The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

INVESTIGATING ASSOCIATION OF PREDISPOSITION GENES AS A RISK OF DEVELOPING MULTIFACTORIAL DISEASES

Assel Arginbekova

The thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Poster presentations

Assel Arginbekova and John Armour "INVESTIGATING ASSOCIATION BETWEEN AMYLASE COPY NUMBER VARIATION AND PANCREATIC DISEASES". Presented at the 7th Joint Dutch/UK Clinical Genetics Societies and Cancer Genetic Groups meeting on 10-11 June 2020, Cambridge, UK. Sanger Institute, Hinxton.

Assel Arginbekova and John Armour "EXPLORATION OF DEFA1A3 AS A RISK FACTOR IN PERIODONTITIS". Presented at the 7th Joint Dutch/UK Clinical Genetics Societies and Cancer Genetic Groups meeting on 10-11 June 2020, Cambridge, UK. Sanger Institute, Hinxton.

Assel Arginbekova and John Armour "INVESTIGATING ASSOCIATION BETWEEN AMYLASE COPY NUMBER VARIATION AND PANCREATIC DISEASES". Presented at European Human Genetic Conference on 12-15 June 2021.

Acknowledgement

I would like to express my gratitude to everyone who helped me and contributed to the research for my thesis.

Firstly, I would like to thank my supervisors, John Armour and David Brook, for their guidance and support in my PhD journey.

I want to dedicate special thanks to Queen's Medical Center, Nottingham University Hospitals NHS Trust, particularly the Gastrointestinal Surgery Unit team and Prof. Dileep Lobo for providing biological material for the research. Also, to Twin UK Research & Genetic Epidemiology, King's College London, and Dr Claire Steves for collecting and providing data for the periodontitis study. My thanks to DeepSeq Labs and Matthew Carlile for technical assistance.

I want to thank Professors Robert Layfield and Sara Goodacre for their support and advice in my PhD study.

I am grateful to the Ministry of Education and Science of Kazakhstan and the Center for International Programs "Bolashak" for the scholarship and funding my PhD study.

Abbreviations

AD	Alzheimer's disease
ADGC	Alzheimer Disease Genetics Consortium
ADH	Alcohol dehydrogenase
AGPAT5	1-Acylglycerol-3-Phosphate O-Acyltransferase 5
ALDH	Aldehyde dehydrogenase
AMY1	Salivary amylase
AMY2	Pancreatic amylase
ANOVA	Analysis of Variance
AP	Activation peptide
APOE	Apolipoprotein E
ATP	Adenosine triphosphate
BLK	Tyrosine-protein kinase BLK (Lymphocyte kinase)
BMI	Body mass index
BRCA	Breast cancer
CASR	Calcium-sensing receptor
CBD	Common bile duct
CCL3L1	C-C Motif Chemokine Ligand 3 Like1
CEL	Carboxy ester lipase
CEL VNTR	Carboxyl-ester lipase variable number of tandem repeats
CEL-HYB	Carboxyl-ester lipase Hybrid allele
CELP	Lipase gene pseudogene
CFTR	Cystic fibrosis transmembrane protein regulator
CGH array	Comparative genomic hybridisation
CHD	Coronary heart disease
CLDN2	Claudin 2
CMT	Charcot-Marie-Tooth
CNA	CNV alteration
CNTN6	Contactin 6
CNV	Copy number variation
CNVR	Copy number variable regions
COPD	Chronic obstructive pulmonary disease

CVS	Cardiovascular system
DEFA	α-Defensin
DEFB	β-Defensin
DM	Diabetes Mellitus
DTRGE	Department of Twin Research & Genetic Epidemiology
DZ	Dizygotic twin
ENT	Ear, nose and throat
ERCP	Endoscopic retrograde cholangiopancreatography
ESR1	Estrogen receptor 1
EUROPAC	European Registry of Hereditary Pancreatitis and Pancreatic Cancer databases
FISH	Fluorescence in situ hybridisation
FRET	Förster resonance energy transfer
GCK	Glucokinase
GEMMA	Genome-wide Efficient Mixed Model Association
GLP	Glucagon-like peptide
GnomAD	Genome Aggregation Database
GSV	Consortium for Genome Structural Variations
GWAS	Genome-wide association studies
GWAS Central	Human genetic material and a unique database
HapMap	International Haploid Genome Project
HFE	Hemochromatosis
HLA	Human leukocyte antigens
HNF	Hepatocyte nuclear factor
HOMA-IR	Homeostatic model l-insulin resistance
HP	Helicobacter pylori
HPA	Human pancreatic amylase
HPV	Human papillomavirus
HRC	Human Random Control
IL	Interleukin
IL-1RN	IL-1 receptor antagonist
TNF-α	Tumour Necrosis Factor-alpha

INS	Insulin
ITGB2	Integrin beta chain-2
KASP	Competitive Allele-Specific PCR
KLF11	Kruppel-like factor 11
LD	Linkage Disequilibrium
LD	Low dNTPs
M-CSF	Colony-stimulating factor, macrophage-specific
MAF	Minor allele frequency
MAPH	Multiplex amplification and probe hybridisation
MCPH1	Microcephalin 1
MFD	Multifactorial diseases
MFN	Mitochondrial membrane protein
MLPA	Multiplex ligation-dependent amplification
MODY	Maturity-onset diabetes of the young
MRI	Magnetic resonance imaging
MZ	Monozygotic twin
NAD	Nicotinamide adenine dinucleotide
NAPS2	North American Study of Pancreatitis
NCD	Neurocirculatory disorders
NEUROD	Neurogenic differentiation 1
NGRC	NeuroGenetics Research Consortium
NGS	Next Generation Sequencing
NK	Natural killer
OPG	Osteoprotegerin
PAX	Paired-box-containing gene 4
PC	Pancreatic cancer
PCR	Polymerase chain reaction
PDAC	Pancreatic ductul adenocarcinoma
PDX	Pancreatic and duodenal homeobox
Pi	Protease Inhibitor gene
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein
PLA2	Phospholipase A2
PMP22	Peripheral myelin protein 22
PPAR	Peroxisome-proliferator-activated receptors
PRC	Polygenic risk score

PRSS1	Genes of cationic trypsinogen
PRT	Paralogue ratio test
PTK2B	Protein tyrosine kinase 2
PYY	Peptide tyrosine tyrosine
QMC	Queen's Medical Centre, Nottingham
QTL	Quantative trait loci
RANKL	Receptor activator of nuclear factor kappa-B ligand
RR	Relative risk
SD	Standard deviation
SERPINA	Serine protease inhibitor
SLC35C1	Solute Carrier Family 35 (GDP-Fucose Transporter), Member C1
SNP	Single nucleotide polymorphisms
SPINK1	Trypsinogen inhibitor
STR	Short tandem repeats
TCF	Transcription factor
TCF7L2	Transcription factor 7-like 2
TGF-β	Transforming growth factor
TREM2	Triggering receptor expressed on myeloid cells
TSPAN 8	Tetraspanin 8
TspRI	Thermo Scientific TscAI restriction enzyme
TYROBP	Protein Tyrosine Kinase Binding Protein
UBR1	Ubiquitin Protein Ligase E3 Component N-Recognin1
UCSC	Genome Browser
VDR	Vitamin D receptor
WTCCC	Welcome Trust Case Control Consortium

Abstract

Over the past decade, the aetiology and pathogenesis of multifactorial diseases have been studied. The emergence of multifactorial diseases is based on complex mechanisms of engagement between genetic and environmental factors and the search for candidate genes that predispose to diseases. The fundamental results of studying the foundations of multifactorial diseases are based on the analysis of genetic associations. The motivation for our study was to improve the predictive ability of multifactorial diseases and the use of genetic studies in clinical practice.

This thesis explores directions that explain the genetic mechanisms of predisposition to multifactorial diseases. In this regard, two common and severe diseases, periodontitis and pancreatic diseases, belonging to the group of multifactorial diseases, were studied. Our study demonstrated an association between clinical status and copy number variation and single nucleotide polymorphisms in the two diseases.

The first project aimed at determining the association of the *AMY* gene with pancreatic diseases. It should be noted that there have been no previous studies to investigate the association between the *AMY* copy number variants and pancreatic diseases, which prompted the search for this association. Research on *AMY* copy number variants has mainly been in the area of population differences and obesity. It is assumed that copy numbers of the *AMY* gene are associated with the occurrence of multifactorial diseases. Multiple pancreas diseases, such as acute and chronic pancreatitis or cancer, can be considered multifactorial diseases. Therefore, for the first time in the project, the copy number variants of the *AMY1, AMY2A* and *AMY2B* genes and the role of the genes in the development of diseases associated with the pancreas have been investigated. We have studied the possible role of the CNV amylase gene by laboratory testing of biological material (blood) from Queen's Medical Centre patients. The study found evidence of a positive association between the amylase gene and pancreatic diseases. And also, the clinical and genetic features of various forms of pancreatitis were studied. The range of copy numbers of the *AMY* gene was determined specifically for each form of pancreatic disease. In addition, the influence of individual factors (age, gender and BMI) on developing the risk of pancreatic diseases was studied. Clinical and genetic features of various forms of pancreatitis were also investigated.

The presence of N34S mutations in the *SPINK1* gene was determined in the development and progression of pancreatitis. The frequency of occurrence of the N34S mutation in the *SPINK1* gene was identified among patients with pancreatic diseases in Nottingham, UK. A comparative analysis of the frequency of occurrence in other populations was also carried out. A positive association between N34S mutation and pancreatic disease was identified.

The second periodontal disease project focused on the relationship between clinical factors and the presence of SNPs and flanking α-defensin genes. The study identified the association between the periodontitis phenotype and the *DEFA* gene. Previously, combinations of clinical signs of periodontitis and the *DEFA* gene have not been sufficiently studied. Therefore, the research was expanded to investigate the relationship between the main clinical symptoms of periodontitis with specific polymorphisms of the *DEFA* gene, which is responsible for the immune response and the body's defences. The study identified seven specific SNPs associated with periodontitis's four most important clinical characteristics.

In the thesis, the first attempt was made to determine the associations of the *AMY1, AMY2A* and *AMY2B* genes with the development of diseases and the formation of clinical forms. Certain polymorphisms of the *DEFA* gene have been established as associated with the disease, taking into account clinical features.

The development of any process in multifactorial diseases is not determined by the entire genome but only by predisposition genes, which makes it possible to study specific genes. The conducted studies make it possible to assess the occurrence of pathologies, possible interventions in the clinical conditions and their use in the early diagnosis of diseases. The thesis studied the direction of transforming genetic information into clinical practice, for an individual approach to the patient, based on an interpretation of the results of a genetic study and their comparison with clinical phenotypes.

Table of Contents:

CHAPTER I . GENERAL INTRODUCTION

The role of genetic factors in multifactorial diseases.

Multifactorial diseases (MFDs) represent various groups of diseases and a significant part of the entire human somatic pathology. MFDs are characterised by high rates of increase in morbidity, mortality, and disability among able-bodied people in modern populations. Carrying out therapeutic and preventive measures does not cope with the current situation. Traditional approaches for treating common MFD lead to a significant increase in economic costs and very modest results. The low effectiveness of therapeutic and preventive measures is associated with the lack of their aetiological focus due to a poor understanding of the critical mechanisms for forming most MFDs. Difficulties in determining multifactorial diseases were associated with the inability to isolate a pathogenetically significant genetic complex, which, in combination with unfavourable environmental factors, would lead to the onset of the disease.

The occurrence of MFD is formed under the influence of various factors, both genetic and environmental, the latter in the form of lifestyle[1,2]. It is based on complex mechanisms in the field of genetic components and environmental influences (Figure 1). Unlike chromosomal and monogenic diseases, MFD has a high incidence and morbidity[3].

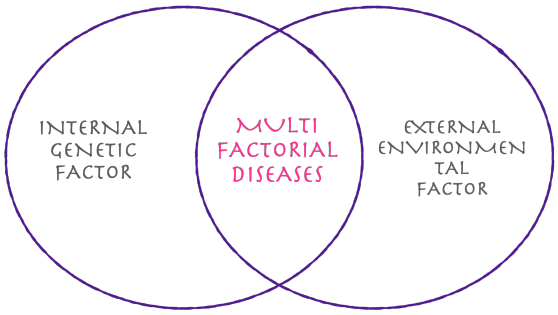


Figure 1. Mechanism of MFDs onset.

Predisposition genes can be defined as hereditary factors suggesting compatibility with life. However, exposure to adverse environmental factors can cause various diseases, such as arterial hypertension, diabetes mellitus, pancreatic diseases, bronchial asthma, atherosclerosis, diseases of the brain and teeth, some forms of cancer, congenital malformations, schizophrenia, etc. Most of these diseases are multifactorial, meaning that clinical symptoms manifest only due to the combined action of several genes and environmental factors. Such diseases are called multifactorial diseases or diseases with a hereditary predisposition.

Predisposition genes are genes in which different allele variants predispose to certain diseases[4].

There is the concept of polygenic inheritance, which can be formed in families with high heritability. They do not obey the Mendelian inheritance laws, as they are influenced by several genes. Monogenic diseases are characterised by the involvement of one mutated gene with a dominant inheritance pattern; recessive requires two copies of the gene. Monogenic diseases are rare in populations and represent a small percentage of the overall disease burden. These models determine the inheritance pattern in families and causal mutations with high penetrance[5].

It is believed that MFD may have a mechanism that includes threshold and quality traits, as well as complex predisposing factors of adulthood in the form of lifestyle, diets and various concomitant diseases. Threshold traits occur when a certain level is reached, which combines the genetic basis and the external environment. An example would be neural tube defects and congenital malformations. Quality traits include a combination of several genes and the environment. An example could be human features such as facial features, height, weight, and pressure. Complex disorders of adulthood are manifested by an apparent clustering in families and a significant influence of the external environment where the genetic background has less of an impact[6].

The main reasons for the difficulties in identifying genes that contribute to the genetic predisposition to MFDs are 1) the absence of the Mendelian type of inheritance in most diseases; 2) they have pronounced genetic heterogeneity (each clinical form is a group of hereditary defects with the same manifestation); 3) insufficient knowledge of the pathogenetic mechanisms of diseases[7].

The active implementation of the Human Genome program at the end of the 20th and the beginning of the 21st century and the

development of molecular technologies in genetics opened up broad prospects for studying the role of individual genes and groups of genes in the formation of MFD. This project helped expand the professional knowledge of genetics[8]. Based on advances in the human genome study, it has been possible to identify genes that cause hereditary diseases and mutations predisposing to the most common multifactorial disorders. The data obtained because of these studies on the involvement of certain molecular genetic factors in the development of multifactorial pathology will make it possible at the preclinical stage to form risk groups for the onset of the disease. In these groups, it is more practical to carry out preventive measures and also to predict the nature of the clinical course and the effectiveness of therapy for the disease. This will improve the diagnostic and treatment process for each patient.

For the preclinical diagnosis of MFD, the study of the role of genetic factors will allow researchers and clinicians not only to identify intergenic interactions that form a predisposition to the disease, to assess the possible pathogenetic mechanisms of its development but also to identify the range of possible environmental factors that can provoke the occurrence and exacerbation of a particular pathology[9]. Research in the field of MFD formation will not only provide fundamental data on the population genetic and molecular genetic factors of MFD and use these data to conduct genetic testing of patients in practical medicine. The most common diseases have genetic, behavioural, and environmental risk factors and are of relevance in genomic research[10].

Genomic studies of predisposition to multifactorial diseases.

In human genome research, considerable attention is given to disease risk prediction using a genomic database[11]. Human genomes have 20000 protein-coding genes. About 5000 genes are associated with diseases[12]. Technological development has made a much more accurate measurement of genetic variation possible, already leading to the identification of hundreds of predisposing genes. However, questions remain unexplained in the field of the genes functioning underlying the realisation of a polygenic predisposition to this class of diseases. Increasing progress in genetic methods is leading to a significant development of the research perspective[13].

The study of MFD includes several stages: determining the degree of heritability of a particular disease, segregation analysis, and examining the relationship between the basis of the disease and genetic localisation[2]. Mapping is used to identify causal variants. To study common diseases, methods have been developed to examine human populations' genetic changes. The most significant for the genome is the HapMap and 1000 Genomes projects. These projects helped to use genotyping technologies with high accuracy. All this allowed research in the field of associations between the genome and the most common diseases[14]. Various projects have made it possible to conduct countless studies in the field of associations throughout the genome, particularly MFDs. The understanding of genetic information obtained during the research is one of the critical components of the development of public health[15]. The study of the association of MFD with specific genes assumes that if a particular gene product is involved in forming a predisposition to MFD, one of its alleles should be found in patients much more often than in the general population. The frequency of occurrence of other alleles of this gene in patients should be lower than in the population.

In most cases, alleles of such genes which may be involved in the development of the disease, are chosen as a test marker. If an increased frequency of occurrence of the studied marker in patients compared with the control is found, it can be concluded that there is an association with the disease[16]. Since genetic risk factors influence the pathogenesis of a particular disease, it is an essential direction for studying their development principles. An example of this is specific polymorphisms and CNVs that create a genetic predisposition, which in turn interact with other risk factors leading to the development of the disease. Understanding the complex mechanism of interaction between genes and environmental factors can increase the motivation of ordinary people to change unhealthy lifestyles in the presence of a genetic predisposition[17].

Multiple factors can be demonstrated in families with a specific and consistent pattern of inheritance[18]. A study of families and twins has shown a genetic basis, but the external environmental contributions are also significant. One example is hyperlipidemia, which leads to atherosclerosis and coronary heart disease. This disease can be caused both by a genetic factor of a single gene, such as familial hypercholesterolemia and by multifactorial processes, where the effects of genes and the environment in the form of lifestyle and diet are combined, it can occur in isolation[19]. With a predisposition to diabetes,

the external factors of the diet for non-insulin-dependent and viral infections in the insulin-dependent type are included. Congenital malformations, viral diseases and folic acid deficiency in the pregnant mother are external factors, but also a 4% risk of a genetic factor as a risk for siblings[20]. Some neuropsychiatric diseases and congenital malformations occur as complex MFDs, and their risk may be higher in men or women[21].

Any MFD is based on genetic and non-genetic factors. Among a variety of biological markers, an important place in the diagnosis of diseases belongs to genetic markers. With the development of genetic technologies, it has become possible to detect various mutations and variants in the genome. Therefore, it becomes relevant to study the use of genetic markers as a predisposition to detect the development of multifactorial diseases.

For an accurate assessment of MFD risks, studies of numerous SNPs and CNVs are needed, which should be considered in combination with non-genetic factors. Currently, such information is generally unavailable, and such studies are scarce. At the same time, finding these missing interactions can be helpful for more accurate risk prediction[22]. For most common MFD, hereditary factors are not strictly determined[5]. In numerous clinical studies, the disease risk for MFD has been identified with a high degree of certainty. Testing for common MFDs is desirable to be introduced into healthcare practice to prevent the development of diseases. For this, genetic research in the field of MFD is of great importance.

GWAS considers a method of inheritance for complex diseases, including MFD, using variants of a low minor allele frequency (MAF), which varies in the range of $0.5<MAF<5\%$. $MAF<0.5\%$ is considered rare. They are not often found and do not have a sufficient impact. Nor can they be discovered by classical analysis of familial inheritance. With a low MAF, association detection becomes impossible. Rare alleles are more often associated with the Mendelian inheritance pattern and are identified by linkage analysis. Such alleles may have a larger effect size, along with rare allele frequencies. This type of allele can cause massive disturbances and changes in the encoded protein, which triggers the onset of the disease. These alleles can have different effects on the phenotype. For this reason, this category is determined by linkage analysis and is recognised as having a familial type of inheritance. Unlike MFDs, Mendelian diseases have a rare allele frequency and high penetrance. Low-frequency variants increase the risk of diseases to a

large extent (2-3 times). These variants do not show Mendelian inheritance and therefore contribute to the lack of clear inheritance. GWAS considers MAF≥1 variants where an association with diseases is found; these are common variants with medium effect size (Figure 2).[23-26]

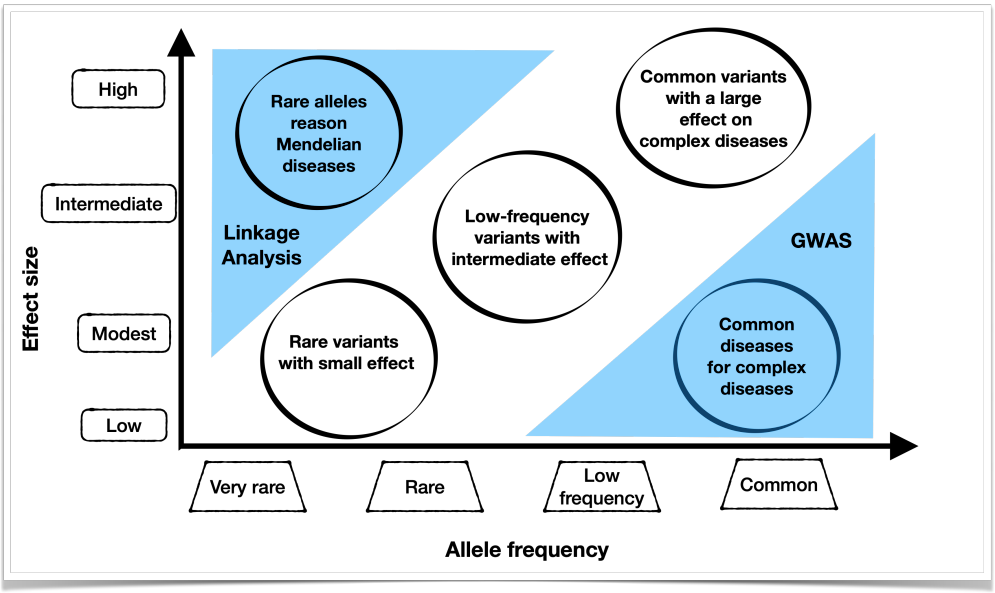


Figure 2. Determination of variants of genetic markers by the frequency of allelic risks and the severity of genetic predispositions. Adopted from Manolio et al. (2009)[23] and McCarthy et al. (2008)[24].

There are several studies illustrating different approaches in the field of MFD. One example is breast cancer, whose genetic nature is being actively studied. Mutations in the *BRCA1* and *BRCA2* genes have been found to significantly affect the development of diseases such as breast and ovarian cancer, which define them as predisposition genes. In studies, screening of the *BRCA1* and *BRCA2* genes has been shown to be an essential marker for the early detection of people at risk of disease and family history[20,27-29].

Diseases such as cardiovascular disease and dyslipidemia are widespread among MFDs, the pathogenesis of which depends on multiple interactions of a large number of genetic polymorphisms, as well as on their interactions with environmental and lifestyle factors. More than 160 genes have been identified by studying the GWAS responsible for the development of these MFD diseases[30].

Hereditary hemochromatosis is an autosomal recessive disorder that affects about 1 in 200 to 250 people. The most common mutation in the *HFE* gene, encoding a protein and transporting iron into the cytoplasm of cells, leads to excessive iron absorption by cardiomyocytes, pancreatic cells, and hepatocytes[31]. Also, mutations of genes Hemojuvenilin (*HJV*), *HAMP* (the gene encoding Hepcidin), Transferrin receptor 2 (*TfR2*) and Ferroportin (*FP*) cause the disease[32].

Without drug therapy, chronic liver diseases develop fibrosis, cirrhosis and hepatocellular carcinoma. There are invasive and non-invasive diagnostic methods, including MRI, to quantify the level of iron in the liver and its changes[31]. However, most methods are costly and time-consuming.

One of the reasons for developing chronic obstructive pulmonary disease (COPD) and pulmonary emphysema is the deficiency of *α1-antitrypsin*, which mutates in the *Pi* (Protease Inhibitor) gene of the *SERPINA1* gene. It is characterised by a low level of serum *α1-antitrypsin*[33].

Genetic testing methods in the case of Diabetes Mellitus (DM) can also help establish the cause of the disease and its prognosis. A severe and aggressive type of diabetes is known as MODY (Maturity-onset diabetes of the young). A feature of MODY is the development of the disease at an early age and an autosomal dominant type of inheritance. This causes a high recurrence compared with other types of diabetes. MODY is a genetically determined form of diabetes. It differs from diabetes in treatment tactics and prognosis. MODY is a polygenic and heterogeneous disease. Verification of the diagnosis of MODY based on genetic methods plays an important role in family history. Diagnosing MODY in 80% of cases is often challenging, as it coincides with type 1 and type 2 diabetes, thereby complicating treatment and its complications[34]. There are currently 13 MODY subtypes. Genes that cause MODY are hepatocyte nuclear factor 4α (*HNF4A*; MODY1), glucokinase (*GCK*; MODY2), *HNF1A* (MODY3), pancreatic and duodenal homeobox 1 (*PDX1*; MODY4), transcription factor 2 (*TCF2*) or *HNF1B* (MODY5), neurogenic differentiation 1 (*NEUROD1*; MODY6), Kruppel-like factor 11 (*KLF11*; MODY 7), carboxyl ester lipase (*CEL*; MODY8), paired-box-containing gene 4 (*PAX4*; MODY9), insulin (*INS*; MODY10), B-lymphocyte kinase (*BLK*; MODY11), adenosine triphosphate (ATP)-binding cassette, sub-family C (*CFTR/MRP*), member 8 (*ABCC8*; MODY12), and potassium channel, inwardly rectifying subfamily J, member 11 (*KCNJ 11*; MODY13). Only

molecular genetic studies can reliably confirm MODY. Clinical manifestations and complications in different types vary. Thus, the most common MODY2 subtype, which develops due to a mutation in the glucokinase (*GCK*) gene, is often a mild type of DM. Hyperglycemia, in this case, is eliminated with the help of diet. Mutations in the hepatocyte nuclear factor 1 α (*HNF1A*) gene are a more severe form. However, such patients demonstrate an excellent hypoglycemic effect when prescribed oral hypoglycemic drugs and sulfonylurea derivatives. In this case, insulin treatment may be delayed. MODY1 is clinically similar to MODY3. The subtype of MODY5, caused by mutations in the hepatocyte nuclear factor 1 β (*HNF-1B*) gene, is characterised by a rapid decrease in the number of β-cells and requires immediate insulin administration. This highlights the need to recognise subtypes of DM to optimise their treatment[35-37].

Type 1 diabetes is also a multifactorial disease in which autoimmune damage to pancreatic β-cells plays a major role in its pathogenesis[38]. The HLA class II genes at 6p21 are the main loci. Since the genetic factor is an important component, the risk of type 1 diabetes is 30–50%[39]. There are also other loci, such as non-HLA, but they have a lower genetic risk. The loci include the insulin gene (*INS*) on chromosome 11p15[40], the polymorphic, cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*) gene on chromosome 2q33[41], the protein tyrosine phosphatase, non-receptor type 22 (lymphoid) (*PTPN22*) gene on chromosome 1p13[42], the interleukin 2 receptor, alpha (*IL2RA*) and interferon induced with helicase C domain 1 (*IFIH1*) genes[43]. These loci were discovered using GWAS studies[38].

Type 2 diabetes is a polygenic metabolic disease characterised by hyperglycemia, which occurs due to impaired insulin secretion as well as insulin resistance. Among the various genetic factors associated with type 2 diabetes, transcription factor 7 (*TCF7L2*), *KCNQ1* and *KCNJ11* are at the highest genetic risk. Of note, the mechanisms by which *TCF7L2* predisposes to DM are unknown[44,45].

Alzheimer's disease (AD) is caused by mutations in the apolipoprotein E (*APOE*) gene (alleles - *APOE ε 4*) and the *TREM2* gene. There are three isoforms of this protein (*APOE2, APOE3* and *APOE4*), which differ from each other in amino acid residues. The *APOE3* gene allele encodes a protein with the amino acid cysteine at position 112 and arginine at position 158. The *APOE2* gene allele encodes cysteine at positions 112 and 158. The *APOE4* gene allele has arginine at both positions. The risk of Alzheimer's disease in carriers of *APOE4* mutations is higher compared to patients who do not have this

mutation. Patients with the *APOE4* genotype respond worse to treatment with the drugs. *APOE ε 4* is the main genetic risk factor. The risk of AD increases ~3-fold if one is inherited and ~12-fold if two copies are inherited. *TREM2* is a myeloid cell surface receptor. When combined with *TYROBP* (Protein Tyrosine Kinase Binding Protein), activated *TREM2* creates a variety of variants that aid cell survival, proliferation, chemotaxis, and phagocytosis. They are involved in vital functions. Pharmacogenetic testing will allow individualised treatment of patients who have symptoms of AD[46-48].

Thus, introducing screening genetic technologies for genes of common MFDs is the most promising. A genetic screening test for MFD-causing mutations can provide early detection, treatment, and targeted surveillance, which will effectively prevent unwanted disease complications. In this regard, research is needed to identify the further impact of genetic factors on the development of common MFD. The results of genetic testing of common MFD will help suggest preventive strategy options related to lifestyle, drug therapy options and timely monitoring (such as determination of biochemical parameters in the blood, organ function, etc.)[49].

Heritability and the Implications for genetic studies

For MFDs with genetic and exogenous risk factors, the genetic component of a family member with a specific disease with an established phenotype plays an important role in determining the individual risk of the disease, taking into account the proband's genotype. Family history is a motivating factor in predicting the risk of diseases with pronounced heredity and considering family components of the environment. In a family model, the critical question is how accurate the genotype is in predicting risk and its association with the disease, as well as its contribution to the disease, taking into account the family history[50].

Heritability is defined as a trait depending on the influence of changes in genetic factors on them and reflects the proportion of phenotypic variability[51]. However, there are different definitions of the interpretation of heritability in the literature. Often, heritability is interpreted as a genotypic factor. In fact, heritability is a population parameter dependent on factors such as allele frequencies, the effects of

gene variants, and variation caused by environmental factors[52,53]. It should be noted that heritability has a relative contribution to the population. This suggests that if the exogenous factor significantly increases the disease risk, then the relative genetic contribution will be less[54]. To determine the genotypic value of the progeny, it is necessary to determine the influence of a particular allele in a population on a trait. Thus, in some studies, the average effect of the allele and the dependence of the trait on the frequency of the allele in the population was determined[55]. The sum of alleles' average effects is represented as individuals' additive genetic value, called the phenotypic value. Individuals in a population have different phenotypic traits due to environmental and genetic factors and their unequal interactions. Except for identical genetic factors (for example, monozygotic twins (MZ) in humans), humans in a population differ in genotypes, genes located at loci that affect certain features. All locus effects combined, with putative allelic engagement with and among loci, is genotypic importance. The importance builds genetic variability in a population by varying among individuals[52].

To assess heritability, both genetic factors and the environment and their influence on the phenotype are used. This is mainly due to the expected similarity between relatives and experimental data. Usually, heritability is assessed by the scheme of correlation of progeny and parental phenotypes and the determination of the ratio of MZ and dizygotic (DZ) pairs of twins[52]. Thus, heritability is assessed by conducting family or twin studies[56]. Currently, genome-wide association studies (GWAS) are used in heredity analysis[54]. The genetic contribution of GWAS is used in research both in physical characteristics[57] and various diseases[58].

Family studies rely on additive genetic factors. Additive genetic traits are passed from parents to progeny. In a twin study, similarities are assumed between MZ and DZ pairs of twins due to similar shared environments. Within the family, there are differences in identity, and differences can be determined using genetic markers. This difference occurs during chromosomal divergence in meiosis. The average sibling identity is 50%, with a standard deviation of 4%. Therefore, some full siblings share 40% or 60% of their genome by descent. Siblings with an above-average common genome are phenotypically similar compared to siblings with a smaller common genome, where the similarity is determined by a genetic factor[59].

In family analysis, multiple diseases show a cluster pattern, which is one of the essential factors in predicting risk for clinicians. Family associations are determined using family relative risks (FRRs). FRR refers to the disease risk of an affected family member compared to the risk in the general population. Carrying out a family history of the disease helps identify a predisposition to diseases[60]. Relatives are at higher risk of disease because of genetic, epigenetic, common environmental factors, or a combination of these[61]. An example would be a *BRCA* mutation in the family, which greatly increases the risk of breast cancer. However, this concerns only a limited number of cancer cases[62]. With unknown causes of diseases and the absence of an obvious external factor, research using family aggregation helps to determine the risk of disease[63]. Family associations are characteristic of many MFDs[64]. For example, if a first-degree relative has diseases such as breast, colon, or prostate cancer, then his family members have a double risk of such diseases[64-67]. When several family members are involved, the risk increases many times over. The same familial pattern is observed in some autoimmune and neurodegenerative diseases[68,69]. Atopic dermatitis is characterised by an *FLG* mutation; risk in family studies increases the incidence by about two times, and in case-control studies by almost five times[70]. The twin method allows studying of the causes of genetic and environmental variability of MFDs. The basis of the method lies in the fact that MZ twins have similar genes and environments, while DZ twins have 50% of the common segregating genes. The phenotypic differences in MZ twins are due to differences in the environment, while in DZ twins, the differences are due to differences in genetics and the environment. Given this, MZ twins have greater similarities in MFDs comparing to DZ twins. However, when analysing the results of a study of twins, it is necessary to use the phenomenon of epigenetics and transcriptional dynamic changes. These factors alter gene expression, which demonstrates phenotypic differences in both DZ and MZ twins. Asthma is an example of a complex aspect of phenotypic manifestation[71]. For example, the Danish twin study found an association between age and the early stage of asthma that was more in MZ twins compared to DZ twins. The heritability of the age in the early stage of asthma was 35%. These results showed that the degree of genetic relationship prevails over the onset of the disease. However, a genetic predisposition to disease with early onset disease in one twin greatly increases the risk of disease in the other twin. In the late development of asthma, status asthmaticus has less dependency on the twin's asthmatic features[72]. Considering the clinical data and the

symptomatic severity of MFD, a more significant association was found between MZ, in contrast to DZ. The study's results revealed that the genetic factor accounted for 24% of the variations in the severity of the symptoms of the disease, and 76% was environmental[73]. This suggests that random variations are caused by specific environmental influences[71].

Thus, comparing the similarity between two types of twins helps to assess the genetic and environmental influences. This helps to evaluate differences in genetic contributions. Because any similarity in MZ twins that is not due to genetic contribution is due to environmental effects. Therefore, a similar environment and equal conditions are important for assessing genetic contribution.

Another type of twin research is when there is a clear separation of gene and environmental influences: twins reared apart (TRA). This study includes MZ twins who were separated and raised in different environments. This type of study helps to understand the importance of genetic contribution and the influence of an external factor. Since the upbringing was different, the external factor disappears as a cause of similarity. This method helps to assess heritability directly.

There are limitations in the twin methodology, which is the Equal Conditions Assumption (EEA). Its essence lies in the similarity of the environment in which the MZ and DZ are located. With the same environment for finding twins, the only difference will be genetic similarity. Since MZ twins have a similar environment, the behaviour factor depends not on the genetic factor but on the environment[74]. The same pattern can be observed with MFD. Therefore, in this research method, the most important is the study of twins and the influence of an external factor.

In studies with different samples, a relative risk (RR) indicator is used. This indicator is used to determine the frequency of occurrence of diseases in individuals exposed to a risk factor in relation to people who have different alleles of the similar candidate genes[75]. In some studies, for known marker genes associated with MFD, the RR value does not exceed 1.5; more often, it is in the range of 1.16–1.2. Thus, unfavourable allelic variants of the *PTPN22* candidate gene in diabetes and systemic lupus erythematosus and the *NOD2* gene in Crohn's disease increase the RR of the disease only 2–3 times. The majority of other associations RR between 1.1 and 1.5 [75,76].

There is also the concept of polygenic risk. The Polygenic Risk Score (PRS) uses only genomic information, which estimates an individual's likelihood of having or acquiring a specific disease. Without

consideration of environmental or other factors, the human PRS is a statistical calculation that relies on the existence or lack of numerous variations in the genome. A high PRS indicates a high risk. An important detail for the analysis of PRS is the choice of the phenotype or outcome under trait. Reproducing a PRS requires a list of SNPs, an effect allele for each SNP, an effect size for each SNP, and a genome build. The PRS is useful for checking and reproducing a relationship concerning some feature[77].

Genetic factors have a significant influence on a wide range of diseases and diagnoses. However, in many cases, the causative variants are difficult to identify. This lack of causation has been termed "missing heritability"[53]. At the same time, many studies have shown that about 50% of coronary heart disease risk and 80% of autism risk are associated with a genetic contribution[23,78,79]. To identify the influence of genetics on the phenotype, it is necessary to analyse the heritability of the progeny, as well as the same phenotype as their parents. Conducted family studies have determined the existence of a gap between the alleged heredity and genetic variants. The presence of phantom heredity was determined, the cause of which is epistasis[80].

Thus, any disease must include different levels of genetic polymorphism and their complex interactions. In this case, the cause of the disease can be not only mutations but also unfavourable combinations of polymorphic variants of the genome, as well as errors in the regulation of gene activity and the interaction of their products[81].

It should be noted that genetic predisposition to diseases is essential to investigate the totality of environmental risk factors. Studies have shown the role of gene-gene and gene-environment engagements in the heritability[82,83]. For example, different gene variants play an essential part in the formation of obesity depending on external factors (for example, nicotine withdrawal, pregnancy, stress and antidepressant use, etc.)[81,84].

Copy number variations as a factor in forming the phenotypic variability

Understanding the human genome variations has become one of the most critical aspects of medical genetics. For the further development of medical genetics, copy number variation (CNV) is being studied intensively since CNV can considerably contribute to understanding phenotypic variability and the occurrence of human diseases. It is known that aneuploidy of chromosomes, a change in the copy numbers (CN) of all genes located on a deleted or extra chromosome, has critical consequences for the phenotype.

In 2004, papers were published in which a further source of genome variability was studied — variations in the CN of DNA segments [85,86].

CNV is a type of genetic polymorphism in which an individual genome differs in the number of copies of chromosomal segments longer than 1kb[86]. Structural types of CNV include deletions, insertions, duplications, translocations, and complex multi-site variation[87]. According to different authors, they can occupy from 4.8% to 9.5% of the autosomal genome of a healthy person[88]. According to McCarroll et al., approximately 80% of the observed CNVs between any two individuals occur with a frequency of>5%, where 99% of them are inherited [89]. The human genome may contain 1,000 or more CNVs longer than 50 kb, affecting a total of about 4 million base pairs[90,91]. Thus, CNV loci, including genes, affect the function and expression of genes, which can subsequently lead to a phenotype important from both evolutionary and medical points of view [92-94].

CNV has a quantitative characteristic involving the gain or loss of genetic material. They help designate CNV-polymorphism, changes in the number of copies that do not exceed twofold relative to the reference genome. When overcoming this relative threshold, CNVs are referred to as “amplification” or “deletion” and as “CNV alteration” (CNA)[87]. Sebat et al. suggested that around half of the CNV in the human genome consists both whole and part of the genes, thereby changing both the structure and the gene dosage, that is, the quantitative content of gene copies in the genome under study[86,95].

The location of CNV in the genome can be very diverse. Most copy variable genes are implicated in the control of cell adhesion, in the processes of olfactory reception, in neurophysiological processes, in the

implementation of the body's immune defence and others[96]. Approximately 15% of CNVs usually have a significant clinical effect, which also includes changes in the structure and dosage of genes, gene dysregulation and the expression of recessive alleles[97,98].

The consequence of the gene dose effect is the genetic disease Charcot-Marie-Tooth disease (CMT), with an autosomal dominant pattern of inheritance[99]. Duplication of a large region on chromosome 17p12 of the *PMP22* gene is a common cause of CMT (approximately 70-80% of cases)[100,101]. Individual mutations affect the *MFN2* gene. Some sets of genes like *MFN2* are found in the nuclear genome, but their products are active in mitochondria. The mutated *MFN2* gene causes the formation of a large mitochondrial cluster or bundle that cannot travel down the axon to the synapses, which in turn disrupts their functionality[102].

There is another concept called “genomic sister disorders”; these are opposite structural chromosomal aberrations, which occur as diseases caused by duplications as opposed to deletions in the exact region of the genome. Genomic sister disorders have information about the genomic development of diseases. These reciprocal CNVs can lead to diametrically opposed changes in gene expression or transcription processes[103]. For example, in cells of stomach adenoma and adenocarcinoma of the intestinal type, identical disorders were found in the chromosomal regions 17p and 20q13. However, such aberrations were detected only in tissues taken from the same patient, while not detected in tissues from different individuals of this pattern[104]. In addition, tumour tissues have a higher level of chromosomal abnormalities, unlike pre-tumour changes.

Failure to recognise a molecular duplication has been shown to lead to misinterpretation of marker genes in affected individuals, as well as misidentification of false recombinants and inaccurate localisation of the genetic cause of the disease[99].

The determination of CNV action is influenced by clinical variability, low penetrance, and variable gene expression[105, 106]. The study of CNV in the examined families with established diagnoses showed that a significant part of the variations manifests phenotypically. It is possible to interpret the variable severity of genomic disorders and phenotypic polymorphism. First, the length of a deletion or duplication may vary when several dose-dependent genes are involved[98]. Examples can be Smith-Magenis[107], Potoki–Loopski[108], Williams-Beuren syndrome[109], and microdeletion syndrome 15q24.3[110]. Mapping of such CNVs allows the

identification of candidate genes linked with various diseases, where examples are the *JAG1* gene in Alagille syndrome[111], the *RAI1* gene in Smith – Magenis syndrome[112] and the *LIS1* gene in Miller–Dieker syndrome[113].

Previous studies showed that CNV is associated only with highly penetrant effects in genomic and psychiatric diseases. For example, for diseases like autism, microdeletions and duplications can be observed instead of 100% Mendelian penetrance. This may cause abnormalities with high penetrance that present as severe phenotypic manifestations of alleles across entire populations[114,115]. Further research from scientists has shown a link to multifactorial diseases, such as psoriasis (*DEFB* genes) and sporadic prostate cancer (*DEFB* genes)[116-118].

CNVs may be one of the reasons for sporadic disorders caused by genomic changes[119] and hereditary diseases in the form of monogenic diseases[100].

Genetic methods of studying copy number variations in the genome.

The human genome includes a substantial amount of submicroscopic variation in the copy number of DNA sites in chromosomal segments[85,86].

Using various methods, it is possible to identify CNV in mammals and compare different populations of people. CNV is a large group of structural variations with a wide range of dimensions (from one thousand base pairs to one million), affecting a large number of genes and causing several phenotypic manifestations. During cell division, it was determined that the frequency of formation of CNV is higher by a few orders of magnitude than SNP[120]. CNV comparing SNP has higher mutation rate[121]. Lynch et al., in their study, showed that the mutation CNV rate of somatic cells during the cell division is 4-25 times higher comparing the germline[122].

Several studies have examined the relative contribution to the phenotype of SNP and CNV. Thus, in a study by Stranger et al., they compared SNP data with CNV. As a result, the minimal overlap between the contribution of CNV and SNP was determined (from 8.75% to 17.7%) to explain the variation in gene expression; 92.5% to 83.6% was

explained by SNP[123]. However, Zhang et al. consider the overall contribution of CNV to level expression are insufficiently understood, and additional research is required to show the contribution of CNV to the human phenotype[121].

In 2015, as a result of research, CNV of the *CNTN6* gene was determined in 3724 patients with clinical features such as bipolar disorders, behavioural anomalies, developmental delay and intellectual disabilities, as well as with autism spectrum diseases and epileptiform seizures. A total of 14 CNVs have been identified that affect the *CNTN6* gene. Seven of the subjects had complete or intragenic deletion. Intragenic duplication was diagnosed in 5 subjects, and duplication of the *CHL1/CNTN6* and *CHL1/CNTN6/CNTN4* genes was detected in 2. The results showed that the incidence of CNV in *CNTN6* was 4:1000 (0.4%)[124,125].

Mapping of CNVs showed that there are "Copy number variable regions" (CNVR) containing large groups of genes that are mainly involved in interaction with the environment, for example, in the immune response - genes encoding defensins and immunoglobulins[126-129].

In 2006 Redon et al.'s studies investigated genome mapping in 270 participants from different populations. They discovered new types of variation at 1500 CNV sites using array-CGH methods. The variants differed in length and included deletions and duplications. The length of DNA involved was up to 12% of the entire genome. It was suggested that these areas might contain numerous amount of genes that may be associated with various human diseases. Additionally, it was suggested that mutations could either be inherited or occur *de novo* [96].

Modern molecular research methods to detect CNV across the genome are divided into two main groups: technologies using microarrays and DNA sequencing technologies (Next Generation Sequencing, NGS). NGS is based on the simultaneous parallel sequencing of millions of short DNA fragments, making it possible to identify variations and mutations across the whole genome or its individual parts. In clinical practice, two methods of NGS are used: targeted genetic diagnostics and full-exome or whole-genome sequencing. Targeted diagnostics undertakes a study looking for specific genetic variations, DNA polymorphisms or mutations at particular locations. Its primary advantage is the ability to scan the types of mutations across numerous candidate genes in each patient (for example, in studying the Ehlers–Danlos syndrome) separately. The latter circumstance is significant given the pronounced clinical

polymorphism[130]. Genomic sequencing conducts a simultaneous analysis of all known genes in the human genome. It is used to diagnose genetic diseases with a non-specific clinical picture (non-syndromic) or to identify mutations in several genes[131,132]. Direct exon sequencing is relevant for genetic diagnosis where only the coding sequence for protein is sequenced, which is approximately 8 million nucleotides compared to the total genome (3.2 billion nucleotides). This makes it possible to identify mutations in unknown genes, which is essential in the detection of rare hereditary diseases[132,133].

