotype-specific and clade-specific patterns of variation, because the LD-selection algorithm reduces the set of all sites to bins of sites with similar patterns of genotypes (15).

In practice, for genotypic data where the haplotypic phase is unknown, LD between a pair of loci is tested for genotypic data using a probability-ratio test, whose empirical distribution is obtained by a permutation procedure. The probability of the data assuming linkage equilibrium is computed from the fact that, under this hypothesis, the haplotype frequencies are obtained as the product of the allele frequencies. The probability of the data not assuming linkage equilibrium is obtained by applying the EM algorithm to estimate the haplotype frequencies. This test of LD assumes the Hardy–Weinberg proportions of the genotypes (rejection of the test could be also due to a departure from the Hardy–Weinberg equilibrium). Minor allele frequency can therefore be used as an approximate measure of relative utility of each marker for genetic application in humans, as SNPs with high MAF are likely to be more useful for LD studies than those with low MAF.

We calculated the LD among five SNPs (MAF > 1%) in *TGFB3* gene. Both maximum probability estimates of LD, D' and r^2, were checked. Significant pairwise LD (significance level = 0.05) was calculated and the permutation test was performed using the EM algorithm (ARLEQUIN Ver. 2.000 (Table 4) and SeattleSNPs vg2 program packages (Fig. 1)).

The strongest pairwise LD was estimated only among the SNPs rs2359992, rs3917177, rs3917168 using both programs (Table 3, Figure). This was not unexpected, because the measure of LD is sensitive to allele frequency.

To optimise the subset of SNPs for assay using the observed patterns of LD, a lot of SNPs are necessary. A rational selection of a subset of sites that provides maximum information about common variation in the region is then possible on the basis of the observed patterns of LD among common SNPs (>10% MAF) (15). The number of sites required for genotyping in this study design depends on the strength and extent of LD. For regions with a strong linkage disequilibrium or a few haplotypes only, a number sites are required to represent or "tag" the region. However, if the genomic region contains many haplotypes indicating low levels of linkage disequilibrium, many more sites will be required for an association study of sufficient power.

Nevertheless, the knowledge of LD patterns revealed through SNPs within a candidate gene is useful. Under certain circumstances, LD may be a useful way to reduce the complexity of candidate-gene association analysis and selecting valuable SNPs. Otherwise, this study structure is acceptable but not ideal. Several limitations of this study should be acknowledged. This approach is not statistically strong and has no clear endpoint: true associations may be missed because of the incomplete information provided by individual SNPs; negative results do not rule out association involving other nearby SNPs; and positive results do not indicate the discovery of the causal SNP but simply a marker in LD with true causal SNP located some distance away.

In our study, we used an efficient algorithm to identify an optimal set of tagged SNPs that describes common SNPs and LD patterns. In summary, we verified true SNPs in Lithuanian population samples. Our results showed that only five SNPs of all selected nine ones are true biallelic polymorphisms. Out of them, two SNPs are common sequence variations in the candidate-gene *TGFB3*. Two SNPs were observed to be monomorphic. Two SNPs failed the genotyping and were excluded from the study.

For the first stage we selected a limited number of SNPs in the *TGFB3* gene, which brought some difficulty to define the significant LD characteristics as only two common SNPs were observed. Actually, we identified the potential LD in the *TGFB3* gene but only among few SNPs. We hope that more densely selected SNPs will reflect all the diversity in the gene and will be reliable for association studies. Therefore, we should use larger studies with a new higher-density collection of SNPs for objective results to identify all the common SNPs and to define significant LD patterns. That is to say, on the second stage we should check almost all known SNPs in the *TGFB3* gene or resequence an appropriate number of samples.

Nevertheless, we conclude that the regenotyping of a number of samples must be done before using SNPs in association analysis in order to define all true SNPs in a candidate gene, as well as the patterns of LD among these SNPs.

ACKNOWLEDGEMENT

We thank Dr. D. Steponavièiûtë for critical reading of the manuscript.

