



THE RATIONALE, DESIGN AND METHODS OF NEW STUDIES

DOI: <https://doi.org/10.20883/medical.78>

Searching for new genes and loci involved in cleft lip and palate in the Polish population – genome-wide association study

Adrianna Mostowska¹, Kamil K. Hozyasz², Piotr Wójcicki³, Barbara Biedziak^{1, 4}, Joanna Wesoły⁵, Anna Sowińska⁶, Sylwia Matuszewska-Trojan¹, Paweł P. Jagodziński¹

¹ Department of Biochemistry and Molecular Biology, Poznan University of Medical Sciences, Poland

² Department of Paediatrics, Institute of Mother and Child, Warsaw, Poland

³ University Clinic of Medical Academy in Wroclaw and Department of Plastic Surgery Specialist Medical Center in Polanica Zdroj, Poland

⁴ Private Orthodontic Practice

⁵ Laboratory of High Throughput Technologies, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland

⁶ Department of Computer Science and Statistics, Poznan University of Medical Sciences, Poland

ABSTRACT

The project “Searching for new genes and loci involved in cleft lip and palate in the Polish population – genome-wide association study” is a case-control study in a group of unrelated subjects with non-syndromic cleft lip with or without cleft palate (NSCL/P) and healthy individuals with no family history of clefting or other congenital disorders. The overall goal of this grant proposal is to identify novel genetic factors, which can play a significant role in the pathogenesis of orofacial clefts in the Polish population. To accomplish the proposed aim, a two stage genome-wide association study will be performed. In the first stage, Illumina's HumanOmni Express BeadChips arrays will be used to genotype over 700,000 polymorphisms in NSCL/P patients and controls. In the second stage, SNPs showing the most compelling association with the risk of orofacial clefts will be tested in an independent sample set using standard genotyping methods. This research project is expected to be completed in July 2015.

Keywords: genome wide association study, cleft lip and palate, risk factors, polymorphisms.

General information

The project “Searching for new genes and loci involved in cleft lip and palate in the Polish population – genome-wide association study” was awarded by the Polish National Science Center (NCN) under project number: 2012/07/B/NZ2/00115 (OPUS 7 competition). *The duration of the grant is 24 months, from 2nd July 2013 to 1st July 2015. The contract between NCN and*

Poznan University of Medical Sciences (PUMS), Poland, was signed on 2 July 2013.

Management

The Principal Investigator of the grant is Dr Adrianna Mostowska, Assistant Professor in the Department of Biochemistry and Molecular Biology at PUMS. Main Co-Invetigators in the grant are Prof. Paweł P.

Jagodziński from PUMS and Prof. Konrad K. Hozyasz from the Institute of Mother and Child in Warsaw. In addition, Co-Investigators are: Prof. Piotr Wójcicki (*University Clinic of Medical Academy in Wrocław and Department of Plastic Surgery Specialist Medical Center in Polanica Zdrój*), Dr Barbara Biedziak (PUMS), Prof. Joanna Wesoły (Department of Human Molecular Genetics, Institute of Molecular Biology and Biotechnology in Poznań), Dr Anna Sowińska (PUMS) and MSc Sylwia Matuszewska-Trojan (PUMS).

Ethics

Bioethical Committee at PUMS on 3rd November 2011 accepted all project's protocols and forms, including information for patients form and consent form for participation in a research study (the permission number 891/11).

Finance

The total amount of grant funding is 1,475,140 Polish Zlotys (about 352,000 Euro). The amount of funding in the first year is 1,300,420 Polish Zlotys (about 310,000 Euro). Grant funds were designed to purchase all reagents, disposable laboratory equipment and software products for the conduct of the proposed research and analyse received data. In addition, grant funds were designed to cover personnel costs of project participants and costs associated with the dissemination of research results, i.e., through publications in peer reviewed journals and presentations at professional scientific conferences.

Research basic concept and objectives

Orofacial clefts are one of the most common congenital anomalies in humans (OMIM #119530; OMIM #119540). According to the *Polish Registry of Congenital Malformations*, the prevalence of these birth defects in Poland ranges from 1/500 to 1/1,000 births [www.rejestrwad.pl]. Orofacial clefts represent a significant public health problem since their treatment requires comprehensive surgical, orthodontic, phoniatric and psychological management strategies [1]. In addition, long term outcomes of individuals born with these anomalies may include higher overall death rates, mental health problems, alternations of childbearing patterns and a higher risk of cancer (particularly breast cancer) in both affected individuals and their family members [2].