Significant progress has been made in the technological ability to test humans for genetic variation[134]. High-density SNP arrays can be used to identify structural variation additionally to other genomic signs. These additional functions of GWAS involve analysis of population structure and ancestry, analysis of linkage disequilibrium (LD), and assessment of the heritability of features. LD patterns interpret association studies. A method based on the degree of LD identifies SNPs having an association with one or more quantitative trait loci (QTLs)[135]. SNP arrays can also detect copy number changes, especially in cancer research. SNP microarray analysis helps to detect the loss of heterozygosity and determine the change in the number of copies of the genome in various types of cancer. The technology can be used from SNP array data to study allelic association in cancer and to identify cancer predisposition genes, oncogenes and tumour suppressor genes in certain types of tumours[136]. Appropriate algorithms for determining the number of copies allow the detection of CNV in germline DNA[137]. SNP arrays perform genotyping with >99% accuracy on about a million human SNPs in a single analysis[136]. As a consequence, they have the potential for MFD risk assessment, diagnosis, prognosis, and treatment selection[138].

Most CNVs are phenotypically neutral. For MFDs, which are based on a complex genetic structure with different frequencies and effects of allele combinations, the determination of CNV is important. When analysing CNV-associated diseases, it is important to isolate pathogenically significant CNVs to determine the risk of disease manifestation. However, the mechanisms of clinical severity of CNV and low penetrance remain poorly understood[139].

The role of the single nucleotide polymorphisms spectrum in the development of multifactorial diseases.

Exploring the molecular genetic mechanisms underlying MFD will largely determine the improvement of new strategies for the prevention and treatment of pathological conditions[140].

Developing high-throughput methods for genome analysis has led to the creation and use of panels based on SNPs (Single Nucleotide Polymorphisms). Despite the exceptional importance of functional areas, they are not identical to a certain extent in different people, resulting in various phenotypic manifestations, hereditary diseases, and various genetically determined individual predispositions. These differences are mainly due to variations in SNPs. SNPs are considered substitutions that occur in the observed population with a frequency of at least 1%[141]. SNPs are broadly distributed in the human genome. The occurrence of SNP is approximately once every 300 base pairs. On the order of 3 million or more is suggested as the total number of relatively common SNPs[142]. The SNP itself is the result of point mutations. A genome-wide search for links correlates a specific genome variant, characterised by a unique set of SNPs, with the presence of disease, as an analysis of genetic predisposition to diseases and corresponding prediction.

With this frequency, SNP best meets all the requirements that apply to markers for large-scale linkage studies[141]. SNPs are one of the many forms of genome variation broadly used to investigate genetic and non-genetic diseases. SNPs are involved in the transformation of the genomic sequence of the coding (exons), intergenic or non-coding (introns) region[143-145]. Ensuring high confidence in individual identification based on SNP requires more loci than microsatellites STR (short tandem repeats). In this context, getting the same power as one short STR multiplex is comparable to 50 loci. At the same time, arrays with two alleles are more useful in distinguishing related individuals[146].

STRs were the main type of markers used for linkage analysis, as they are evenly distributed throughout the genome. However, SNP markers reflect the strongest influence of LD on linkage results, i.e. significantly increase linkage information, since they are only diallelic. For SNP, the average information content is 61%; for microsatellite markers, the average information content is 41%[147]. Although in other studies: 61% for SNP compared to a mean of 84% for STR reported by

Middleton et al. [148]. The difference between SNPs and STRs is the greater number and density of SNPs. The advantage of STRs is that they are polymorphic and much larger than diallelic SNPs. The advantage of SNP is that the SNP map is denser and potentially more informative than the microsatellite map[149]. SNPs are much larger and create haplotypes that contain more linkage information and function as "super" alleles[147].

In some cases, the SNP may be the mutation itself, causing a hereditary disease[141]. The use as markers mainly of SNPs expressed as single nucleotide substitutions in the DNA sequence helps in assessing the genetic components of predisposition to MFD.

At the same time, there is still a significant gap in the study of the genetic components of MFDs, since the work carried out in this direction is isolated and, to some extent, has ambiguous results. Presumably, this may be due to population characteristics, heterogeneous study design, and statistical approaches used in the meta-analysis of the results. Previously, it was assumed that SNPs greatly contribute to genome variability. However, it is now recognised that structural variations have a powerful, at the same time, poorly understood, contribution to human genetic diversity.

The search and identification of SNPs based on data analysis make it possible to obtain a genetic set of polymorphisms and can determine the association between differential expression patterns and phenotypic variability[145,150]. The presence or absence of specific SNP markers in the patient's genome is associated with the human's health, longevity and quality of life[151]. Therefore, it remains relevant to ensure the possibility of evaluating the manifestation of each of the 10 billion human SNPs in practice[142]. SNPs have shown an association with some forms of cancer[152]. It should be noted that the accuracy of bioinformatic predictions for SNPs is still below the threshold of their applicability in clinical practice[153].

Most SNP genotyping methods can be grouped into four groups relying on their molecular mechanism of action: allele-specific hybridisation, primer extension, oligonucleotide ligation, and restriction enzyme analysis[154]. Studies of candidate SNP markers make it possible to determine clinical data on the physiological signs of human pathologies that correspond to changes in gene expression.

Significant differences in the frequency of specific polymorphisms between two groups (case-control) suggest that the candidate gene is essential in determining disease susceptibility[155]. Suppose the association of the allele with the disease is found. In that case, the interpretation of

the result obtained is either not predisposing to the disease or is predisposing to the disease. There are several interpretations of the results obtained in the literature[156]:

- The associated allele is or, conversely, is not a predisposing allele for diseases.
- The associated allele is in LD with the specific locus that leads to the disease.
- The correlation demonstrates the characteristics of samples due to the division by race, ethnicity, gender, etc.
- The apparent association is a chance occurrence or an artefact (possibly statistical).

Through research and significant projects involving SNPs, in-depth analysis of each individual's genome is possible, as well as efficient mapping of genes and allelic variants[157]. All of these studies provide information about the specific location and association of an individual SNP in the human genome with phenotypic features.

Association study designs in genetic research.

Earlier methods for analysing populations using genetic studies have been based primarily on the analysis of linked inheritance in families. Before the structures and mechanisms of DNA were described, Mendelian research patterns were used to define a genetic disorder[158]. As practice has shown, they turned out to be ineffective in identifying the causes of complex diseases. Therefore, after decoding the human genome, the “statistical” method became an alternative to the “family method”. There are several large-scale genetic studies where the search for disease-associated factors in the human genome was carried out. One of them is GWAS.

Direct association research methods are associated with specific phenotypes, processing the genotype of each participant originating from an exact genetic location. In contrast to these techniques, GWAS examines genotypes across the entire genome for association with different phenotypes[158]. It is supposed to use the information about the detected genetic associations to develop more effective strategies for treating and preventing complex diseases.

GWAS is a method used to study the genetic predisposition to polygene-based complex diseases, involving the rapid scanning of the entire set of DNA or the genome of many people for markers to look for alleles associated with a particular disease, that is, this is the direction of genetic research, which establishes a relationship between any traits and genetic markers. GWAS uses the definition of single loci, thereby examining the association of each SNP and CNV with clinical characteristics. This method has proved helpful in finding a large number of genetic variants associated with human characteristics[11,159,160].

The goal of association analysis is to search for specific variants that contribute to the disease. A positive association between a disease and even a single allele within a LD block immediately identifies the genome region containing the linked disease allele. Therefore, this region will be the place to look for an allelic variant that is functionally involved in the disease process. Based on the search for LD to reduce the number of polymorphic alleles used in association analysis, this strategy has become the primary motivation for creating the International "Haploid Genome Project" (HapMap) and its subsequent refinement in more comprehensive data sets like the 1000 Genomes Project sequences or the aggregated resources of gnomAD[161-163].

Another resource contributing to the success of the GWAS study is HapMap. The LD blocks defined by the HapMap project enable a full-scale genomic association test that helps to detect significant associations across the genome. The HapMap discovered and validated LD blocks, including several million SNPs. The project aimed to draw up a genetic map considering the identification of SNP. The essence of the project was that the found SNPs are inherited in blocks, which are a set of alleles of several loci located on the same chromosome and act as associated molecular markers. By finding an association between SNP and phenotype, the most precise location of candidate genes and mutated polymorphisms associated with MFDs can be identified. Such an SNP block is a haplotype — a set of alleles of several loci located on the same chromosome. Knowledge of these blocks also informs SNP genotyping array design so that each of the mapped SNPs acts as an independent molecular marker. By finding the association of such SNP markers with the studied trait (disease, symptom), the most probable localisation sites of candidate genes, mutations (polymorphisms) of which are associated with one or another MFD, are determined[158,161,164-166].

The main criteria for the applicability of GWAS are the presence of a representative sample (usually with a large number of participants)

and the possibility of identifying a relationship (association) between the genotype and phenotype.

The principles of work in the GWAS are:

1. Scanning full-genome DNA markers in many people's genomes to look for genetic variations associated with a particular disease.

2. Identification of new genetic associations.

3. Using the information to develop the best strategies for detecting, treating and preventing the disease.

The GWAS process consists of the following steps:

1. Collection of phenotypic information from numerous individuals.

2. Extraction of DNA and genotype of at least 500,000 SNPs; typing 500,000 SNP is expected to yield 25,000 outcomes at $p \leq 0.05$ at random.

3. In identifying the association and calling the genotype through the use of various programs.

4. Analysis of results and identification of at least one signal associated with a p-value that is less than $P<10^{-8}$ [159]. Only results identified and replicated in a duplicate cohort are considered significantly positive (Figure 3). A database GWAS Central is in the public domain[167], the complete collection of well-known phenotypic associations of SNPs. To visualise the data in GWAS, the so-called Manhattan plots are used, where the degree of association is displayed as a negative logarithm of the p-value depending on the locus, that is, the genomic coordinates of the snapshot (Figure 4)[168].

Thus, the GWAS is based on the search for genetic variants. In this connection, the principle of its work is to compare the frequencies of alleles of common genetic variants between cases and controls and the search for these variants.

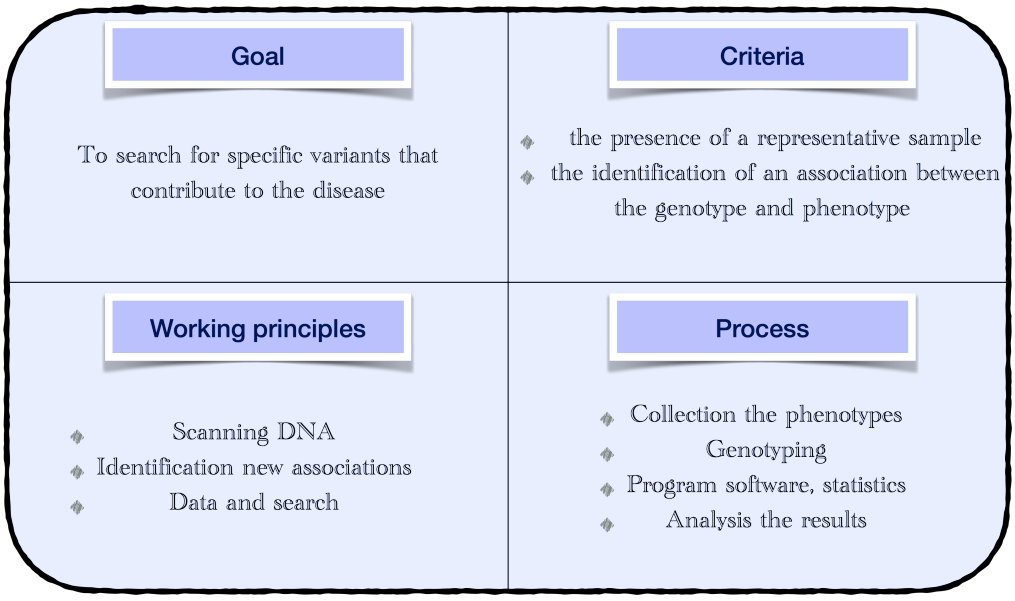


Figure 3. GWAS working principles.

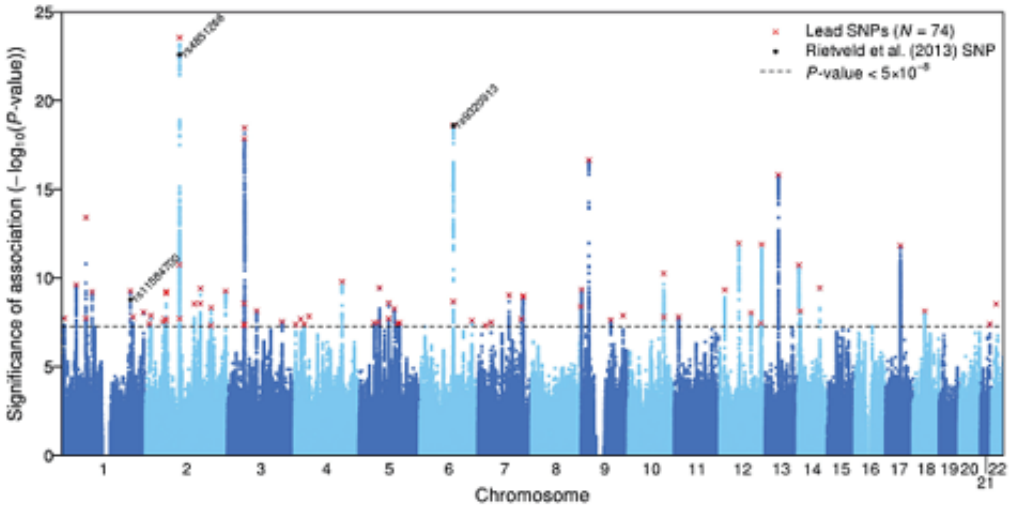


Figure 4. Example of the Manhattan plot that is used in the GWAS study. The chromosomal position is shown on the x-axis, and the indicator on the scale -log10 is shown on the y-axis. The genome-wide significance level (5×10^{-8}) is represented by the black line. Independent genome-wide significant connections (known as "lead SNPs") are indicated by reds x's. The rs numbers of the SNPs are written on the black dots. Extracted from Okbay, 2016[168].

Genetic association studies are carried out using two fundamentally different approaches. One of them is based on the fact that an allelic variant of a specific target gene, the protein product of which is included in the pathogenesis of the disease under study, acts as a genetic marker. That is, the choice of a gene and its polymorphisms is determined based on a specific hypothesis for the development of the disease, and this gene is called a "candidate gene" for susceptibility to MFDs. The search for candidate MFD genes was carried out in association and linkage analyses.

When selecting candidate genes for research, the requirement is the involvement of this gene in the mechanism of development of the studied MFD. For many MFDs, candidate genes have already been found, the involvement of which in a specific pathology has been confirmed by the studies. For example, Alzheimer's disease (*APO ε 4*)[46], DM2 (*PPARG, TCF7L2, KCNJ11, MTNR1B, SLC30A8*)[169-171], macular degeneration (*CFH*)[172], DM1 (*IL2RA, CD25, PTPN22, IFIH1*)[173-176], autoimmune thyroiditis (*CTLA4*)[177], Hirschsprung's disease (*RET*)[178], Crohn's disease (*NOD2/CARD15*)[179,180], rheumatoid arthritis (*PTPN22*)[181], obesity (*SH2B1, NPC1, ADCY3, FTO*)[182-186], and inflammatory bowel disease (*TNFRSF6B*, *PRDM1, CARD9, IL23R, RNF186*)[187-189]. By using gene sequencing, a hypothesis was proven that the genes identified by GWAS might be suitable as candidate genes for detecting mutations in hereditary diseases and MFDs[81,190].

Disorders occurring in some candidate genes show low penetrance and do not match the expected heritability. The GWAS approach is thought to be the most efficient method for discovering candidate genes. It is supplemented by sequencing of linked loci and the study of the expression of its constituent genes[191]. The GWAS tests a large number of SNPs to identify the risk of disease. In identifying candidate genes, the combined effect size of all significant SNPs is about 20% of estimates of heritability in a population using additive models[192,193].

First, the basis of the association method is the search for molecular markers closely associated with MFD. Based on the pathophysiology of MFD, the allele frequency of the putative candidate gene variant in patients and the control group is compared. The study aims to obtain evidence of its primary product's involvement in the disease's development and to measure (estimate) the strength of its effect. The long time and high cost of the method are insufficient to identify allelic differences in the pathogenesis of the disease. Some

important genes may be missing, and clinically different disease forms may have different candidate gene patterns.

Second, the linkage method is based solely on positional mapping of the locus without predicting the pathophysiology of the disease. This effectively analyses families with several siblings with disease and the presence of extensive pedigrees. It aims to identify molecular markers in patients transmitted from parents to diseased offspring. All this makes it possible to reveal the linkage of MFDs with a sufficiently extended (1–10 Mb) chromosome region and identify the causative gene, which is sometimes an unsolvable task[75,194]. The "candidate" approach was the first variant of work in genetic research. Due to that, the hypothesis of the pathogenetic basis of the possible connection between the genes and disease is tested[157,195].

The key difference between association and linkage analyses differs in the approach to conducting the methods themselves. An association analysis looks at the link between a particular allele and the disease or feature in a population. At the same time, linkage analysis aspects consider the link of a locus with a disease or a feature within a family[196].

Comparison of association and linkage methods for disease gene mapping.

Table 1.

Genetic association analysis	Genetic linkage analysis
✓ Association analysis establishes a link between the specific allele that functionally affects the disease. ✓ Controls the altered frequencies of specific alleles or haplotypes in patients compared to controls in a population. ✓ Identifies variants of alleles that can functionally influence the development of the disease; often determines the haplotype associated with the disease. ✓ Based on the search for a set of alleles, including a	✓ Determines the transfer of the locus from one pedigree member to another within the family and relationship of the disease. ✓ Searches for regions of the genome containing disease alleles; uses polymorphic variants only to mark some region of the genome inherited from parents. ✓ Not designed to search for the specific variants responsible for disease predisposition. Defines the location of the

predisposing variant passed through many generations due to the small number of recombination events between markers. ✓ Can be performed both in case-control studies and population samples. ✓ Sensitive to stratification of the population and compensated for by the appropriate selection of a control group or use of methods based on family analysis. ✓ Useful in studies with low effect size and common allele frequency predisposing to complex traits.	variant (usually within one or more megabases). ✓ Based on recombination events occurring in families over only a few generations, which measures the genetic distance between the disease gene and polymorphic markers on the chromosomes. ✓ Requires a survey of the whole family (trio). ✓ Primarily used to map pathological mutations with sufficiently pronounced effects. ✓ In linkage analysis, there is no problem of population heterogeneity since relatives have the exact ethnic origin.

The GWAS is used to find genetic factors related to the risk of disease development, clinical phenotypes and the effectiveness of treating these diseases. The study compares the genome of a group of people with a given trait or disease with a control group. There is an opinion that GWAS is one of the primary methods for determining genes related to certain of human diseases. GWAS analysis can define the approach to studying the basis of genetic disorders. These studies have recommended careful quality control in GWAS trials[197,198]. GWAS can provide information about the phenotype, genotype, and causes of genetic diseases, which can help further treatment and prevention[199].

Thus, complex genetic tests based on GWAS studies make it possible to predict the risks of developing common MFDs. This can lead to the following:

- Development of a program of preventive medicine for people with a genetic predisposition to the occurrence of a particular disease, especially MFD.
- Assessing the risk of MFD, considering the obtained genetic markers, and conducting pathogenetic treatment.

- Development of drug therapies based on the understanding of gene products involved in pathology ("drug targets")
- Formation of the assumption of the construction of pathology, considering the genetic risks in various diseases and determining their complications.

GWAS is impressive in finding a large number of risk loci, thus proving the importance of understanding the genetic basis of MFDs. In particular, this is shown by association and linkage analysis research, as well as the crucial role of candidate genes. However, only a small part of the genetic variability and their architecture are explained by the identified loci. Associations with non-coding variants make interpreting the study results as causative genes difficult. This makes it challenging to explain their molecular mechanism since the genes of non-coding proteins do not undergo a transcription process. Because of this, some find GWAS controversial in terms of identifying causal genes[23,200]. An early design of the GWAS focused on identifying haplotype-labelled SNPs with the findings of causal variants based on gene loci. Subsequent GWAS studies began using SNP arrays to capture a wide range of genomic variation. This approach revealed additional risk genes. However, the use of chips of even very high density does not guarantee the detection of all SNPs associated with the disease[53].

Overall, GWAS limitations are:

1. The first limitation is the determination of the level of significance in genetic testing. The GWAS method does not allow identifying the main marker genes associated with MFD if the frequency of their rare alleles does not exceed 0.5%. The limitation is exacerbated by the increased use of WGS[201].
2. GWAS explains only a small part of the missing heritability. GWAS SNPs may suggest heritability for only some complex traits. GWAS is unable to identify every genetic factor to complex traits[57,202,203].
3. GWAS does not necessarily identify causal variants and genes.
4. GWAS have mostly failed to identify epistasis in human.
5. SNP array-based GWAS are unable to find extremely rare mutations that cause the disease.
6. GWAS based on SNP arrays cannot detect ultra-rare mutations contributing to the disease.
7. The use of the standard variant of GWAS with a standard sample-free evaluation of the results does not allow one to reveal the metabolic pathways of the pathogenesis of a particular MFD, to

understand the nature of the intergenic interactions underlying it[81].

8. When evaluating the results of GWAS and forming panels of MFD candidate genes, the role of environmental factors in developing the disease is not considered in any way.
9. The use of the standard GWAS variant with a standard non-selective evaluation of the results does not allow one to reveal the metabolic pathways of the pathogenesis of a particular MFDs, to understand the nature of the intergenic interactions underlying it[204,205].

The GWAS methodology made it possible to combine data from different platforms into one, which helped analyse one of the most significant projects. Before a gene is linked to a disease, candidate genes are first identified and then reproduced in independent samples. All this allows for a meta-analysis, increasing the statistical power to identify additional options. Meta-analysis will enable us to find out if there is an association between a disease and its effect on disease risk. To determine the factors underlying an association in GWAS, researchers carry out various methods of analysis, including in vivo and in vitro, as well as the use of NGS, WES and others (Figure 5, central part). These studies help better understand the mechanism of occurrence of various diseases, including MFD, and a personal approach to diagnosing and treating diseases for each person. (Figure 5, lower part)[200].

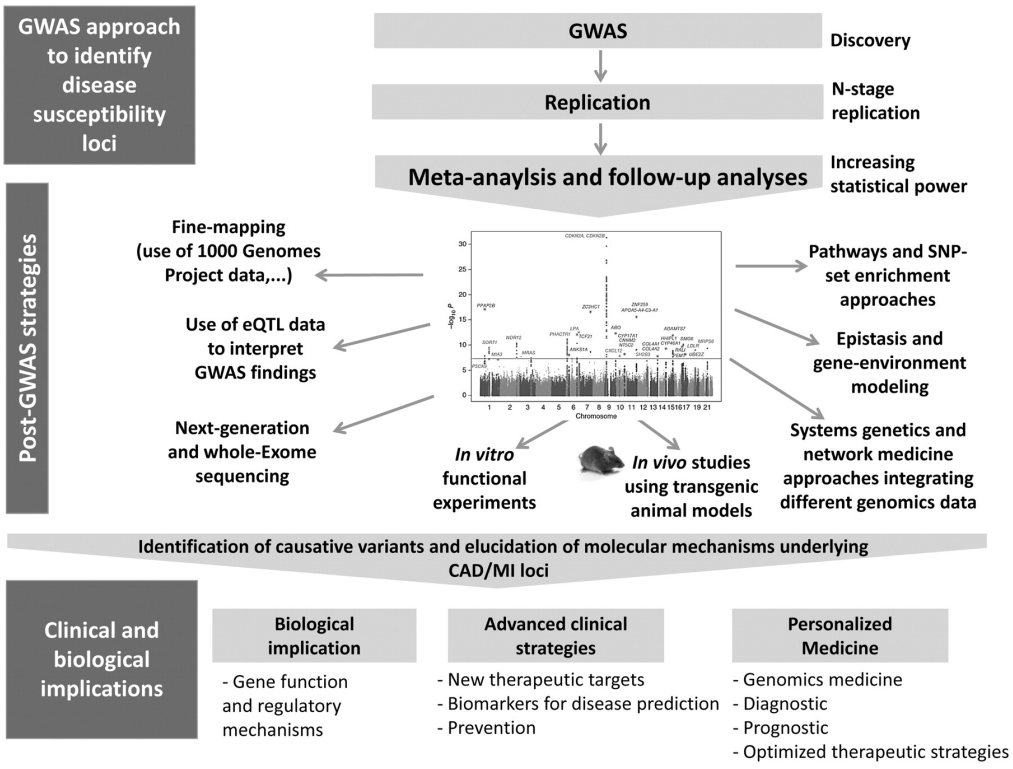


Figure 5. Strategies for postgenomic investigation of GWAS associations with biological and clinical implications of GWAS results. Extracted from Maouche et al. 2012[200].

As an example of a post-GWAS approach to an MFD, the Brains for Dementia Research project was created to study brain tissue from neuropathological samples[206,207]. This project examined 19 analyses of established GWAS index SNPs for Alzheimer's. DNA was collected and isolated from 600 postmortem brains, where dementia was 68.9%. All diagnoses were confirmed by neuropathologists. The cohort consisted of 315 cases and 149 normal controls. For cases, the average age at the time of death was 82.9 (±8.7); for the control group, 83.6 (±8.7) years, the proportion of women (was 49.2% and 47.9%). Both groups did not differ between sex and age. The analysis revealed the importance of the association of rs28834970 (*PTK2B*), rs10792832 (*PICALM*), and rs1476679 (*ZWCPW1*). As expected, a highly significant association of the *APOE ε 4* allele ($P=3.99x10^{-12}$) was found[207].

The WTCCC was founded to study and analyse GWAS research. Its structure consists of more than 50 British research groups. These studies were conducted in various ways to explore aspects of human

disease genetics, clinical experience, genotyping, statistical analysis, and informatics[197,198].

In 2007, the WTCCC described a collaborative study with GWAS of 2,000 British people based on seven diseases, totalling 14,000 cases and 3,000 controls. Comparison of control and cases revealed 24 independent association signals at $p>5x10^{-7}$; one for coronary heart disease, three for rheumatoid arthritis, one for bipolar disorder, and ten in DM, namely type 1 was seven, type 2 was three, nine for Crohn's disease. All 17,000 samples were genotyped and contained 500,568 SNPs. All SNPs were used in the association analysis. The study was conducted for genetic analysis of the disease and to determine the allelic structure. These signals may reflect actual susceptibility effects. It was determined that all the studied diseases have many further signals leading to additional susceptibility at the loci. In the study, the issues of choosing the quality control of genetic studies, improving algorithms in the analysis of genotypes, building analytical methods, and determining the degree of population structure were highlighted[160].

In 2010, a more extensive WTCCC study was carried out. Its essence was to conduct a genome-wide study to determine the association between CNV and the occurrence of common human diseases. Genetic polymorphism is essential in genetic susceptibility to diseases, and its central part is CNV. The study already included 19,000 individuals in separate classes with polymorphic CNVs[208].

The project "Additive Effects of Multiple Genetic Variants on the Risk of Coronary Artery Disease" presents the effect of summing the risks of harmful alleles in the development of cardiovascular disease and genetic risk. The GWAS method within the WTCCC project using the Affymetrix Gene Chip 500K Mapping Array analysed 1988 patients with cardiovascular diseases and 5380 healthy controls. As a result, 18 SNPs associated with the risk of developing cardiovascular diseases were identified. The fluctuations in the relative risk value for SNP risk alleles ranged from 1.12 to 1.47. At the same time, nine SNPs could be associated with lipid metabolism, and nine were independent risk factors for cardiovascular diseases, particularly variants in the *MIA3, WDR12, MRAS, PHACTR1, MTHFD1L* genes, *CDKN2A/2B, CXCL12, SMAD3, SLC5A3*. A linear relationship was established between SNPs and the risk of developing cardiovascular diseases, with an increase in the risk of developing the disease by a factor of 1.18 for each risk allele[160,209].

The analysis showed that using a common control group leads to a practical approach in the study of various phenotypic diseases in the

GWAS analysis, as well as creating a complete genomic information base for further research[160].

GWAS has a case-control design to determine the association between a genotype of interest and DNA variants[210]. The WTCCC case-control comprised 19,050 analyses, where 2,000 for each of the eight diseases and 3,000 total controls were added to 270 HapMap samples and 610 control duplicates. Of the samples used in the 2007 WTCCC SNP GWAS, approximately 80% of the samples were used. Three loci were identified as significantly associated in testing, where *HLA* for Crohn's disease, *HLA* for rheumatoid arthritis and type 1 DM, and *TSPAN8* for type 2 diabetes were all associated with CNV. The identified CNVs are associated with diseases that were previously identified using SNPs. They showed the association and replication of Crohn's disease ($p=1.2x10^{-5}$), type 1 diabetes ($p=8x10^{-153}$), and rheumatoid arthritis ($p=1.4x10^{-39}$). CNV HLAs associated with autoimmune diseases are signals for different haplotypes[208].

A CNV typing array was developed with the Consortium for Genome Structural Variations (GSV) to measure the copy number for compiled CNV stocks[211]. An array was used to identify genotype 3000 controls and 2000 cases for each disease (bipolar disorder, Crohn's disease, coronary heart disease, breast cancer, rheumatoid arthritis, type 1 and type 2 diabetes, and hypertension)[208]. These eight diseases are considered among the major diseases affecting human health[212] and have been closely analysed based on SNP by GWAS and WTCCC[160]. The relevant results have contributed to the methodological analysis of CNV and the use of resources for human genetic research[208].

The results obtained in the course of meta-analyses, combining data from molecular and clinical studies conducted on a case-control basis, were used in medical practice. Since the start made by WTCCC studies, many much more extensive studies have expanded on the foundations provided by this initiative and provided a very large volume of results derived from GWAS.

Due to major projects, DNA samples are widely available in the public database, consistent with clinical and phenotypic data[210,213]. The results entered into the common database have helped with other genetic studies and have been one of the primary sources for the formation of the database[160]. MFD mapping, determination of the main genes and modifier genes in them, and identification of the association of SNPs with diseases will help to develop a set of measures for patients in the field of predictive medicine.

Aims and design of the research

The study's general aim was to develop an improvement in the diagnosis of diseases using genetic methods, namely the analysis of predisposition genes and their relationship with multifactorial diseases.

To form a risk group of patients, biological material (blood) was used with the extraction of human DNA obtained from volunteer patients suffering from various diseases of the pancreas. Modified PRT-based DNA amplification methods were combined with capillary electrophoresis of fluorescently labelled products to identify the CN of the genes. PCR amplification, Sanger sequencing and KASP genotyping were used for genotyping the studied genes. Statistical data processing and calculation methods were carried out using various software (PLINK and GEMMA). Statistical methods were carried out to elucidate the relationship between the values and test the hypothesis about the relative risk of diseases.

This project aimed to establish the relationship between clinical manifestations and genetic prerequisites for the development of diseases, namely the study of the association between amylase genes and pancreatic diseases such as acute and chronic pancreatitis, as well as pancreatic cancer. *AMY1, AMY2A* and *AMY2B* genes were considered genetic markers in patients with pancreatitis. The CNs of *AMY1, AMY2A* and *AMY2B* genes were measured in cases and controls. Also studied were the clinical and genetic features of various forms of pancreatitis.

For the *SPINK1* gene, the study aimed to investigate the frequency of N34S mutation in patients with pancreatic diseases.

In a project on periodontal disease, questionnaire and SNP data from British twins were used as material to study the relationship between dental pathology, namely periodontitis, and human defensin gene polymorphisms.

In this study, we examined the contribution of CNV and SNP to the development of multifactorial diseases, such as diseases of the pancreas and periodontitis. As well as the possibility of considering candidate genes for predisposition to the diseases.

In subsequent chapters, research methods and analysis of the significance of susceptibility genes will be considered using the example of pancreatic diseases and periodontitis.

CHAPTER II . INVESTIGATION ASSOCIATION BETWEEN COPY NUMBER VARIATION OF AMYLASE GENES AND PANCREATIC DISEASES

2.1 Introduction.

Description of the pancreas. Aetiology. Pathogenesis. Factors in the development of pancreatic diseases.

The pathology of the pancreas tends to grow while remaining difficult from the standpoint of diagnosis and interpretation of clinical symptoms, as well as data from laboratory and instrumental research methods. Due to the peculiarity of the anatomical location, the pancreas is quickly involved in any pathological process in the gastrointestinal tract. The degree of this involvement varies, influencing the difficulty of accurately verifying the diagnosis. Among the variety of pancreas diseases, pancreatitis and cancer have the leading places.

Pancreatitis is a polyetiological acute aseptic inflammation of the pancreas, which is based on the production and activation of digestive enzymes in the pancreas, which leads to necrobiosis of pancreatocytes and enzymatic autoaggression with subsequent self-digestion, necrosis and dystrophy of the glandular infection. Therefore, pancreas diseases significantly affect the body as a whole and can lead to severe complications.

The pancreas is a unique human organ, as it is the only gland in the human body that has both essential exocrine and endocrine functions. It produces hormones (insulin, glucagon, somatostatin, pancreatic polypeptide) and a number of highly potent digestive enzymes. The exocrine part of the pancreas is a complex alveolar-tubular gland, divided into segments by thin connective tissue septa. Acinus is the structural and functional unit of the exocrine part, which includes the secretory section and the insertion duct. Acinar cells synthesise and secrete protein secretion into the cavity of the acinus, 98% of which consists of enzymes. The endocrine portion of the gland exists as separate endocrine cells, small clusters of endocrine tissue and islets of Langerhans. One of the complications of pancreatic diseases can be the development of acute and

chronic pancreatitis. Acute and chronic pancreatitis are considered as multifactorial diseases since they are polyaetiological and polypathogenetic diseases.

Despite improvements in diagnostic and treatment procedures, morbidity and mortality from pancreas diseases remain high. The annual incidence rate for pancreatic cancer is approximately 8 cases per 100,000, with a mortality rate of 7 per 100,000; for acute pancreatitis, it is 34 cases per 100,000, with a mortality rate of 2 per 100,000; for chronic pancreatitis, 10 cases per 100,000, with a mortality rate of 0.09 per 100,000 for developed countries. In European countries, the rates of chronic pancreatitis are significantly higher than in acute pancreatitis or pancreatic cancer. However, the frequency of acute pancreatitis is higher in the American region. Southeast Asia is characterised by low mortality from pancreatic cancer. At the same time, in the American region, this mortality is higher than in the European one[214].

Inflammation of the pancreas can have varying clinical manifestations, from mild dysfunction to a severe destructive process. It can also have different outcomes, from the occurrence of reversible changes to irreversible loss of function. In acute pancreatitis, the organ can restore its functions if the main cause of the lesion is identified and eliminated. In chronic pancreatitis, there is a slow and irreversible destruction of the acinar part of the pancreas. In recent decades, acute pancreatitis has remained one of the most urgent conditions in emergency abdominal surgery. The prognosis of pancreatitis relies on the degree of organ failure development and the addition of a secondary infection, which can lead to pancreatic necrosis. About 20% of patients suffer from complications of necrosis of the pancreatic tissue or peripancreatic tissue with acute pancreatitis. This increases the mortality rate up to 20-40%[215]. For example, one of the severe complications of pancreatitis is pancreatic necrosis, which is a dangerous surgical condition. Mortality in pancreatic necrosis can range from 11 to 30%, and in large focal pancreatic necrosis - more than 70%[216-218].

It is believed that various enzymes have a major part in the occurrence of diseases of the pancreas in the self-digestion of the pancreas under the influence of external factors (such as smoking, alcohol, and unhealthy diet)[219]. Among the many aetiological factors that cause pancreatic pathology, the role of heredity is of greater interest. The risk group includes people with a hereditary predisposition, as well as with disorders of lipid and carbohydrate metabolism.

The leading causes of acute pancreatitis are primarily alcohol abuse (the frequency is about 25-35%) and cholelithiasis (with a frequency of about 40-70%)[217,220-223]. Other causes may be hypertriglyceridemia, parasites (roundworms), mechanical trauma to the pancreas (including surgical operations), gastric ulcer and duodenal ulcer (penetration of the gastroduodenal ulcer into the pancreas), pancreatic tumours, hypoparathyroidism, viral infections, severe food allergies, tissue ischemia (atherosclerotic vascular disease), chemotherapy (drug pancreatitis), autoimmune factors, instrumentation via endoscopic retrograde cholangiopancreatography (ERCP), and hypercalcemia. A risk factor such as obesity increases the likelihood of developing severe pancreatitis[222-226]. An additional factor can also be the abuse of fatty and spicy foods, leading to excessive activation of pancreatic enzymes.

Special forms of hereditary pancreatitis are distinguished separately: pancreatitis with hyperaminoaciduria (urinary excretion of amino acids as a result of a genetic tubular defect), cystic fibrosis with an increase in electrolyte excretion with sweat (encoded by *CFTR*), and pancreatitis due to defects in calcium metabolism (encoded by (*CASR*) calcium-sensing receptor gene). They are due to different genetic mechanisms[227] and are primarily associated with smaller risk.

Environmental conditions and genetic constitution can contribute to the occurrence of various diseases in the human body. Despite the enormous influence of external and internal factors on the development of pancreatic diseases, the significance of genetic components for multiple forms of diseases and population frequency is essential in understanding the mechanisms of the development of pancreatitis and cancer. Thus, various pancreas diseases, such as acute and chronic pancreatitis and cancer, can be considered multifactorial diseases. This approach provides a more accurate understanding of the pathogenesis and course of the disease in time in a particular patient. It makes it possible to evaluate the effectiveness of treatment[228].

The role of enzymes and genes, and their influence on developing pancreatic diseases.

Presently, the role of genetic influence on the development of pancreatitis of several genes has been determined, such as mutation of the genes of cationic trypsinogen (*PRSS1*), trypsinogen inhibitor (*SPINK1*), cystic fibrosis transmembrane protein regulator (*CFTR*) and lipase gene (*CEL*).

The studies of Whitcomb et al. (1996) demonstrated the involvement in hereditary pancreatitis of mutation of the gene encoding trypsinogen[229]. Inheritance of specific markers is linked to the development of hereditary pancreatitis. Hereditary pancreatitis association occurs with mutations in the trypsinogen (*PRSS1*) gene on the long arm of chromosome 7q35. In addition, predisposition to pancreatitis may arise due to additional mutations of genes responsible for the synthesis of alcohol dehydrogenase, alpha-1 antitrypsin[230], iron metabolism (in the case of hereditary hemochromatosis) [231], *PPARG-alpha* in lipid metabolism and gamma (affecting the development of diabetes, atherosclerosis, obesity, cancer), *α 1-antitrypsin* gene mutation, *CEL* lipase genes and *CELP* pseudogene.

These mutations are only part of the causes of pancreatitis. In general, genetic defects in pancreatitis can occur due to the dysfunction of pancreatic enzymes. One of the main enzymes is trypsin. The trypsin molecule consists of two subunits connected by a polypeptide chain. The active site is located between the two trypsin subunits. It has the ability to recognise arginine and lysine and carry out lysis of the polypeptide chain at the junction of these amino acids[232]. Trypsin has a special place in the pancreas, as trypsin has an autoactivation activity and can activate other proteolytic enzymes. The principle of autoactivation is that an N-terminal hexapeptide is cleaved from trypsinogen. The enterokinase of the small intestine is the endopeptidase that causes the oligopeptide to be cleaved from trypsinogen. Consequently, trypsinogen is converted into trypsin, which then converts pancreatic proenzymes into active enzymes (with the exception of amylase and lipase). The conversion of trypsinogen to trypsin is normally strictly controlled. With mutations in the gene encoding trypsinogen, excessive formation of this proenzyme is possible, leading to excessive trypsin synthesis, self-digestion of pancreatic cells and pancreatitis[233,234] (Figure 6). One of the examples of mutation is when cationic trypsinogen has the mutation R122H, arginine is replaced with histidine. This mutation is associated with excess enzyme

activity because the replacement of arginine by histidine blocks a site for the inactivation of trypsin by proteolytic cleavage. Another gene, *PRSS1*, can alter the expression of the initial trypsinogen gene, namely, the one encoding cationic trypsinogen, and *PRSS2* encodes anionic trypsinogen. The mutation of N29I and R122H trypsinogen leads to approximately 80% risk of chronic pancreatitis[234]. The mutation of anionic trypsinogen G191R (*PRSS2*) in some studies was presented as a factor protecting against the onset of pancreatitis[235].

The primary function of the trypsinogen inhibitor (*SPINK1*) is to coordinate the activity of trypsin. Thus, *SPINK1*, as an inhibitor of trypsinogen activation, binds serine in trypsin with lysine of its active centre. *SPINK1* is synthesised 20 times less than the amount of trypsinogen produced by the pancreas. In this regard, *SPINK1* can completely inhibit trypsin in the tissue of an organ only when the level of trypsin activity is low. In these cases, *SPINK1* prevents the subsequent autoactivation of trypsinogen and blocks the cascade of pancreatic enzyme activation and pancreas autolysis. With increased trypsinogen activity, *SPINK1* cannot inactivate it. The power of *SPINK1* in these cases is not enough to block the autoactivation of trypsinogen, and the cascades of pancreatic enzyme activation and pancreas autolysis continue[227,234] (Figure 6). With *SPINK1* mutations that reduce its function, the degree of inactivation of trypsin decreases, and when exposed to a powerful, provocative factor, hereditary pancreatitis also develops. Mutation does not lead to enhanced activation of trypsin but only violates one of the protective mechanisms of acinar cells[234,236]. For the onset of the disease and clinical symptoms, initiating factors are needed to trigger this activation.

In separate studies, the *SPINK1* mutation was found in 15% of patients with chronic alcoholic pancreatitis[236,237], whereas the *CFTR* mutation was detected in 22% with pancreas divisum[238]. *SPINK1* gene of N34S mutation occurs in 5-10% of patients with pancreatitis[239], with 1% allele frequency in the general population[240,241]. In heterozygous *CFTR*, the risk of developing pancreatitis by 80-90% in the asymptomatic variant. In homozygous for *CFTR* risk of developing pancreatitis by 20% that have symptoms[242]. The autosomal dominant mutation of cationic trypsinogen (N29I or R122H) approximately occurs only in 75% of patients[243]. *SPINK, CFTR*, and *PRSS1* genes increase the risk of disease in patients of the sporadic nature of the occurrence. Several families without distinct autosomal dominant inheritance were observed to have

the N34S mutation of the *SPINK1* gene and also the *PRSS1* gene variants A16V, N29I and R122H[244].

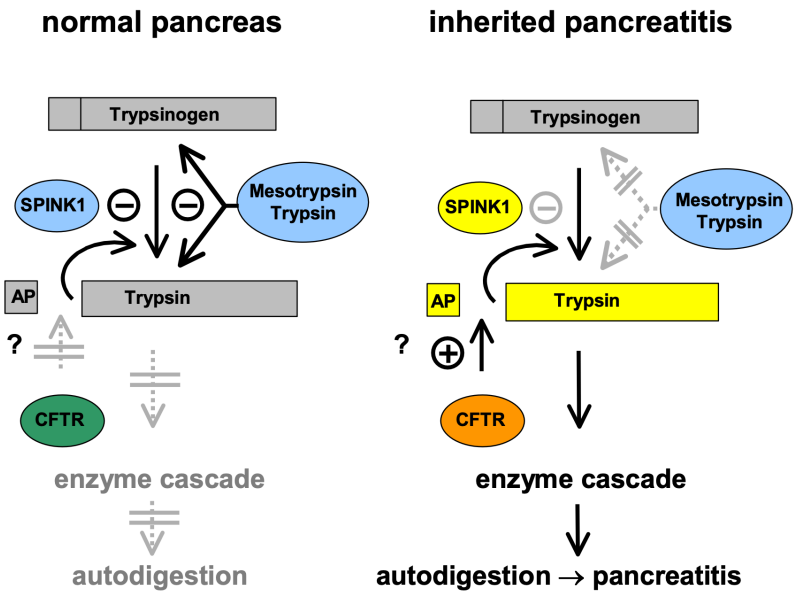


Figure 6. Mechanism of occurrence of hereditary pancreatitis.
On the left is the principle of the work of enzymes in the pancreas without pathology. *SPINK1* and trypsin/mesotrypsin inhibit trypsin activation from trypsinogen, thereby preventing pancreatic self-digestion. On the right is the mechanism in hereditary pancreatitis, where mutations in the *PRSS1* or *SPINK1* genes cause disruption of the work of proteases and their inhibitors. This process leads to self-digestion of the pancreas and the occurrence of pancreatitis. AP = activation peptide. Extracted from Rosendahl et al., 2007[234].

