

PERSPECTIVE

Pitfall of identifying a disease locus by using low-resolution SNP arrays

Hulya Gundesli¹, Sebahattin Cirak², Pervin Dincer^{1*}

¹Department of Medical Biology, Hacettepe University Faculty of Medicine, Sıhhiye 06100, Ankara, Turkey, ²UCL Institute of Child Health, The Dubowitz Neuromuscular Centre, London, UK

*Correspondence to: Pervin Dincer, E-mail: pdincer@hacettepe.edu.tr, Tel: +90 312 3052541, Fax: +90 312 3096060

Received 05 October 2011, Revised 03 November 2011, Accepted 05 November 2011, Published online 22 November 2011

J Mol Genet Med, 2010, Vol 5, 264-265

© Copyright The Authors: This is an open access article, published under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>). This license permits non-commercial use, distribution and reproduction of the article, provided the original work is appropriately acknowledged with correct citation details.

Until recently, a whole-genome scan using a set of 300-400 polymorphic DNA microsatellite markers was considered the most favoured strategy for gene mapping in Mendelian disorders. In many situations, however, it has not been possible to unravel significant disease loci by using DNA markers spaced across the genome at 10 cM intervals. Nowadays, to facilitate gene mapping or genome-wide association studies, several high-density SNP genotyping arrays with different SNP marker densities have been developed. These technologies offer highly-automated and rapid methods of genotyping as compared with genotyping microsatellites by PCR. For disease locus identification these data could be analysed by various methods, such as parametric multipoint linkage analysis and, in particular, homozygosity mapping for autosomal recessive inherited disorders in consanguineous families (Hugun et al, 2010). Homozygosity mapping is a powerful tool to detect disease loci, particularly in consanguineous pedigrees. Here, we have discussed a pitfall of this approach, namely the incapacity to detect the disease locus by the use of an Affymetrix 50K *Xba* I SNP array (50K array) in a highly-informative pedigree.

Our recently reported study of a Turkish consanguineous family with three affected individuals led to the discovery of a novel locus responsible for autosomal recessive limb-girdle muscular dystrophy [LGMD2Q, OMIM 613723] (Gundesli et al, 2010). After genotyping using a 50K SNP array, quality control of SNPs was performed by checking the relationships within the pedigree (*GRR* software) (Abecasis et al, 2001), the removal of Mendelian inconsistencies using Mendelian test (*PedCheck*) (O'Connell and Weeks, 1998). Unlikely genotypes on the basis of double recombination, which were detected by a Non-Mendelian test as an option of *EasyLINKAGE* plus

v5.052 (*EasyLINKAGE* program) were also removed (Hoffmann and Lindner, 2005). Subsequently, multipoint linkage analysis was performed using Merlin v1.0.1 through the *EasyLINKAGE* program assuming autosomal recessive inheritance with 100% penetrance (Hoffmann and Lindner, 2005). Due to the large number of SNP markers we used variable sliding windows analysis using panels of 100-150 SNPs. However, no reliable significant linkage interval could be detected. Some marker subsets showed possible linkage regions (chr3: 3p24-LOD score: 3.07, 8qtel-LOD score: 2.1 and 15q11-q14-LOD score: 2.2) with the given LOD scores (Gundesli et al, 2010). To evaluate these chromosomal regions in detail, homozygosity mapping was performed using *VIGENOS* software (Hemasoft, Ankara) (Gundesli et al, 2010), which facilitated the visualization of shared homozygous as well as heterozygous regions of the genome. Interestingly, when we compared all patient haplotypes, no regions of shared homozygosity common between all three affected individuals could be detected. Thus, the 50K array did not allow us to define neither a significant linkage interval, nor a shared homozygous region for all affected individuals within the family.