The aetiology of non-syndromic cleft lip with or without cleft palate (NSCL/P) is heterogeneous and both genetic and environmental factors affect the risk of this developmental malformation [2]. Research studies of NSCL/P using a wide range of methodological approaches have discovered a number of genes (e.g. *IRF6*, *BMP4*, *FGFR2*, *FOXE1*, *MSX1* or *MYH9*) and chromosomal regions underlying this structural anomaly [2, 3]. In addition, four independent genome wide association studies (GWAS) for NSCL/P have successfully identified several loci (8q24.21, 1p22.1, 10q25.3, 17q22 and 20q12) at which nucleotide variants influence the risk of NSCL/P [4–7]. However, all mutations and/or polymorphisms of candidate genes and chromosomal loci characterized so far may still explain only a fraction of the inherited contribution to NSCL/P aetiology. This can be due to the locus and allelic heterogeneity among populations (the existence of different candidate genes and many different disease causing alleles at a given locus, respectively), incomplete penetrance, as well as complicated epistasis and gene-environmental interactions. In addition, in most cases, the functional role of identified polymorphic variants in the pathogenesis of orofacial clefts is unknown. Therefore, the overall goal of this grant proposal is to identify novel genetic factors, which can influence the risk of NSCL/P in the Polish population. To accomplish the proposed aim, a two stage genome-wide association study will be performed using Illumina's HumanOmni Express BeadChips arrays. GWAS is defined by the National Institutes of Health as a study of common genetic variation across the entire genome designed to identify genetic associations with observable traits. GWAS have greater power than linkage studies to detect small to modest effects, even with a strict alpha-level for statistical significance ($\alpha = 5.0 \times 10^{-8}$) [8]. Moreover, by casting a wide net of genetic markers across the entire genome, this approach does not require to select candidate genes for study and examines much of the common variation across the genome. To date, GWAS have detected hundreds of variants associated with a large number of diseases [9]. Many of these findings are novel, since identified SNPs were not previously recognized as disease risk factors.

Research plan

The proposed project will be composed of two stages: Stage 1

- selection of individuals with orofacial clefts and a suitable comparison group (number of cases =

288, number of controls = 576), DNA isolation and quantification;

- high-throughput genotyping and data review to ensure high genotyping control;
- statistical tests for associations between SNPs passing quality thresholds and orofacial clefts.

Stage 2

- selection of additional individuals with orofacial clefts and a suitable comparison group (number of cases = 400, number of controls = 500), DNA isolation and quantification;
- replication of identified associations in an independent population sample;
- examination of functional implications using various research databases and tools.

Research methodology

Sample collection

Peripheral blood samples from 688 (288 for the first stage and 400 for the second stage of the study) unrelated subjects with NSCL/P will be collected in the Departments of Paediatrics and Paediatric Surgery at the Institute of Mother and Child in Warsaw, as well as in the Department of Plastic Surgery at the Medical Academy in Wrocław. To avoid sample heterogeneity, individuals with cleft palate only will be excluded from the study prior to genotyping. Also patients with accompanying congenital malformations, chronic medical conditions, or vaguely described dysfunctions (syndromic forms of orofacial clefts) will not be included in the analyses. Eligibility to the patient group will be ascertained from detailed medical records. In addition, 1076 (576 for the first stage and 500 for the second stage of the study) healthy individuals with no family history of clefting or other congenital disorders will be used as controls. Both patients and age- and gender-matched control samples will be Caucasians of Polish origin recruited from the same geographic regions. DNA will be extracted from peripheral blood samples using standard salt extraction procedure. DNA concentrations will be adjusted to 50 ng/μl and verified using PicoGreen dsDNA Quantitation kit (Molecular Probes, Invitrogen, CA, US).

Genome-wide association study

High-throughput genotyping of all samples will be carried out using the Illumina's HumanOmni Express *BeadChips* (Illumina, CA, US) according to manufacturer's

protocol. These 12-sample *BeadChips* feature 730,525 strategically selected markers to capture the greatest amount of common SNP variation (> 5% minor allele frequency). The SNPs selection is based on linkage disequilibrium (LD) between nucleotide variants. A high r^2 between two SNPs indicates high correlation, making these SNPs good proxies (tag SNPs) to each other. At a maximum r^2 of 1, two SNPs are in perfect LD and can serve as pure proxies. Thus, only one SNP needs to be genotyped to know the genotype of the other. Illumina DNA Analysis products offer unparalleled genomic coverage using tag SNPs with the highest average r^2 values in the industry (<http://www.illumina.com/>).