One of the major diseases in Europe is cystic fibrosis, with an autosomal recessive mode of inheritance and multiorgan manifestation. In case of cystic fibrosis or genetically determined pancreatic insufficiency[227], mutation of the transmembrane conductivity regulator (*CFTR*) gene occurs. The *CFTR* gene is located in the q31 region of chromosome 7's long arm, has a length of approximately 250,000 base pairs, and includes 27 exons[245]. *CFTR* is a chloride channel that regulates water balance across epithelia. Mutations lead to structural and functional changes in the calcium-dependent regulatory protein localised in the apical part of the membrane of the ductal epithelium of the pancreas[234]. This leads to progressive destruction and atrophy of the

acinar and ductal epithelium with its replacement by fibrous tissue and lipid deposits[246]; as a result, the secretions of the glands become viscous, and their secretion becomes difficult, and inflammatory processes develop in various organs, including lung and pancreas. The primary mutations in the *CFTR* gene are del21kb, delF508, delI507, 1677delTA, 2143delT, 2184insA, 394delTT, 3821delT, G542X, W1282X, N1303K, L138ins, R334W and 3849+10kbC>T. The del F508 deletion is one of the most common in European populations, which accounts for approximately 75% of genetic defects resulting in cystic fibrosis[245]. In cystic fibrosis, two mutated copies of the *CFTR* gene are inherited. If these mutations are the same, they are homozygous; if they are different, they are considered heterozygous. People with homozygous mutation tend to develop pancreatic insufficiency[247]. Cystic fibrosis is one of the first diseases targeted by gene therapy. The problems associated with gene therapy include a low level of gene expression and its transient nature, the development of an immune response to a vector protein, and local and systemic inflammatory reactions[248].

The term "hereditary pancreatitis" is used in families with already two or more members with pancreatitis. The appearance of multiple affected individuals in single families may be due to environmental factors. In separate articles, studies describe hereditary pancreatitis as having pancreatitis in only two patients of the same pedigree or more than two patients in the same generation. Because of shared environmental factors, this can correspond to the presence of actual hereditary pancreatitis in less than 5% of patients[249]. On the other hand, in Keim's studies, 35% of families identified with an autosomal dominant trypsinogen mutation (*PRSS1* N29I or R122H) had a pattern of inheritance not meeting the definition of "hereditary pancreatitis"[244]. If the causes of chronic pancreatitis are not determined, then with available evidence of hereditary pancreatitis in relatives of the first or second degree, hereditary pancreatitis can be diagnosed.

Rare mutations of the autosomal dominant type of inheritance are N29I and R122H from *PRSS1*, as well as N34S from *SPINK1*, which can occur in childhood or adolescence, and 80% of patients are up to 20 years old, whereas, with *PRSS1*, it can be detected in 50% in hereditary pancreatitis[244]. Cystic fibrosis due to mutations in *CFTR* is associated more with chronic pancreatitis, approximately in 90% of cases[250] and inherited in an autosomal recessive manner[227]. In addition, with hereditary chronic pancreatitis, a high level of superoxide dismutase is

noted, and the content of antioxidants, such as vitamin E and selenium, on the contrary, decreases[234].

There are several syndromes related to pancreatic diseases. The primary syndrome is Shwachman-Diamond syndrome (inherited in an autosomal recessive manner, the responsible gene is *SBDS* which is mapped in the long (q) arm of chromosome 7 at position 11.21), in which there is pancreatic hypoplasia, lipomatosis, early development of exocrine insufficiency[251]; sideroblastic anaemia with exocrine pancreatic insufficiency[252]; Clarke-Hadfield syndrome[253]; enterokinase insufficiency syndrome[254]; macroamylasemia[255] and isolated insufficiency of individual pancreatic enzymes (lipase, amylase, trypsinogen).

Hereditary chronic pancreatitis should be distinguished from congenital anomalies of the pancreas, such as longitudinal "cleavage" or doubling, with the annular or aberrant pancreas. An aberrant pancreas can undergo inflammation, necrosis, or even perforation. With Johanson–Blizzard syndrome (inherited in an autosomal recessive manner - the responsible gene is *UBR1* which is mapped in the long (q) arm of chromosome 15 at position 15.2q)[256], there is a congenital deficiency of the main pancreatic enzymes[234].

One of the major studies in the pancreatic field examined more than 1,000 people with acute and chronic pancreatitis in the North American Study of Pancreatitis 2 (NAPS2), which included both phenotypic data and biological samples with controls[257,258]. The criterion for the selection was the presence of symptoms of pancreatitis arising from the use of large amounts of alcohol. In the first stage, 758 samples were taken from the NAPS2 research program, and 493 controls, as well as 4076 cases and control samples from Alzheimer Disease Genetics Consortium (ADGC). In the second stage, 343 samples with chronic pancreatitis and 627 samples with recurrent acute pancreatitis were used. For the control group, 3986 cases from NeuroGenetics Research Consortium (NGRC) and 205 from NAPS were involved. Data analysis was performed using PLINK. NAPS2 and ADCG were genotyped using Illumina Human OmniExpress BeadChips, and NGRC samples were genotyped using Illumina Human 1M-Duo DNA Analysis BeadChip. The purpose of the NAPS2 was to collect data from all 1000 subjects for further study. All collected data were organised and classified for easy access. They were investigating hereditary pancreatitis and gene mutations. It was a two-stage general GWAS. The study identified significant genome-wide association at two loci: *PRSS1-PRSS2* ($1x10^{-12}$, rs10273639) and X-linked *CLDN2* ($p<1x10^{-21}$, rs12688220)[259]. The

functions of *CLDN2* include control of tissue-specific physiological features of tight junctions[260] and association with atypical localisation of claudin-2 in pancreatic cells. The study used individuals with European ancestry. At the first stage, 676 cases and 4507 controls with the genotypes of 625739 SNPs were identified. In the second stage, 910 cases (of which 331 had chronic pancreatitis and 579 recurrent acute) and 4170 controls with genotypes of 625,739 SNPs were determined. The study showed that rs10273639 is directly associated with *PRSS1-PRSS2* and had a risk of developing chronic and acute recurrent pancreatitis. The *CLDN2* locus showed a more significant association with chronic pancreatitis in comparison to acute recurrent pancreatitis. *CLDN2* appears as a disease modifier and accelerates the alteration from acute to chronic pancreatitis. Variants of this gene may be associated with the risk of alcohol consumption. These results also explained the difference in alcohol consumption between males and females. In men, there is a high incidence of alcoholic pancreatitis since the frequency of hemizygotes was 0.26, compared with women, where the homozygote frequency was 0.07[259].

The other studies were conducted based on the International Hereditary Pancreatitis Study Group and the European Register of Hereditary Pancreatitis and Pancreatic Cancer (EUROPAC - European Registry of Hereditary Pancreatitis and Pancreatic Cancer) databases. Most often, pancreatic cancer develops in families with mutations R122H and N29I or without mutations of the *PRSS1* gene but with an apparent phenotype of hereditary pancreatitis[227].

Basically, in a European cohort of 584 patients with non-alcoholic chronic pancreatitis (NACP) and 6040 healthy controls with various types of pancreatitis, GWAS discovered genome-wide risk loci for chronic pancreatitis and studied the underlying mechanisms driving GWAS. It was found in the mutation risk loci of the *CTRC* ($p=1.22x10^{-21}$) and *SPINK1* ($p=6.59x10^{-47}$) genes that reached genome-wide significance in NACP. *CTRC* observed a decreased expression and showed a pathogenic effect. At the same time, due to the increase in the expression of *CLDN2* (rs12688220), a modification of the risk of ACP (alcoholic chronic pancreatitis) and NAC was observed[261]. Rosendahl et al. showed during GWAS that inversion of the *CTRB1-CTRB2* locus modifies the risk of developing chronic pancreatitis, and inflammatory diseases are involved in common pathomechanisms[262].

Other genes like *CEL* lipase genes and *CELP* can also predispose to the development of pancreatitis, mostly the chronic variant[263]. The

CEL gene encodes the pancreatic enzyme carboxyl ester lipase involved in digestion. Carboxyl ester lipase is located in the pancreas, namely in acinar cells and in lactating mammary glands[263,264]. This enzyme is involved in the absorption of lipid-soluble vitamins and cholesterol and is activated in the duodenum by bile salts. The *CEL* gene, including the neighbouring *CELP* pseudogene, is mapped on chromosome 9q34.3[263]. *CELP* has a deficiency of exons 2-7 in *CEL* and contains stop-codons in its second exon. These characteristics distinguish *CELP* from the *CEL* gene since, according to other characteristics, the genes can be very similar to each other[265,266]. *CEL* may additionally be the cause of the maturity-onset of diabetes in young, type 8, called MODY 8, resulting from a single-base deletion in *CEL VNTR*. MODY 8 is a hereditary disease of the pancreas with an autosomal dominant mode of transmission, characteristics of the onset of diabetes, and morphological and structural changes in exocrine pancreatic dysfunction[267,268]. Additionally, a hybrid allele *CEL-HYB* formed from an unequal crossover between *CEL* and *CELP*. Fjeld et al. conducted research defining the function of *CEL-HYB* in the genetic aetiology of chronic pancreatitis. *CEL-HYB* in cell models revealed a disorder of pancreatic secretion, a decrease in lipolytic activity, and intracellular accumulation with induced autophagy. Studies have been shown on several familial cases in which chronic pancreatitis was observed. In these studies, *CEL-HYB* was detected in 14.1% (10/71) of pancreatitis patients and 1.0% (5/478) control subjects. In non-alcoholic chronic pancreatitis observed in three replication studies, the *CEL-HYB* frequency was 3.7% (42 / 1,122) in cases compared with 0.7% (30 / 4,153) controls. *CEL-HYB* was significantly more common among persons with chronic pancreatitis. In 57 of 1122 individuals, rare variants of *PRSS1* were found, which in turn is associated with hereditary chronic pancreatitis. The *PRSS1* variant carrier status of the *CEL-HYB* variant carriers was not identified. Additionally, the group with chronic pancreatitis of alcoholic aetiology was tested, and the presence of *CEL-HYB* was also detected, with a frequency of 1.8% compared with the control group of 0.8%. In positive *CEL-HYB* individuals in 5 cohort studies, heterozygous carriage of alleles was found. In two *CEL-HYB* family members, the pedigree showed an autosomal dominant mode of pancreatitis inheritance. During the *CEL-HYB* variant analysis of 228 parents (parents of subjects), six heterozygous parents were detected. All of them gave their children the allelic risk of pancreatitis[269].

Another aspect of the development of pancreatitis is the influence of two enzymes of alcohol metabolism, alcohol dehydrogenase (*ADH*) and aldehyde dehydrogenase (*ALDH*), which is a contributor to the pancreatitis of alcoholic aetiology[249]. The existing allelic variants of the aldehyde dehydrogenase gene *ADH2*2* and *ADH3*1*, which encode highly active isoforms and low activity encoding *ALDH2*, can help protect the body from the development of alcoholism[270,271]. The *ADH* family has four genes divided into two classes *I* (*ADH 1-3*) and *II ADH* (*ADH4*). These genes are mapped in the long arm of chromosome 4 in 4q21-23. The *ADH2* and *ADH3* polymorphisms encode alcohol dehydrogenase subunits. The allelic variants *ADH2*1, ADH2*2* and *ADH2*3* encode the β 1, β 2, and β 3 variant subunits to be detected in the *ADH2* locus. *ADH3*1* and *ADH3*2* encode the subunits γ 1 and γ 2 are located in the *ADH3* locus. Aldehyde dehydrogenase is the second enzyme after alcohol dehydrogenase in ethanol metabolism. Four NAD-dependent aldehyde dehydrogenase isoenzymes were identified and named *ALDH I, II, III*, and *IV*, characterised by a decrease in electrophoretic mobility, as well as an increase in the isoelectric point. The results revealed that the allelic frequency of *ADH2*2, ADH3*1*, and *ALDH2*2* in alcoholics is lower in comparison with healthy individuals. *ALDH2*2* may contribute to the protective properties of alcoholism due to the fact that this genotype was not found among people suffering from alcoholism. Features of *ALDH2*2/*2* homozygotes are sensitivity to low doses of alcohol and long-term accumulation of acetaldehyde occurs in the blood[272,273].

Unlike chronic alcoholic pancreatitis in chronic pancreatitis, the progression of the disease occurs slowly, and acute pancreatitis occurs even less frequently. After 10–15 years, more evidence of chronic pancreatitis may be found.

One of the other genetic reasons for the development of pancreatitis may include a mutation of the *α 1-antitrypsin* gene[249]. *α 1-antitrypsin* is a protein initially formed in the liver, and along with additional protein, β 2-macroglobulin binds the enzymes of the pancreas during admission to the blood or peritoneal fluid[274]. The antitrypsin gene is mapped on chromosome 14 (14q32.1) and exists as two genes: *α 1-antitrypsin* and *α 1-antichymotrypsin*[249].

Several meta-analyses have been conducted to identify genetic associations with predisposition to pancreatic diseases. Several databases were used (MEDLINE, Embase, BIOSIS, Web of Science, Cochrane Library and PubMed). Ninety-six studies were identified, reporting 181

variants in 79 genes. Significant associations were found with genes like *SPINK1* (OR 2.87, 95% CI 1.89 to 4.34, $p=6.3x10^{-7}$) and *ALDH2* (OR 0.48, 0.36 to 0.64 $p=3.3x10^{-7}$)[275]. The *PRSS1* gene was significantly associated with total pancreatitis (OR: 10.799, 95% CI: (5.489-21.242), p<0.000). European subgroup (OR: 9.795, 95% CI: (2.923-32.819), p<0.000) and Asian subgroup (OR: 11.994, 95% CI: (5.156-27.898), p <0.000) had the association of *PRSS1* with both pancreatitis[276]. 11 analysed studies of the meta-analysis showed estimation effect of 2.07 (95% CI: 1.36-2.78) for acute pancreatitis and pancreatic cancer risk. Five cohort studies reported 103961 patients with acute pancreatitis, relative to 1442158 subjects in the control group, with RR of 7.81 (95% CI: 5.00-12.19)[277]. Forty-six studies with 2,341,007 patients of acute pancreatitis in 36 countries were analysed. The 95% CI for acute biliary pancreatitis was 42 (39-44%), for alcoholic pancreatitis 21 (17-25)% and for idiopathic pancreatitis 18 (15-22)%. The results showed that globally, acute pancreatitis has gallstones as its primary aetiology, and this factor is twice as common as the next most common factor[278]. The meta-analysis showed that these genes are associated with a predisposition to pancreatitis and its severity and are also very important as candidate genes.

Thus, genetic studies help to understand the causes of pancreatitis. New research in the field of *PRSS1, CFTR, SPINK1*, and *CEL*, as well as other genes, proves a genetic link with pancreatic diseases.

Basically, the studied genetic defects in pancreatic diseases are associated with an interruption of the system of safe activation of pancreatic enzymes. Indeed, trypsin starts the activation of all proteolytic enzymes in a cascade. However, this cannot explain all genetic mechanisms of predisposition to pancreatic diseases, and sensitivity to environmental factors differs among candidate genes for the role of predisposition.

Genetic aspects of pancreatic cancer

Pancreatic cancer (PC) is a malignant disease that develops from the epithelium of the glandular tissue or pancreatic ducts. About 90% of pancreatic neoplasms, where 80% are invasive adenocarcinomas, are ductal origin. The other 10% of pancreatic neoplasms are islet cells and cystic tumours[279]. The estimated leading cause of death from cancer in Europe is in fifth place[280,281]. Worldwide, around 450,000 cases of pancreatic cancer and approximately 430,000 deaths in 2018 were reported[282]. In 2020, 146,063 deaths from pancreatic cancer were recorded in Europe and the United States[281]. By 2030, it is anticipated that PC will be the second greatest cause of cancer-related death in the US[283,284].

The development of invasive cancer in PC occurs in stages. It is at least ten years from the earliest genetic changes in cells to precancerous changes in the pancreas and pancreatic ductal adenocarcinoma (PDAC) with a metastatic process[285]. Precancerous conditions of the pancreas include pancreatic intraepithelial neoplasia (microscopic precancerous lesions in the pancreatic ducts of small calibre (less than 5 mm), which lead to the development of PDAC), intraductal papillary mucinous tumour (macroscopically visible cystic neoplasms that occur in the main pancreatic duct or one of its branches), mucinous cystadenoma (the rare precancerous condition, microscopically visible cystic neoplasms that do not connect with the ductal system of the pancreas)[286]. Pancreatic cancer is considered the most fatal disease, from which about 95% of all patients die. By the time of cancer diagnosis, approximately 80% of patients have metastases, and the 5-year survival level is fewer than 5%[287].

The aetiology of PC is still unclear; more than 90% of patients acquire it sporadically. About 10% of patients have incidence of PC in family history[288]. Comparing first-degree relatives with pancreatic cancer to the general population, those with one first-degree relative have a 2-fold greater chance of having the disease, those with two first-degree relatives have a 6-fold increase, and those with three or more first-degree relatives have a 14-32-fold increased risk[289]. Some genetic syndromes are also associated with the development of hereditary PC: hereditary breast and ovarian cancer, familial atypical multiple melanomas, Peutz-Jeghers syndrome, and hereditary pancreatitis[290].

Biological, lifestyle and environmental factors have a significant impact on the risk of PC since the incidence of PC varies in different countries, and the risk differs from 2 to 13 times. The incidence of PC

links with age, with the average age at developing the disease being over 70. Gender and lifestyle are also essential and cause the incidence of pancreatic cancer. For example, alcohol abuse and smoking are more often in men and are considered risk factors. About 20-25% of pancreatic cancer is caused by lifestyle and environment. Associated diseases such as chronic pancreatitis, diabetes, *Helicobacter pylori* (HP), vitamin D deficiency and occupational exposure greatly raise the risk of PC. At the same time, metformin used to treat diabetes and atopic allergies were linked to lowered risk[291].

Many genes have been discovered as candidate genes, affecting the risk of developing pancreatic cancer *BRCA1, BRCA2, PALB2, CDKN2A, ATM, TP53, STK11, MLH1, MSH2, MSH6, PMS2*, and *EPCAM*. The most significant are the *KRAS, p16/CDKN2, TP53* and *SMAD4/DPC4* genes, which are also called the "big four" [292,293].

KRAS is a proto-oncogene belonging to the Ras protein family, located on chromosome 12p, and encodes a guanosine triphosphate-binding protein. *KRAS* gene is most commonly mutated in the pancreatic cancer oncogene[294]. Mutation in the *KRAS* gene is found in diseases such as pancreatic[295], colon[296] and lung cancer[297]. In a study by Biankin et al., *KRAS* mutations were detected in 93% of 142 patients with PDAC[298]. Krasinskas et al. found that *KRAS* copy number alterations are present in 90% of ductal adenocarcinomas and undifferentiated pancreatic cancers[299].

P16/CDKN2A is the following gene that contributes to PC development. On the short arm of chromosome 9 contains the tumour suppressor gene *CDKN2A* which engaged in 80–85% of cases of PDAC[300]. *CDKN2A* variant increases the development of PDAC risk 12.3-fold[301].

On chromosome 17p, the tumour suppressor gene *TP53* encodes the p53 protein, which is critical for controlling the cell cycle and triggering apoptosis, among other functions[302]. Cancer Genome Atlas pancreatic cancer study (TCGA) revealed that *TP53* had been found in pancreatic cancer in 60% of cases[303], and the MSK-IMPACT study with the GENIE project identified it in 70% of cases[304].

On chromosome 18q21.1 located the tumour suppressor gene *DPC4/SMAD4* which encodes a protein related with the transforming growth factor β (TGF-β) by signalling pathway and is involved in 55% of PC through intragenic mutations and homozygous deletion[305].

Mutations in p16 are found in 30-50% of cases with pancreatic cancer[306,307]. p16 is an essential tumour suppressor gene[308] and is

associated with poor prognosis[309,310]. The inactivation of the p16 gene appears in loss of heterozygosity (40%), homozygous deletion (40%), and methylation-associated transcriptional silencing (15%)[309] and occurs in 95% of cases of pancreatic ductal adenocarcinomas[311-313].

Currently, the association of genetic risk factors with individual variable risk factors for PC identified using GWAS is being actively studied. Overall, 21 susceptibility loci for PC have been identified by GWAS[314]. Two extensive studies have been conducted in the field of studying the association of gene mutation and pancreatic cancer, where five new areas of association were identified.

The first project of PanScan (Pancreatic Cancer Cohort Consortium) and PanC4 (Pancreatic Cancer Case-Control) used samples of 9040 patients and 12,496 controls of European origin. As a result, an association with the rs78417682 locus (7p12/*TNS3*, $p=4.35x10^{-8}$) was revealed.

The second PANDoRA (PANcreatic Disease ReseArch) project used data from 2737 patients and 4752 controls. As a result, five new association areas were identified. Loci rs13303010 at 1p36.33 (*NOC2L*, $p=8.36x10^{-14}$), rs2941471 at 8q21.11 (*HNF4G*, $p=6.60x10^{-10}$), rs4795218 at 17q12 (*HNF1B*, $p=1.32x10^{-8}$) and rs1517037 in 18q21.32 (*GRP*, $p=3.28x10^{-8}$).

In additional studies involving PanScan, PanC4 and PANDoRA, more candidate SNPs were identified that showed an association with PC in rs4626538 (7q32.2), rs7008921 (8p23.2) and rs147904962 (17q21.31) with ($p<10^{-5}$), as well as a possible relationship between rs36018702, *BCL2L11* and *BUB1* expression[281].

In another project, three latest PC signals were found based on the data of PanScan I-III GWAS (5107 cases and 8845 controls), PANDoRA, and PanC4 (6076 cases and 7555 controls). SNP rs2816938 on chromosome 1q32.1 *NR5A2* ($p=4.88x10^{-15}$), rs10094872 in 8q24.21 *MYC* ($p=3.22x10^{-9}$) and rs35226131 in 5p15.33 *CLPTM1L-TERT* ($p=1.70x10^{-8}$). These SNPs are new susceptibility patterns that each independently can be a risk of PC[315].

In other studies of pancreatic cancer GWAS involving a cohort of Jewish origin (Ashkenazi), 406 cases with PC and 2332 controls were taken. The results showed a significant association in the loci of chromosomes 19p12 (rs66562280, $p=10^{-7.6}$) and 19p13.3 (rs2656937, $p=10^{-7}$). SNPs rs66562280 and rs2656937 are located in *ZNF100-like* and *ARRDC5* introns, which may be carcinogenic. An increased risk of

PC of 50–80% is identified in the Jewish cohort compared to the non-Jewish cohort[316].

The study of PC GWAS confirms the importance of studying the genetic factors influencing the risk of developing the disease.

Amylase gene. Features: salivary amylase and pancreatic amylase.

CNV may be associated with the occurrence of various multifactorial diseases. Specifically, the CNV of *AMY1* and *AMY2* may play a significant role in developing diseases associated with the pancreas. In some previous research, the correlation between the *AMY* CNV gene and other conditions such as obesity and insulin sensitivity[317-319], metabolic syndrome[320] has been demonstrated, which shows that the *AMY* CNV can form a particular phenotype[321]. However, in other research, no single association between *AMY* CNV and pancreatic diseases has been examined, prompting a search for this connection. So far, the precise definition of copy number for *AMY* remains challenging.

Amylase is a hydrolytic enzyme that splits oligosaccharides, such as starch and glycogen, into maltose, maltotriose and dextrin[322]. Mainly amylase is formed in the salivary glands and pancreas, enters the oral cavity or the lumen of the duodenum and is involved in the digestion of carbohydrates. With impaired amylase synthesis, the organism's ability to absorb starch-containing products is reduced or lost. However, some researchers claim that salivary amylase makes an insignificant input to total starch digestion, complementing the pancreatic amylase mechanism[323,324]. At the same time, studies of the enzyme process have demonstrated the presence of a large family of plant proteins that inhibit α-amylase[325-327]. The specificity of amylase is in its connection with N-linked oligosaccharides, which are located on the border of glycoproteins, which in turn, can characterise the activity of amylase[328]. The function of oligosaccharides is to protect salivary amylase from the aggressive effects of gastric hydrochloric acid. Maltose binds to the receptors of the upper part of the stomach to, subsequently, give a signal to GLP-1 (Glucagon-like peptide-1), PYY (peptide tyrosine tyrosine) and pancreatic amylase for their secretion[322,329]. All three are pancreatic peptides. Human pancreatic α-amylase is synthesised as a 57 kDa protein, for which 512

amino acids of protein are determined by cDNA[330]. Amylase is represented by two isoelectric forms, 7.2 and 6.6, called HPA I and HPA II[331]. As a pancreatic enzyme, the quantity of amylase is adjusted by the volume of food substrate[332-334].

The amylase genes are mapped on chromosome 1p21 and consist of two types; the *AMY1* gene encodes salivary amylase, and pancreatic amylase is encoded by the *AMY2A* and *AMY2B* genes. *AMYP* is a pseudogene and is a deactivated (truncated) duplicate sequence from *AMY2A*. *AMY1* and *AMY2* gene products are involved in absorbing carbohydrates by digesting complex sugars and starch (Figure 7)[335].

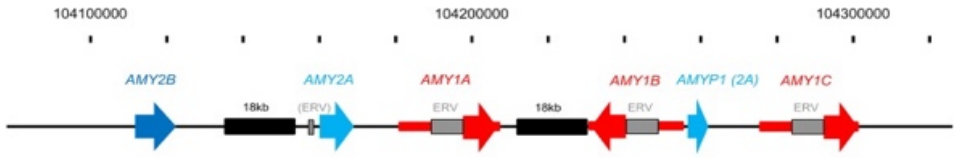


Figure 7. Overview of the *AMY* gene region (human assembly hg19). Here are two copies of *AMY2* (*AMY2A/AMY2B*) and three copies of *AMY1* (*AMY1A/1B/1C*, where *AMY1B* is in the opposite direction). at the telomeric end of the same cluster, *AMY2A* and *AMY2B* appear. The arrow specifies the genes' orientation and position. The red rectangle indicates the sequence identity around the copies of *AMY1*. Black boxes show two identical regions with similar sequences to other parts of the locus. Extracted from Carpenter et al., 2017[335].

Amylase genes are located in a set of loci, with inversions, deletions and duplications[336]. *AMY1* encodes one of the main proteins of saliva secreted by highly differentiated epithelial acinar cells of the salivary glands after activation of β-adrenergic receptors. Each person has at least two copies of this gene, but the number of copies of the gene *AMY1* varies from 2 to 18[335]. The amylase gene *AMY1* includes three repeats on the human genome reference assembly in the form of *AMY1A, AMY1B* and *AMY1C*, where *AMY1B* is inverted in orientation (Figure 7). All these three copies are nearly identical to each other (sequence identity 99.9%), and their length is 26.5 kb[335].

Amylase in the pancreas is encoded by the *AMY2* genes, which can range from 0 to 8 copies[337]. *AMY2A* is found in the range of 0 to 4 copies, and *AMY2B* is 2 to 6 copies[338]. Genes encoding human pancreatic

amylase (*AMY2A* and *AMY2B*) are mapped on the telomeric site of the reference sequence cluster. Their sequences are 94% identical to each other; identity with *AMY1* is about 93% (Figure 7).

Researchers have determined that sequential copying within the locus of the number of these genes leads to their individual variability. The greater the number of copies in the genotype, the higher the production of the enzyme in this individual for *AMY1*[335]. However, variations in CNV are challenging to measure[339], and studies of the CNV association often include approximate estimates of the number of copies, which may be attributed to technical factors[340-342].

The CN of *AMY1* may correlate with the quantity and activity of salivary α-amylase. However, the physiological importance is still unclear. The *AMY1* has significant variation in the number of diploid copies[343-345]. People with a high content of CN in saliva have a higher concentration of amylase protein, which leads to faster digestion of starch in vitro[344].

Carpenter et al., in their work, showed a structural difference in haplotypes in the main classes. Haplotypes differ between those with even and odd numbers of the *AMY1* gene. The majority of human amylase haplotypes have an odd CN of *AMY1*, one copy of each of the *AMY2B* and *AMY2A* genes. It has been determined that even-numbered haplotypes of *AMY1* are associated with a change of arrangement, thereby leading to CN of *AMY2A* and *2B.* Their study showed that the *AMY1* gene has an even number of copies in most (diploid) individuals, while *AMY2A* and *2B* each have two copies in 60% of the population. Deletions and duplications of the *AMY2* gene lead to an even number of the entire *AMY* gene[335]. Three major haplotype series can account for 98% of *AMY1*, *AMY2A* and *AMY2B* CN combinations[346]. Other research based on array-comparative genomic hybridisation (CGH) or real-time polymerase chain reaction (PCR) confirmed the presence of CN and the link with the expression of the enzyme. Still, it did not demonstrate genotype prevalence with even *AMY1* numbers. In some of these studies, structures with 12 *AMY1* replicates in both tandem and inverted pairs in the direction were shown in fibre-FISH images[343,347]. Data from HapMap samples showed two relationships among individuals, with four times more often even (2, 4, 6, etc.) numbers of copies of *AMY1* than odd (1, 3, 5, etc.)[348].

Studies by Usher et al. showed that most chromosomes of European origin (89%) had an odd CN of *AMY1* and were equated to an even total CN. *AMY1's* odd haplotypes included one CN of *AMY2A*.

Even haplotypes of *AMY1* had two copies or did not have copies of *AMY2A*. In this way, *AMY2A* and *AMY1* CNs show parity with each other[346]. *AMY1* can segregate on many identical SNP haplotypes. Haplotypes that had an even number of copies of *AMY1* segregate on different backgrounds of SNP haplotypes as a result of mutational mechanisms. Individual SNPs located close to *AMY1* genes cannot fully correlate with the *AMY1* CN. Deletion of *AMY2A* can be related to the SNPs rs1930212 and rs72694406. The *AMY2A* region had many SNPs that had minor allele frequencies of more than 0.3 (one of the examples is rs42443712)[346]. However, Hasegawa et al. identified that *AMY2A* and *AMY2B* genes have larger numbers of SNPs compared to *AMY1*[349].

Other researchers used a technique where the amount of CNV was determined using real-time duplex quantitative PCR (qPCR) with two TaqMan assays, pre-isolating DNA from peripheral blood. Digital PCR-based methods as digital droplet PCR, are suggested as an exact method but also have a small degree of uncertainty. The key difference between qPCR and PCR is that it tracks the amplification of the target DNA in real-time and amplifies the product by a fluorescent signal[346]. Estimates of the relative copy number of *AMY1* shown by Falchi et al. using qPCR were strongly correlated (R^2=0.95; $p<2.20\times10^{-16}$) with estimates obtained using digital PCR[350]. However, the same analysis and qPCR calibrators in Falchi et al. [350] showed the various distribution of the CN, which also reflects the differences between populations[343,351]. The comparative analysis used a certain number of copies in *AMY1* as normal and elevated ranges, with cut-off points at the median CNV *AMY1* CN<6 against CN>6[352], unlike Falchi et al., where CN<4 vs CN>10 was used[350]. The number of *AMY1* genes ranged from 2 to 19. In adults, the presence of 4 copies gave a high risk of obesity, and six copies showed a borderline. As a result, studies have shown that only *AMY1* CNV shows a relationship with body weight[352]. The indirect association of *AMY1* with BMI was examined by research by Usher et al. Association analysis was performed by directly measuring the number of copies of *AMY* genes with the high-resolution method in three cohorts of European origin. When using ddPCR to gain integer copies of amylase genes, no association was observed between obesity and the *AMY* genes (p=0.70, p=0.31, p=0.54 etc. for *AMY1* CNV; p=0.6-2.0 for SNP *AMY1*; p=0.80 for *AMY2A* and *2B*). Analysis of the CN of the *AMY* showed enrichment of even compared with odd numbers of copies[346]. Using different methods can be one of the reasons for differences in results. For example, in research by Leon-Mimila et al. for the CNV research in *AMY1*, qPCR

was used[352]. In studies by Markovecchio et al., the CN of the *AMY1* was estimated using qPCR, including two TaqMan assays (Life Technologies), one for *AMY1* (Hs07226361_cn, labelled FAM) and one particular for the control gene (RNase P, VIClabeled)[353]. Methods based on a practical analysis of TaqMan show a variation of the *AMY1* gene number from 2 to 19. Using the appearance of fibre-FISH or read-depth of whole-genome sequencing, control sample NA18972, in different studies, had a different number of copies (ranging from 14 to 20)[335,343,346,354]. In studies by Carpenter et al., fibre-FISH analysis was used to verify the haplotype structure and the number of copies. The result confirmed the number of copies of haplotypes 13 and 5, giving a diploid sum of 18 for the sample NA18972. Such a comparison of CN results demonstrated less accuracy with qPCR and more accuracy for the Paralogue ratio test (PRT) (pairwise associations among repeat measurements for qPCR (R^2=0.782) were less compared to PRT (R^2= 0.893). In the studies of Carpenter et al., combined microsatellite analysis used profiles and PRT data. As a result, conforming estimates were obtained for the number of *AMY1* copies for 749 unrelated samples consisting of 269 HapMap phases I and II and 480 UK samples from ECACC HRC1-5. The study used the likelihood ratio score (LRS) as a maximum likelihood method to assign each sample's most likely integer value for the *AMY1* copy number. The locus-specific analysis confirmed the accuracy of the PRT_ref12 measurements for the entered samples. The second analysis of PRT_ref1 showed a match with the data from PRT ref12. However, some samples reproduced a greater or smaller visible number of copies of *AMY1*. This confirmed the CNV in the *AMY2A* control locus and also demonstrated the presence of four different classes of CNV affecting the pancreatic gland genes (*AMY2*)[335].

A study was conducted with the participation of Mexicans to determine the risk of obesity in adults and children. A positive correlation of *AMY1* gene number was reported, as well as the prevalence of *Prevotella*. The interest towards *Prevotella* is in their enzymes and gene clusters that can decompose and utilise complex polysaccharides[352]. Provetella can be found more in people who eat more carbohydrates and fibre in food[355,356]. Moreover, the metabolism of *Provetella* enzymes can be indirectly linked to the *AMY* gene. In this study, five candidate genes were used that are presumably related to obesity and analysed the association of *AMY1* CNV was with intestinal bacteria *Enterobacteriaceae* and *Provotellaceae*, especially their role in obesity in Mexican adults and children. In the project, 921 unrelated Mexican

children between the age of 6 and 12 were used, of which 485 had normal weight and 436 obesity. The adult part consisted of 920 unrelated Mexicans from 18 to 75 years old, where 536 had average weight and 385 obesity. The CN of *AMY1* was determined through real-time duplex qPCR with two TaqMan assays, pre-isolating DNA from peripheral blood. In addition, diet and intestinal analysis were evaluated. From 5 CNV *NEGR1*- neuronal growth regulator 1, 1p31.1; *GPRC5B* - G Protein-Coupled Receptor Class C Group 5 Member B, 10q11.22; *NPY4R*- Neuropeptide Y Receptor Y4, 16p12.3 did not have a positive connection with obesity. *OR4P4/OR4S2/OR4C6*- olfactory receptor genes (p=0.047) and *AMY1* (p=0.064) demonstrated a relationship with body weight among Mexican children (higher risk with low CN) but not in adults. Also, a positive correlation between the CNs of *AMY1* and intestinal microbiota was highlighted[352].

The high CNV of *AMY1* is linked with a lower risk of insulin resistance in humans. In Korean studies, Choi et al. analysed the connection between *AMY1* CNV and the assessment of the homeostatic model l - insulin resistance (HOMA-IR). Subjects were selected from 1257 men aged 20 to 65 years, not have specific symptoms. For the study of CNV in *AMY1*, quantitative real-time PCR was selected using a TaqMan qPCR assay. The researchers concluded that in Korean men without specific symptoms of hyperglycemia, low CNV *AMY1* values correlated with high insulin resistance (p=0.008). Additionally, the risk of smoking and drinking was investigated. In smokers and non-alcohol users, an association with CNV *AMY1* was not detected (p=0.28 for smokers and p=0.13 for non-alcohol users). Research into risk factors for developing the disease and CNV *AMY1/AMY2* is of paramount importance. Different studies have been conducted to define the risk factors for diseases, the relationship of amylase towards them, and the amount of CNV. Some samples were determined by the FISH method[357].

Hasegawa et al. found that *AMY1* and *AMY2* CNs are found to be positively correlated with diabetes ($p=2.69x10^{-2}$, $1.89x10^{-2}$, $1.71x10^{-2}$). In their study, individuals with diabetes were 4/338 (low group) and 10/240 (high group) in *AMY1* CNs first criteria, 0/90 (low group), 10/417 (medium group), and 4/71 (high group) in *AMY1* CNs second criteria, and 2/276(low group) and 12/302 (high group) in *AMY2A* CNs[349].

As already known, the level of amylase in the serum and saliva is associated with the CNV *AMY1* gene[343,345,358]. Some studies have shown that with high amylase activity in saliva and high *AMY1* CN, after meals, people showed low levels of blood glucose, as well as high levels of insulin

after starch consumption[358]. Other studies have found that an increased risk of diabetes and a high level of insulin resistance has been correlated with CNV values for salivary *AMY1*[359,360]. Atkinson et al. used real-time qPCR assays to determine *AMY1* CN and its relationship with glucose tolerance, insulin sensitivity, salivary α-amylase activity, BMI, and consumption of macronutrients. As a result, *AMY1* CN was associated positively with salivary α-amylase (R^2=0.62, p<0.0001, n=201). However, there was no positive correlation with BMI, glucose tolerance or insulin sensitivity. The correlation of copy numbers with normalised glycemic reactions in starchy foods except sucrose and fructose was shown[361].

The productivity and activity of amylase are impacted by external factors such as the nature of nutrition, stress, and body hydration. Hancock et al. found that environmental parameters are associated with SNPs localised in genes whose products are involved in energy metabolism. Individuals from populations with starch-containing products in their diet have, on average, extra copies of the *AMY1* gene than those with a low proportion. In the conditions of modern nutrition, it is suggested that genotypes with a small number of *AMY2* gene repeats, and carriers of its deletions become maladaptive[362]. The incidence of different *AMY2* variants has differences among the number of groups[335,346].

Inchley et al., in their studies, systematised data on the relationship of the *AMY* gene with habitat factors. The population coverage achieved by researchers makes it possible to estimate the distribution of *AMY* variants. In populations of the indigenous inhabitants of Northeast Asia, some other variants of genetic polymorphism are highly prevalent. In the study, it was shown that the northern people had the maximum rates of deletion of the *AMY2A* (frequency 52%), as well as the minimum number of copies of this gene in combination with a reduced CN of the *AMY1*[363]. In the East Asian populations used in the 1000 Genomes project, a small number of variations of *AMY2* are observed. In African samples, duplication of *AMY2A/AMY2B* occurs[335]. The high prevalence of the *AMY2A* gene deletion in Northeast Asia can be attributed to the deficiency of starch and disaccharides in the traditional diet[363]. Studies have shown that oligosaccharide deficiency can be attributed to the high prevalence among the indigenous people of Far North Asia and America. Thus, in the Canadian Arctic locales, the frequency of deletion of dinucleotide AG in the rs781470490 locus of the sucrase-isomaltase gene (*SI* gene) is

17%[364], in Greenlandic Inuit 39%[365]. The deletion in the rs781470490 locus of the *SI* gene leads to the premature termination of the synthesis of *SI* and, as a result, to the total lack of this enzyme. The carriers of this genetic variant develop sucrase-isomaltase deficiency. Thus, the high frequency among the indigenous population of Northeast Asia of individuals who cannot digest specific carbohydrates may be linked with a weakening of adverse selection for the *SI* and *AMY2A* genes, which arose as a result of a deficiency of carbohydrate substrates in the past.

The level of α-amylase in the saliva is also investigated as an indicator of stress[366]. A study revealed substantial growth in the activity of α-amylase in saliva in pregnant patients placed under stressful operating conditions. The study showed that mean amylase levels correlate with verbal subjective assessment of anxiety at each sampling stage. As the intensity of the pain increases, so does the concentration of amylase. However, with a high intensity of pain, for some reason, the concentration of amylase falls, resources for its synthesis are depleted, or metabolic pathways are switched to the synthesis of something else. The average increase in the level of amylase during labour is statistically significantly higher compared with the average increase in the level of amylase after birth. However, after birth, the level of amylase still remains, on average, significantly higher than the initial level (at rest). It is possible that a subsequent decrease in amylase activity illustrates adaptation to stress. With high pain intensity in the early postoperative period, the amylase concentration no longer reflects the patient's degree of pain. Still, it continues to correlate with the subjective assessment of anxiety[367].

Amylase is also increased in various surgical, traumatic and neoplastic diseases[368-370]. Kidney disease may decrease the filtration of the amylase and increase the level of plasma. However, the rise is commonly minor because of glomerular filtration, as it is a tiny part of amylase clearance[369]. With recurrent chronic pancreatitis or alcoholic pancreatitis, the disease occurs without an increase in amylase, and there is an increase in lipase levels[371].

Conducted studies by Gallassi et al. showed that with diseases of the pancreas, an enhancement in the activity of amylase is slight. Thus, in 183 patients with benign pancreatic hyperenzymemia, 74.9% had elevated levels of lipase and both amylase isoenzymes, 7.2% only had lipase, and 6.3% only amylase and isoamylase were elevated[372]. Studies conducted by Amodio et al. included 160 patients in whom, for a long time (3.3-3.9 years) was an increase in the activity of pancreatic enzymes

(hyperamylasemia in 32%, hyperlipasemia in 6% and an increase in both enzyme levels in 62%) in the absence of clinical manifestations[373]. However, chronic pancreatitis has a loss of functioning in the pancreatic tissue. Therefore, around 90% of carbohydrates can be digested and absorbed, and inhibition of amylase reduces this to 80-85%[374]. Up to 20% of carbohydrates achieve the ileum, and the growth of plasma GIP and C peptides from pro-insulin is blocked[375]. Studies have shown that affected individuals who had isolated amylase deficiency have impaired carbohydrate absorption and the obtained symptoms[376,377].

The risk of developing MFD in modern populations is influenced by nutritional factors. One of the areas of genomic research on amylase productivity is the epigenetic modification of the genome as the cause of the MFD. Epigenetic modification is an inherited change in the phenotype or expression of genes caused by mechanisms other than DNA sequence changes. Genes can act as intermediaries for epigenetic modifications of other genes[378,379]. Extreme lifestyle in the past has an impact on genetic change in populations. In current conditions, the risk of developing metabolic and other diseases increases when there is a carbohydrate deficiency in the supply in the region. In case of inconsistency like nutrition with the genetic needs of the organism, various pathological diseases can arise. With age, epigenetic changes occur, which explains the development of MFD in older people. Research by Christensen et al. revealed about 300 loci, which are characterised by age-dependent methylation, and many of these genes have associations with MFDs in the elderly (for example, for the *ESR1* gene with cancer and obesity)[380]. Throughout almost the entire life of an adult person, the rate of alpha-amylase is not prone to change, and only in older people does its interval slightly expand. The lower limit decreases, and the upper increases.

The studies mentioned above show that amylase is one of the essential components of the human body. Because of this, research in the field of CNV amylase is relevant.

General review of methods used to determine *AMY* CNV

There are various research methods for identifying CNV (Table 2). These methods are used to detect DNA sequences, gene copy number changes and gene deletions/duplications. Table 2 shows the advantages

and disadvantages of the methods used in genetic research[381]. At the same time, there are three main locus-specific methods used to determine the CNV genotype more accurately in a large number of samples: MLPA, qPCR, and PRT. There are often inconsistencies between these methods[321].

The MLPA method can cover a large number of loci in one test, including target and reference, in one reaction. A multiplex PCR assay called MLPA uses up to 40 probes. Each probe analyses the corresponding CN in each DNA sequence and is exclusive to a particular DNA sequence. This method analyses up to 50 DNA sequences in a single reaction. Additionally, it can detect CNV, including minor intragenic rearrangements[381]. The method of multiplex ligation-dependent amplification is usually used in clinical genetics to screen disease genes for exons and can determine an integer number of copies even at complex multicopy loci[382]. However, this is a relatively expensive method per sample and requires much time, including overnight hybridisation steps. It also forces you to use a large amount of DNA samples[383].

Most studies of associations between *AMY* CNV and disease have been conducted primarily by qPCR[317,384]. qPCR is used to type CNV using fluorescence reporter systems such as TaqMan. However, the method is sensitive to changes in DNA quality and sometimes gives inaccurate results at multicopy loci[339]. qPCR is very sensitive to the quality of template DNA, causing systematic distortions that can result in false-positive or negative connections of the disease[321]. The advantage of qPCR is the versatility and simplicity of the technique[385]. PRT, unlike other methods, is known for its simplicity and the ability to use different designs[385]. A feature of PRT is, in most cases, the reliability of the results. The PRT method is capable of detecting low-frequency sequence variations[321]. PRT is based on CNV typing studies using small amounts of DNA with high throughput[386]. The PRT was developed initially to eliminate different PCR heterogeneities between reference and target genes[387].