We hypothesized several reasons why no unique disease locus was detected in this initial study. Firstly, under the assumption that all affected individuals should be homozygous for a single mutant allele in this multiple consanguineous family, we have hypothesized that a shared homozygous region might have been very small and could have been missed due to insufficiently informative SNPs on the array to detect this small locus. According to our calculations about the genome coverage of 50K and 250K (250K *Nsp*I SNP array, Affymetrix) SNP genotyping arrays, we found that the 50K array

consists of 58,960 SNPs with an average physical distance of 5.7 kb (<http://www.affymetrix.com/>, NetAffyx Annotation Files). Moreover, the GeneChip Mapping 250K array genotypes 262,264 SNPs in a single experiment and the average physical distance between them is 1.2 kb (<http://www.affymetrix.com/>, NetAffyx Annotation Files). The distance between markers in the 250K array is approximately 5-fold less than that in the 50K. Furthermore, the second reason might be allelic or locus heterogeneity despite multiple loops of consanguinity because LGMD2 is a genetically heterogeneous disorder (Fendri et al, 2006). There are many examples for the potential pitfalls of homozygosity mapping as a result of locus and allelic heterogeneity even in consanguineous families (Lezirovitz et al, 2008; Benayoun et al, 2009).

To test the first hypothesis, we decided to genotype three patients and their parents with the high-density 250K SNP array in order to check if we might have missed any smaller linkage region with the 50K array. After homozygosity analysis (*VIGENOS* software), we observed a single homozygous segment of approximately 3.7 Mb on the telomer 8q24 (Gundesli et al, 2010). All affected individuals shared a common haplotype across this critical interval, and no additional overlapping homozygous haplotypes were observed throughout the genome. Remarkably, only one SNP marker (rs2717594) was included in the 50K array within the relevant chromosomal region whereas the 250K array contained 110 consecutive homozygous SNP markers in the identified locus. In addition, the average physical distance between SNPs covering chromosome 8 in the 250K is 4-fold shorter than in the 50K array (250K: 1 kb and 50K: 4.2 kb) (<http://www.affymetrix.com/>, NetAffyx Annotation Files). Consequently, as the linkage region was only covered by a single SNP in the low density SNP array, we were not able to discover the disease locus initially. Subsequently, DNA sequencing analysis of the *PLEC* as a positional candidate gene revealed a homozygous deletion mutation in all patients (Gundesli et al, 2010).

Our study emphasizes the potential pitfall in the usage of low-resolution SNP arrays for linkage analysis and homozygosity mapping, especially if the mutated gene is located in a low-coverage region of these arrays such as the telomeres. In our experience, approximately 250000 to 300000 SNPs spanning the genome should provide sufficient resolution and power to detect also smaller loci. Although, high-density SNP arrays generate massive data,

which may cause false-positive linkage results or multiple homozygous regions, their use is valuable in order to provide the ability to analyze all chromosomal regions in more detail. In conclusion, a combination of exome sequencing and linkage analysis with high-resolution SNP arrays provides a valuable approach for disease gene identification, and to help filter the variants found in the exome data (Bilgüvar et al, 2010).

COMPETING INTERESTS

None declared.

REFERENCES

- Abecasis GR, Cherny SS, Cookson WO and Cardon LR. 2001. GRR: graphical representation of relationship errors. *Bioinformatics*, 17, 742-743.
- Benayoun L, Spiegel R, Auslender N et al. 2009. Genetic heterogeneity in two consanguineous families segregating early onset retinal degeneration: The pitfalls of homozygosity mapping. *Am J Med Genet*, 149A, 650-656.
- Bilgüvar K, Öztürk AK, Louvri A et al. 2010. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*, 467, 7312, 207-210.
- Fendri K, Kefi M, Hentati F and Amouri R. 2006. Genetic heterogeneity within a consanguineous family involving the LGMD 2D and the LGMD 2C genes. *Neuromus Disord*, 16, 316-320.
- Gundesli H, Talim B, Korkusuz P et al. 2010. Mutation in exon 1f of *PLEC*, leading to disruption of plectin isoform 1f, causes autosomal recessive limb-girdle muscular dystrophy. *Am J Hum Genet*, 87, 6, 834-841.
- Hoffmann K and Lindner TH. 2005. easyLINKAGE-Plus-automated linkage analyses using large scale SNP data. *Bioinformatics*, 21, 17, 3565-3567.
- Huqun, Fukuyama S, Morino H et al. 2010. A quantitatively-modeled homozygosity mapping algorithm, qHomozygosityMapping, utilizing whole genome single nucleotide polymorphism genotyping data. *BMC Bioinformatics*, 11 Suppl 7, S5.
- Lezirovitz K, Pardon E, de Mello Auricchio MT, et al. 2008. Unexpected genetic heterogeneity in a large consanguineous Brazilian pedigree presenting deafness. *Eur J Hum Genet*, 16, 89-96.
- O'Connell JR and Weeks DE. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*, 63, 259-266.