Received 9 November 2004
Accepted 15 March 2005

References

1. Beaty TH, Wang H, Hetmanski JB, Fan YT, Zeiger JS, Liang et al. A case-control study of nonsyndromic oral clefts in Maryland. Ann Epidemiol 2001; 11: 434–42.
2. Botstein D, Rich N. Discovering genotypes underlying human phenotypes: past successes for mendalian dise-

ase, future approaches for complex disease. Nat Genet Suppl 2003; 33: 228–37.
3. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004; 74: 106–20.
4. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloguing human DNA sequence variation. Science 1997; 278: 1580–1.
5. Cordon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003; 19: 135–40.
6. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B et al. The structure of haplotype blocks in the human genome. Science 2002; 296: 2225–9.
7. Jiang R, Duan J, Windemuth A, Stephens JC, Judson R, Xu C. Genome-wide evaluation of public SNP databases. Pharmacogenomics 2003; 4: 779–89.
8. Kaartinen V, Voncken JW, Shuler C, Warburton D, Bu D, Heisterkamp N et al. Abnormal lung development and cleft palate in mice lacking TGF-beta3 indicates defects of epithelial-mesenchymal interaction. Nat Genet 1995; 11: 415–21.
9. Lidral AC, Romitti PA, Basart AM, Doetschan T, Leysens NJ, Daack-Hirsch S et al. Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. Am J Hum Genet 1998; 63: 557–68.
10. Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF et al. Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. Am J Med Genet 1997; 73: 337–44.
11. de Massy B. Distribution of meiotic recombination sites. Trends Genet 2003; 19: 514–22.
12. Pritchard JK, Przeworski M. Linkage Disequilibrium in Humans: Models and Data. Am J Hum Genet 2001; 69: 1–14.
13. Proetzel G, Pawlovski SA, Wiles MV, Yin M, Boivin GP, Howles PN, et al. Transforming growth factor-beta3 is required for secondary palate fusion. Nat Genet 1995; 11: 409–14.
14. Risch N. Searching for genetic determinants in the new millennium. Nature 2000; 405: 847–56.
15. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 2001; 29: 308–11.
16. Wyszynski DF, Beaty TH, Maestri NE. Genetics of nonsyndromic oral clefts revised. Cleft Palate Craniofac J 1996; 33: 406–17.

Auðra Morkûnienë, Vaidutis Kuèinskas

VERIFIKAVIMAS BEI INFORMATYVIAUSIO VIENO NUKLEOTIDO POLIMORFIMØ RINKINIO PARINKIMAS *TGFB3* GENE

Santrauka

Vieno nukleotido polimorfizmai (VNP) yra vieni plaèiausiai naudojamø DNR þymenø tiriant daugiaveiksnës patologijos genetines prieþastis tiesioginës ir netiesioginës aleliø asociacijos analizës metodu. Tokiø tyrimø rezultatai itin priklauso nuo pasirinktøjø VNP informatyvumo ir nuo to, ar yra ir kokio masto nepusiausviroji sankiba tarp polinká daugiaveiksnei patologijai lemianèio dar nenustatyto funkcinio alelio ir aplinkiniø VNP þymenø.

Ðiame straipsnyje apraðomo tyrimo tikslas buvo verifikuoti pasirinktø VNP rinkiná *TGFB3* geno srityje bei atrinkti informatyviausius þymenis, kurie vëliau bûtø panaudoti tiriant nepusiausviràja sankibà tarp *TGFB3* geno aleliø ir lûpos ir/ar gomurio nesuaugimo fenotipo. Pagal nekoduojanèiose *TGFB3* geno dalyse pasirinktus devynis VNP buvo genotipuota 50 negimimingø asmenø imtis ið bendros Lietuvos populiacijos. Gauti rezultatai parodë, jog penki VNP (rs2359992, rs3917194, rs3917177, rs3917168, rs3917147) yra polimorfiðki: nustatytas retojo alelio santykinis daþnis buvo >1%. Dviejø ið VNP, rs3917177 ir rs3917168, nustatytas retojo alelio santykinis daþnis buvo >10%. Du þymenys (rs3917211 ir rs7158338) buvo monomorfiniai, taigi jie neatitinka VNP keliamø kriterijø. Tarp trijø bimorfiniø VNP (rs2359992, rs3917177 ir rs3917168) nustatyta nepusiausviroji sankiba.

Raktaþodþiai: vieno nukleotido polimorfizmai (VNP), *TGFB3*, nepusiausviroji sankiba