PRT is a comparative PCR technique designed to determine the CN in DNA based on the amplification of dispersed repetitive sequences. The essence of the method is based on the construction of exact primers, where one set of primers is used for both reference and test loci, amplifying two similar sequences[387]. Since a matching pair of primers is used to amplify the test and reference regions with the same sequence, they target more segmental duplications or deletions and repeats.

Amplification products of test and reference loci have similarities in sequence and consequently similar amplification properties[388]. In the first step, the CNV region is divided into overlapping segments. For each segment, a variety of primer pairs are selected, regardless of the type of sequence. The amplicon is then aligned to the genome in order to choose the one that matches the sequences in the genome twice (CNV locus and reference sequence)[321,387,389,390]. This leads to improved accuracy in the measurement of the CN. With PRT, PCR is performed with suitable cycle numbers to remain in the growth phase of amplification, in which quantitative relative yield is expected. The sizes of products and their yield are measured by capillary electrophoresis. The area and peak ratio of the samples obtained are compared using fluorescent dyes. In the PRT technique, this ratio of the test and reference loci makes it possible to calculate the number of gene copies. This technique is better suited for assessing complex CNV sites. Studies have noted a high speed of CN detection using PRT of 93% for the beta-defensin locus and 85% for *CCL3L1*[387,391]. PRT costs are relatively low, especially in extensive studies determining association and CN. PRTs do not show this systematic bias and determine the copy number in the correct range; therefore, it is more accurate for measuring the copy number. Comparing PRT results with earlier studies based only on qPCR, there was an inaccuracy in CN. Another advantage is the low consumption of DNA samples (up to 20 ng). In terms of accuracy, the PRT shows good results, as well as the MLPA and MAPH (multiplex amplification and probe hybridisation) methods, which have low error and high accuracy of the results[387].

It should be noted that when using PRT, the reference loci only sometimes distinguish the CN, making it challenging to find paralogous sequences beyond the region of study. The drawback is technical limitations that affect the differences in sequences in areas with high sequence identity. An additional limitation is a low throughput of locus multiplexing compared to other methods (MLPA and MAPH)[389]. Another disadvantage of PRT is a slight underestimation of CN per unit[321]. However, accurately calibrated PRT can help with precisely detecting CNV. Another disadvantage is that it does not allow for determining the structure and origin of CNV.

Method	Advantages	Disadvantages
MLPA	Detects small rearrangements Up to 40 targets High throughput Low cost	Cannot detect copy neutral loss of heterozygosity. May have problems with mosaicism, tumor heterogeneity, or contamination with normal cells.
FISH	Detects balanced rearrangements Detects mosaicism Detects tumor heterogeneity Can quantify multiple copies	Cannot detect copy neutral loss of heterozygosity. Cannot detect small rearrangements (e.g., deletions <100 kb or duplications >500 kb). Limited number of targets and throughput.
Quantitative/Sq-PCR	Detects small rearrangements and even point mutations Can quantify multiple copies Low cost	Test optimization and efficiency is a concern. Limited number of targets. May have problems with mosaicism, tumor heterogeneity, or contamination with normal cells.
Southern blot	Detects small rearrangements Detects mosaicism	Cannot detect copy neutral loss of heterozygosity. Not quantitative. Laborious and time consuming Limited number of targets and throughput.
CGH array	Can detect very small rearrangements Can probe entire genome Low cost per data point	Cannot detect copy neutral loss of heterozygosity. Costly equipment and reagents Low throughput
SNP array	Can detect copy neutral loss or heterozygosity Can probe entire genome Low cost per data point	Cannot detect small rearrangements (e.g., deletions or duplications <100 kb). Costly equipment and reagents Low throughput

Table 2. Advantages and disadvantages of different methods used in genetic analyses. Extracted from Stuppia et al[381].

2.2 Aim

The aim of the study was to investigate the association between the CNV of human pancreatic amylase genes and pancreatic diseases, such as pancreatitis and pancreatic cancer, in particular by measuring the CN of the *AMY1, AMY2A* and *AMY2B* genes in cases and controls.

2.3 Materials and Methods

Study Design

In the study, we proposed to examine the connection of *AMY1* and *AMY2* CNV with pancreatic diseases such as acute and chronic pancreatitis or pancreatic cancer by measuring the CN of the *AMY* gene in cases and comparing them with unaffected controls. Blood samples were used from 124 selected patients with verified diagnoses and receiving treatment. These samples were obtained from the Division of Gastrointestinal Surgery of QMC (Queen's Medical Centre, Nottingham), and was accompanied by additional information, including

diagnosis, gender, age, and body mass index (BMI). All patients were of European race. (Supplementary material, table 1).

In addition, we compared our results with data from 418 samples for *AMY2* and 472 samples for *AMY1*, which were used as a control of healthy people of European descent from the UK. Data from panels 1-5 of the European Collection of Cell Cultures (ECACC) for Human Random Control (HRC) were used (https://www.phe-culturecollections.org.uk/collections/ecacc.aspx). To assess the quality of measurements and comparative analyses for *AMY1* and *AMY2A/B*, 2704 measurements were performed.

The UK control DNA samples were taken from the HRC project as they were better suited for our study. Genotyping among healthy individuals (HRC) was carried out by the control groups in the studied region, which was characterised by significant variability depending on geographic and ethnic conditions. The concentration of DNA samples was diluted to a volume of 10 ng/μl. To measure the CN of the *AMY* gene for cases and controls, we used the PRT method with the selection of primers for the studied loci, which has been previously used in other studies[335,392].

DNA extraction

To obtain a DNA sample, 7-8 ml of blood was received from the Division of Gastrointestinal Surgery. A volume of 2-3 ml was mixed with solution A (consisting of 10 mM Tris HCl, pH 8.0, 320 mM sucrose, 1% Triton X-100) in a centrifuge tube buffer and adjusted to 15 ml. The use of solution A is necessary for the gentle lysis of cells to keep the nucleus intact for further collection by centrifugation. Centrifugation occurs at 1500 rpm (RCF 1700g) for 5 minutes. In order to avoid the dispersion of the nuclear pellet, the supernatant was carefully removed. Solution B (consisting of 400 mM Tris HCl, pH 8.0, 60 mM EDTA, 150 mM NaCl, and 1% SDS) in a volume of 350μl was added to the resulting precipitate. Solution B is used to release DNA from proteins and solubilise nucleic acids. Then 100μl of 5M sodium perchlorate was added. After stirring, 500μl of phenol/chloroform/isoamyl alcohol was added. The function of phenol is to denature the protein, but not DNA, allowing phase separation in which DNA remains in an aqueous solution. Still, denatured proteins are found at the chloroform/water interface.

The sample was centrifuged for 5 minutes at a speed of 13K. Then the upper (aqueous) layer of the liquid where the DNA was located was removed. 900 μl of 100% undiluted ethanol was added to the isolated aqueous portion. Then the samples were centrifuged again for 3 minutes. Washing of the DNA pellet was carried out using 1 ml of 80% ethanol after the preliminary removal of the liquid. After drying in the last step, 100 μl of distilled water was added. The extracted DNA was stored in a refrigerator at 4-8 ℃ for 48 to 96 hours to allow full dissolution. All extracted DNA samples were tested for concentration on Nano-drop spectrophotometry and 1% agarose gel electrophoresis. In spectrophotometry, an absorption coefficient of 260/280 was used to assess DNA purity. An A260/A280 ratio estimated in the range of 1.7-1.9 is considered to indicate satisfactory purity. DNA concentrations were maintained between 10 and 40 ng/μl. The final working sample was a DNA solution diluted to 10 ng/μl.

PRT assays with *AMY1*

AMY1 assay is based on distinguishing between *AMY1 A/B/C* variants. In this assay, two groups of primers were used: PRT1 and PRT2. Three pairs of primers for *AMY1*, targeting different repeat regions, were combined in one PCR reaction. A total of 6 primers were used, 3 for each group. Products for amplifying one group of primers (PRT1) were used AMY1C-F, AMY1C-R-Hex and forward primer NF2; for the second group (PRT2) AMY1C-F, AMY1C-Rb2-Fam and NF5 forward primer were used[335,392]. These primers amplify products from each copy of *AMY1A* and *1B*, using *AMY1C* as the reference locus.

The products were amplified using 1 μM of each of the primers AMY1C-F, AMY1C-R-Hex, NF2, AMY1C-Rb2-Fam and NF5. Primers AMY1C-F and AMY1C-R-Hex create the starting products from *AMY1A* and *AMY1B* at 633bp and *AMY1C* at 624bp, after which NF2 and AMY1C-R-Hex create *AMY1A/B* products at 427bp and 436bp from *AMY1C*. AMY1C-F and AMY1C-Rb2-Fam have initial longer products at 872bp from *AMY1A* and *AMY1B* and 884bp from *AMY1C*, and then NF5 and AMY1C-Rb2-Fam create 344bp products from *AMY1A* and *AMY1B* and 357bp from *AMY1C*. For each reaction, the following components were used separately for NF2 and AMY1C-R-Hex, NF5 and

AMY1C-Rb2-Fam (Figure 7A). For them, for each sample, the mix contained 10 × LD mix (1x LD mix contains 50 mM Tris–HCl (pH8.8), 12.5 mM ammonium sulphate, 7.5 mM 2-mercaptoethanol, 125 μg/ml BSA, 1.4 mM $MgCl_2$, and 200 μM each dNTP)[393] - 1 μl, (10 μM) AMY1C-F 1 μl, Taq polymerase (5U/μl) 0.1 μl, H_2O 4.9 μl. The process was carried out with initial denaturation at a temperature of 95 ℃ for 5 minutes, then 15 cycles were used, consisting of a stage of denaturation at a temperature of 95℃ for 30 seconds, annealing at a temperature of 61℃ for 30 seconds and elongation at a temperature of 65℃ for 2 minutes.

The second stage of the analysis was moderated by cycles 14 for NF2 and AMY1C-R-Hex, and 13 for NF5 and AMY1C-Rb2-Fam. The denaturation step was carried out at a temperature of 95℃ for 30 seconds, annealing at 54℃ for 30 seconds, and elongation at 65℃ for 1 minute. The last extension was carried out at a temperature of 72℃ for 50 minutes. To remove excess primer products, an AMPure purification step was performed, where all four PCR products resulting from *AMY1* and *AMY2A/2B* analyses were combined using a volume of 2.5 μl of each, which totalled 10 μl. Then, 18 μl of AMPure product was added. The resulting precipitated DNA was washed twice with 70% freshly prepared ethanol. In the end, it was dissolved in distilled water in an amount of 10 μl.

PRT assays with *AMY2A* and *2B*

AMY2 analysis was performed by amplifying the duplication junction and *AMY2A:2B* ratio fragment. For the duplication junction assay, the products were amplified using primers AMY2B2F, FAM-AMY2B2R and AMY2B2D[392]. For the *AMY2A:2B* ratio, the primers FAM-AMY2A2B-F and AMY2A2B-R[335] were used. For our assay, in contrast to the previously described methods, the cycles were adjusted to improve the quality of the result: 24 for duplications and 27 for the ratio. The technique was modified by identifying the correct numbers of PCR cycles. To analyse the duplication junction, the products were amplified using primers AMY2B2F, FAM-AMY2B2R and AMY2B2D, which amplify a 192bp product upstream of *AMY2B*, corresponding to primers AMY2B2F and FAM-AMY2B2R. The primers used for the duplication product at 176bp are FAM-AMY2B2R and AMY2B2D. FAM-AMY2B2R

is used to identify and label both products, as this primer is fluorescent (Figure 7A). The following components are prepared for 1 PCR sample: 10 x LD PCR mixture (1x LD mix contains 50 mM Tris–HCl (pH8.8), 12.5 mM ammonium sulphate, 7.5 mM 2-mercaptoethanol, 125 μg/ml BSA, 1.4 mM $MgCl_2$, and 200 μM each dNTP)[393] - 1 μl, (10 μM) FAM-AMY2B2R -1 μl, (10 μM) AMY2B2F - 1 μl, (10 μM) AMY2B2D - 1 μl, Taq polymerase (5U/μl) 0.1 μl and H_2O, 4.9 μl. The resulting product was distributed in wells of 9 μl, and 1 μl of 10ng/μl of DNA was added. Several cycle numbers were tested in the thermocycler, and the definitive process finally used 24 cycles. The denaturation step was preliminarily carried out before the start of the cycles, which were kept at 95°C for 5 minutes. Then there were the three-cycle stages of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds and an extension of 65°C for 1 minute. The process was completed with the final extension step at 72°C for 50 minutes.

To analyse the ratio *AMY2A:2B*, the primers FAM-AMY2A2B-F and AMY2A2B-R[335] were used. A 163bp product from *AMY2A* is amplified, and a product of 167bp is amplified from *AMY2B*. The fluorescent label here is FAM (Figure 7A). The following was used to prepare the product: a mixture of 10 x LD - 1 μl, (10 μM) FAM-AMY2A2B-F - 1 μl, (10 μM) AMY2A2B-R - 1 μl, Taq polymerase (5U / μl) 0.1 μl, H_2O 5,9 μl; 9 μl was also dispensed into each well, and 1 μl of 10ng/μl of DNA was added. The cycles were analysed by various options. As a result of the analysis, the most effective cycles were selected, which made it possible to obtain a result for *AMY2*. For the analysis of the ratio between *AMY2A* and *2B*, cycles were set in 27. Then, a basic denaturation step was carried out at 95°C for 30 seconds, annealing at 58°C for 10 seconds, and an extension of 61°C for 30 seconds with a final extension at 72°C for 50 minutes.

DNA fragments were separated by capillary electrophoresis. Fragment analysis was performed by electrophoresis on an ABI 3130 xl 36 cm capillary, where 170 μl of HiDi and 2 μl of ROX-500 were taken for 16 samples (15 DNA samples and 1 control) were mixed and 10 μl distributed into each sample. The resulting solution was mixed with 2 μl of the product after purification with AMPure, then denatured at 96 °C for 3 minutes.

Measurement of *AMY1* and *AMY2* copy numbers.

Based on the ratio of reference and test products, the CN was assessed in the test region. Fragment assays of fluorescently labelled PCR products were carried out by electrophoresis on an ABI3130xl 36 cm capillary using POP-7 polymer with an injection time of 30 s at 1 kV. Capillary electrophoresis for product imaging is marked with a 5'-fluorescent label with the appropriate FAM or HEX primers. In the results obtained, the peak areas were analysed using the PeakScanner 2 software (Applied Biosystems) to represent them in the form of the ratio of corresponding peaks.

The final assessment of copies for *AMY1* was carried out by measuring two ratios *AMY1AB:AMY1C*, where *AMY1C* is present in two copies.

The CN estimation of peaks for *AMY2A* and *2B* was made by measuring the duplication sequence of *AMY2A* and *2B* with an additional measurement of the *AMY2A:AMY2B* ratio. For CN *AMY2B* of 2 copies, the duplication sequence is supported by more than one likely number of copies. The calculation of the peak ratio of *AMY2A* to *AMY2B* was made in the area of 163bp and 167bp. The analysis for *AMY2A:2B* duplication junction assay was analysed by the ratio in the areas of the constant 192bp and the duplication-specific 176bp product peaks.

An analysis of clinical diagnoses was carried out by considering the results of CNV.

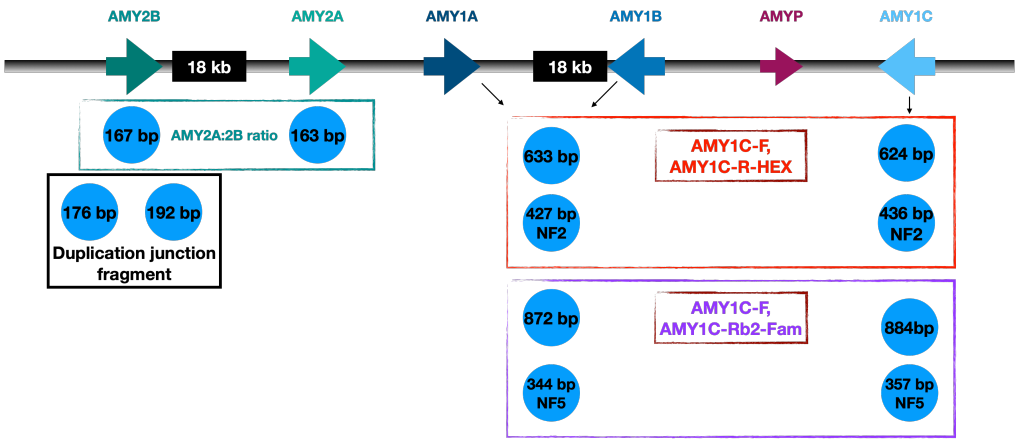


Figure 7A. The diagram shows the arrangement of primers for the amylase gene CN assays. Boxes show the amplification of a group of primed products for *AMY1*, PRT1 and PRT2. PRT1 is shown as a red

box, which includes primers AMY1C-F, AMY1C-R-Hex and forward primer NF2. The second group is PRT2, marked with a purple box, which includes primers AMY1C-F, AMY1C-Rb2-Fam and forward primer NF5. The *AMY2* assay consists of the amplification of the duplication junction fragment when present and the *AMY2A:2B* ratio, which is green in the box. For the *AMY2A:2B* ratio, primers FAM-AMY2A2B-F and AMY2A2B-R were used. The black box indicates the primers for assays of the duplication junction fragment. Product length is marked for each gene.

Statistical analysis

To test the correlation relationships between the values: Pearson's correlation coefficients, simple regression, and one-way and multi-factor dispersion analysis were used. A p-value at a significance level of 0.05 was taken as an indicator of the statistical reliability of the results. To verify the correspondence of the empirical frequency distribution of genotypes to the theoretically expected χ^2 test of goodness of fit. To determine a statistically significant difference between two or more categorical groups simultaneously and to calculate the correlation between them, an interaction test analysis of variance (ANOVA) was used. If more than two variables from the same sample are considered, repeated measures ANOVA is usually used. The main reason why using ANOVA is preferable to re-comparing two samples at different factor levels using a t-test is that ANOVA is more efficient and, for small samples, also more informative.

2.4 Results

General review of the PRT results

CNV shows correlated measurements for *AMY1* and *AMY2* in the studied cohort of people. The indicated CN for *AMY1* and *2A/B* were based on the conclusions drawn from PRT analysis. To assess the measurement of ratios and CNV, Peak Scanner software was used to represent it in the form of corresponding peaks (Figures 8 and 9).

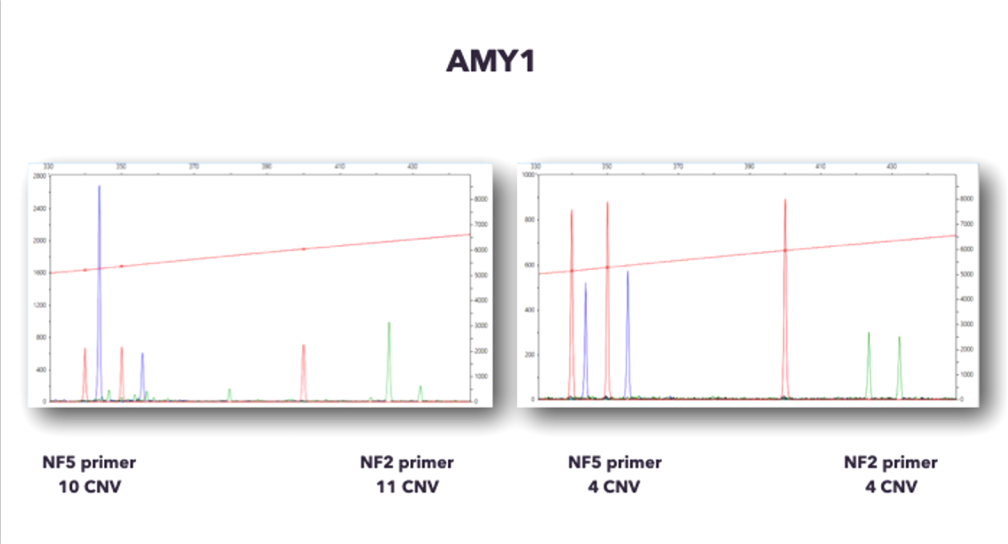


Figure 8 shows the amplification of *AMY1*. The NF5 nested primer produces the product *AMY1C* at 357bp, as for *AMY1A* and *AMY1B* at 344bp, which are marked in blue. The second primer, NF2, produces the product *AMY1C* at 436bp and *AMY1A* and *AMY1B* at 427bp, which are marked in green. Red peaks are fluorescent size markers. The number of copies was calculated by determining the ratio between the peaks.

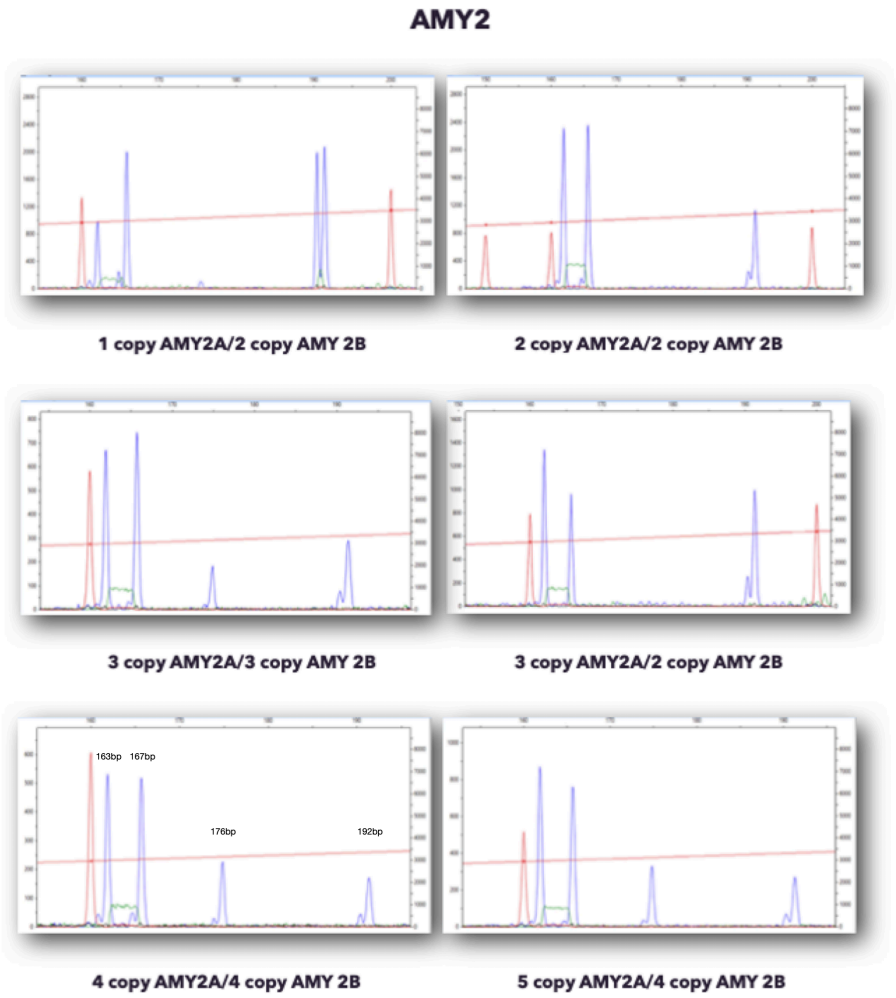


Figure 9 shows the product amplification from *AMY2A* at 163bp and for *AMY2B* at 167 and 192bp. A fragment of a duplication junction is shown at 176bp, indicating duplication of *AMY2B*. To visualise *AMY2* products, a fluorescent FAM (blue) is used. The number of copies is calculated by determining the ratio between the peaks. The number of copies is indicated below each figure for an individual product.

Cluster analysis of *AMY1*

To assess the accuracy of the *AMY1C* PRT measurements and compare the results of the NF2/NF5 primers, a Cluster analysis was used, which subsequently served to determine a more accurate calibration for the CNV measurement. To plot the graph, a total of 1379 measurements of peak ratios from case samples and controls were used. Both primer systems are used to evaluate the outcome of the *AMY1C* PRT assay. The ratio of two amplified peaks, one corresponding to *AMY1A + AMY1B* and the other to *AMY1C*, was taken from the PRT analysis for each sample with the corresponding primer (NF2 or NF5). These ratios were used separately for each of the primers, NF2 and NF5. The consistency of the relations between results from the two primers was shown in the correlation between the NF2 and NF5 ratios and the integer values represented in the form of clusters which are assumed to centre on integer copy number values (Figure 10). The process revealed a pattern between the results of PRT *AMY1C* and CNV analysis and demonstrated the resulting regression line. The constructed linear calibration indicates a positive regularity between the PRT *AMY1C* and CNV analysis data. The analysis showed the correlation between the two measurements, which amounted to R^2=0.83. The values from 0 to 1 are explained by the fact that the closer the coefficient is to 1, the stronger the dependence.

By cluster calculation, it was determined that the results of both primers give a small difference in their individual calibration, which can be estimated as one common result because, for any sample, there will be a single correct integer value. This technique helped to improve the calibration method and derive a specific formula for the CNV measurement equation with minimal errors by rearranging the regression line equations with the best fit.

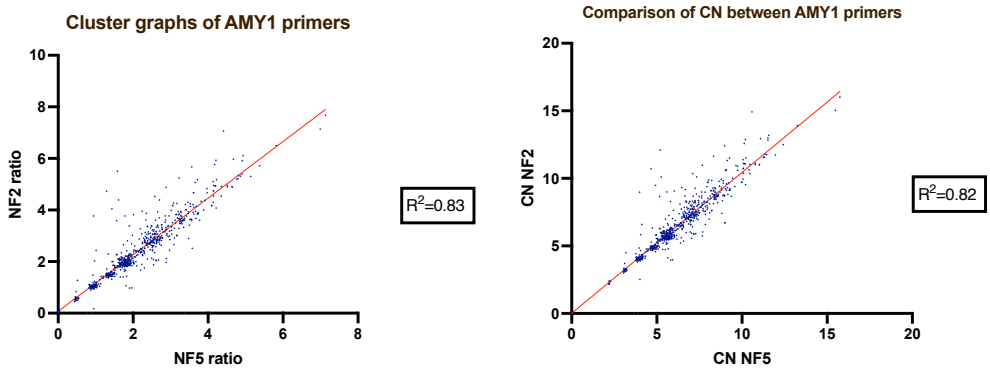


Figure 10 shows the calculations using cluster analysis. (A) shows the comparison of ratio measurements between the two primers for 1379 measurements. (B) shows a comparison of the *AMY1* CN of two primers after calibrating the values. Clear clusters are observed up to CN of 7. Scattering measurements start a from high copies (7 CN and above). The coefficient of determination amounted to R^2=0.83, which shows the correlation of CN between the two measurements.

AMY2A and *2B* measurements distribution

To calculate the distribution of the ratio of *AMY2A/2B* and duplication junction fragments, normal distributions were fitted to the mean and standard deviation of the two main clusters of ratio and duplication junction fragments assay results from the study (Figure 11) used for *AMY2A/B* ratios and *AMY2B* duplication measurements, where the total number of measurements was 1325. From this histogram, we can understand how the values of the *AMY2A* and *2B* samples are distributed relative to the average values. The SD (relative to the mean) in the two subset histograms is 0.05. This indicates a small variability in the data, and the values are considered consistent and positive. This method reveals the degree of variability of measurements of samples.

To test the correlation relations between the measured values, a histogram was used. In the following histogram (Figure 11A), the mean value was 0.55 (N=173). For cluster B, the mean value was 1.05 (N=567) (Figure 11B). In addition to the normal distribution of measurements, a 1.05 (Figure 11B) may indicate equal numbers of *AMY2A* and *2B*. A 0.55 (Figure 11A) may indicate the deletion of *AMY2A* (1 copy *AMY2A*/2 copy *AMY2B*). Around 1.5 means one more copy for *AMY2A* than *AMY2B* (3 copies of *AMY2A* and 2 copies of *AMY2B* or 4 copies for *AMY2A* and 3 copies for *AMY2B,* etc.*)*. In samples with an independent duplication of *AMY2A* and one less copy for *AMY2B*.

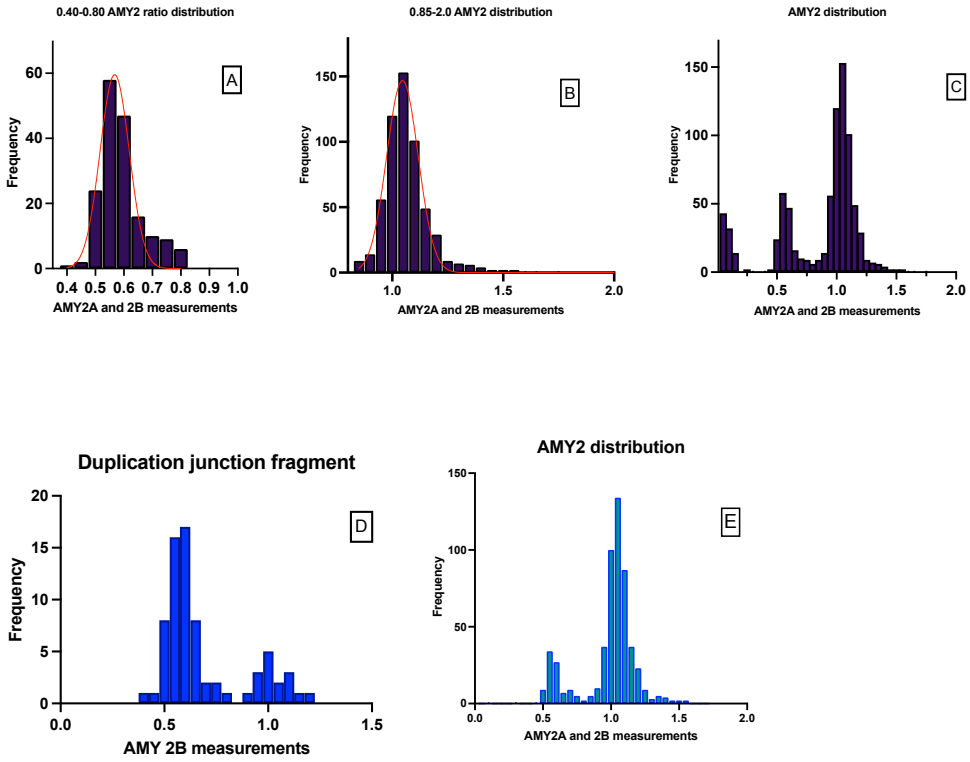
The graph shows that for the samples, the ratio of equal copies and heterozygous deletions of *AMY2A* prevails. The construction of the curves was based on the dependence of the average value and their deviation. The red curves in both histograms also show the normal distribution. The general histogram (Figure 11C) shows the combined data obtained from *AMY2A* and *2B* (N = 1325). Plotting does not include data corresponding to 0 measurements (N = 493) related to the

minimal 2 copies of *AMY2B*. Results close to 0 show a homozygous deletion of *AMY2A* (N = 92).

Duplication junction fragment analysis associated with *AMY2B* showed measurements with two mean values, around 0.5 and 1. The histogram (Figure 11D) shows a mean value of around 0.5 (N = 56), which corresponds to 3 copies of *AMY2B* and additionally gives 1 copy for *AMY2A*. A mean value of around 1 (N = 16) corresponds to 4 copies for *AMY2B* and an additional 2 copies for *AMY2A*. Additional copies for *AMY2A* and *2B* are duplications.

This graph shows the distribution of *AMY2A* and *2B* using data from 124 samples of patients with pancreatic disease and HRC controls 418 samples (no duplicate assays). Plot E shows a difference from Plot C, showing the homozygous deletion of *AMY2A* (N=2).

In general, the graphs show that the distributions of the samples in each cluster correspond to the normal distribution curve.



Figures 11 represent histograms of the measurement frequencies. Graph data were analysed by fitting normal curves (red) with the same mean and standard deviation. Figure A shows the distribution between 0.40-0.80 measurements. Figure B shows the distribution between 0.85 and

2.0 measurements. C shows the overall data set of the measurements (N=832). Figure D represents the distribution of the measurement's histogram with duplication junction fragment analysis. Figure E shows the distribution of the measurement's histogram from the affected patients with pancreatic diseases and HRC control.

Evaluation of *AMY1* CNV results

In accordance with the purpose of the study, 124 samples for cases and 472 samples for controls were used to assess the *AMY1* CNV. The *AMY1* study revealed a variation from 2 to 16 copies (Figure 12). Two to five copies were identified in cases ranging from 3 to 10 samples and 8 to 91 in controls. Six to 16 copies were identified in cases ranging from 1 to 34 samples and 1 to 123 samples in controls. Figure 12 shows the gradual growth of low copies (4 and 5 copies) in controls and a decrease in the number of high copies starting from 6. In these cases, we observed a shift towards higher CN. Comparative analyses used specific copy numbers as low and high ranges for cases and controls. This comparison of the SN results allowed us to determine the threshold value. As a cut-off for grouping representing low and high CN, a copy number of 6 copies was determined.

Comparative analysis of the two groups in cases showed a 3.5 times increase between low (2 to 5) and high (6 or more) CN (from 22% to 78%). Analysis of the distribution of CNVs by groups showed that the group with control samples split was 38% to 62% (a difference of 1.5 times). In the first group of copies (2-5 CN), the difference between the case and the control changed upwards by 1.7 times in the control, compared to the cases. In the second group (6–16 CN), the number of cases was 1.3 times higher than in the control (Figure 13). The correlation of CNV *AMY1* between case samples and controls was assessed using the chi-square test of independence. The analysis showed that the *AMY1* gene CN is significantly associated with the risk of pancreatic diseases ($p=1.5x10^{-2}$).

The correlation between low and high *AMY1* CN between cases and controls showed $p=6x10^{-4}$ (OR=0.4475, 95% CI 0.05798 to 0.1893, RR=0.5192), meaning the higher the CN, the higher the risk of disease.

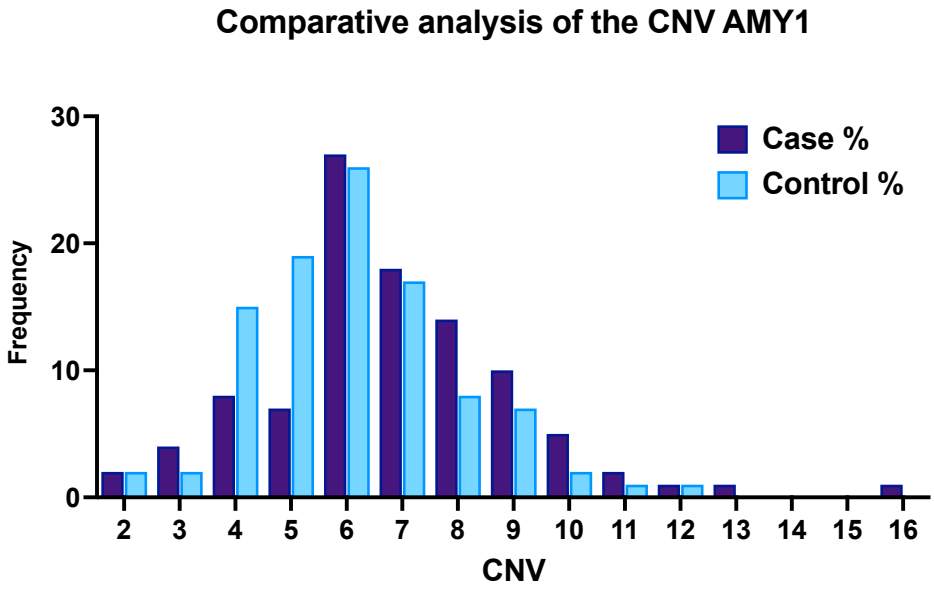


Figure 12. Distribution of *AMY1* variation in cases (124) and control (472) samples. CN ranged from 2 to 16. The difference starts from the 6 CN (34 samples - 27% for cases and 123 samples - 26% for controls). The highest 16 copies are found only in cases.

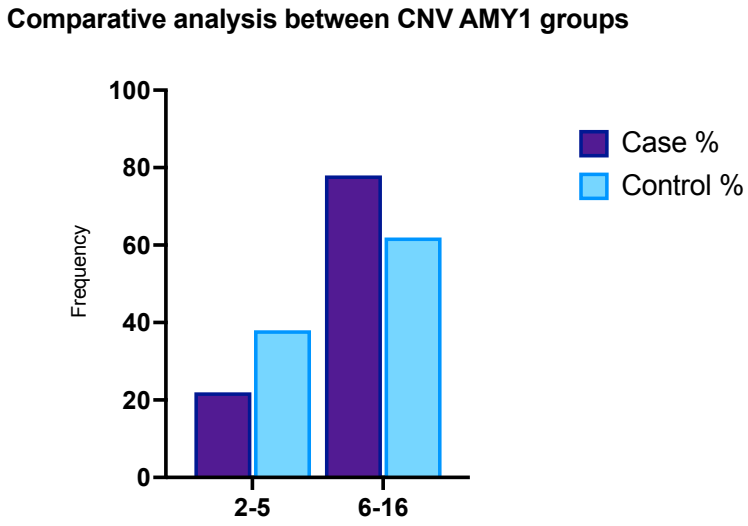


Figure 13. Comparative analysis revealed an increase in the number of *AMY1* CNV in case samples and a decrease in controls. The graph demonstrates a significant increase in CN frequency with a high CN (≥6) in sample cases.

Evaluation of *AMY2* CNV results

To assess the effect of *AMY2* CNV on the susceptibility to pancreatic diseases, 124 cases and 418 controls from healthy subjects were studied. The study revealed that the total number of *AMY2A/B* ranged from 0 (for *AMY2A*) and 2 (for *AMY2B*) to 5 copies for *AMY2A* and 4 for *AMY2B* (Figure 14). Deletions could be homozygous, leading to the complete absence of *AMY2A* (0:2) or heterozygous in the form of 1 copy (1:2) for *AMY2A* with 2 copies of *AMY2B*, or duplication of 3 copies for *AMY2B* (1:3). The most common result was 2 copies for *AMY2A* and *AMY2B*, found in 80 (64.5%) cases and 295 (71%) controls. A distinctive feature is an absence in cases of 2:3 copies (2 copies of *AMY2A*:3 copies of *AMY2B*) compared to the identified six samples in controls. There are also no examples of results in cases such as 0:2 (0 copies for *AMY2A*:2 copies for *AMY2B*), 3:4 (3 copies for *AMY2A*:4 copies for *AMY2B*) and 1:3 (1 copy for *AMY2A*:3 copies for *AMY2B*) compared to controls where they are found in small numbers. It is possible that these copies are not typical for diseases of the pancreas and indicate the carriage of special copies or characteristics exclusively for healthy individuals. Both cohorts had 3:3 duplications (single heterozygous for duplications) (3 copies each for *AMY2A* and *2B*), with 11.3% for cases and 7% for controls and homozygous 4:4 duplications or heterozygous for triplication (4 copies each for *AMY2A* and *2B*), 1.6% in cases and 1% in controls. Duplication of *AMY2A* alone to produce 3:2 copies (3 copies of *AMY2A*:2 copies of *AMY2B*) occurs in just over 7% of cases compared to 2% of controls. 5:4 (5 copies of *AMY2A*:4 copies of *AMY2B*) occurs in 2% versus 0.2% in controls.

When calculating the ratio of duplications and deletions of *AMY2A* and *2B*, it was found that an increase in the number of duplications was found in cases (22%) compared with the controls (12.7%) and a decrease in the number of deletions (14%) in cases against (16%) in controls. The ratio between duplications:deletions for cases was 1.6 and in controls 0.8. This illustrates the difference between the two indicators separately for cases and controls. The ratio between case duplications and control duplications was about two times. The ratio between case deletion and control deletion was 0.9. This suggests that the difference in duplications is higher between cases and controls. The lower deletions in cases and controls compared to duplications indicate the significance of the duplications for pancreatic disease. (Figure 15). Comparative analyses were made for *AMY2* as a whole. Individually,

AMY2A duplications versus deletions showed significance ($p=1.7x10^{-2}$, OR=0.4228, 95% CI 0.2111 to 0.8527, RR=0.5355). At the same time, *AMY2B* has at least two copies, so the significance of the deletion/duplication cannot be analysed separately. In this connection, χ^2 is calculated independently for *AMY2A* and not for *AMY2B*.

To confirm the hypothesis of an increase in the number of copies and an association with pancreatic diseases, a statistical analysis of the correlated relationship between the measured values was performed. *AMY2A* copy numbers were significantly higher in affected patients compared to the healthy group ($p=2.8x10^{-2}$), with no significant correlation in *AMY2B* (p=0.3346).

The chi-square test shows a significant correlation between 2:2 copies, deletions (0:1 and 1:2 copies) and duplications (all copies above 2 to 5 copies) with a p-value of $4.6x10^{-3}$. Copies 2:2 and duplications for *AMY2A* showed $p=4x10^{-3}$. *AMY2* calculated by ANOVA showed significance $p=2x10^{-4}$.

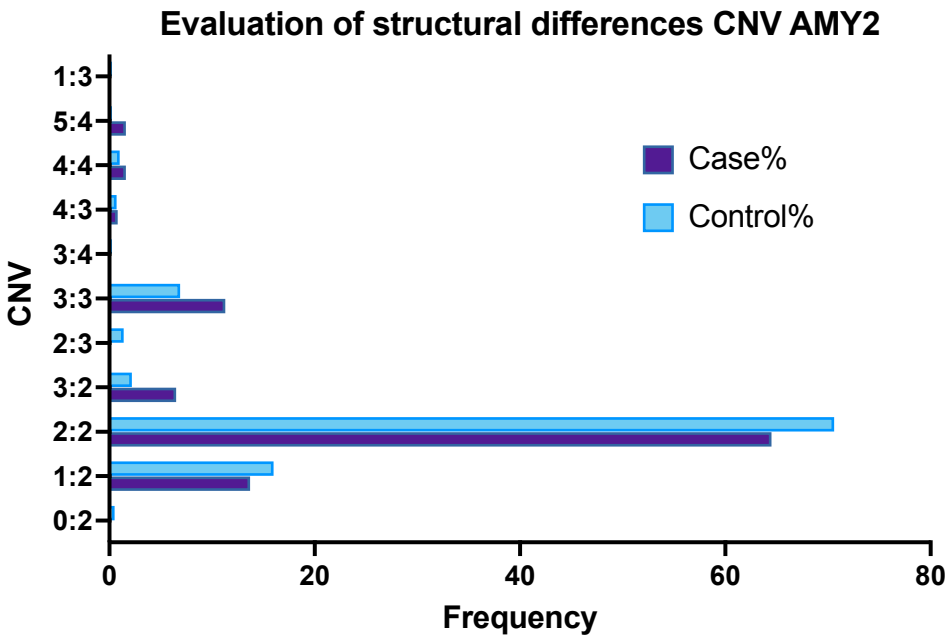


Figure 14. CN for *AMY2A* and *2B* range from 0 to 5 copies. The graph shows the frequency of *AMY2A* deletion (0 or 1 copy for *AMY2A*:2 or 3 copies for *AMY2B*), duplications only for *AMY2A* (3:2; 3 copies of *AMY2A*, 2 copies of *AMY2B*), duplications only for *AMY2B* (2:3; 2 copies of *AMY2A,* 3 copies of *AMY2B*), duplications of *AMY2A/B* (3:3, 3 copies of *AMY2A/B*; and 4:4, 4 copies of *AMY2A/B*), uneven

duplication of *AMY2A/B* (3:4, 3 for *AMY2A* and 4 for *AMY2B*; 4:3, 4 for *AMY2A* and 3 for *AMY2B*; 5:4, 5 copies of *AMY2A*, 4 copies of *AMY2B*). Most of the patients were carriers of 2 copies of *AMY2A* and *2B*. Copies 0:2, 1:3, and 2:3 are found only in controls. Copy rates for 3:4 and 5:4 are low in controls and close to 0.

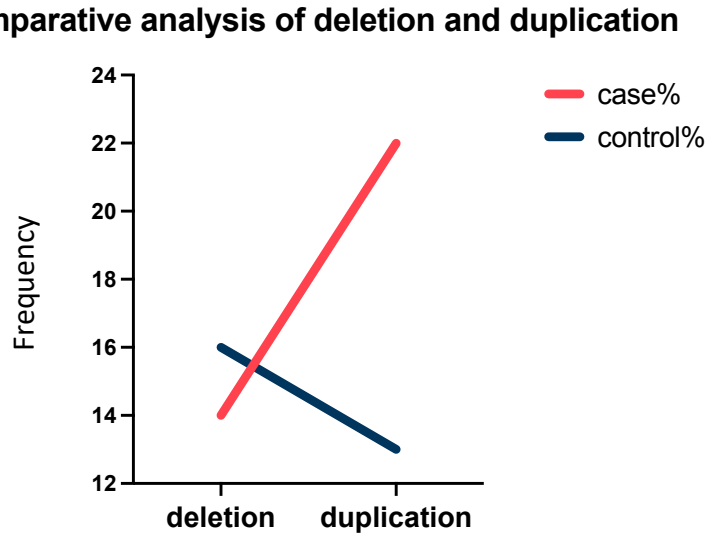


Figure 15. The graph shows the overall *AMY2A* and *2B* deletion and duplication in cases and controls. In cases, there are 27 (22%) duplications and 17 (14%) deletions. For controls, duplications were 53 (12.7%), and deletions were 67 (16%). The ratio of deletion/duplication in cases was 1.6 and in controls 0.8. The occurrence of duplication shows a characteristic increase in the case of samples compared to control samples and opposite numbers with a deletion.