Statistical analysis of GWAS results

Genotypes for all arrays will be calculated using *BeadStudio's* genotyping module (v2.0, Illumina, CA, US). Stringent quality control will be applied to the genotyping data in order to exclude experimental errors and low-quality SNPs. All statistical analyses will be performed using *PLINK* version 1.06. The association of SNPs with NSCL/P will be tested by a 1-degree-of-freedom Cochran-Armitage trend test. The p-value below 5×10^{-8} will be considered as statistically significant [10]. Statistical analysis of obtained results will also include the determination of the differences in allele and genotype frequencies between cases and controls using the standard χ^2 and Fisher exact tests and the calculation of the Odds Ratio (OR) with corresponding 95% confidence intervals (95% CIs). For nucleotide variants showing the strongest associations further computational analyses will be performed in order to determine if these nucleotide variants may be real, etiological polymorphisms involved in NSCL/P aetiology. PolyPhen online database (<http://genetics.bwh.harvard.edu/pph/>) will be used for prediction of functional consequences of the non-synonymous SNPs on protein structure and function.

Replication study

The SNPs showing the most significant associations with NSCL/P risk in the GWAS will be retested in a subsequent replication study conducted in an independent set of cases and controls (400 patients *z* NSCL/P and 500 healthy individuals). Genotyping will be carried out either by PCR followed by digestion of the amplified products with the appropriate restriction enzyme (PCR-RFLP), high-resolution melting curve analysis (HRM) or using TaqMan assays (*Applied Biosystems*; CA, US).

Statistical analysis of replication study results

For each SNP, the *Hardy-Weinberg equilibrium* will be assessed by Pearson's goodness-of-fit Chi-square *statistic*. The differences in allele and genotype frequencies between cases and controls will be determined using standard Chi-square or Fisher tests. SNPs will be tested for association with NSCL/P using the Cochran-Armitage trend test. The OR and associated 95%CI will also be calculated. Haplotype analysis will be performed using the UNPHASED 3.1.5 program. Gene-gene interactions will be evaluated using the nonparametric and genetic model-free Multifactor Dimensionality Reduction (MDR) approach (MDR version 2.0 beta 5).

Expected results

The presented project will allow the identification of novel genes and chromosomal loci, the nucleotide variants of which are significantly associated with the occurrence of NSCL/P in the Polish population. *This will enable a more complete understanding of the highly complex etiology of this common developmental anomaly, which can in the future help to improve the current treatment methods and to design the programs of primary prevention of craniofacial abnormalities.* The results obtained during this project may also contribute to deepening the knowledge of the molecular mechanisms involved not only in face development, however, also in the development of the whole embryo. Development of the head and face comprises one of the most complex events during embryonic development, coordinated by a network of transcription factors and signalling molecules together with proteins conferring cell polarity and cell-cell interactions [11]. The results of this project should also help us to design our future research plans. Markers selected in the genome-wide association study will be further evaluated in fine mapping and next generation sequencing studies to identify causal variants, and in functional studies to understand the biological mechanism of the observed associations with the non-syndromic orofacial clefts in the Polish population.

Conflict of interest

The authors declare that they have no competing interests.

References

1. Mostowska A, Hozyasz KK, Wojcicki P, Biedziak B, Paradowska P, Jagodzinski PP. Association between genetic variants of reported candidate genes or regions and risk of cleft lip with or without cleft palate in the Polish population. *Birth Defects Res A Clin Mol Teratol.* 2010;88: 538–545.
2. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet.* 2011;12:167–178.
3. Leslie EJ, Marazita ML. Genetics of cleft lip and cleft palate. *Am J Med Genet C Semin Med Genet.* 2013;163: 246–258.
4. Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet.* 2009;41:473–477.
5. Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *JPediatr.* 2009;155:909–913.
6. Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MFB and ACA4. *Nat Genet.* 2010;42: 525–529.
7. Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet.* 2010;42:24–26.
8. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273:1516–1517.
9. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci US A.* 2009;106:9362–9367.
10. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437:1299–1320.
11. Stanier P, Moore GE. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet.* 2004;13 Spec No 1:R73–81.

Correspondence address:

Adrianna Mostowska, PhD
Department of Biochemistry and Molecular Biology
Poznan University of Medical Sciences
6 Swiecickiego Street
60-781 Poznan, Poland
phone: +48 618546511
fax: +48 618546510
email: amostowska@wp.pl