Age estimation in cases and controls

The "Global Burden of Diseases, Injuries, and Risk Factors Study" meta-analysis found that age rates of disease at 44-49 for men and 60-64 for women[394]. Considering the average age of cases and controls in our study, the average value for the age threshold was taken. Thus, to identify age characteristics and the likelihood of developing the disease risk, the age threshold was taken from 40 and older.

In our study, the samples of cases, there is a change from 27% to 73% in the direction of increasing age; the difference was 2.7. For controls, these indicators ranged from 54% to 46% in the direction of decreasing age; the difference was 1.2. The mean age in cases was 53 (SD=16.5) years, and in controls was 38 (SD=8.9). SE=0.34, SP=0.38. (Figure 16,17). The correlation dependence between ages (case and control) was $p<0.0001$ ($\chi^2=18.82$, RR=0.50, OR=0.32, 95% CI 0.19-0.54).

A separate comparative analysis of the correlation between age and *AMY1* CNV in cases and controls was also carried out. In the group of patients, the level of CN is significantly higher than in control. There was an increase in indicators in cases in the group with low CN (1–5) by 2.4 times; in the group with a high CN (6–16), three times in the direction of increasing age (≧40). The opposite tendency is observed in the control samples, respectively, 1.4 and 1.2, toward decreasing age (<40). The correlation dependence in the group with low CN was $p=3.2x10^{-3}$ ($\chi^2=6.2$); in the group with high CN, it was $p<0.0001$ ($\chi^2=16.04$, OR=0.2815). At the same time, the presence of a high copy number and an increase in age significantly increases the risk of morbidity (Figure 18).

AMY2 is characterised by age-related changes with deletions, as well as in the range of two and three copies (Figure 19).

Calculations by ANOVA test indicate a correlation between *AMY1* and *AMY2* CNV and age ≧40 ($p<0.0001$).

As a result, we observe a correlation associated with age 40 and above.

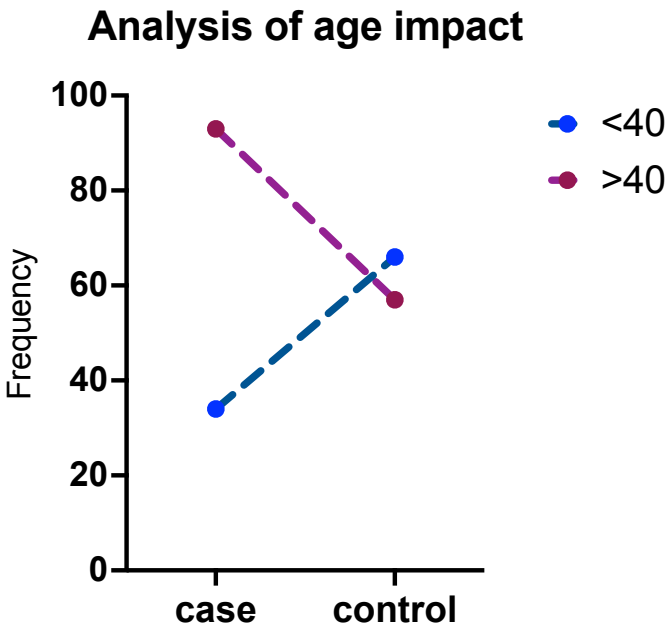


Figure 16 shows the analysis of age indicators for two groups. The threshold value was taken as an average age of 40 years. The dotted lines show an increase in indicators ≥40 years in cases and an increase in <40 years in controls.

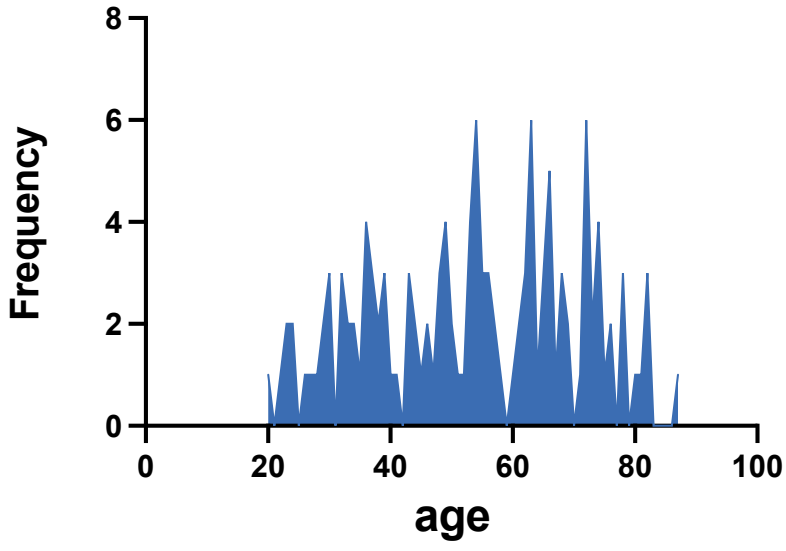


Figure 17. The graph shows the distribution of ages among patients with pancreatic diseases. The highest incidence of the disease occurs in the age category, around 50-70.

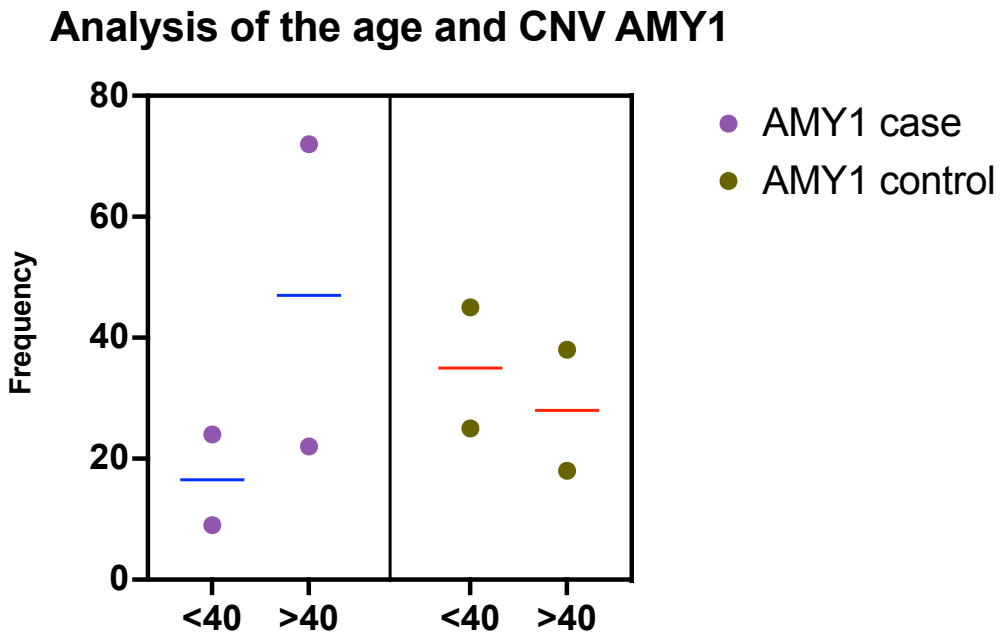


Figure 18. The graph shows the direct and reverse tendencies of the relationship between age and *AMY1* CN. The bottom point shows low copies (2-5 CN), and the top point shows high copies (6-16 CN). The graph shows the distribution of age for cases with low CN is 9 for <40 and 22 for >40 age, with high CN is 24 for <40 and 72 for >40. In the control group, with low CN is 25 for <40 and 18 for >40, with high CN is 45 for <40 and 38 for >40.

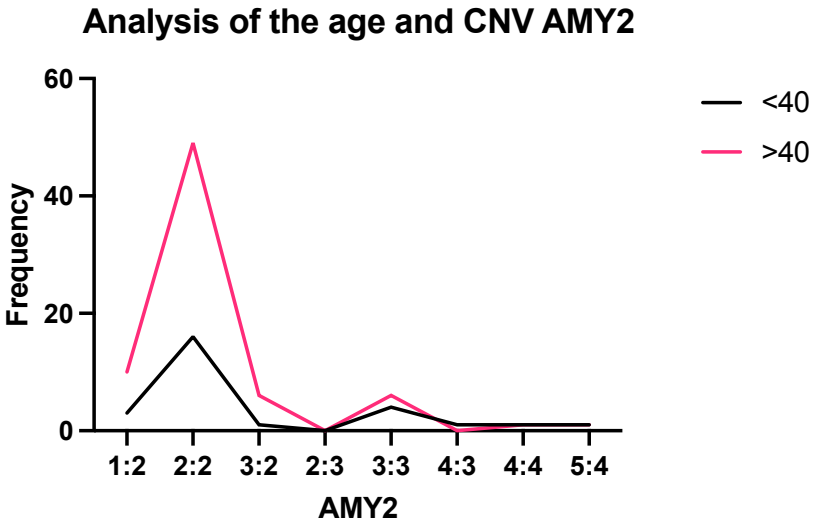


Figure 19. The graph shows the age difference among copies in *AMY2* in cases. In cases with *AMY2*, there is an increase in the age numbers

starting with deletion and delivering a peak in 2 copies of *AMY2A* and 2 copies of *AMY2B*, subsequently decreasing in duplication for patients of both age groups. A significant difference is shown between the two copies (16 in below 40 years, 49 in above 40 years). In the deletions (3 in below 40 years, 10 in above 40 years) and 3:2 copies (1 in below 40 years, 6 in above 40 years), the difference is minor.

Gender analysis in cases and controls

Analysing the frequency of pancreatic diseases by gender, we observed that women (61%) appeared at a higher frequency among cases than men (39%). The gender numbers in the controls were 49.8% for males and 50.2% for females (Figure 20). The association of gender between cases and controls was significant ($p=2.6x10^{-2}$, RR=0.78, OR 0.63, 95% CI 0.60-0.97). For diseases of the pancreas, the influence of gender is significant. The correlation coefficient was made between variables according to their absolute value in order to establish the tightness of the relationship.

Analysis of the data on the distribution depending on the gender of patients showed that women prevailed over men both in the group with a low CN of 3.8 and in the group with a high CN of 1.4 times. In cases, the incidence of *AMY1* with low CN for men was 4%, and for women, 15%. For *AMY1* with high CN, it was 33% for men and 45% for women. High copy rates (12,13 and 16 CNs) were found only in women (Figure 21).

AMY2 is characterised by female-related changes with deletions, as well as in the range of two, four and more copies. Males have fewer deletions and more 3 copies. Three copies of *AMY2A* and two copies of *AMY2B* are similar in both genders. Figure 22 shows that women tend to develop pancreatic disease more often than men.

Calculations by ANOVA test indicate a correlation between *AMY1* ($p=8x10^{-4}$), *AMY2* CNV ($p<0.0001$) and gender. Pancreatic diseases in females are more prevalent, and association is more significant than in men.

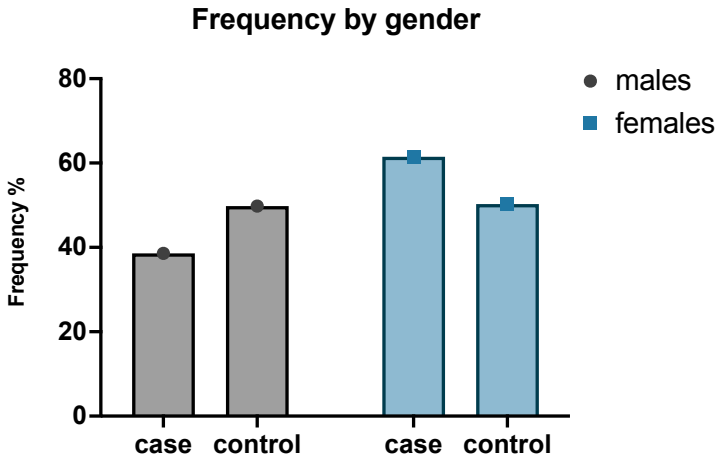


Figure 20. The graph shows the occurrence of individuals with pancreatic diseases in women and men, comparing healthy individuals. Y-axis show the frequency of gender in per cent. X-axis show differences between cases and controls.

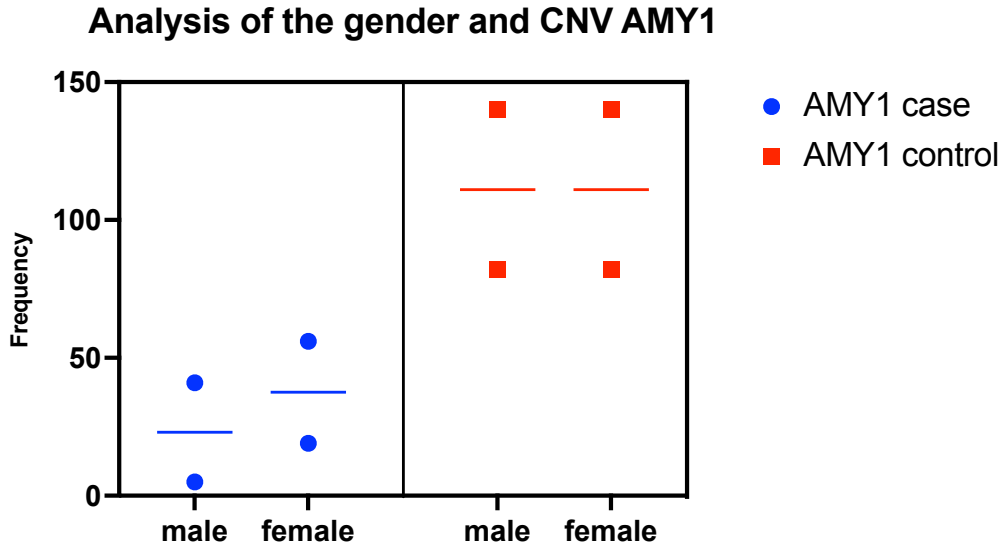


Figure 21. The graph shows the distribution of gender by the CN of *AMY1*. The bottom point shows low copies (2-5 CN), and the top point shows high copies (6-16 CN). In cases of CN *AMY1* in women, the growth and development of pancreatic diseases are more common than in men. The graph shows the distribution of gender data for cases with low CN for 5 men and 19 women, with high CN for 41 men and 56 women. The control group, with low CN for 82 men and women with high CN for 140 men and women.

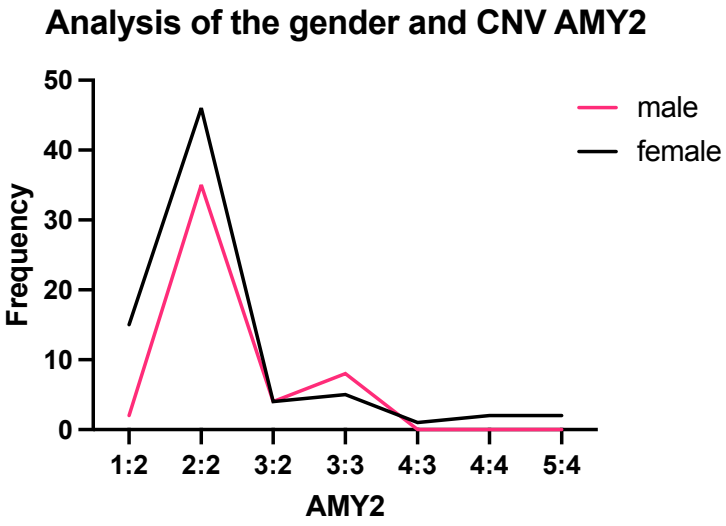


Figure 22. The graph shows the gender difference among copies in *AMY2* in cases. The graph shows the occurrence of copies of *AMY2* in both genders. The 2 copies of *AMY2A* and 2 copies of *AMY2B* showed an advantage in 46 women compared to 35 men. The deletions showed a great difference between 15 women comparing two men. The 3 copies of *AMY2A* and 3 copies of *AMY2B* were found predominantly in 8 men compared to 5 women. Both genders are characterised by the same in 3 copies of *AMY2A* and 2 copies of *AMY2B* (4 each). The 4 copies of *AMY2A* and 3 copies of *AMY2B*, 4 copies of *AMY2A* and 4 copies of *AMY2B* and 5 copies of *AMY2A* and 4 copies of *AMY2B* were found only in women, not men.

BMI estimation in cases

The BMI analysis used 124 samples of individuals with pancreatic disease. Figure 23 shows the frequency of certain BMIs in patients with pancreatic disease. The average values found in patients with a normal BMI (<25) were 25% and obesity 1 level (30–34) 24%, respectively. The most significant percentage is observed in patients with increased body weight (25-29), 33%. Obesity 2 levels (35–39) showed 9%, and 3 (>40) showed 14%. The limitation was that we did not determine the relationship of BMI between cases and controls due to the lack of BMI data in controls of healthy individuals.

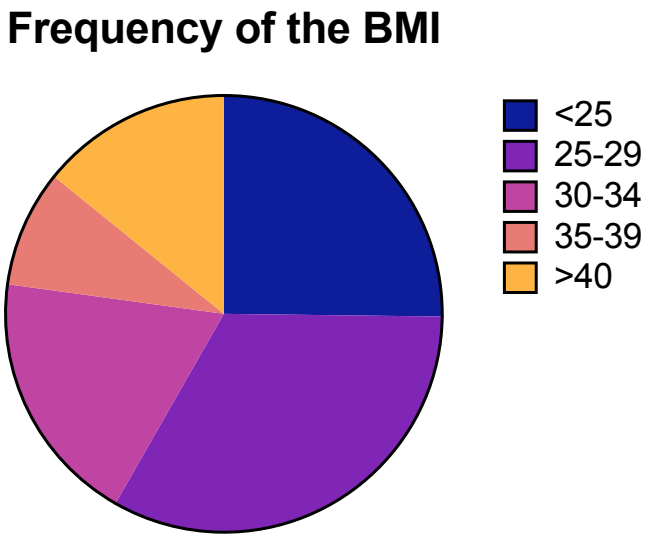


Figure 23. The graph shows the distribution of BMI in patients with pancreatic diseases. BMI was divided into 5 groups. The figure shows the BMI of the patients with pancreatic diseases in each section. The major two sections show the BMI in the <25 and 25-29 categories.

Analysis of the diagnoses of pancreatic diseases according to the study results

Clinical features and the presence of a detailed description of the patient's phenotype can also help correctly interpret the obtained data. The main types of pancreatic diseases are presented in the examined group of patients, namely acute and chronic pancreatitis or cancer. The analysis examined the structural characteristics of pancreatic disease diagnoses (Figure 24).

If, in general, the correlation between pancreatic diseases was $p=6x10^{-4}$ for *AMY1* and $p=2x10^{-4}$ for *AMY2A*, this proves the association between CN and diseases. Types of pancreatic disease were analysed and divided into groups. Correlations were found between acute pancreatitis and *AMY1* with $p=1.96x10^{-2}$, with *AMY2A* $p=2.38x10^{-2}$, with *AMY2B* p value was 0.3222. With acute pancreatitis, we received similar results as with the overall results. The correlation between diseases such as pancreatic cancer and chronic pancreatitis requires more data to obtain results.

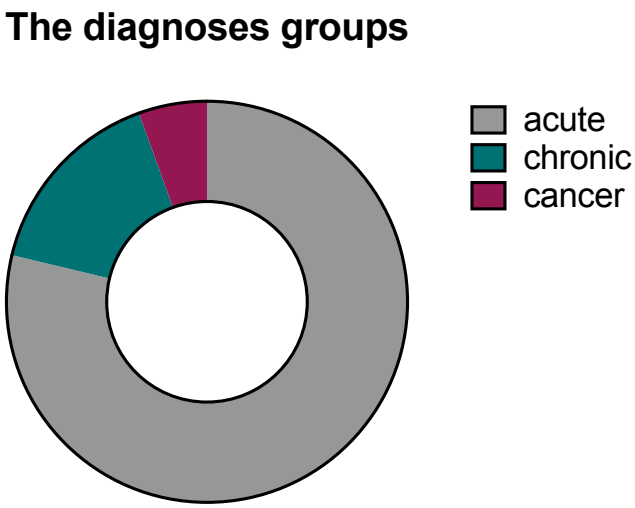


Figure 24. The graph shows the analysis of the distribution of patients depending on the clinical phenotype. Of most of the patients, 79% have acute pancreatitis, whereas chronic pancreatitis, 16%. Cancer cases account for 5% of patients.

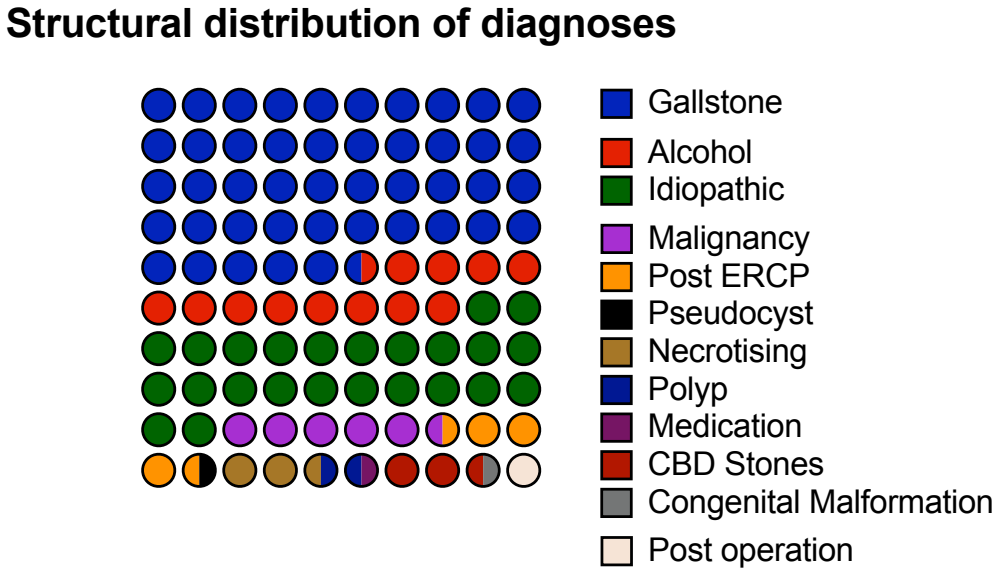
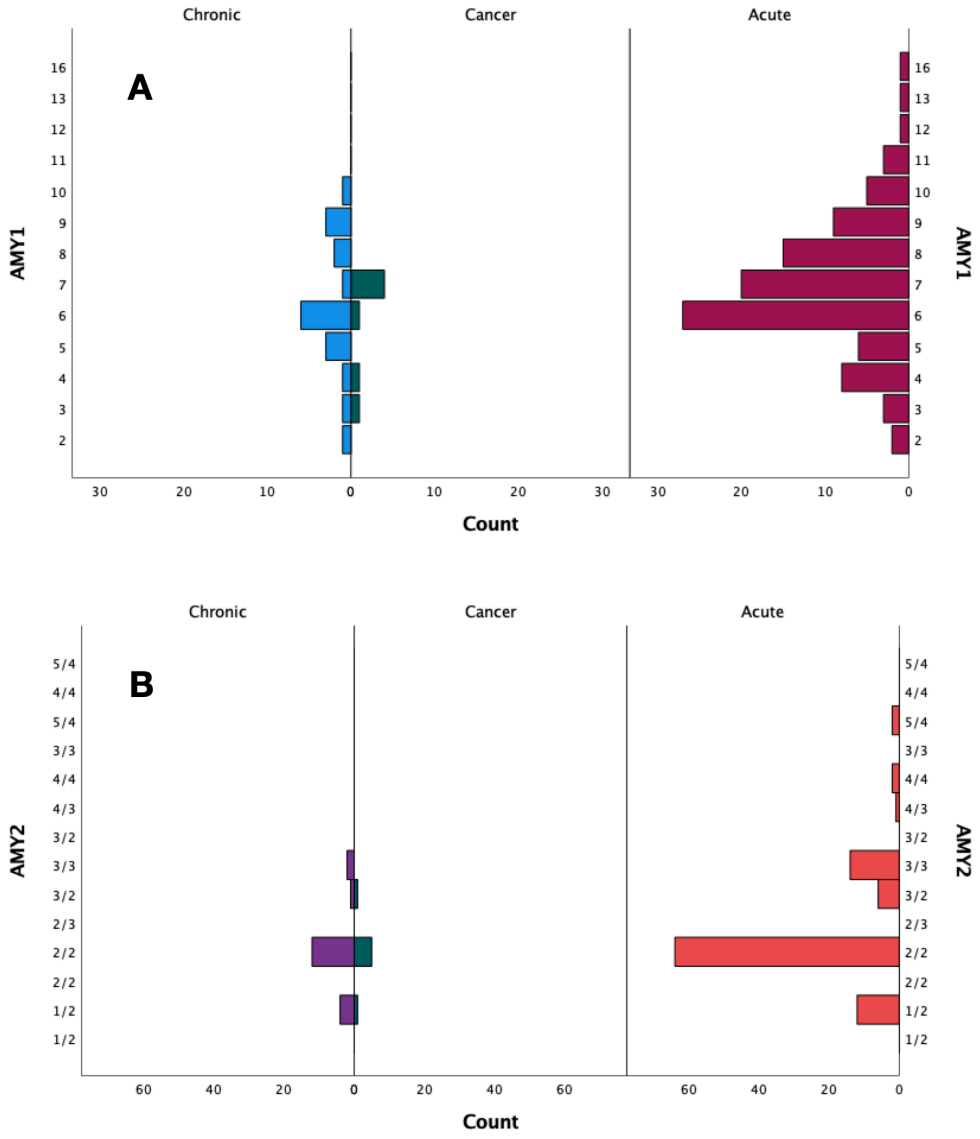


Figure 25 shows a comparative assessment of the structure of diseases of the studied patients.

With a comparative assessment of the disease's structure, a number of diseases can be distinguished (Figure 25). Gallstone developed in 45% of patients, the Idiopathic form of the disease was

observed in 24%, and Alcoholic pancreatitis in 13%. At the same time, among patients with destructive forms of the disease, 5% had a total lesion of the pancreas like Malignancy. Five diseases states occurred at 1% each (Pseudocyst, Polyp, Medication, Congenital Malformation, Post-operation pancreatitis), and three diseases at 2-4% each (Post ERCP, Necrotising, CBD (common bile duct) Stones).

Analysis of the CNV *AMY1* and two genes and their contribution to the development of diseases should help to determine the nature of the disease more reliably. The structure of the groups of diagnoses demonstrated different levels of CN variability in different diseases of the pancreas (Figure 26 A/B and 27 A/B).



Figures 26 A and B show the presence of *AMY1* and *AMY2* CNV in three groups of pancreatic diseases. The obtained results suggest that the development of acute pancreatitis is accompanied by the presence of all CNs for *AMY1* and *AMY2*. The development of chronic pancreatitis involves nine types of CN *AMY1* (2-10 CNV) and four types of CN *AMY2*.

Gallstone disease occurs in all 16 *AMY1* CNV. Most of the case samples are 6, 7 and 8 CN *AMY1*. The majority of the copy numbers for *AMY2A* were 1 (deletions), 2 and 3 (duplications) copies. High copies (5 copies) were identified only in gallstone disease, and four were found in small numbers. For *AMY2B*, 86% had two copies; the rest were 3 and 4. The study showed that the incidence of the idiopathic form of pancreatitis among patients is more in 5-7 CN *AMY1*. For *AMY2*, the CN range mainly was in 2 copies, and equally deletion (1 copy) and duplication (3 copies); for *AMY2B,* the majority were in 2 copies, and some were in 3 and 4 copies. Alcoholic pancreatitis, the third most crucial diagnosis, occurs within 6,8,9 CN *AMY1*. *AMY2A* is present in 1 (deletions), 2 and 3 (duplications) copies; 4 copies are the minority. *AMY2B* is only present in 2 and 3 copies, where 73% are 2 copies. In our studies, the highest indicator for cancer is characterised by 7 CN *AMY1*. Two CNs are typical for *AMY2A* and *2B*. Pseudocyst, Polyp, Medication, Congenital Malformation, Post-operation pancreatitis, Post ERCP, Necrotising, and CBD stones are all presented in copies of higher 6 (Figure 27 A and B).

The limitation of our study is that all other diseases (such as Pseudocyst, Polyp, Medication, Congenital Malformation, Post-operation pancreatitis, Post ERCP, Necrotising, and CBD stones) occurred in small numbers of samples. This made it challenging to complete an analysis of pancreatic disease structures.

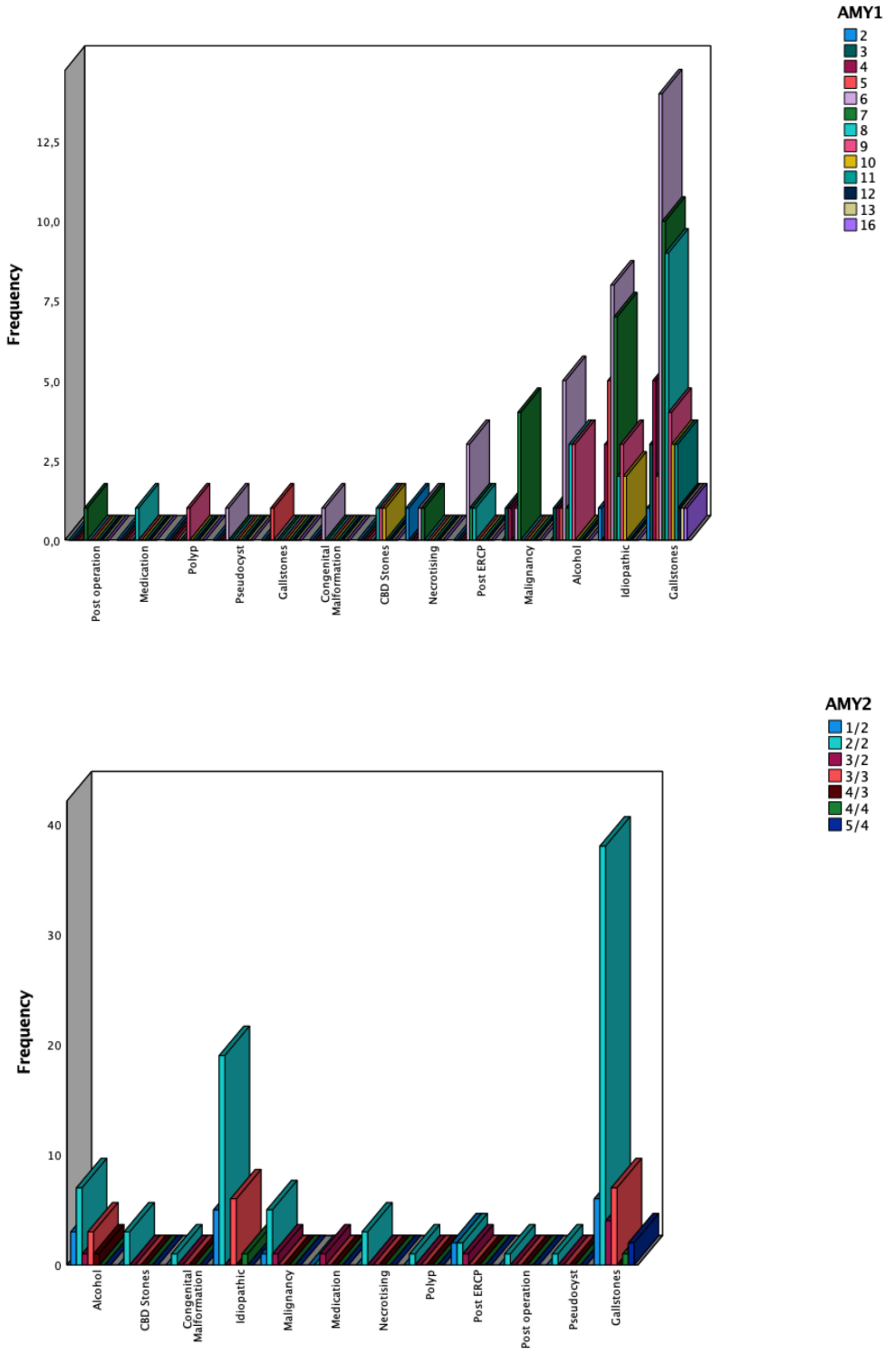


Figure 27 A and B. Two graphs show the CN of *AMY1* and two by disease structure. Figure A shows the distribution of *AMY1* CNs in each disease. The major CN is six. The 6 CN presents in six diseases. The second major CN is 7 presents in six diseases. The highest 16 CN is presented in the Gallstone. Figure B demonstrates the distribution of *AMY2* CNs in each

disease. The highest distribution of 2 CN of *AMY2A* and *2B* presented in all diseases. The highest 5 CN of *AMY2A* and 4 CN of *AMY2B* are presented in the Gallstone.

2.5 Discussion of the association and prognosis of pancreatic diseases based on the study's results

Genetic variants of susceptibility to multifactorial diseases can be represented not only by SNPs but also by larger structural variants of the genome, such as CNV. The frequency of changes in the CN of DNA regions is 3-4 orders of magnitude higher compared to single nucleotide substitutions[120]. The phenotype of CNV amylase is not well investigated. CNV analysis has its advantages and limitations. Often, the formation of individual responses in an organism is determined not only by changes in the genes themselves but also by the number of copies of this gene. So, it was hypothesised that an abnormal loss or gain of copies of some important genes could provoke various diseases[343,395]. The study of CNV in patients with pancreatic diseases contributes to expanding the understanding of pancreatic morphogenesis, as well as related disorders. The influence of exogenous or random factors can be combined, modified or cumulative, leading to a change in the risk associated with a specific CNV.

Our study used data from 124 patients to look for evidence of CN at the amylase gene locus. The analysis showed an association between higher CNV and pancreatic disease. The study determined that in the blood of patients with pancreatic diseases, the *AMY1* and *2* CNV were increased compared to the control group.

The study's results showed a variation in *AMY1* CNV from 2 to 16. The 16 copy was in a 56-year female patient with acute gallstone pancreatitis and BMI 27, with two copies of *AMY2A* and *2B*.

Comparative analysis showed an increase in the number of copies of *AMY1* in the case of samples compared to control samples ($p=6x10^{-4}$). These values indicate an association between the risk of pancreatic disease and *AMY1* in general.

For *AMY2A/2B*, CN from 0 to 5 was found in the case and control samples. The pancreatic amylase gene *AMY2A and 2B* was also significantly associated with pancreatic diseases ($p=2x10^{-4}$ between cases

and controls). *AMY2A* showed $p=2.8x10^{-2}$, *AMY2B* p=0.3346. *AMY2A* duplications are more significantly associated with the disease compared to deletions ($p=1.6x10^{-2}$). For the association of the disease with *AMY2A* and *2B* deletions and duplications, significance was $p=4.6x10^{-2}$; *AMY2A* duplications are more significant in cases compared with two copies ($p=4x10^{-3}$). Some CNVs were exclusively found only in control samples (0:2 copies, 0 copies of *AMY2A* and two copies of *AMY2B*; 3:4 copies, three copies of *AMY2A* and four copies of *AMY2B*) and 1:3, one copy of *AMY2A* and three copies of *AMY2B*) which is typical on our assumption for healthy people. Duplication cases were more common compared to the control group. According to the study, it is assumed that the risk threshold for developing pancreatic disease for *AMY1* is six copies or more. The study showed that high copy numbers of *AMY1* and *AMY2* were associated with the risk of developing pancreatic diseases.

A CNV range has been determined for each disease, which may assist in the prognostic diagnosis of pancreatic diseases. Redon et al. found that low proportion of deletions than duplications are linked with disease-related genes[96]. In our study, a RR calculation was performed to estimate the effect size of the determining influence of high and low CN. For *AMY1* RR=0.5192. Pancreatic disease with a high CN has a risk twice of pancreatic diseases than with a low CN. For *AMY2A* RR=0.5355, duplications prevail over deletions.

Comparative analysis of control samples and cases showed the development of pancreatic diseases for 40 years and above ($p<0.0001$). Analysis between gender and pancreatic diseases showed a correlation ($p=2.6x10^{-2}$). Females tend to be more susceptible to pancreatic diseases. According to calculations of the frequency of BMI for patients with a pancreatic disease, the occurrence of normal and slightly increased body weight is more characteristic.

Of particular importance in the development of diseases of the pancreas is the presence of a large number of samples of high (6-9) CN *AMY1*; at the same time, *AMY2* is characterised by a wide variability of copies. Accurate measurement of the *AMY* CNV is a difficult task since the difficulty in identifying the CNV *AMY1* gene lies in their potentially large number of copies (up to 18) and the presence of similar paralogues (between *AMY1, AMY2A,* and *AMY2B*)[335].

One of the genetic predisposition factors for pancreatitis and pancreatic cancer may be a change in the number of copies of the amylase gene. However, detailed studies of amylase and pancreatitis are not found in modern studies.

Previously, certain studies were carried out mainly in the field of studying interpopulation differences, where, on average, the number of repeats of the amylase gene was determined. In the research, it was found that the number of repeats of the *AMY1* and *AMY2A* genes significantly correlates with the geographical latitude of the localisation of populations (for *AMY2B*, no reliable relationship was found). Thus, the further north the range of the group, the fewer copies of the amylase gene in the population gene pool on average. Another important observation of intergroup features is the frequency of deletion of the *AMY2A* gene, which manifests itself in the phenotype in the form of insufficient synthesis[363]. In European populations, there are two common gene CN variants for *AMY2:* one carries a deletion of the gene *AMY2A,* and the other duplicates both *AMY2A* and *AMY2B.* There is an association between *AMY1* and *AMY2A* CN, where *AMY1* has a minor effect on the phenotype[335,396]. Studies by Carpenter et al. have shown that the *AMY2A* deletion allele has an incidence of nearly 7% in Europeans (1 in 200 people will be completely lacking in this gene product) and more than 60% for 2 copies of *AMY2A* and *2B*[335]. In our study, *AMY2A* deletions showed around 16% and 65% for 2 copies of *AMY2A* and *2B* in Europeans. The higher-order *AMY2A* and *AMY2B* amplifications were identified in Carpenter et al. studies mainly in the African population, which may also have duplication of *AMY2A* without *AMY2B*[335].

Genes *AMY1* and *2* have been well characterised in the population, and their role in starch digestion and influence on the organism has been shown in some research[335,363]. It should be noted that the results of general genomic associative studies have shown that contradictory results were found in different populations[397]. A significant association of the *AMY1* gene with populations with a diet rich in starch was shown by Perry et al. In this study, the qPCR measurement method and the CGH array were also used[343].

They are some GWAS studies in the field of the CNV *AMY* human amylase gene. For example, Nayema et al., in a study, showed an association between serum amylase levels in 814 Japanese with *AMY1* diploid CN ($p=1.89x10-19$). *AMY2A* CN ($p=0.54$) and *AMY2B* CN ($p=0.15$) showed no association. Diploid CN *AMY1* and BMI also showed no association ($p=0.14$)[398]. In studies by Falchi et al., it was shown that the magnitude of the effect of CNV *AMY1* on the risk of obesity was significantly increased comparing other genetic variants, such as gene polymorphisms associated with fat mass and obesity. Falchi et al. reported that BMI is inversely proportional to the CN of *AMY1.* A

high number of copies of *AMY1* is correlated with a low BMI, and a low number of copies is a contributing factor to obesity[350].

The contribution of the number of copies of *AMY1* to BMI has produced conflicting results. Studies conducted in large adult populations of Abruzzo, a region in the centre of Italy, showed a clear link between the low number of *AMY1* copies and an increased BMI in boys, not in girls, which was considered a gender-specific influence[353]. In other studies of three different adults groups, the association was not confirmed[346,399]. However, data from children's studies on variations in the number of copies of *AMY1* are different. A case-control study by Mejía-Benítez et al. conducted among Mexican children showed the effect of *AMY1* on obesity ($p=4.25x10^{-6}$), characterised by the influence of a large number of copies (>10) on reducing the risk of obesity[319]. Another study conducted on young Finnish patients with an early onset of severe obesity proved no relationship between the number of *AMY1* copies and obesity in young women[400]. Some studies have shown that obesity can have between 40% and 80% heritability. For the Meta-analysis were taken studies of the US or European versus East Asia. Meta-analysis identified 88 estimates of BMI heritability from twin studies (a total of 140,525 twins) and 27 estimates from family studies (42,968 family members). Twin studies showed that BMI heritability was from 0.47 to 0.90, and family studies showed 0.24–0.81. Meta-regression analysis found that the heritability of BMI from the twin study was 0.07 (P = 0.001), which was higher in children compared to adults. Children showed heritability estimates (0.11, 95% CI: 0.06–0.17, p=0.001) and adults (0.13, 95% CI: 0.008–0.26, p=0.038)[401]. In the Caucasian population, a relationship was found between BMI and 90 SNPs [402-404], which accounts for less than 5% of the observed variability[404]. The research carried out only one of the first attempts to systematise data on the relationship of the *AMY* gene with environmental factors. Several research studies have reported that a decrease in the *AMY* CN is linked with increased overweight and obesity[317]. It should be noted that other studies have demonstrated an association of low *AMY1* CN with the risk of obesity in European and Asian populations[353,400,405]. However, some studies have not identified an association of the CNV *AMY1* gene with obesity and disordered glucose metabolism in other European and East Asian populations[335,346,406]. Thus, some studies found a negative relationship between *AMY1* CNV and BMI. *AMY1* CNV (range 4 to 14) was shown to be inversely correlated with HbA1c (r=-0.36, P=0.003 and r=-0.30, P=0.02)[407]. Other studies have shown that the CN of *AMY1* is

lower in overweight children than in children with normal weight due to differences according to race and ethnicity. This effect is more pronounced in African Americans[408]. At the same time, research has shown that high CN of *AMY1* affects glucose metabolism and energy[319].

Also, the deletion of *AMY2A* and *AMY1* in 24% shows that the gene is (or linked to) a suppressor gene of carcinoma of the stomach since deletions of the chromosome region containing the *AMY2A* and *AMY1* gene were detected with high frequency in the tumours of Korean patients[409]. Thus, the loss of these genes in the indigenous population of Northeast Asia may indicate a genetic risk of stomach cancer.

CN *AMY1* has been implicated in a relationship with the activity of brain α-amylase. High activity of the brain α-amylase is more typical for women than for men[410]. Individual studies were considered depending on the factors of smoking and alcohol consumption. Low CNV *AMY1* values were associated with high insulin resistance in asymptomatic Korean men[318]. Other researchers did not associate high *AMY1* CN levels with insulin resistance and could not replicate the result[400,411]. No association was found between *AMY1* CNV and HOMA-IR (homeostasis model assessment-insulin resistance) among French adults[405], Australian adults[361] and the Finnish population[400]. However, another study found an association between high fasting insulin, HOMA-IR, and the risk of type 2 diabetes with a very high CN of *AMY1* (10 copies or higher)[338]. In the gendered study characterising pancreatic diseases, it was found that the rate of chronic pancreatitis in men is two times higher than in women. However, the incidence of pancreatic cancer and acute pancreatitis showed no difference between gender[214].

One of the reasons for the difference in the research results can be the use of different designs and research methods. Differences in research results can also be explained by using different CNV technologies for measurements. The second reason may be a difference in ethnicity and gender. Since previously reproducible studies with other methods, different age groups and populations gave different results[353,400,405]. The reason for this may be the distribution in populations of different variations of genetic polymorphism. Also, the reason may be the complexity of quantification in multi-allele loci, as indicated by differences in the evaluations of the number of copies of various laboratories in some research for the Japanese HapMap sample NA18972 with a large CN[335,343,346,354]. Characterisation of complex amylase loci, which mutate with high frequency in such a way that causes large-scale alters in the quantity and gene expressions, is necessary to determine the

genetic variability of a person and its relationship with phenotypes. *AMY1* gene locus is capable of rapid evolutionary adaptation. CNV of *UGT2B17*, α-globin, and *CCL3L1* are also shown evolution adaptation[121,412,413]. The study revealed the combined rate of all mutations, which amounted to $1.8x10^{-8}$ per generation. The rate change of variation coefficient was calculated for 20 loci and was 0.53. The link between mutation rates and loci was also determined[414]. Structurally complex loci can reflect mutations and also can be better characterised by combining different methods of various analysis, including direct molecular analysis[346].

2.6 Conclusion

Differences in the results of many studies can also be associated with the individual lifestyle, genetic origin, the prevalence of diseases, different statistics of risk factors and their influence on various diseases, habitat and living conditions, and diet[338]. Our study showed that a combination of factors such as age 40 years and above and high copy numbers, especially for women, has a more significant impact on the development of pancreatic morbidity.

In our opinion, the change in the *AMY* CNV can be considered part of the aetiology and pathogenesis of pancreatic diseases. The results confirming the involvement of CNV in the *AMY* gene in the development of the disease can help in further clinical developments in the field of pancreatic diseases. The use of genetic association studies can lead to identifying new opportunities in exploring new risk factors and targets. These studies can influence effective tactics of treatment for diseases of the pancreas.

CHAPTER Ⅲ. MUTATION OF N34S OF THE *SPINK1* GENE WITH PANCREATIC DISEASES

3.1 Introduction

The functional role of *SPINK1* is to defend the pancreas from excessive trypsinogen activation. It shows that in the majority of patients with *SPINK1* mutations, N34S acts as a modifier of the disease or as part of a multi-gene model along with other unidentified genetic, exogenous, and ecological co-factors of pancreatitis (Section 2.1). Some studies have proven that the N34S mutation of the *SPINK1* gene is an important risk factor for developing pancreatitis and, therefore, clinically important[415-418]. Mostly N34S mutation is linked with familial pancreatitis[419], idiopathic pancreatitis[241] and alcoholic pancreatitis[236,420]. Other mutation variants of the *SPINK1* gene are not so significant compared to N34S mutation[421].

3.2 Aim

The study aims to investigate the clinical and genetic features of the various form of pancreatitis and determine the significance and frequency of occurrence of N34S mutations in the *SPINK1* gene in the development and progression of pancreatitis.

3.3 Materials and Methods

The *SPINK1* gene (N34S mutation) design and sequencing

Two types of experiments were obtained in the N34S of the *SPINK1* gene mutation analysis. In the first experiment, the materials for the study included extracted DNA from the 1000 Genomes Project subjects, which had five samples (NA12347, NA18975, NA18940, NA18870, NA11930) with proven results of N34S mutation of the *SPINK1* gene and were used as a test sample and quality control to create primers and establish the method. In the second experiment, the materials were used for 124 unrelated patients with various pancreatic diseases from the European population. Blood samples were taken from patients with various pancreatic diseases from the Division of Gastrointestinal Surgery of QMC (Queen's Medical Centre, Nottingham). Genomic DNA was extracted from the whole blood samples (Section 2.3). The diagnosis was based on clinical presentation, disease history and proven examination tests.

PCR primers for both experiments of the N34S of the *SPINK1* gene were designed using the human reference sequences from the UCSC Genome Browser February 2009 assembly and Primer3 software:

SPINK1Forward TGGTCACTGAGGAAATGAAGC
SPINK1Reverse CAACACCCGACAATGCCTAG

These primers avoid common SNPs and amplify a 983bp product at hg19: chr5:147,207,173-147,208,155 centred on rs17107315 in exon 3 of *SPINK1*. To validate the mutation, all samples were subjected to PCR amplification followed by a Sanger sequencing reaction. The results were evaluated by using BioEdit software ABI chromatogram. The software displayed chromatograms of sequence alignments. Another method of identification of N34S mutation used *Tsp*RI (CASTG) restriction enzyme to replace the sequencing method to allow the assay higher throughput. The N34S variant creates a site for the restriction enzyme *Tsp*RI. The primers SPINK1F and SPINK1R amplify a 983bp product; *Tsp*RI cuts reference product into fragments of 12, **767**, 146, 16 and 42bp, but the S34 (C) variant adds a site to give fragments of 12, **499, 268**,146, 16 and 42bp. This method was chosen for its value in terms of time reduction during the assay process and the relatively low cost of the kit.

N34S detection assay

To make five samples with 20µl PCR reactions, a master mix was made, including *Taq* DNA polymerase enzyme 1 µl (5U), 10x "LD" PCR mix 10 µl, H_2O 74 µl, and 10µM SPINK1 forward and reversed primers (5 µl each). For each sample, genomic DNA (1 µl of 10ng/µl) was added to a 19µl portion of this master mix. 1 x LD mix contains 50 mM Tris–HCl (pH8.8), 12.5 mM ammonium sulphate, 7.5 mM 2-mercaptoethanol, 125 μg/ml BSA, 1.4 mM $MgCl_2$, and 200 μM each dNTP. PCR used a denaturation step at 95°C for 30 seconds, an annealing step at 60°C for 30 seconds, and an extension step at 70°C for 1 minute 30 seconds. Amplification used a total of 37 cycles. Amplified DNA fragments were separated using a 1.7% agarose gel in 0.5 x TBE in the presence of 0.5µg/ml ethidium bromide and 100bp DNA marker ladder (New England Biolabs). PCR products were purified using PEG/NaCl precipitation facilitated by magnetic beads (AMPure purification), checked on a 1.7% agarose gel, and cycle-sequenced using BigDye mix (Applied Biosystems). At this stage, 5 µl of purified PCR product was added to 5 µl of the prepared master mix; for five samples, this master mix included 5.5 µl BigDye, 11µl 5X sequencing buffer (containing Tris-HCl and $MgCl_2$), 8.3µl distilled water and 2.8µl of forward SPINK1 primer. Sequencing was done with 25 cycles of 96°C for 30 seconds, 50°C for 15 seconds, and 60°C for 4 minutes. Sequencing products were subject to CleanSEQ purification (ethanol precipitation) to remove unincorporated fluorescent terminators. 5 µl of CleanSEQ solution and 70µl of 85% ethanol were added to the DNA samples. After that, samples were washed twice with 100µl of 85% ethanol. In the end, ethanol was removed, and the dried beads were processed for capillary electrophoresis in the DeepSeq facility (University of Nottingham).

*Tsp*RI digestion of *SPINK1* gene PCR products

The restriction enzyme *Tsp*RI cuts at the sequence |NNCASTGNN^ (where | and ^ represent the bottom and top strand cut sites, and S = C or G). At the N34S site in *SPINK1*, the reference (N34) sequence is NNCATTGNN (AAT = Asn on the bottom strand, which does not cut with *Tsp*RI) and the variant (S34) sequence is NNCACTGNN (AGT = Ser on the bottom strand), which does cut.

After receiving the PCR products with the *SPINK1* gene, digestion with the *Tsp*RI enzyme was the next step. A 15µl master mix included all components except the PCR product, which was added in the amount of 5µl of each PCR product. The master mix contained 10x CutSmart buffer 2µl, *Tsp*RI enzyme (10U/µl) 0.2µl and H_2O 12.8µl in a total amount of 20µl for each product. Then the obtained samples were incubated at 65°C overnight (usually 22-24 hours). A 5µl loading dye for each sample and 2% agarose gel were used to check the product fragment lengths by electrophoresis.

Statistical analysis

For statistical calculation of the correlation between N34S mutations and diseases of the pancreas, the χ^2 test with p-value<0.05 was applied. As a control group, the results of population samples in the UK TWINS cohorts were used, which are subsets of the UK10K survey. A frequency of 1 (3574) was used for the calculation, specifically for the UK population (https://www.ensembl.org).

3.4 Results

PCR and *TSP*RI enzyme analysis of the N34S mutation of the *SPINK1* gene

In our experiment, 124 individuals from the QMC with various pancreatic diseases were analysed, in which four of them were identified as carriers of rs17107315 (N34S).

At the beginning of the study, all samples were used to amplify the *SPINK1* gene using the PCR method. The optimal annealing temperature for the *SPINK1* primers was found to be 60°C. The *SPINK1* product has a length of 983bp. The gel lanes at each side show 100bp ladder size markers. The *SPINK1* PCR product was detected in both N34S carriers and non-carriers (Figure 28).

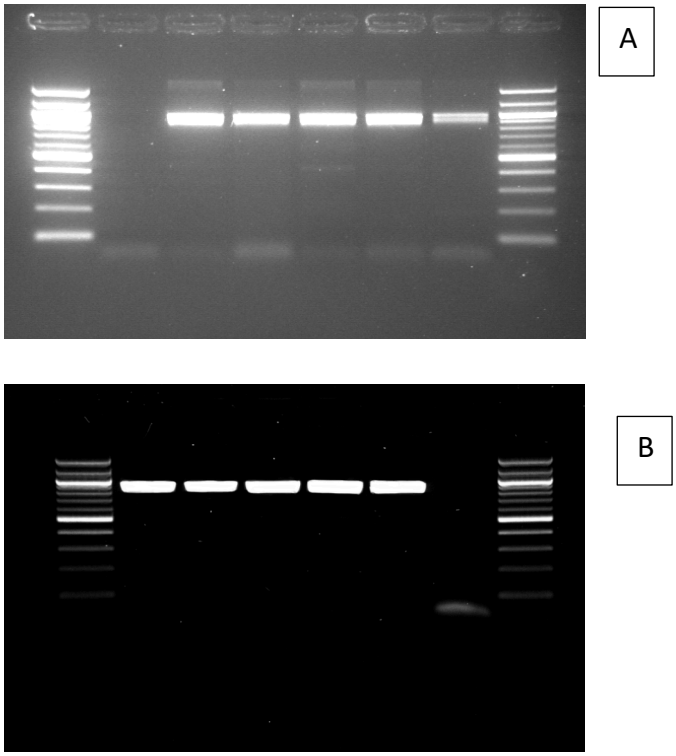


Figure 28. PCR analysis of *SPINK1* gene. Picture A of PCR products from experiment 1. The first lane is a no-DNA control, from second to fourth are carriers, and the last two are non-carriers. In picture B, PCR products from experiment 2. The first four lanes are the carriers, 5th is non-carrier, and the last lane is a no-DNA control.

The *Tsp*RI restriction digests were used to identify the presence of the N34S mutation. Figure 29 shows an analysis of the N34S mutation by using the *Tsp*RI restriction enzyme method. The figure shows seven enzyme products. The final lane contains a control sample with no DNA used. The first two lanes of the gel come from carriers of the N34S mutation, and the following four enzyme products have samples of individuals with no carriers of the mutation. In the first two enzyme products, *Tsp*RI cuts the products into four visible fragments at sites 767, 499, 268, and 146 bp. Enzyme cutting to create the additional (268bp/499bp) fragments occurs only in the presence of the N34S mutation.

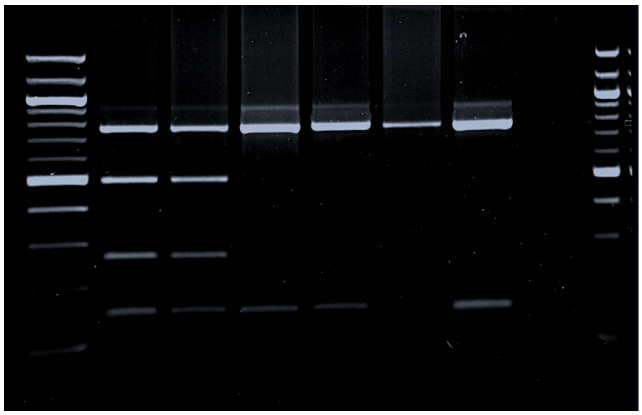


Figure 29 shows carriers of the N34S mutation. The restriction site for *Tsp*RI is produced in the presence of the N34S mutation (first and second lanes). The last blank lane is no-DNA control. The lanes from third to sixth are non-carriers of the N34S mutation.

Sequence analysis of the N34S mutation of the *SPINK1* gene

The Sanger sequencing technique identified and confirmed four individuals with the N34S mutation. The chromatogram of DNA sequencing of the human *SPINK1* gene PCR product in the third exon is shown in Figures 30 and 31. The results show amino acid replacement in position 476 of the trace for individuals with rs17107315 carriers. This mutation leads to an amino acid replacement from Asparagine to Serine (N34S). Sequence analysis of the gene revealed heterozygosity for the expected N34S mutation. This sequencing shows the complement of the reverse strand. Relevant codons are AAT for Asparagine and AGT for

Serine (Figures 30 and 31, top strand). The reference allele is T, minor allele C on the top strand in the sequence AYT (reference ATT, variant ACT); on the coding (bottom) strand, this corresponds to AAT (Asn) to AGT (Ser). dbSNP151 suggests that MAF is in the 0.6-1% range. Figure 30 demonstrates results from the first experiment using the 1000 Genome project samples. Next, Figure 31 presents the results from the second experiment by using 124 affected patients' QMC. In Figures 30 and 31 (A) and (B), there is a double peak at base 476, which indicates heterozygosity (A) which was found in carriers and (B) shows homozygosity which was found in non-carriers. The results in both experiments are similar, which confirms the validity of the second experiment.

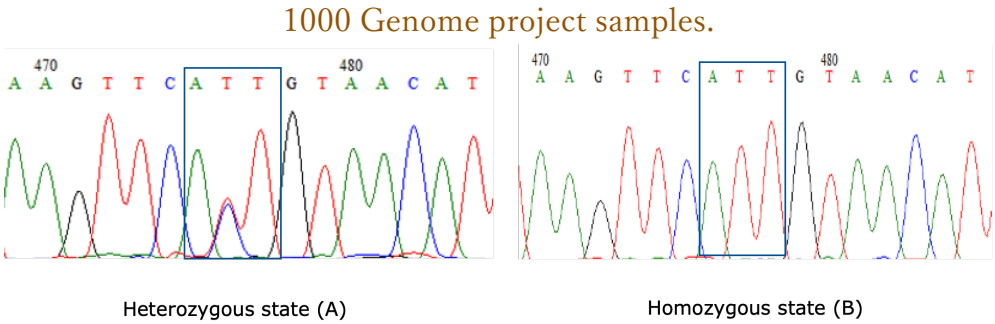


Figure 30. The DNA sequencing of the N34S mutation. The samples from 1000 Genome project.

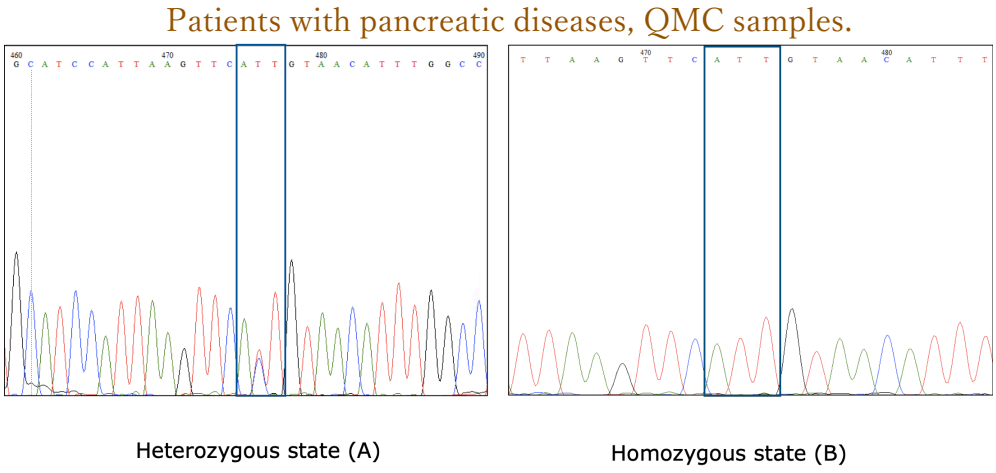


Figure 31. The DNA sequencing of the N34S mutation. The samples from QMC.

The table 3 shows that the occurrence of mutations appears more in women (3) compared to men (1). The study revealed the occurrence of the mutation is all above the age of 40. In addition, the dependence of the N34S mutation of the *SPINK1* gene variant on different types of pancreatitis was separately investigated, which may be one of the criteria for distinguishing types of pancreatitis. The identified mutation appeared in three patients with Gallstones and one with Idiopathic pancreatitis. The mutation was not identified in patients with pancreatic cancer. Furthermore, the N34S mutation occurrence was analysed and compared with the comparison to other studies.

Table 4 shows the incidence data and the calculation of statistical significance between different studies associated with the N34S mutation. The calculation was presented statistically using an χ^2 test to determine the difference in the N34S mutation frequency in other populations. For this, five different studies were analysed[422], including indicators of the occurrence of the N34S mutation in pancreatic diseases.

Four of 124 (3.3%) affected individuals carried the N34S mutation, and all were heterozygous.

In our study, N34S carriers mainly had duplications of *AMY2A* and *2B* (5 copies of *AMY2A* and 4 copies of *AMY2B*; 3 copies of *AMY2A* and 2 copies of *AMY2B*), 2 copies of *AMY2A* and 2 copies of *AMY2B* and one deletion (1 copy of *AMY2A* and 2 copies of *AMY2B*). It should be noted that for N34S carriers among cases, a high copy number for *AMY2* was revealed. However, for *AMY1*, this feature is not observed. CNV was identified in 3,4,8 and 11 copies.

1 patient (female, 43 years, Gallstone, 8 CN *AMY1*, 5:4 CN *AMY2*), 2 patient (female, 87 years, Gallstone, 3 CN *AMY1*, 1:2 CN *AMY2*), 3 patient (male, 56 years, Idiopathic, 4 CN *AMY1*, 2:2 CN *AMY2*) 4 patient (female, 82 years, Gallstone, 11 CN *AMY1*, 3:2 CN *AMY2*).

Gender	N	Age	N	PD*	N
Females	3	>40	4	Gallstones	3
Males	1	<40	0	Idiopathic PD	1

*PD=pancreatic diseases.

Table 3 of the structural distribution of the detected *SPINK1* gene in the study.

Population	N34S (n, %) *	p-value comparison**
Nottingham, UK	4/124 (3.3%)	$1.5x10^{-2}$
Lempinen et al, Finland	14/116 (12%)	$7.6x10^{-3}$
Witt et al, UK, Germany, Switzerland	16/274 (5.8%)	0.2909
Perri et al., Italy	1/45 (2.2%)	0.7405
Bernardino et al, Brazil	0/64 (0)	0.1530
Schneider et al., USA	2/32 (6.3%)	0.4492

* Mutation carrier frequency in different countries.

**The difference between Nottingham, UK study patients and other patients, in the statistical calculation, p-value.

Table 4 shows the frequency of the N34S mutation in the *SPINK1* gene in patients with pancreatitis. The occurrence of pancreatitis and comparison between different populations.

Our study determined an association between N34S mutations in the *SPINK* gene and pancreatic diseases with a p-value of $1.5x10^{-2}$ (RR=3.156, OR=3.226, 95% CI 1.191-8.735).

A comparative analysis of the N34S mutation frequency and p values in pancreas diseases was conducted with some studies (Table 3). Research comparing the mutation occurrence in different populations showed statistical significance ($p=7.6x10^{-3}$) with the Finnish patients, and in the other analysed studies, the significance was not found. The Finnish patients have a higher predisposition and frequency of the N34S mutation in the *SPINK1* gene than the patients in our experiments. The mutation's prevalence and association with pancreatic diseases in the other four countries are relatively similar.

3.5 Discussion

There are publications in the scientific literature on the prognostic value of genetic features in terms of identifying various diseases and assessing the clinical significance of variation in the pancreatic *SPINK1* gene[423]. *SPINK1* is the gene that leads to chronic pancreatitis and acute pancreatitis formation. The most common variant of the *SPINK1* gene was the N34S missense mutation in exon 3[424]. The N34S mutation is

located at the border in the protein, and its side sequence is oriented further than trypsin[425]. The N34S mutation plays a substantial role in the practical action of the gene. *SPINK1* inhibits about 20% of the total amount of trypsin by blocking its synthesis[426,427].

The study aimed to determine the presence and occurrence of the N34S mutation in the *SPINK1* gene in individuals with various pancreatic diseases. The N34S mutation in the *SPINK1* gene was analysed in all 129 individuals, where five samples were from the first experiment, and 124 samples were from the second experiment with the affected individuals. To detect the N34S mutation of the *SPINK1* gene as a potential risk factor for pancreatitis, PCR followed by Sanger sequencing and the *Tsp*RI restriction method (CASTG) as a test method for N34S mutation were used, as described above.

The mutation was found in 4 out of 124 patients in the Pancreas Research Project. The individuals obtained with positive results were carriers of rs17107315 SNP. All positive samples highlighted by restriction enzyme analysis showed a double peak of heterozygosity on sequencing. It was found that the frequency of mutations in our study was 3.3%, which is higher than the population frequency of 1% (human assembly hg19, https://genome.ucsc.edu). This indicates the association and high incidence of the N34S mutation of the *SPINK1* gene and rs17107315 SNP with pancreatic diseases (p-value $1.5x10^{-2}$). Since approximately 20% of trypsin activity can be blocked by *SPINK1*, it is likely to be associated with pancreatic diseases and a higher occurrence in affected individuals[428,429].

In similar studies in different countries, the frequency of the N34S mutation averaged appeared as 0-6%[236,420,430,431]. At the same time, a low incidence of the N34S mutation is shown in various countries[432]. The highest reported frequency of discovery of the N34S mutation in patients with alcoholic chronic pancreatitis is observed in India, where it reaches 26.8%[433,434]. At the same time, in South Korea, the frequency of detection of this mutation in patients with alcoholic chronic pancreatitis is only 2.4% and was entirely absent in a sample of 64 subjects from Brazil[431,435].

Our comparative analysis showed that, in general, most studies are characterised by relatively the same level of N34S mutation frequency in different populations (Table 3). Various studies have shown that sampling genetically similar populations may be the reason for the higher incidence of N34S mutation[435,436]. Another reason for the discrepancy in the data on the occurrence of N34S mutations may be the genetic heterogeneity of the population in different regions of the world, the

heterogeneity of the form of pancreatitis itself, and the criteria for diagnosing this disease used in a particular study[437].

The incidence of chronic pancreatitis is approximately 1:10,000, while the *SPINK1* N34S haplotype occurs in 1-3% of the population[249,419]. The odd ratio of *SPINK1* N34S ranges from non-significance[438] to approximately 80[439]. It is hypothesised that the high-risk haplotype of *SPINK1* N34S mutations may arise due to distinctions in the part of subjects with pathogenic pathways connecting exogenous factor stressors to pancreatic cell activation via the trypsinogen reactivation mechanism, as well as *SPINK1* inhibition of trypsin[432]. In studies of Kereszturi and Sahin-Toth, the N34S haplotype was determined to affect the deterioration of enzyme activity and is a predisposition to chronic pancreatitis. The correlation of the N34S haplotype with pancreatitis diseases is based on a pathogenic mechanism[440].

Several studies have identified numerous of the *SPINK1* gene sequence variants other than N34S, except for their clinical significance, which has lesser certainty[241]. Chen et al. found an N34S mutation located in the third exon, c.101A>G[441]. However, the search for the N34S mutation in further investigations led to a significant association between pancreatic diseases and the N34S mutation in the *SPINK1* gene. One of the crucial discoveries of the link was in the studies of Witt et al. In their studies, Witt et al. proved a significant correlation of the N34S mutation with one of the types of pancreatitis, an idiopathic form[241]. In the outcome, they emphasised genetic causes, as well as the transferring of the disease by inheritance. At the same time, Pfützer et al. suggested that N34S mutations only modify the disease and not causing its occurrence. The reason may be a decrease in the pancreatitis threshold due to genetic and environmental factors [419], which supports the MFD theory. Schneider et al. reported a very high frequency of N34S mutation of the *SPINK1* gene in 55% of fibrocalculous pancreatic diabetes. Also N34S mutation was found in 20% of tropical calcific pancreatitis, and 14% in non–insulin-dependent diabetes mellitus[420]; 46% in tropical calcific pancreatitis and fibrocalculous pancreatic diabetes were reported by Chandak et al[442].

A meta-analysis was carried out on the effect of the N34S mutation *SPINK1* gene variation in Caucasic patients with chronic pancreatitis, as well as the frequency of the N34S variation in alcoholic and idiopathic chronic pancreatitis. 25 studies of 8800 participants (2981 cases and 5819 controls) were analysed. The presence of the N34S mutation increased the risk of chronic pancreatitis by nine times (OR 9.695, 95%

CI 7.931–11.851). The influence of *SPINK1* in idiopathic pancreatitis (OR 13.640, 95% CI 8.858–21.002) is greater than in alcoholic pancreatitis (OR 5.283, 95% CI 3.449–8.092)[443]. In another meta-analysis, tropical chronic pancreatitis (OR 19.15, 95% CI 8.83–41.56) was additionally analysed, in which the effect of the *SPINK1* N34S mutation gene was higher along with idiopathic pancreatitis (OR 14.97, 95% CI 9.09–24.67) than alcoholic pancreatitis (OR 4.98, 95% CI 3.16–7.85). Familial chronic pancreatitis showed high heterogeneity I^2=80.95%. The analysis revealed an association between the large effect sizes of *SPINK1* N34S as a candidate gene and several aetiological pathways, which may lead to a common clinical characterisation[432].

A connection between the N34S mutation and idiopathic chronic pancreatitis has been reported in various studies[236,419]. In our research, N34S was also detected in idiopathic pancreatitis. Other groups are studying the function of the mutations in different types of pancreatic diseases. In many studies, the N34S mutation is associated with alcoholic chronic pancreatitis[444]. In our research, N34S has not been detected in patients with alcoholic pancreatitis. Some studies have shown that the N34S mutation may predict the development of acute pancreatitis, particularly in alcohol abusers, and contribute to a more severe course of the disease[445]. Moreover, these mutations may be a factor in persistent incidents of acute pancreatitis[446]. Some studies have shown the relationship of N34S mutations with cholelithiasis within 5-10%[422,447]. In our research, the N34S mutation was found in three patients with cholelithiasis, and we assume there is no specific form of pancreatic disease for the N34S mutation.

3.6 Conclusion

Many studies are proving that the N34S mutation is one of the major variants of the *SPINK1* gene, which is 0.4-1.7% in the world population[236,241,246,419,448-450]. In view of the fact that the *SPINK1* mutation frequently appears in the population, it is considered a pancreatic disease modifier gene[241]. However, because pancreatic diseases are MFD, the activation of the process requires the involvement of an external factor, which also determines the severity of the disease course.

The occurrence of the N34S mutation determines the acceleration of the progression of pancreatitis, determines the earlier development of the disease, a greater likelihood of pronounced structural transformations in the parenchyma and ductal system of the pancreas and the development of complications, increases the risk of a severe course and a worse response to conservative treatment. The wide occurrence of the N34S mutation with the increasing availability of genetic research methods will make it possible to identify patients with pancreatitis, carriers of this mutation, characterised by a less favourable prognosis and an increased risk of complications.

Understanding the genetic mechanisms will allow for preventing the phenotypic expression of pancreatitis in people with a genetic predisposition, as well as understanding the aetiology and possibly developing new approaches to treatment and prevention.

CHAPTER IV. KASP GENOTYPING WITH *AMY2A* and *AMY2B*

4.1 Introduction

The previous chapters described various studies used to determine the association between pancreatic diseases, cancer and the *AMY1* and *AMY2* genes. Molecular genetic analysis for determining polymorphic variants of genes is carried out by a wide range of methods used to extract SNP genomic data.

One of the possible efficient and fastest ways to genotype SNP markers is KASP technology (Competitive Allele-Specific PCR, LGC Biosearch Technologies). Genotyping is carried out to determine single nucleotide polymorphisms.

The advantage of this method lies in the complete automation of the diagnostic process and the ability to analyse a large number of samples simultaneously[451]. In various scientific sources, KASP is highly sensitive in determining SNP genotypes, a fast and straightforward method, and cost-effective, which suggested the possibility of using KASP to study the CNV at the *AMY* gene[452].

The KASP genotyping technology was developed and is applied by LGC Genomics. The method relies on extension from allele-specific oligonucleotides and fluorescent resonance energy transfer (Förster resonance energy transfer, FRET) to generate a signal[453,454]. The FRET process is used to quantify amplification from each allele of a single SNP. The process involves using the initial stage of PCR, where primers enhance the target region of SNPs. A mixture of SNP-specific primers and a 2-fold reaction mixture universal for genotyping (master mix) are added to the DNA sample. Then a polymerase chain reaction is carried out, followed by reading and generating a fluorescent signal at the endpoint. Biallelic discrimination is achieved by competitively binding two allele-specific forward primers (Figure 32)[451]. There will only be one of two possible fluorescence signals produced if the genotype at each SNP is homozygous; when the genotype is heterozygous, a mixed fluorescent signal is generated[452]. Using KASP genotyping, it is possible to determine the allelic SNP and insertions/deletions of any type of sample; in practice, it has been applied mainly in studies of animals, plants and microorganisms[455].

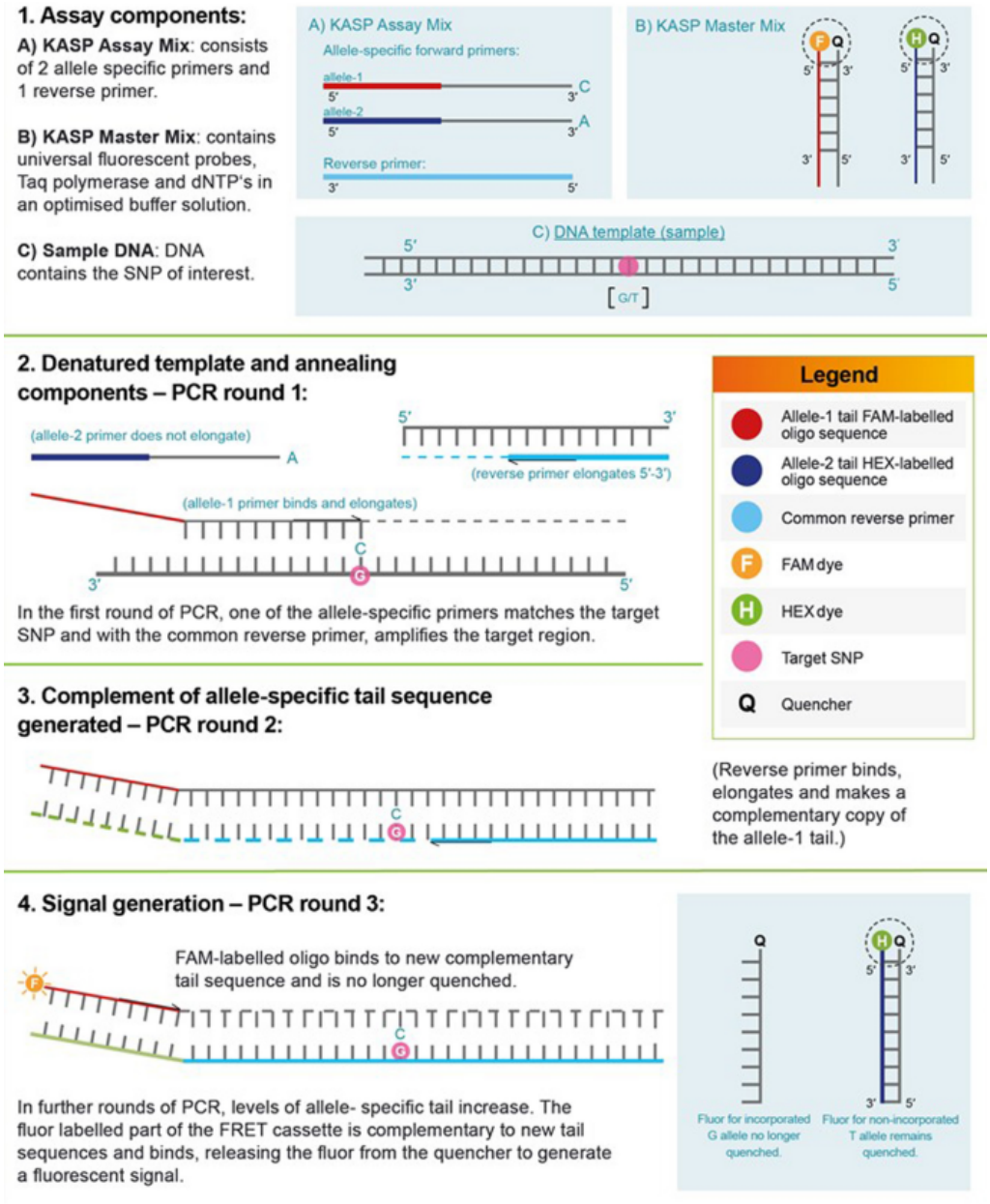


Figure 32. The working principle of the KASP technology in a schematic representation. Extracted from He et al., 2014[451].

Various studies have shown a positive outcome of the KASP method, for example, in the detection of SNPs associated with effective drug use[456]. Other studies have demonstrated the use of KASP in SNP detection to genotype candidate variants associated with, for example, *G6PD* deficiency[457]. Various studies have shown errors around 0.7–1.6% in KASP genotyping. In studies using recursive marker-assisted selection, KASP genotyping costs were 7.9-46.1% cheaper than those of the BeadXpress and GoldenGate platforms[458].

In our association studies of amylase gene copy number with clinical states (Chapter **II**), the rate of progress was limited by the time required for PRT assays to produce accurate measures of gene copy number. KASP technology is usually applied to genotyping simple biallelic SNPs, for which determining the presence or absence of signal from each allele in a sample is sufficient to define the genotype. Our study attempted to test the KASP technology for genotyping *AMY2* gene blood DNA samples by testing the accuracy of KASP analogues of the PRT assays we used for *AMY2*.

4.2 Aim

The study aims to evaluate the possible usefulness and effectiveness of the KASP method in the region of the *AMY2* gene by determining *AMY2A:2B* ratios. Because the *AMY2* measurements had already been determined by the PRT method, results could be used to evaluate the accuracy of the KASP data.

4.3 Material and Methods

Genotyping was performed by LGC using technology based on KASP genotyping on DNA samples dried in 96-well format.

For the control DNAs, the biological material from the five plates of HRC cell lines from healthy people was used; on each 96-well plate, there were two empty wells used as controls, so there were 470 samples overall. Patients with various pancreas diseases were taken for cases, which amounted to 94 blood-derived DNA samples from the QMC, with two blank control positions on the 96-well plates (Section 2.3).

A series of DNA dilutions were used for samples with a known genotype to determine the detectable amount of genetic material in the PCR reaction. The tested amount of DNA in the PCR reaction was 100 ng/reaction with a concentration of 10 ng/µl. Genotyping by the KASP technology was carried out according to its protocol by LGC Biosearch Technologies, UK[455].

In the study, primers were used to compare the dosage of the two *AMY2A* and *2B* genes and their ratio. Primer Allele 1 corresponded to the *AMY2B* locus, and primer allele 2 corresponded to the *AMY2A* locus. The common reverse primer was designated C1 (Figure 33).

```
A1:         GAAAATAGAGAATTTAGGAAATAAAGTCGAC
A2:         TAGAAAATAGAGAATTTAGGAAATAAAGTCGAT
C1:         GATTTTTAATCAATACACATTTGC

A1:         GAAAATAGAGAATTTAGGAAATAAAGTCGAC
A2:         TAGAAAATAGAGAATTTAGGAAATAAAGTCGAT
C1:         GATTTTTAATCAATACACATTTGC

AMY2B       TTAAGAAACTCAAATCCATATTTAAAAAC-TTAAATATAGATTTAAGATTTTTAATCAA
AMY2A       TTAAGAAACTCAAATCCATATTTAAGAACTTTCAAATATTGATTTAAGATTTTTAATCAA
AMY1        TTAAGAAACTCAAATCCATATTTGAAAACCTTTAAATATTGATTTAAGATTTTTAATCAA

AMY2B       TACACATTTGCCCACTTTTAAAAAGCTCCCAACCAATTAAAAAGCTCGTCGACTTTATTT
AMY2A       TACACATTTGCCCACTTCTAAGRAGTTCCC----AATTAAAAATCTCATCGACTTTATTT
AMY1        TACACATTTGTCCACTTTTAAAAAGCTCCCAACCAATTGAAAAACTCATCGACTTTATTT

AMY2B       CCTAAATTCTCTATTTTCTATTAGAAAATATTTCCAAGATACATCTATAGTAGAATGTGA
AMY2A       CCTAAATTCTCTATTTTCTATTAGAAAATATTTCAAAGATACATCTGTAGTAGAATGTGA
AMY1        CCTAAGTTCTCTATTTTCTATTAGAAAATATTTCCAAGATACATCTATAGTAGAATGTGA
```

Figure 33: The alignment shows the regions of *AMY2B* and *AMY2A* that form the basis of the *AMY2A:2B* KASP ratio assay, aligned with the corresponding sequence from *AMY1*. To avoid interference from products derived from *AMY1*, the common forward primer C1 (yellow) is designed with a 3' terminal mismatch to *AMY1* (green). The binding sites for the A1 primer matching *AMY2B* and the A2 primer matching *AMY2A* are highlighted in cyan and magenta, respectively; these primers distinguish the single base (A/G on the alignment) at which *AMY2B* and *AMY2A* differ. The red highlighting shows the 4bp deletion used to distinguish *AMY2B* and *AMY2A* in the PRT *AMY2A:2B* ratio assay (Section 2.3). The KASP assay, therefore, samples a substitutional (A/G) variant that is 13bp away from the 4bp indel used in the PRT assay.

After PCR and the available results were completed, each sample was assigned values corresponding to allelic intensity data. In the overall KASP analysis, the X-signal allele pattern corresponded to the *AMY2B* gene and the Y-signal of the *AMY2A* gene.

4.4 Results and Discussion

In accordance with the aim of the study, the analysis involves checking the sufficiency of the KASP genotyping method and assessing the applicability of the KASP technology on extracted DNA. Assessment of the KASP genotyping was carried out to calculate the measurement of the ratios of the *AMY2A* and *2B* genes. The current result is the first demonstration by the KASP of performing ratio measurements for *AMY* genes.

The obtained relationship between *AMY2A* and *2B* signals is presented in the form of clusters (Figure 34). A total of 576 ratio measurements for cases and controls were used for this plot after subtraction of the background signal, using results from zero-DNA controls. The process revealed a pattern between the results of the KASP genotyping analysis and demonstrated the main group of results approximating the regression line shown. (Figure 34 A).

The expected graphical results showed five main cluster groups of sample type. The selected clusters were marked with different colours. The yellow cluster shows stronger signals for *AMY2B* but very low signals for *AMY2A*. Homozygous deletions of *AMY2A* are located along the horizontal axis. The green cluster, which is close to 0, is DNA-free control. Heterozygous deletions are marked in the black cluster. Similar signals are characteristic for the majority of samples for *AMY2A* and *2B*, presumably having a ratio of 1:1 and highlighted in red. The brown cluster demonstrates a higher signal strength from the *AMY2A* than *2B* genes (Figure 34 B).

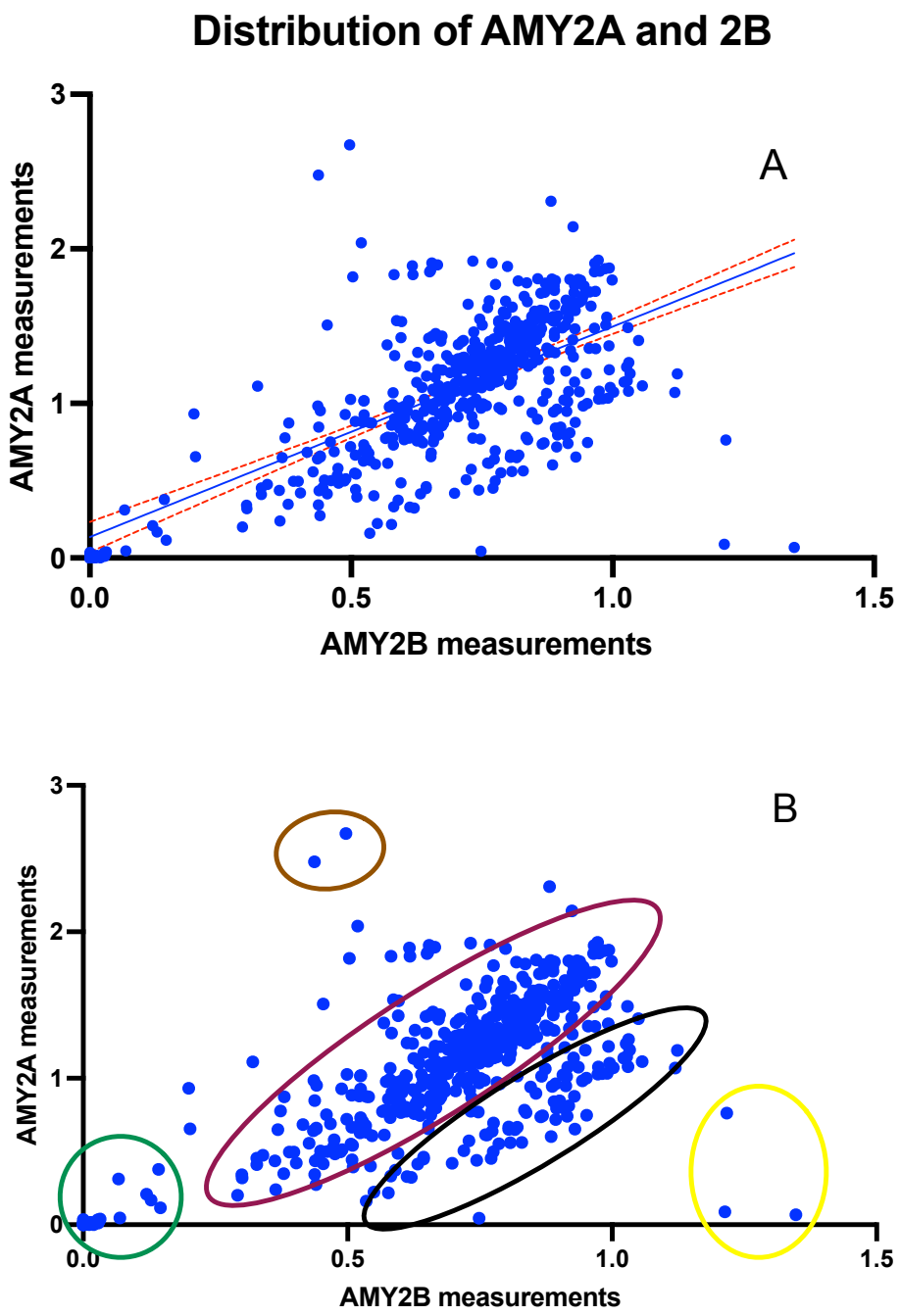


Figure 34 shows the results of the KASP genotyping of raw data using a cluster plot. The graph shows the relationships of measurements between the two genes, *AMY2A* and *2B*. Five main cluster groups have been identified. Clusters are circled in different colours. The yellow cluster shows strong signals for *AMY2B* and weak signals for *AMY2A*. The green cluster shows the group with no DNA. Brown cluster has strong signals for *AMY2A* and weak signals for *AMY2B*. The red cluster shows the same signals for *AMY2A* and *2B*. Heterozygous deletions are noted in the black cluster.

To improve the calculation, a calibration was used. Calibration included the exclusion of erroneous measurements and low-frequency signals. The calibration method was to introduce reproducibility (how close independent measurements are to each other) and accuracy. As a result of the adjustment, DNA samples with very low signals were excluded. The *AMY2A:2B* ratio was calibrated as 1.0 for the central cluster. Linear calibration was used to determine the calibration lines. The resulting linear calibration indicates a positive pattern between the *AMY2A* and *2B* data.

The analysis was carried out twice using the same DNA samples to exclude randomisation of the results and determine the reproducibility. The comparability of the two results with their good correlation was revealed to assess statistical differences. We observe the internal consistency of each sample and sufficient correlation (p value<0.0001). Both repeats appear to show good measurements of *AMY2A:AMY2B* ratios. Repeated measurements helped to distinguish the measurement error. Slight deviations from zero were observed in the direction of increasing or decreasing measurements. However, the most significant data match between the two repeats is at zero point (in the form of a lead cluster). The data flipped horizontally at point zero shows a good correlation between the two repeats (Figure 35). This plot showed that the results of both data measurements give a slight difference, which can be estimated as one overall result since there will be only one correct integer value for any sample.

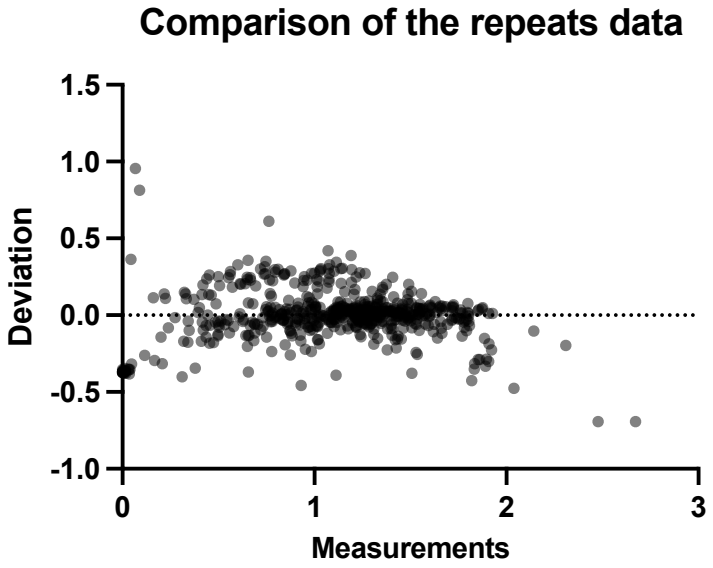


Figure 35. The plot shows the measurements of the two repeats from the data. The results of both measurements show a slight difference.

Next, the distribution of the ratios of *AMY2A* and *AMY2B* genes was analysed, as shown in Figure 36. To check the correlation relationships between the measured values, a histogram was used, which showed four main peaks. The median peak demonstrated the highest frequency. A ratio near 1 indicates the presence of equal copies for both the *AMY2A* and *AMY2B* genes. These results are found in most of the European population (2 copies of *AMY2A* and 2 copies of *AMY2B* or 3 copies of *AMY2A* and 3 copies of *AMY2B*). The value of the second peak indicates a heterozygous *AMY2A* deletion. Minor changes are also revealed at the beginning of the histogram. Results close to 0 indicate a homozygous deletion of *AMY2A*. An increase in the direction of 1.2 suggests a higher copy number for *AMY2A* than *AMY2B*, from three *AMY2A* copies and above.

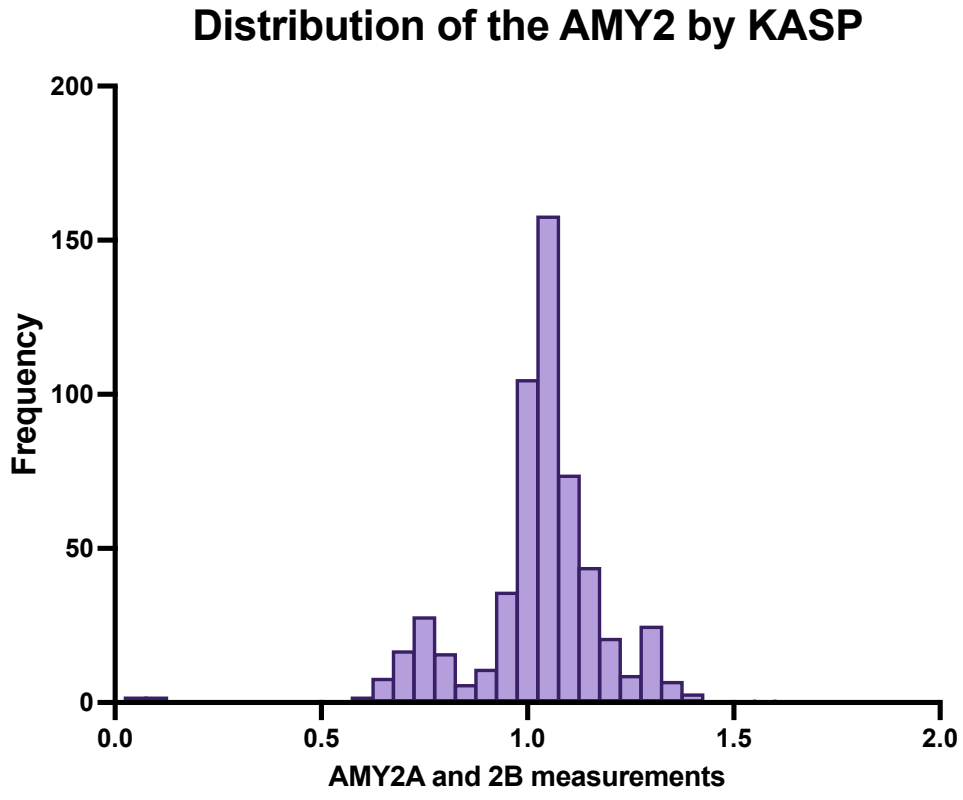


Figure 36 demonstrates the measurement of the *AMY2A:2B* ratios from KASP genotyping data by showing peaks. The central peak shows a ratio of about 1.05. A smaller peak of samples at 0.75. Minor numbers of samples with a ratio near zero are presumably homozygous for *AMY2A* deletions. Thus, the majority group of samples has a peak ratio of about 1; the smaller group is about 0.6-0.8. At the beginning of the histogram, there are several samples with a value near 0. Individual samples above a value of 1.2 indicate more copies of *AMY2A* and *AMY2B*.

Initially, we tested the sufficiency of the KASP data to calculate the ratios between the *AMY2* genes by comparing the results of the KASP method with the results of the PRT. We investigated whether there is a dissonance between the KASP genotyping and the results obtained with PRT. As a result, a slight discrepancy in genotypes was found between the two methods. Due to the fact that the KASP method for calculating the ratios between genes was used for the first time, the PRT method was used as a quality control when checking the KASP method. The resulting comparative analysis of the KASP and PRT data is shown in Figure 37. The graphs showed a slight difference between the PRT and KASP data in the deletions and duplications; in the homozygous deletions of *AMY2A* and 2 copies of *AMY2A* and *2B*, the ratio results were identical. According to the genotyped KASP data, the measurement distributions are slightly higher in the deletions and the duplications with some measurements than when using data from the PRT method. It is important to note that the PRT data are more constant and have less scatter. In the analysis of PRT (N=556) and KASP (N=576) without junction fragment was used.

The data quality assessment presented in Figure 37 showed that the analysis options of the KASP study to obtain results were sufficient to study the ratio of the amylase gene *AMY2A* and *AMY2B*. However, in our research, slight upward deviations of measurements were observed in the analysis of the KASP in comparison with the PRT.

In general, the measurements of the complete KASP data are consistent with the genetic datasets of the PRT; and match the expected values.

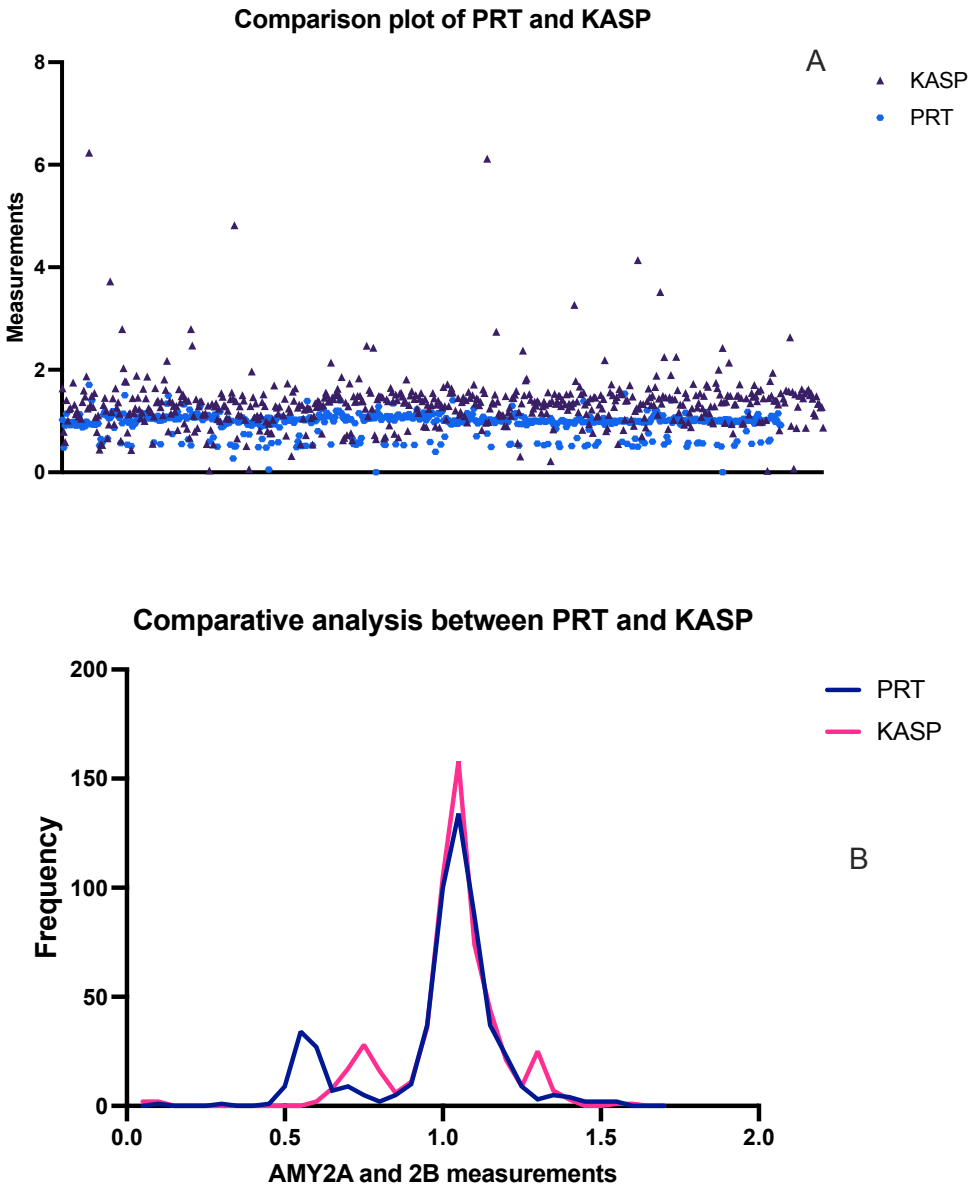


Figure 37 shows the KASP data compared to the PRT data. Figure 37 A with the initial data, Figure 37 B after the calibration. Results for the central peak were similar for the two methods (Figure 37B). Figure 37 A shows the initial results without the use of calibration. Figure 37 B shows distributions between measurements between 0.9 and 1.60 for PRT (N=455), for KASP (N=495); the distribution between measurements was 0.40-0.85 for PRT (N=99), KASP (N=77), measurements close to 0 were for PRT (N=2), for KASP (N=4).

4.5 Conclusion

Data analysis using the KASP genotyping methods showed the applicability of this technology in calculating ratios for the amylase gene, in particular *AMY2A* and *AMY2B*.

Following the results, it was found that the KASP method is sufficient for analysing the ratios, as well as PRT. However, the PRT method is more promising since it does not require the use of additional manipulations in the form of calibration. Analysis using the KASP technology is limited in applicability to detecting CNV and their study in various diseases. At the same time, the results of the KASP showed the expected genotypes, as did the PRT method. Most studies have shown that this method is mainly used for detecting and genotyping SNPs[459].

The current result is the first demonstration of analysing the amylase gene *AMY2A* and *AMY2B* ratio using the KASP genotyping method. A limitation of the present study is the small scale of biological samples.

CHAPTER V. EXPLORATION OF *DEFA1A3* AS A RISK FACTOR IN PERIODONTITIS

5.1 Introduction

Aspects of periodontitis disease and its manifestation

Periodontitis is a chronic inflammatory disease leading to bone destruction and loss of the gingival insert, which is an important cause of tooth loss[460]. In 2019, the Global Burden of Disease Survey estimated that 3.58 billion people suffer from oral disease; 743 million people suffer from severe periodontitis, which is approximately 14% of the total population[461]. Periodontitis among them is in 11th place in the ranking of the most common diseases[462], and severe periodontitis is in sixth place among common diseases[463].

Data and information on the role of genetic factors in the aetiology of periodontal diseases in the world scientific literature are incomplete[464], and genetic predisposition to periodontitis is still a topical and complex problem in periodontology. Periodontitis is a multifactorial disease where the inheritance pattern cannot be explained by simple Mendelian rules and in which the phenotype is also determined by the action of environmental factors. The aetiopathogenetic basis of such diseases is formed by variants of specific genes as susceptibility genes, the damaging effect of which occurs against the background of the influence of adverse environmental factors (Figure 38).

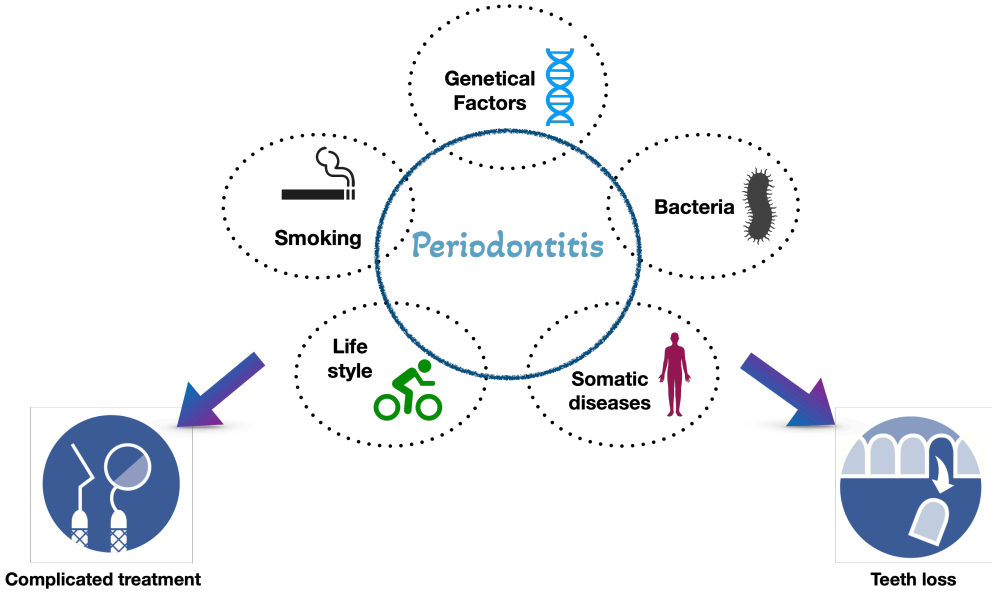


Figure 38. Risk factors that form periodontitis and its consequences.

Genetic conditioning of the periodontal environment should be considered from the point of view of the heritability of physicochemical, anatomical, and morphological features of tooth tissues.

The scientific literature highlights the inflammatory concept of periodontitis, defined as the result of "pathogen-host" interaction. According to this idea, the protective phagocytosis of the body, designed to eliminate microbial agents and restore balance in the oral cavity, becomes a pathological mechanism that destroys the periodontal tissue itself. The presence of complicated complexes of periodontal pathogens colonise the tooth's surface and participate in the formation of plaque[465]. The basis of the clinical manifestations of periodontitis is the process of tissue reaction in the oral cavity, triggered by bacterial colonisation and accompanied by changes in metabolic processes at the level of target cells. All this leads to the activation of proteinases, which subsequently damage the periodontal fibres. This process increases the spread of bacteria along the gums. This process is fundamental in the development of gum disease. However, the microbiome of the oral cavity plays an essential part in the inflammatory process[466]. External and internal factors in the form of a genetic predisposition of certain genes are essential for changing the structure of the dysbiotic microbiome. These factors aggravate the course of the disease (Figure 39). However, the presence of bacteria alone is insufficient for developing periodontitis. Over the past few decades, research has shown this is an organised, active process[467].

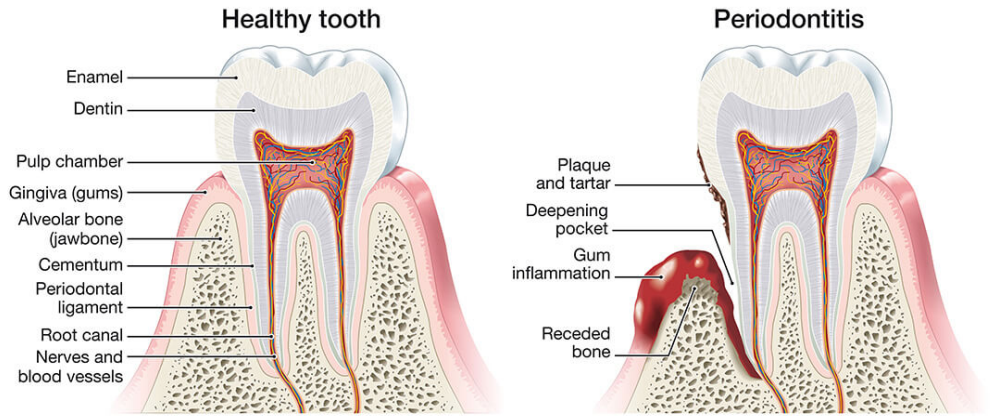


Figure 39. Comparison between healthy teeth and periodontitis. Mechanism of developing periodontitis. Extracted from www.theperiospecialists.com.au.

Periodontitis has the following clinical forms[466]:

1. Necrotising periodontitis
2. Chronic periodontitis
3. Aggressive periodontitis
4. Periodontitis as a manifestation of systemic diseases.

According to the degree of distribution, there is a local form, when only a separate area is affected and a generalised form, when the pathological process covers a large area of the periodontium. The severity of the periodontal disease is determined by three main criteria: depth of the periodontal pocket, level of bone destruction, and degree of tooth mobility.

Aggressive periodontitis is an inflammatory and destructive disease characterised by the rapid destruction of complex periodontal tissues and the increasing collapse of the tooth bone of the alveolar process. This destruction is typical for people over 40 years of age[468] and is the result of the body's reaction to the invasion of periodontopathogens. Chronic inflammation leads to damage to the gum tissue and usually persists for a long time[469]. In chronic periodontitis, colonisation of microorganisms in the gingival sulcus occurs, which activates the mechanisms of innate immunity in periodontal tissues. The development of periodontitis is also facilitated by diseases of the internal organs, especially the digestive, immune, cardiovascular and endocrine systems, etc[470].

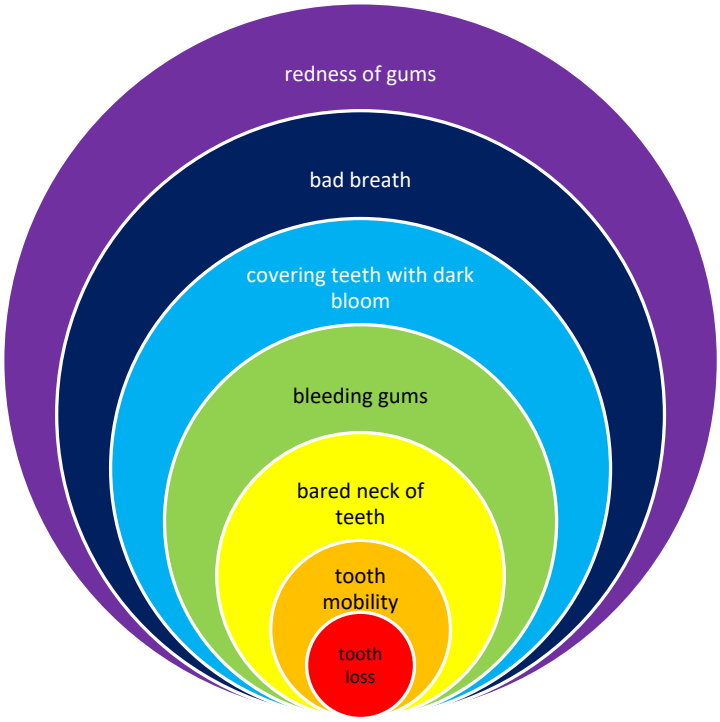


Figure 40. Primary signs of the onset of periodontitis.

Clinical signs of periodontitis are shown in Figure 40. The occurrence and development of periodontitis go through several stages, which determine its clinical picture (Figure 41).

Stage 1 of periodontitis demonstrates an early form, a borderline condition between gingivitis and periodontitis. In this stage, persistent inflammation of the gums occurs. Early diagnosis can still stop the process. These patients are still susceptible to therapy. Early intervention and early diagnosis are cost-effective.

Stage 2 is established periodontitis. At this stage, certain changes and damage are identified. Even at this stage, the treatment is reversible and not complicated. Constant monitoring can help stop the worsening of the disease.

Stage 3 is defined by significant damage to the gums and periodontal ligament. The ligaments of the teeth become loose, and there is a loosening of the teeth, so the risk of them falling out. Periodontal lesions at this stage are deep.

Stage 4 causes significant damage to the periodontal tissue that extends to the apex of the tooth. At this stage, there is a violation of the

chewing function. The dental composition is completely destroyed[466]. This stage is characterised by the loss of a tooth or teeth. This process is completely irreversible. The treatment is not effective.

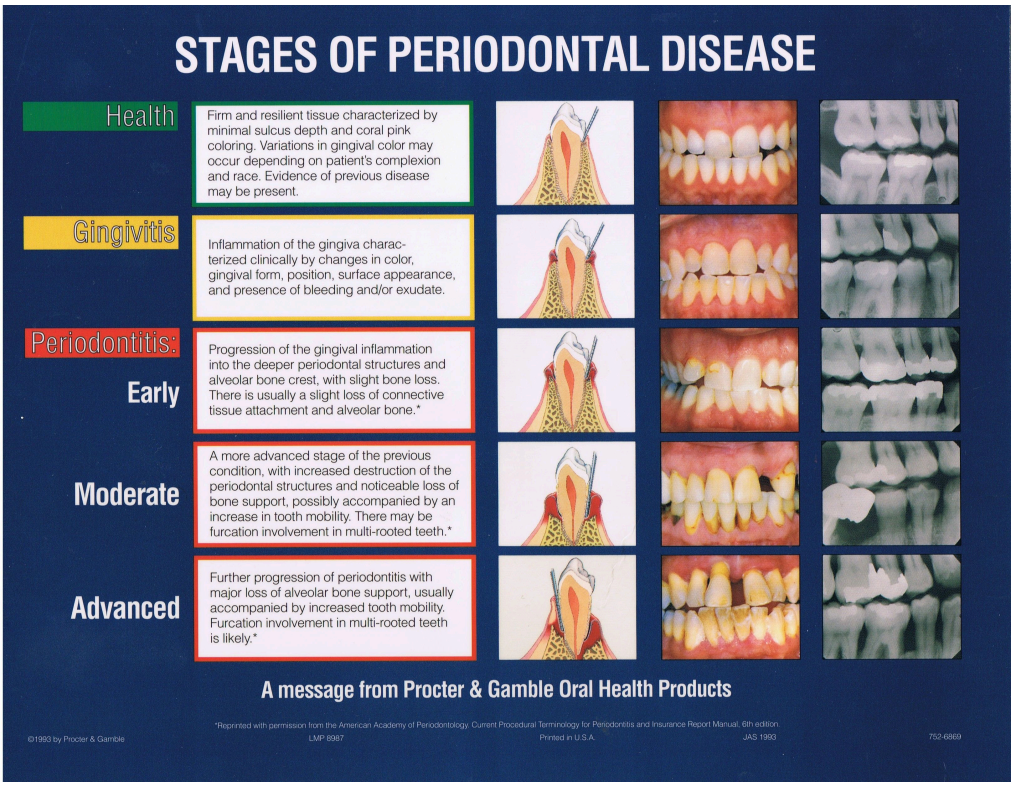


Figure 41. Description of the stages and clinical characteristics of periodontitis. Information from the American Academy of Periodontology. Made by Procter & Gamble (1993).

The balance between bacterial invasion and local oral resistance is currently considered one of the main factors determining clinical manifestations. However, these well-known hypotheses do not consider host genetic factors that significantly influence both plaque composition and disease susceptibility[471]. Tissue resistance factors can also be a set of genetically determined protective mechanisms responsible for immunity to infection, which act as the first protective barrier when exposed to microflora[472]. Periodontitis is a multifactorial disease, in the pathogenesis of which the immune response of the macroorganism to bacterial exposure, the presence of general somatic pathology, as well as such

predisposing risk factors as smoking, poor oral hygiene, and stress, are of greater importance[473,474]. The immune response can be regulated by genetic factors. The nature of the course of periodontitis in each clinical case is determined by the individual characteristics of the human body, which determine its susceptibility to the development of the disease. Genes may be associated with a predisposition to periodontitis, which needs to be determined in each clinical case[475]. Scientific studies have shown that heredity and genetic factors determine 50% of the predisposition to periodontal diseases. Clinical studies have shown that the host response pattern may be an important factor involved in the heritability of periodontitis[155]. Genetic predisposition affects not only the occurrence of periodontitis but also the age at the onset of development and the severity of the course. These patients may have more severe forms of periodontitis. Previously, the classification of periodontitis phenotypes and control subjects was inconsistent in various studies. The studies carried out by Munz et al. and Offenbacher et al. are based (as they point out in their papers) on the exchange of data and patient samples between research groups for further study and discovery of new candidate genes, thereby expanding the results of research related to the genome[469,476]. According to twin studies, the heritability of periodontitis was estimated to be 0.38 (95% CI, 0.34 to 0.43; I^2= 12.9%), 0.15 (95% CI, 0.06 to 0.24; I^2= 0%) in family studies and 0.29 (95% CI, 0.21 to 0.38; I^2= 61.2%) when twin and other family studies were combined. At the same time, the GWAS determined a low heritability estimate of 0.07 (95% CI, -0.02 to 0.15) for periodontitis[54]. In gingivitis, hereditary traits were 0.29 (95% CI, 0.22 to 0.36; I^2= 37.6%)[477,478]. In a study of twins, the assessment of the heritability of periodontitis varies from 0.3 to 0.5 (from 30% to 50%[479-481]. Genetic factors in assessing heritability have been identified from 38% to 82%[482]. Other studies reject periodontitis as a hereditary disease[483]. The low heritability found by the twin GWAS study ranges from 0 for moderate[484] to 0.14 for moderately severe chronic periodontitis[485] and from 0.23 to 0.24 for severe chronic periodontitis[484,486]. Different definitions of periodontitis in various studies have contributed to the low heritability value. However, other studies demonstrate the relationship between gingivitis and periodontitis with a genetic factor[54].

Periodontitis is one of the critical problems of modern medicine, as it is accompanied by severe disorders of the dental system and affects the health of the whole organism. It can occur both as a local disease and as a manifestation of the general condition of the body as a whole. The

outcome and course of periodontal disease can be predetermined not only by the virulence of microbes but also by the genetic polymorphism of the human body. Correct identification of the leading causes and mechanisms of the development of inflammatory periodontal diseases determines the proper search for the most effective methods of treatment and prevention of periodontitis.

The relationship of periodontitis with other multifactorial diseases

According to research to date, cardiovascular and endocrine diseases, pathologies of the liver and kidneys, and genetic predisposition disorders of puberty contribute to the progression of inflammatory and dystrophic processes in periodontal tissues, and the degree of damage to the periodontal complex is the more profound, the more somatic pathology[487]. In modern paper sources, the data presented indicate that impaired metabolic processes in periodontal tissues, resulting from some exogenous and endogenous influences and dysfunction of enzymatic systems, disrupt the periodontal blood supply. In patients with diabetes, periodontitis occurs in almost all cases and is characterised by the aggressive nature of the course. At the same time, the frequency and severity of the pathology are in direct correlation depending on the duration of carbohydrate metabolism disorders[488,489]. In other words, in the occurrence of diseases of periodontal tissues, not only local factors (dental plaque, periodontopathogens, anatomical features of the structure and others) but also systemic ones are significant. Recent studies have demonstrated the pathogenetic relationship between periodontal lesions and diseases of internal organs. The variety of clinical manifestations of periodontal diseases, as well as their close connection with somatic pathology, led to the fact that the problem of their diagnosis, prevention and treatment went far beyond the limits of the dental clinic. Inflammatory diseases of periodontal tissues dominate the overall structure of the pathological processes of the oral cavity, and today they are defined as the result of the interaction of the pathogen-macroorganism system, which has a low level of protective forces, which makes the latter sensitive to this bacterial load. In this case, indeed, it becomes necessary to study the bilateral relationship between the

microflora of the oral cavity, which causes the pathology of periodontal tissues, and chronic systemic inflammation[490].

Aspects of the relationship between somatic and dental diseases are various. On the one hand, the occurrence and severity of diseases in the organs and tissues of the oral cavity depend on the severity of common diseases. On the other hand, an essential role in forming somatic pathology belongs to odontogenic infection. Pathogens that cause inflammation in the periodontium can enter the general circulation, thereby leading to the development of somatic diseases affecting many organ functions. There is evidence that testifies to the negative impact of dental diseases on the severity of general pathology and the development of focal diseases of the body (for example, bacterial endocarditis, rheumatism and others)[491]. In other words, this comorbidity factor has a negative, mutually potentiating effect on the severity of both dental and background systemic pathology. As a confirmation, according to the papers, patients with chronic generalised periodontitis who had gastritis and pancreatitis demonstrated a direct relationship between the severity of periodontal tissue diseases and the degree of compensation for gastrointestinal disease[492]. Periodontal diseases often precede the recurrence of conditions in the digestive system; by being a focus of chronic infection, they can lead to developing or exacerbating chronic diseases, including the gastrointestinal tract. Microorganisms of periodontal pockets can cause sensitisation of the body and violate the immunological status.

Separate studies have revealed the relationship between pathological signs of various diseases (diabetes mellitus, diseases of the pancreas, nervous system, etc) with periodontal disease. The development of symptoms of chronic periodontitis was also observed in diabetes mellitus. Periodontitis is more common in people with gastrointestinal tract pathology than in people without concomitant pathologies. The deterioration of the protective functions of the immune system is associated with the development of dysbacteriosis to complex forms of periodontitis[492-495].

The pathogenetic community is characteristic in the diseases of the oral cavity and ENT organs (ear, throat, nose). In particular, it decreases the oral cavity's barrier functions in upper respiratory tract diseases (tonsillitis and bronchial asthma). Inhaled glucocorticoids which are used in bronchial asthma treatment, lead to severe oral dysbiosis[496,497,498].

Pathology of blood microcirculation plays an essential role in developing periodontal disease[499]. To a sufficient extent, they influence the occurrence of pathogenic factors, further causing clinical symptoms[500]. Violation of blood factors gives rise to periodontal disease. Microcirculatory changes lead to the development of thrombosis, a decrease in oxygen to cells and malnutrition, which causes necrosis of periodontal tissues[501].

Diseases of the endocrine system (pituitary adenoma, Cushing's syndrome, parathyroid disease) can trigger the development of periodontitis. For example, glucocorticoid hormone leads to incomplete destruction of bone tissue. Other hormones, such as cortisol and parathyroid hormone, can reduce the functional ability of young bone tissue cells that synthesise intercellular substances of the alveolar bone, thereby leading to the destruction of collagens. This process leads to the massive destruction of periodontal bone tissue. The lack of estrogen leading to osteoporosis also affects the occurrence of periodontitis[502,503].

Numerous clinical observations indicate that pathological changes in periodontal tissues accompany gastrointestinal tract diseases. Oral pathology is diagnosed in 92-100% of patients with diseases of the gastrointestinal tract, and, mainly, it is represented by periodontitis. There are also studies confirming the existence of a link between certain types of oral bacteria (for example, *Granulicatella*) and the development of pancreatic cancer. Chronic gastritis, gastric ulcer, duodenal ulcer, chronic pancreatitis, chronic hepatitis and cirrhosis of the liver, inflammatory bowel diseases are often associated with gingivitis, periodontitis, or aphthous lesions of the mucous membrane of the cheeks, lips and tongue[504]. At the same time, a distinctive feature of the severity of inflammatory and destructive periodontal diseases in patients against the background of the pathology of the digestive system is its early onset, compared with those without background pathology[492], the generalisation of the pathological process, a tendency to relapse, short remission and the formation of resistance to standard pharmacotherapy. In the early stages, the periodontium of all teeth is involved in the pathological process, and more intense inflammation phenomena are observed, often accompanied by the festering of periodontal pockets[498]. There is a significant decrease in local immunity of the oral cavity with the development of generalised periodontitis in the background of pancreatic adenocarcinoma or chronic pancreatitis[505,506]. In the case of pancreatic lesions, changes can result from secondary hypovitaminosis and the involvement of other digestive system organs in the pathological

process[507]. As the data of paper sources show, the pronounced interrelation of dental diseases and diseases of the digestive organs is explained by the fact that dysfunction of chewing due to carious tooth decay, dentofacial anomalies[508,509], and periodontal diseases lead to ingress into the stomach that is not crushed and not adequately treated food, which becomes a significant factor affecting the development of diseases of the digestive system in children and adults.

On the other hand, chronic gastrointestinal diseases (gastritis, peptic ulcer and 12 duodenal ulcers, hepatitis, pancreatitis, etc.) are accompanied by a deficiency in the body of vitamins, minerals, proteins, carbohydrates, and impaired immune regulation. All this leads to the development of functional and organic disorders, inflammatory and dystrophic changes, and chronic diseases of all parts of the oral mucosa (recurrent aphthous stomatitis, angular cheilitis and others.). Background pathology weakens the body's defences and creates conditions for reducing resistance in periodontal tissues against dental plaque bacteria and activating the periodontal pathogenic microflora[510-512].

The nervous system plays a vital role in periodontitis. Various factors can trigger impaired function of the nerve supplying the nervous system. The action of free radicals has a negative effect on nerve cells, which affects the destruction of periodontal tissue. The whole process causes an imbalance of periodontal metabolism, namely oxidative degradation of lipids[503,513].

With HIV, one of the symptoms is various oral cavity diseases. Some studies found the connection between oral complications with HIV[514-517]. One of them is periodontitis[500]. CD4/CD8 plays one of the leading roles in the development of periodontitis. Active reduction of CD4 cells about CD8 leads to a great complication for various microorganisms[518]. HIV infection depletes CD4, which in turn causes CD8 dysfunction[519]. CD4/CD8 is of great importance compared to the absolute number of CD4 cells[520]. Several studies have investigated the relationship of the CD4/CD8 ratio with oral manifestations[521-524]. For example, oral candidiasis results in low CD4/CD8 and high CD8 lymphocytosis[521-524].

In chronic nephropathy, inflammation of the periodontium is observed during active treatment[525]. Complications of somatic and chronic diseases lead to the worsening of periodontitis.

The role of cytokines in the development of inflammatory processes and regulation of the immune response

The Human Genome and Human Diversity projects identified gene variations that lead to hereditary human diseases or increase the risk of developing multifactorial human diseases. Thus, polymorphisms of cytokine genes that regulate inflammation (IL-1, TNF α, IL-4, IL-6 and others) have been identified as important in the risk of periodontitis[526-530]. In these cases, the presence of the variant gene does not cause the disease, but the production of cytokines changes, which, consequently, can determine an overreaction to the bacterial load and clinically manifest exacerbations of periodontal disease. Critical genes of inflammatory cytokines, regulators of the immune response, have been studied as candidate genes that influence the nature of the course, outcome, and prognosis of periodontal disease.

One of the key inflammatory cytokines is Interleukins (IL), which comprise a large group of cytokines of different biological activity. IL-1 has significant biological and physiological characteristics. Interleukins (IL) make up a large group of cytokines of different biological activity. IL-1 is often not detected in the blood plasma of healthy people. IL-1 is an inducible protein. Its products are directly proportional to the amount of inducer. In humans, IL-1β is the main form of secretory IL-1. IL-1 is produced by various cells of the organism, which include macrophages and keratinocytes, which are activated by B-lymphocytes, as well as fibroblasts. IL-1 has an extensive range of functions in the immune system, initiates and regulates immune processes, and participates in the development of acute and chronic inflammation and bone tissue resorption. The IL-1 family includes three homologous proteins: IL-1 alpha and IL-1 beta (IL-1α and IL-1β), which are pro-inflammatory, and IL-1RN (IL-1 receptor antagonist), which has a molecule anti-inflammatory action. The balance between expression and inhibition of IL-1 synthesis is decisive in the development, regulation and outcome of the inflammatory process[531].

The IL-1 cluster polymorphism is presented as an independent modifying factor since IL-1 activates the deterioration of the extracellular structure of bone and periodontal tissues and contributes to an increase in the level of prostaglandin E2 (PGE2) which regulates the inflammatory response with the activity of vascular smooth muscle[532] and potentiates the formation of tumour necrosis factor-alpha (TNFα)[533]. Genomic studies have shown that the presence of polymorphic alleles of

the IL-1 responsible for the primary response to a microbial infection affects the nature of the immune responses[534-536].

IL-6 is an early inducible cytokine. Expression of the IL-6 gene occurs under the influence of viruses, bacteria and their products entering the body. The primary function consists of activating the proliferation of antigen-associated B-lymphocytes and enhancing antibody synthesis. IL-6 enhances the functional activity of fibroblasts and osteoclasts[528]. IL-6 is a cytokine with a dual function (pro-inflammatory and anti-inflammatory). This explains its ability to transfer inflammation from the acute to the chronic phase, attracting mononuclear cells and forming mononuclear granulomas. One of the biological functions of IL-6 is switching protective reactions from nonspecific to specific immune responses with a change in the types of leukocyte infiltrates. IL-6, inhibiting the synthesis of IL-1, TNF and chemokines that attract neutrophilic granulocytes, enhances the migration of monocytes and lymphocytes[537].

With the development of protective reactions to the introduction of pathogens and the occurrence of immunopathological processes, IL-8, which stimulates the migration of neutrophils involved in nonspecific protection responses, is essential. IL-8 enhances cell production of IL-1β, IL-6 and TNF, thereby contributing to the elimination of pathogens and increasing the bactericidal properties of neutrophils after leukocyte migration to the inflammatory focus[538,539]. Additionally, it has immunoregulatory properties, namely in the mechanisms of innate and adaptive immunities, and is the first line of defence against viral and other intracellular infections[540]. IL-4, on the contrary, is a negative regulator of cellular immunity reactions, suppressing cell-mediated activation. The key regulator of the immune response is IL-10, as it inhibits the synthesis of anti-inflammatory cytokines by macrophages and cytokines. Additionally, IL-10 inhibits the synthesis of reactive oxygen and nitrogen species by macrophages and monocytes.

The genes encoding another important pro-inflammatory cytokine, TNF-α is located on chromosome 6 within the limits of HLA class III region[541]. TNF-α cytokines are among the best-known factors associated with the destruction of periodontal tissues. TNF-α plays an essential role in the development of periodontitis; as shown in many studies, increased expression of TNF-α is found in the affected periodontal tissues, and TNF-α expression is significantly increased in the gingival fluid in areas where there is a loss of attachment of the gums and bone tissue[527,542,543].

Liu and Li showed a meta-analysis study of GWAS to determine the link of SNPs cytokines with periodontitis. The genetic relationship between 12 interleukins and cytokines was analysed. A comprehensive molecular genetic analysis of Th1 (IL-2, IFN-γ, and TNF-α), Th2 (IL-4 and IL-13), Th17 (IL-1α, IL-1β, IL-6, and IL-17 and Treg (IL-10 and TGF-β) in patients struggling from inflammatory-destructive periodontal diseases. The study's results showed the relationship between polymorphic variants of a number of genes and manifestations of inflammatory-destructive processes in the periodontium. The significance of these genetic markers for preclinical diagnosis and prognosis of the progression of inflammatory-destructive periodontal diseases is shown. IL-4, IL-13, IFN-γ have no association with periodontitis. IL-17 polymorphism has little association with inflammatory processes. IL-2 and TNF-α, on the contrary, are related with the pathogenesis of periodontitis and the prevention of its risk. IL-1α has an inverse function in the pathogenesis of periodontitis. IL-1β is a biomarker of an anti-inflammatory cytokine in the development of periodontitis and has a strong association with the risk of periodontitis with OR (95% CI). IL-6 may play a protective role in periodontitis inflammation. At the same time, the researchers determined that Treg cytokines like IL-10 and TGF-β perform a protective function at the risk of periodontitis and reduce inflammation. Thus, the genetic factor in the development of periodontitis is a regulator in the autoimmune system and contributes to the development of effective treatment of the disease and protection against inflammation[544].

One of the most studied markers is the *VDR* (vitamin D receptor) gene. The *VDR* gene is located on chromosome 12 (q13-14 segment), and its 60 kb long structure includes 11 exons: the non-coding 5'-end of the *VDR* gene contains exons 1A, 1B and 1C. At the same time, 8 more exons encode its translated product[545]. The influence of vitamin D on the aetiology and pathogenesis of periodontitis is mediated primarily by its calcium regulatory activity and participation in the regulation of bone homeostasis, as well as its immunotropic effects. In bone tissue, vitamin D performs a dual function, affecting the differentiation of osteoclasts and osteoblasts through an increase in the synthesis of RANKL (receptor activator of nuclear factor kappa-B ligand) and M-CSF (colony-stimulating factor, macrophage-specific) and reduced OPG synthesis (osteoprotegerin). Thus, calcitriol is involved in the resorption and bone formation processes. *VDR* is widely represented among immune cells, such as antigen-presenting cells, NK (natural killer), T-cells, and B-cells;

1,25 (OH) 2D3 has important antiproliferative, differentiating and immunomodulating functions[546-548]. Calcitriol activates NK cells and increases the phagocytic activity of macrophages. Immune effects of 1,25 (OH) 2D3 are fundamentally associated with dendritic cells in lymph nodes; it modulates antigen-specific immune responses *in vivo*[549]. Innate immune reactions of the body against infections of the mucous membranes are closely associated with antibacterial peptides. 1,25 (OH) 2D3 induces β-defensin four expressions. β-defensin has antimicrobial activity against oral microflora, including periodontal pathogenic strains, such as *Actinobacillus actinomycetemcomitans, Porphyromonas gingivalis, Fusobacterium nucleatum, Candida,* and *HPV (human papillomavirus)*[550]. Researchers believe that *VDR* adjusts the transcription of approximately 500 of the 20,488 genes in the human genome. A majority of *VDR*-regulated genes express the distribution of *VDR* and 25 (OH) D3-1 α hydroxylase in many organs[545].

The role of defensins in the pathogenesis of periodontitis.

It should be noted that, along with inflammatory cytokines, there are anti-inflammatory factors, the balance of which largely determines the body's response to bacterial aggression. Defensins play an essential role in protecting the oral cavity. Defensins are cationic peptides of the immune system that are active against bacteria and viruses. They are immunoprotective peptides produced by most multicellular organisms' immune systems. Studies have shown their role as effector molecules of innate immunity[551]. Defensins perform essential functions in many clinically significant pathological processes. During infectious and inflammatory processes, neutrophils are activated, leading to their rapid release, found in plasma and other body fluids[552]. During inflammation, an increase in the concentration of defensins in biological fluids is observed, and the problem arises of preventing a cytopathic effect in relation to one's own cells[551,553].

Defensins are a large family of low molecular weight (4 kDa), amphipathic peptides rich in cysteine residues, represented by three subclasses: α-, β-, and rare θ-defensins that are not synthesised in humans[554] (Figure 42). Defensin molecules are characterised by a substantial presence of amino acids (arginine, lysine, histidine), which

gives them a positive charge. Defensins differ from each other in the structure of cysteine and the location of disulfide bonds[555]. Cytokines and neuroendocrine signals regulate them. Bone marrow neutrophil precursors synthesise neutrophil defensins in promyelocytes and early myelocytes. Defensins are contained in azurophilic granules, which are key components of phagocytes[556]. Defensins have hydrophilic and hydrophobic regions, which facilitates their incorporation into the phospholipid bilayer of microorganisms. Defensins have bactericidal activity on both Gram-negative and Gram-positive bacteria. In addition, they actively have an anti-inflammatory effect by inactivating pathogen-associated molecular structures, adhesins, and toxins of infectious agents, suppressing the activity of induced monocytes, macrophages, and dendritic cells[557-559]. This occurs by inhibiting the entry of the virus into the cell by blocking the synthesis of viral receptors[560]. Some studies have shown that patients lacking α-defensins suffer from frequent and severe bacterial infections[561,562]. They are also found in NK cells, macrophages, B and T-lymphocytes, immature dendritic cells of monocytic origin, monocytes, and epithelial cells. α-defensins are synthesised by Paneth cells (HD5, HD6), thus playing a vital role in the moderation of the quantitative and qualitative composition of the intestinal microbiota[563]. In turn, β-defensins, which combine about 40 peptides, are secreted by epithelial cells of the gastrointestinal, respiratory, and urogenital tract. Keratinocytes are the first barrier to various infections[564].

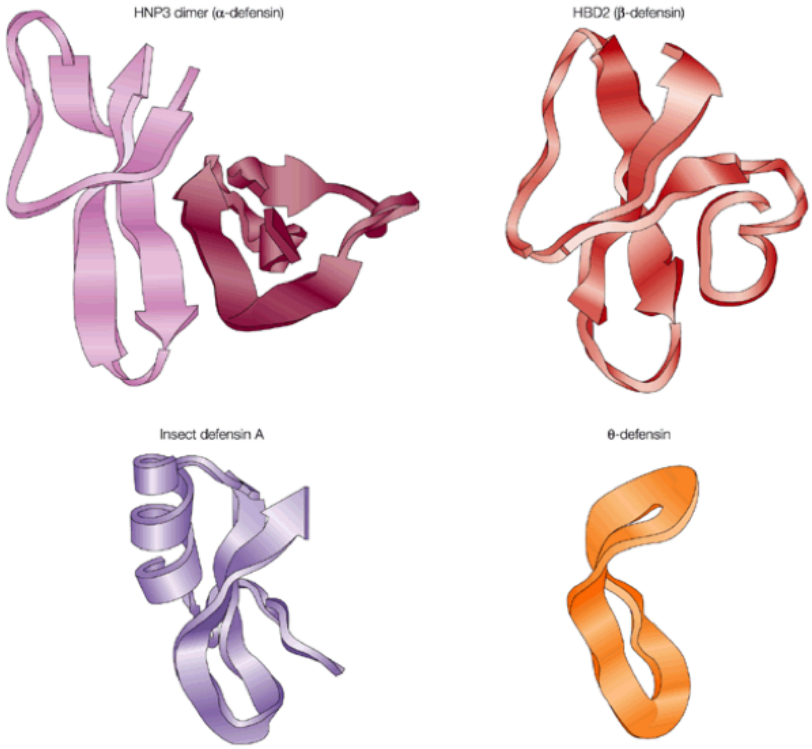
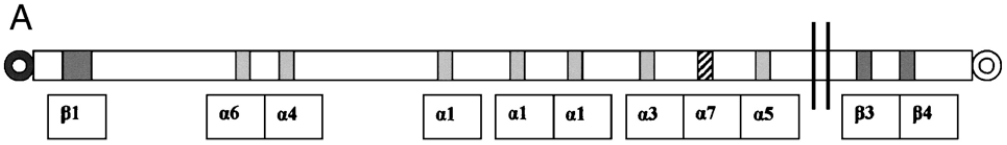


Figure 42. Peptide structure of defensins. Human neutrophil peptide 3 (HNP3, α-defensin) and Human β-defensin 2 (HBD2). The difference between defensins is determined by the size of the lengths of the peptide segments that are connected by disulfide bonds. The θ-defensin has a simple β-sheet with a cyclic structure. Only α-defensin and β-defensin are characteristic of humans. θ-defensin is found in macaque leukocytes. Extracted from Ganz T, 2003[554].

Human defensins have strong immunoadjuvant properties that can enhance the activity of the humoral and cellular immune response against various antigens of infectious pathogens[565]. Moreover, they can influence the processes of inflammation, proliferation, wound healing, production of cytokines, and chemotaxis of immunocompetent cells[566]. Defensins are small structures that freely pass inside the pathogen[567] and inhibit the synthesis of DNA and RNA[568], phospholipase A2 (*PLA2*)[569], bacterial wall repair enzymes, chaperones and ribosomes[570], increasing the mechanism of antibacterial action. Defensins directly affect the adhesion of microorganisms to periodontal tissues and oral mucosa, and consequently to the development of periodontitis and diseases of the oral mucosa[571-573]. It is assumed that among the variety of components of the

innate immune response involved in maintaining periodontal homeostasis, it is defensins that can play a crucial role in providing resistance to inflammatory diseases of the oral cavity[574]. Due to the fact that defensins are antimicrobial peptides and are involved in the immune process, as shown by some studies, in the pathogenesis of periodontitis, the immune process fails both in the oral cavity and in the whole body.
In humans, there are eight genes encoding α and β defensins. The genes are located on chromosome 8 (8p22-23)[556]. The *DEFA1* and *DEFA3* genes encode the HNP-1 and HNP-3 peptides, respectively, and are differentiated by the N-terminal amino acid. There is no separate gene for HNP-2, as it is similar to HNP-1 and HNP-3 in the last 29 amino acids; HNP-2 is presumed to result from proteolytic processing of HNP1 or HNP-3, and the gene cluster designation *DEFA1A3* is used to indicate the genes encoding HNP-1, 2 or 3[575]. Human α-defensin genes consist of three exons; the first encodes a 5'-untranslated region, and β-defensins consist of 2 exons. β-defensin 2 (hBD2) is small, encoded by a gene with an intron of about (2 kb) and is secreted in epithelial cells. The β-defensin 1 (hBD1) gene has a larger intron (approximately 10 kb), and hBD1 is continuously released in various cell types[576] (Figure 43.)

A



B

Label	Gene	Peptide	Peptide Location
α1	DEFA1	HNP-1, HNP-2	Neutrophils
α3	DEFA3	HNP-3 (HNP-2)	Neutrophils
α4	DEFA4	HNP-4	Neutrophils
α5	DEFA5	HD-5	Paneth Cells
α6	DEFA6	HD-6	Paneth Cells
β1	DEFB1	HBD-1	Epithelia
β3	DEFB103A	HBD-3	Epithelia
β4	DEFB4	HBD-2	Epithelia

Figure 43. Structure of the defensin gene cluster at 8p22-23. In Figure A, β-defensin genes are marked in dark grey and α-defensin genes are marked in light grey. The two black lines show the location of the DNA sequencing break between α-defensin and β-defensin. The dashed line

marks the pseudogene α-defensin *DEFA7P*. The two rings show the direction of the 8p telomere (solid) and chromosome 8 centromere. Figure B lists the defensin genes. Extracted from Linzmeier and Ganz (2005)[555].

DEF genes are found in complex genomic regions subject to structural variation, and some members of the defensin family exhibit CNVs[577]. Studies have shown variability in CNV and defensin SNP, which increases the risk of developing various diseases and can lead to the definition of resistance to infections. There are rare human diseases (Chediak-Higashi syndrome and specific granule deficiency) associated with a decrease or absence of neutrophil alpha-defensins, but other components of neutrophil granules are also deficient, making it difficult to attribute these disorders to the defensins themselves[578]. Loss or downregulation of defensin genes is associated with some types of human cancer[579]. Some studies attribute human helicobacteriosis and associated gastritis, dyspepsia, peptic ulcer of the stomach and duodenum to defensin-dependent diseases. Studies have shown substantial increases in the expression of defensin in the gastric mucosa biopsy with HP-associated gastritis in adults, compared with HP-negative patients, while the level of increased peptide synthesis correlated with the histological degree of mucosal inflammation activity[580]. Human defensin (encoded by the *DEFB1* gene) is continuously expressed in the prostate, kidneys, and luminal epithelium of the urogenital tract. Studies show that 82% of prostate cancer cell samples and 90% of clear cell renal cell carcinoma samples have reduced or completely absent expression of *DEFB1*[581]. Some researchers have found that mutations in the *ITGB2* (21q22.3) and *SLC35C1* genes cause severe periodontitis in young patients with leukocyte adhesion deficiency 1 and 2, which is transmitted in an autosomal-recessive pattern[582].

Molecular genetic mechanisms influencing the pathogenesis of periodontitis are currently not fully understood. There is practically no data on markers in the genes of antimicrobial peptides (defensins), which are one of the main factors protecting oral mucosa. Candidate genes were studied in various populations, like *GLT6D1, ANRIL, COX2*, cytokines IL1, IL10, *MMP* and other genes. In various pathological processes, including those in the organs and tissues of the oral cavity, an important research task is a search for genetic predictor markers in innate immunity

genes. The data obtained will allow for predicting the development and severity of inflammatory periodontal diseases.

Genetic mechanisms affecting the pathogenesis of periodontitis are currently not entirely investigated. Various periodontal phenotypes are investigated as different parts of a range of similar factors and modulated as combinations of genetic risk loci. In this connection, various disease manifestations have common alleles and covariates of risk. In the occurrence and progression of periodontal disease, unknown variants of genetic risks play a role. One of the important backgrounds for studying α-defensins in periodontitis is the GWAS study by Munz et al. The Munz et al. study identified SNPs associated with aggressive periodontitis at the *DEFA1A3* locus. The GWAS study analysed German and Dutch case-control samples with characteristics of aggressive periodontitis (896 cases, 7104 controls) and moderate chronic periodontitis (993 cases, 1419 controls). Turkish samples with aggressive periodontitis were also used (223 cases, 564 controls). The study used patients with established diagnoses of both aggressive and chronic periodontitis. As a result, associations were confirmed in the German sample of severe forms of periodontitis with $p=1.09x10^{-8}$(rs4284742) and more moderate phenotype of chronic periodontitis with $p=5.48x10^{-10}$ (rs2738058). The *DEFA1A3* locus showed a connection with both disease phenotypes and was significant across the genome-wide level in its correlation with periodontitis. Of the 16 leading SNPs and nine neighbouring SNPs, the rs2738058 SNP at the *DEFA1A3* locus has a genome-wide significant association with another feature (Ig-A nephropathy, $p=2x10^{-27}$). Associations with cholesterol levels, obesity, heart disease, and bone disease have also been identified. However, no association was found with smoking. This research detects the first common genetic risk loci for aggressive periodontitis and chronic periodontitis with genome-wide significance and points the role of innate and adaptive immunity in the aetiology of periodontitis. The significance of SNPs at the two loci, *DEFA1A3* and *SIGLEC5*, showed a complete association with chronic periodontitis and aggressive periodontitis disease phenotypes. α-defensins are antimicrobial peptides and play a role in phagocyte-mediated host defence. *SIGLEC5*, an association gene involved, inhibits calcium's release (indigestion) to protect against accidental damage of host-cells and is formed in bone-forming osteocytes. It should be noted that variability in the number of copies of the defensin gene was observed, which indicated a difference in individual resistance to infections.

A meta-analysis of aggressive periodontitis and chronic periodontitis (2,067 cases and 8,533 controls for *DEFA1A3*; 2,027 cases and 8,330 controls for *SIGLEC5*) revealed genome-wide significant proof for an association of leading GWAS SNPs in *DEFA1A3* ($p= 2.06x10^{-8}$ and $p= 6.78x10^{-10}$) and *SIGLEC5* ($p=1.34x10^{-8}$) and made it possible to establish the distribution of allele frequencies in the studied sample. These results were positively replicated with aggressive periodontitis in Turkish samples. SNP rs1122900 in *SLC1A3* showed an association with p=0.0137. To confirm the association with aggressive periodontitis, the association level of rs1122900 was $p=8x10^{-7}$, compared with the association of the leading SNP GWAS rs6887423 with $p=1.42x10^{-6}$. However, in the Turkish analysis, none of the other 16 loci revealed a specific link with aggressive periodontitis. Interpopulation features of the prevalence and nature of the course of the disease indicated a differentiated nature of the polygenic predisposition to this disease in a particular population[469].

Studies in the structural chromosomal abnormalities of aggressive periodontitis and chronic periodontitis make it possible to investigate better the genetic mechanisms affecting the pathogenesis of periodontitis.

5.2 Aim

A hypothesis is put forward about a relationship between the identification of carriers of periodontitis and the presence of the protective properties of defensin. The relationship between defensin variability and the risk of periodontitis provides a basis for further research, especially in terms of improving the diagnostics and effective therapy of periodontitis in the future. In this work, we examined the association between *DEFA*-region SNPs and predisposition to periodontitis, using participant (twin) questionnaire responses as a proxy for phenotypic evaluation.

5.3 Materials and Methods for genetic research into periodontitis risk

The characteristic of the observed groups

Data from 3641 individuals were received from Twin UK Research & Genetic Epidemiology, based at King's College London. This unit has recorded the names of over 14274 twins, has organised 76 studies and has been involved in more than 800 publications. Data and participants matched the criteria of the Department of Twin Research & Genetic Epidemiology (DTR) policy.

The subjects of this study were selected twins born from 1924 to 1999 (between 20 and 95 years old). Data were provided in the form of quantitative variable values (identification number, year of birth, type of study, gender, actual zygosity and ethnicity). The survey of oral health was conducted by questionnaire. The questionnaire data were provided by Dr Claire Steves (King's College London) and included 25 questions related to issues with oral health in the study participants (Table 5).

Table 5. The table of the questionnaire.

#	Questions
1	How many teeth do you have? If in doubt, please count your teeth
2	If you have lost any teeth for any reason, how long ago did you lose your last tooth? /Years
3	If you have lost any teeth for any reason, how long ago did you lose your last tooth? /Months
4	If you have lost any teeth for any reason, how long ago did you lose your last tooth? /Don't know
5	Do you wear any removable dentures (false teeth)?
6	Do your gums bleed when you brush your teeth?
7	Do your gums bleed when you brush your teeth? / If yes, how often does this happen?
8	Have you ever been told by a dentist or hygienist that you have gum disease?
9	Do you have any loose teeth?
10	How often do you go to the dentist?
11	Which of the following dental treatments have you had in the past? / Fillings
12	Which of the following dental treatments have you had in the past? /Crowns
13	Which of the following dental treatments have you had in the past? /Veneers
14	Which of the following dental treatments have you had in the past? /Bridges
15	Which of the following dental treatments have you had in the past? /Dentures
16	Which of the following dental treatments have you had in the past? /Professional tooth cleaning
17	Which of the following dental treatments have you had in the past? /Gum treatment (deep cleaning or gum surgery)

18	Which of the following dental treatments have you had in the past? /Root filling / Root canal work
19	Which of the following dental treatments have you had in the past? /Dental implants
20	During the past 12 months, did your teeth or mouth cause any pain or discomfort?
21	How often do you use the following to clean your teeth or mouth? /Manual (ordinary) toothbrush
22	How often do you use the following to clean your teeth or mouth? /Electric toothbrush
23	How often do you use the following to clean your teeth or mouth? /Dental floss
24	How often do you use the following to clean your teeth or mouth? /Antibacterial mouthwash
25	How often do you use the following to clean your teeth or mouth? /Interdental brushes

Study design.

To account for the twins' population background, 3320 participants were selected from 3641 with identification numbers of twins whose ethnicity was described as "White". Those 3320 participants of twins comprised 1228 DZ and 2092 MZ, 2864 females and 456 males. MZ females 1782, males 310; DZ females 1082, males 146 (Figure 44). SNP data from the α-defensin region were available from 5654 twins.

Although a specific diagnosis of periodontitis was not included among the questions asked, four questions were initially identified as most relevant to the study of periodontitis. Considering the crucial clinical symptoms and frequency, these questions were selected as the most common in periodontitis (Figure 45) and renamed questions A-D. Selected question A (loose teeth) corresponds to initial question 9 in Table 5 above, question B (lost teeth) to questions 2-4, question C (dentist advice) to question 8, and question D (bleeding gums) to questions 6 and 7.

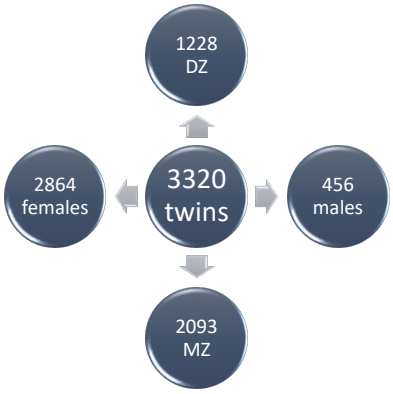


Figure 44.

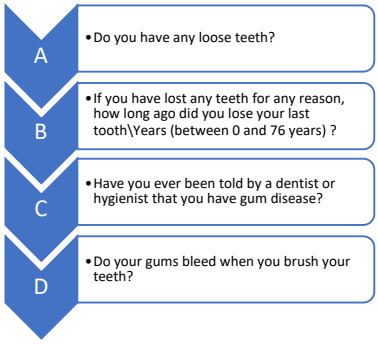


Figure 45. Selected questions.

Methods of statistical analysis.

Correlations between answers to these questions were assessed using a chi-squared test of independence. Measurements were tested in the model using yes=1 and no=0. This measurement is treated as a case-control study, where “yes” or case is affected twin and “no” or control is non-affected every twin. An additional phenotype was allocated, which included the combined values of all three questions used (A, B, D). The fourth group was evaluated on a score scale (A+B+D). A score of 1 was added for each “yes” and 0 for each “no”. Therefore, everyone was given a summarised score between 0 and 3.

The study prioritised identifying the association between the individual and the disease. In the work, a model based on the associative method (case-control) was implemented, determining the relationship between the defensin gene and periodontitis.

To confirm the relationship between α-defensin variability and periodontitis risk factors and evaluate the significance of frequency differences between the questions of twins and SNP sets, a distributed software package on the PLINK platform, version 1.07 (http://zzz.bwh.harvard.edu/plink) and GEMMA version 0.98 was used. In the study, PLINK used 3320 phenotypes and 5654 genotypes. PLINK considered analysis of association with the disease by identifying SNPs for each twin phenotype.

It was assumed that comparing the results of the two methods (PLINK and GEMMA) would allow a more reliable assessment of the relationship between periodontitis and defensin variation. Another goal was to determine the GEMMA software's suitability for analysing the association since this software had not previously been used to identify the link between periodontitis and the *DEFA* gene. GEMMA software analysed the data considering the twins' kinship (MZ and DZ twin pairs). The data used for the association calculation included questionnaire data from 1706 twins and 5654 genotypes at *DEFA1A3* regions SNPs. GEMMA was used to test the association with 5584 SNP α-defensins for each question separately, identifying derived SNP with the lowest p-value (in each cell). Statistical analysis also identified three main clinical characteristics of periodontal disease risk as reflected in questionnaire responses, such as bleeding gums (question D), loose teeth (question A) and tooth loss (question B).

A Manhattan plot was used to visualise both results. The Manhattan plot used the logistic regression data to display the association

between answers to questions and the common SNPs in the population plotted against the base pair position. The negative logarithm (base 10) of the association p-value for each SNP base pair indicator was calculated. Based on the results of the analysis, significant associative relationships were established.

To control the quality of the results, quantile-quantile plots (Q-Q plots) of the distributions of p-values were used. The Q-Q plot was used to compare two probability distributions. In the Q-Q plot, two quantiles of p-values were plotted against each other, comparing the negative logarithm of the observed and expected association p-values.

A Bonferroni test determined adjusted significance thresholds to reduce the chances of obtaining false-positive results. To analyse the heritability of the periodontitis features, concordance rates and Heritability estimates (h^2) were calculated using Holzinger's formula. The study used the UCSC Genome Browser tool to determine genomic coordinates. To specify the SNP coordinates, the assembly database of the February 2009 human reference sequence (GRCh37)/hg19 was used.

PLINK software

PLINK is a software tool designed to analyse genetic associations. The advantage of PLINK is the efficient processing of a large data set and its applicability to GWAS. PLINK examines the phenotype and genotype, as well as the associations between them, and the software package includes multiple command-line utilities helpful in manipulating data files. PLINK gene association analysis is based on the consideration of regression models. When calculating case controls, a logistic regression model is used, which assumes the probability of cases as a linear combination of independent variables. The obtained results are shown as SNPs from significant to least significant. The disadvantage is the lack of ability to analyse multi-allelic variants[583]. The input phenotype file contained two main columns (family ID and individual ID, in this case, the same ID for both columns) and the phenotype columns (which consisted of 4 additional columns, with questions A, B, D and the combination of all 3). Genotype information was in corresponding PED and MAP files. In PLINK software, sex was not specified, and the "yes"/"no" answers were treated as a case-control (logistic regression) test of association by adding relevant commands. The combined

(A+B+D) score was tested as a quantitative variable using linear regression. After entering the data in plain text (PED/MAP) format, PLINK data was converted to the binary (BED/BIM/FAM) file format. The analysis output file contained the association results from the SNP on each row, including base pairs, with ten columns with p-values and odd ratios (Tables 6 and 7).

Table 6. Composition of the final file.

CHR	Chromosome
SNP	SNP ID
BP	Physical position (base-pair)
A1	Minor allele name (based on the whole sample)
F_A	Frequency of this allele in cases
F_U	Frequency of this allele in controls
A2	Major allele name
CHISQ	Basic allelic test chi-square (1df)
P	Asymptotic p-value for this test
OR	Estimated odds ratio (for A1, i.e. A2 is reference)

Table 7. File outcomes.

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
8	8:6705442	6705442	G	0.0307	0.005025	A	20.95	4.71E-06	6.271
8	8:6721816	6721816	C	0.02632	0.005025	A	15.05	0.000104	5.351
8	8:6402671	6402671	G	0.03947	0.01036	A	14.83	0.000118	3.924
8	8:6915330	6915330	C	0.04825	0.01602	G	12.39	0.000433	3.114
8	8:6723173	6723173	T	0.2368	0.1573	C	9.871	0.001679	1.662
8	8:6725694	6725694	TA	0.2675	0.1837	T	9.738	0.001805	1.623
8	8:6723132	6723132	G	0.2368	0.158	C	9.688	0.001855	1.654
8	8:6723162	6723162	G	0.2368	0.158	A	9.688	0.001855	1.654
8	8:6798845	6798845	A	0.1009	0.05214	G	9.66	0.001883	2.04
8	8:6723141	6723141	C	0.2368	0.1583	T	9.597	0.001949	1.65
8	8:6726012	6726012	G	0.2632	0.1822	A	9.16	0.002474	1.603
8	8:6726155	6726155	G	0.2632	0.1828	A	8.997	0.002705	1.597
8	8:6726156	6726156	A	0.2632	0.1828	C	8.997	0.002705	1.597
8	8:6723470	6723470	G	0.2237	0.1501	A	8.788	0.003032	1.631
8	8:6725868	6725868	G	0.2632	0.184	A	8.676	0.003224	1.583

GEMMA software.

In continuation of the study, the Genome-wide Efficient Mixed Model Association (GEMMA) was used to identify associations and to check for a relationship between periodontitis and clinical characteristics. A feature of this software is the use of multidimensional linear mixed models (mvLMM). GEMMA models can consider participants' relatedness and use for complex phenotypes and quantitative traits in humans and model organisms, testing the relationships between SNPs and multiple correlated phenotypes. This is especially true for the analysis of comparisons and traits involving twins. The genetic relationship between the (twin pair) participants can be explicitly modelled in the analysis. Multivariate analysis can increase the power to detect genetic variants that affect only one of the many correlated phenotypes and can be used as a preprocessing step to then perform genetic analysis of the converted values. The calculation time for each SNP increases squared as the number of people grows, which makes them practical for GWAS, in which thousands of people participate[584]. Thus, GEMMA allows the coordination of phenotype transformation during genetic analysis. The likelihood principle can be expanded to objectively evaluate alternative transformations considering the observed genotype and phenotype data. The disadvantage of the software is the complexity in use and the selection of special commands. It requires a large amount of server memory and complicates the one-dimensional analysis of unrelated phenotypes. When calculating, sometimes GEMMA may need additional information[584]. Software is mainly used to evaluate the calculation of dispersive phenotypes, and GEMMA shows the most effectiveness in large-scale GWAS projects.

For GEMMA, the phenotypes file has been modelled. The prepared phenotype file contained six columns where the two main columns (family ID and individual ID for each twin, in this case, it was the same ID for both columns) and phenotype columns (which consisted of 4 additional columns, with questions A, B, D and a combination of all three questions). Gender was excluded. The genotype information was in separate PED files. To run the program, the final file was converted to a *fam file. GEMMA encodes the alleles in 0/1, which were calculated as controls/cases. The output file contained 15 columns, including chromosomes, SNPs and statistical results. 24155 SNPs were extracted from the results. The final file was issued in text format and was converted to excel format for further analysis. (Tables 8 and 9).

Table 8. Output file.

chr	Chromosome
rs	SNP IDs
ps	Base pair position on the chromosome
n_miss	Number of missing values for a given snp
allele1	Minor allele
allele2	Major allele
af	Allele frequency
beta	Beta estimates
se	Standard errors for beta
l_remle	Remle estimates for lambda
p_wald	p-values from Wald test
P_score	p-value results

Table 9. File outcomes.

chr	rs	ps	n_miss	allele1	allele0	af	beta	se	logl_H1	l_remle	l_mle	p_wald	p_lrt	p_score
8	8:6840576	6840576	0	0	C	0.058	8.14E-02	1.81E-02	-2.88E+01	5.22E-01	5.23E-01	7.13E-06	7.08E-06	8.09E-06
8	8:6380095	6380095	0	0	G	0.052	8.36E-02	1.90E-02	-2.93E+01	5.31E-01	5.33E-01	1.20E-05	1.19E-05	1.33E-05
8	8:6913497	6913497	0	0	A	0.022	1.10E-01	2.93E-02	-3.18E+01	5.27E-01	5.28E-01	1.70E-04	1.69E-04	1.80E-04
8	8:6549697	6549697	0	0	A	0.011	1.41E-01	3.75E-02	-3.18E+01	5.67E-01	5.68E-01	1.73E-04	1.76E-04	1.94E-04
8	8:6476734	6476734	0	0	A	0.141	4.29E-02	1.18E-02	-3.23E+01	5.49E-01	5.50E-01	2.97E-04	2.95E-04	3.14E-04
8	8:6439747	6439747	0	C	A	0.057	6.51E-02	1.81E-02	-3.24E+01	5.40E-01	5.42E-01	3.31E-04	3.28E-04	3.46E-04
8	8:6896059	6896059	0	A	C	0.013	1.24E-01	3.44E-02	-3.24E+01	5.54E-01	5.56E-01	3.33E-04	3.33E-04	3.55E-04
8	8:7197937	7197937	0	0	TA	0.075	4.85E-02	1.45E-02	-3.33E+01	5.46E-01	5.47E-01	8.54E-04	8.49E-04	8.87E-04
8	8:6751206	6751206	0	A	G	0.012	1.18E-01	3.62E-02	-3.36E+01	5.22E-01	5.22E-01	1.09E-03	1.09E-03	1.14E-03
8	8:6400911	6400911	0	0	C	0.491	2.64E-02	8.44E-03	-3.40E+01	5.34E-01	5.36E-01	1.78E-03	1.76E-03	1.82E-03
8	8:6901176	6901176	0	0	C	0.066	5.26E-02	1.70E-02	-3.41E+01	5.42E-01	5.43E-01	2.03E-03	2.02E-03	2.08E-03
8	8:6457808	6457808	0	T	A	0.065	5.22E-02	1.70E-02	-3.42E+01	5.50E-01	5.52E-01	2.21E-03	2.21E-03	2.29E-03
8	8:6435943	6435943	0	T	C	0.058	5.46E-02	1.79E-02	-3.42E+01	5.29E-01	5.31E-01	2.31E-03	2.30E-03	2.37E-03
8	8:6911196	6911196	0	0	A	0.015	1.04E-01	3.47E-02	-3.44E+01	5.28E-01	5.29E-01	2.73E-03	2.72E-03	2.80E-03
8	8:6537918	6537918	0	C	T	0.038	6.68E-02	2.23E-02	-3.44E+01	5.34E-01	5.35E-01	2.79E-03	2.77E-03	2.85E-03
8	8:6412406	6412406	0	G	A	0.059	5.28E-02	1.78E-02	-3.45E+01	5.30E-01	5.32E-01	3.04E-03	3.03E-03	3.11E-03

5.4 Results of the genetic analysis for periodontitis

Distribution analysis of the symptomatic characteristics according to participants' responses

Data from 3641 individuals were filtered to 3320 twin individuals to ensure uniform ethnicity. Of the overall 25 questions analysed, four were selected as the main indicators of gum disease. Correlation between answers and questions was tested in the model using yes = 1 and no = 0. The measurement was considered as a case-control study, where "yes" or "case" were treated as affected twins, and "no" or control as non-affected twins. Positive and negative answers were received to these questions. Figure 46 demonstrates that most participants responded negatively to the questions relating to the presence of periodontitis disease. However,

in the twins who responded positively, the majority (40.9%) noted the presence of "gum bleeding". This sign is the initial stage of the onset of periodontitis (stage 1-2 in the pathogenesis of periodontitis). In the second place, the most common symptom was "tooth loss", which is one of the leading clinical characteristics of periodontitis (21%). This symptom is the last irreversible stage of periodontitis. 15.6% related to the “dentist’s advice”; the least positive response, 4.95%, was for the “loose teeth” (Figure 46).

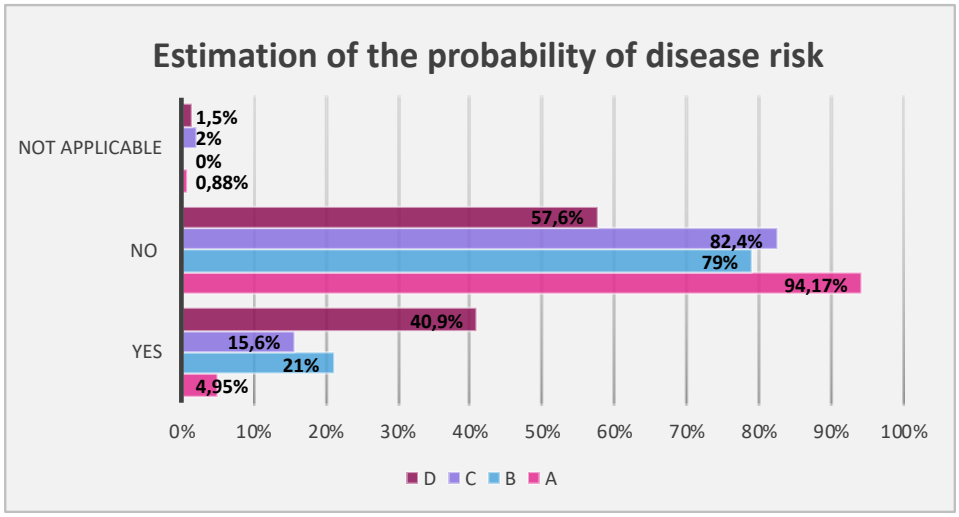


Figure 46. This figure shows the percentage of responses to each question. The highest numbers among negative responses are in question A and the lowest in question D. In the positive answers, the majority said that they had bleeding gums; the lowest response was on the loose teeth.

Calculation of the chi-square and the p-value

The correlation between answers to questions was evaluated by a chi-square test of independence. This chi-square test shows a significant correlation between responses to questions C and D, C and A, and less strongly between A and D. Question C was excluded from the subsequent analysis due to high dependence on other questions. The average correlations show that B and D, A and B, B and C are not strongly dependent on each other (Table 10). All question pairs with high chi-square values have low p-value.

Table 10. Result of chi-square calculation evaluating independence between 4 questions.

Compared questions	Chi-square statistic	P-value	Phi value
C and D	84.5859	$3.67x10^{-20}$	0.1605
B and C	4.4728	0.0371	0.0371
B and A	0.9542	0.3286	0.0169
B and D	0.0815	0.7752	0.0049
C and A	661.127	$8.49x10^{-146}$	0.4486
A and D	15.6895	$7.46x10^{-5}$	0.0687

Dependence between questions

In analysing the numbers of people who answered the two questions in the four possible different ways, there were strong signs of a correlation between questions D and C. Among the 521 people who said 'yes' to the 'dentist' question, 310 of them (59.5%) said 'yes' to the 'bleeding gums' question; among the 2762 people who said 'no' to the 'dentist' question, 1046 (38.4%) said 'yes' to the 'bleeding gums' question. So, dependence between "bleeding gums" and "dentist" was high. Questions D and C were sufficiently correlated, so both should not be used. We also observed a significant correlation between questions C and A. As a result, question C = 'dentist' was excluded (Table 9).

Score calculation

Questions A, B and D were the least correlated of the relevant questions, and all were used in association testing. The next step was to define a composite 'score' combining information from questions A, B and D, assuming that individuals affected by periodontitis would likely answer more than one question positively. This score was created based on the combination of groups A, B and D (A + B + D). When selecting an additional group that included the combined values of all three questions, the fourth group was evaluated on a scale of scores, scoring 1 for each "yes" and 0 for "no", so everyone was given a total score between

0 and 3. As a result, the score on the points answered positively (total score 1-3) amounted to 1843 subjects, and negatively (total score 0) 1477 subjects. Positive responses for all three questions were found in 31 subjects (1%), for two questions in 336 subjects (10%), for 1 question in 1476 (44%) and 0 in 1477 (45%) (Figure 47).

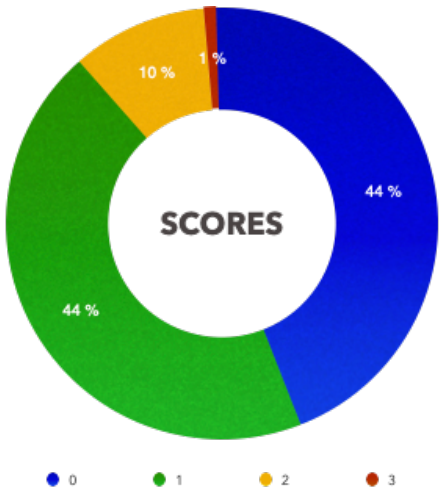


Figure 47. The distribution of scores among individuals. The composition of groups includes the combined values of three questions. The figure shows that negative responses and positive response with one dental feature has equal percentage. People that had all three dental features is 1%.

Assessment of SNP association using PLINK software

SNP association was initially analysed with the PLINK statistical software tool (Section 5.3). All SNP data received from Twins UK were combined with 4 question groups (A, B, D and A+B+D) separately (Table 11). To analyse the association, PLINK used following command:

```
> plink --file twins_DEFA --pheno phenotype.txt --
pheno-name A+B+D (A/B/D) --allow-no-sex --1
```

For all phenotypes, 1706 individuals had both genotype and phenotype data from applied 3320 phenotypes. The 3948 twins were included as they had only genotypes and were treated as “missing”. After frequency and genotyping pruning, the data were available for 24155

SNPs for each question. Composite data from A+B+D were available from a total of 1706 twins PLINK analysed SNPs for 1706 twins (non-pairs).

Table 11. SNP analysis, the outcome results.

Question Group	Cases ("yes")	Controls ("no")	Missing	SNPs
A	114	1592	3948	24155
B	382	1324	3948	24155
D	692	1014	3948	24155

Data analysis by Manhattan and Q-Q plot (PLINK results)

SNP rs56068450 at the 8:6705442 bp position showed a significant association with the response to question A, "loose teeth", in the Manhattan plot with a p-value of 4.71×10^{-6} (OR=6.271). In the question B group, "lost teeth", the strongest association was at rs2741680 at the 8:6794114 bp position with a p-value of 3.39×10^{-4} (OR=1.408). In group D, "bleeding gums", the strongest association was demonstrated by rs200409376 at the 8:6964642 bp position with a p-value of 9.87×10^{-4} (OR=0.6783). The combination of all the signs (A+B+D) showed SNP rs56068450 ($p=4.71 \times 10^{-6}$, OR=6.271). All the SNPs are located in *DEF* gene at 8p23.1 region (Figure 48). The frequency of the minor allele of rs56068450 was approximately 3% among "yes" responders and 0.5% among the others.

According to the Q-Q plot, the distribution of observed p-values with question A deviated from expected p-values, indicating an excess of observed low p-values. Deviation from the baseline is shown in low p-value (Figure 49).

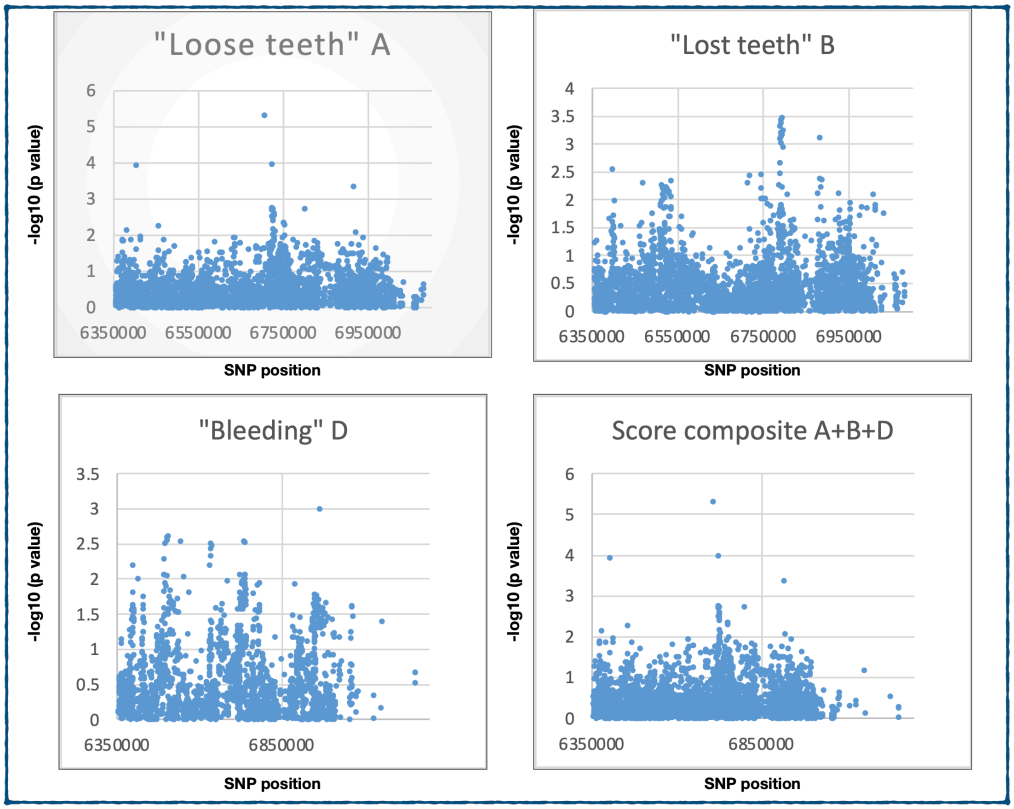


Figure 48. Manhattan plot. The vertical axis is –log10 p-values, and the horizontal axis is the position of the base pair on chromosome 8. Logistic regression data reflect the relationship between *DEFA1A3* SNPs and questionnaire responses.

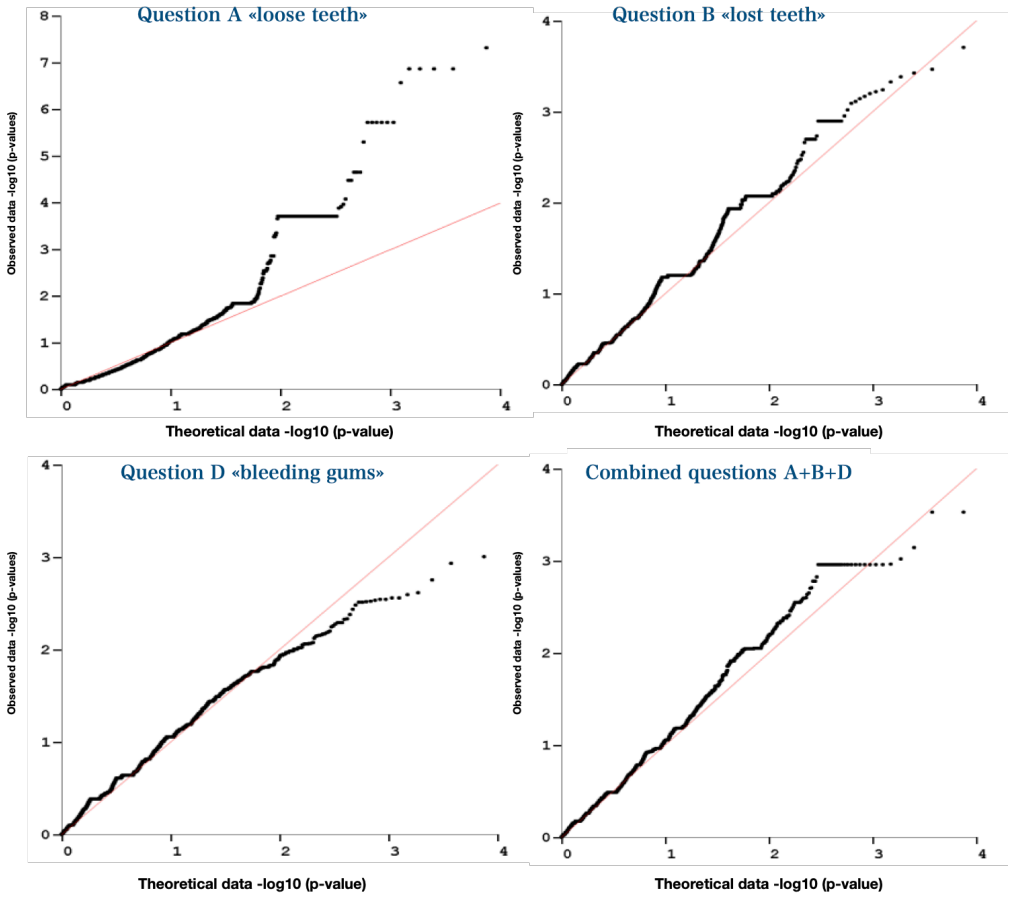


Figure 49. To compare two probability distributions Q-Q plot was used. In the Q-Q plot, two quantiles of p-values were plotted against each other with groups A, B, D and A+B+D, comparing the negative logarithm of the observed and expected association p-value distributions for SNPs around *DEFA1A3*. The vertical axis is observed data -log10 values of quantiles, and the horizontal axis is theoretical of -log10 values of quantiles.

Bonferroni correction and Concordance analysis with heritability

To make the necessary allowance for multiple testing, a Bonferroni value is applied in Table 12. As a result, the Bonferroni correction was 0.0125, where the p-values of all four questions were less than this value. Also, in addition to dividing by the number of hypotheses tested, each of the original p-values was multiplied by the correction Bonferroni. By comparing the obtained adjusted p-values, it was determined that they all corresponded to significant results. Thus, all four questions showed statistical significance according to the Bonferroni correction.

Table 12. Bonferroni test.

Questions	Adjusted p-values	p-value
A rs56068450	$5.89x10^{-8}$	$4.71x10^{-6}$
B rs2741680	$4.24x10^{-6}$	$3.39x10^{-4}$
D rs200409376	$1.23x10^{-5}$	$9.87x10^{-4}$
A+B+D rs56068450	$5.89x10^{-8}$	$4.71x10^{-6}$

An analysis between DZ and MZ twins was used to calculate concordance rates. The concordance was calculated per categorical variables (clinical features: gum bleeding, lost teeth, loose teeth, dentist advice and a combination of three symptoms). h^2 were calculated from concordance rates with Holzinger's formula (concordance MZ – concordance DZ) / (1 – concordance DZ). h^2 for the different parameters ranged from 3-17%, and concordance rates ranged between 0.24 to 0.77 for MZ twins and 0.22-0.75 for DZ twins (Table 13).

Table 13 Concordance and heritability analysis

Questions	Concordance MZ	Concordance DZ	Heritability
A	0.24	0.22	3%
B	0.65	0.6	13%
C	0.25	0.23	3%
D	0.55	0.46	17%
A+B+D	0.77	0.75	8%

Assessment of SNP association using GEMMA software

According to the results of the analysis of traits, the GEMMA model detected significantly associated SNPs. "Loose teeth" (question A) showed an association with rs572069721 ($p=8.09x10^{-6}$), allele C>C/T, which is an unknown variant of the *AGPAT5* gene, located at hg19::chr8:6840576. *AGPAT5* is a protein-coding gene (encoding a member of the 1-acylglycerol-3-phosphate O-acyltransferase family[585]). SNP rs535649221 ($p=2.10x10^{-4}$), associated with the risk of "tooth loss" (question B), allele C>C/T, is an intronic variant of the *DEFA1* gene, location at hg19::chr8:6600401. "Bleeding gums" (question D) shows an association with rs539794186 ($p=4.26x10^{-5}$), allele G>A/G, which is an intron variant of the *MCPH1* gene. Its location is at hg19::chr8:6444168. *MCPH1* encodes a DNA damage response protein. The combination of three questions showed an association with rs1448151183($p=6.78x10^{-5}$), allele A/C/G, an intron variant of the *AGPAT5* gene, located at hg19::chr8:6599161. The results showed that question A, "loose teeth", demonstrated the best-associated SNP compared to the rest of the questions.

Thus, the GEMMA model identified a total of four different associations. PLINK identified only three of these with questions A, B, and D. The last phenotype, a combination of all three questions, completely duplicated the SNPs from question A when calculating with the PLINK tool but was allocated in a separate lead SNP in the GEMMA calculation.

GEMMA identified SNPs in *AGPAT5, DEFA1* and *MCPH1.* All identified SNPs are located in the 8p23.1 region of the chromosome, towards the telomere close to the *DEFA* genes.

Data analysis by Manhattan and Q-Q plot (GEMMA results)

A Manhattan plot was used with the logistic regression data to display the association between answers to questions and the common SNPs in the population plotted against the base pair position. The negative logarithm (base 10) of the association p-value for each SNP base pair indicator was calculated. SNP rs572069721 shows a significant association with "loose teeth" ($p=8.09x10^{-6}$), rs535649221($p=2.10x10^{-4}$)

is associated with “tooth loss” and rs539794186 ($4.26x10^{-5}$) with gum bleeding, and for combined questions was rs1448151183 ($p=6.78x10^{-5}$). (Figure 50).

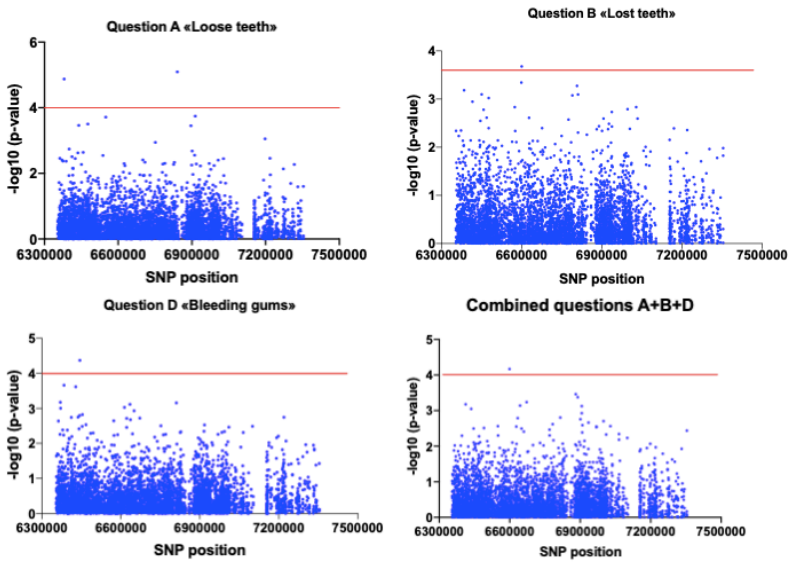
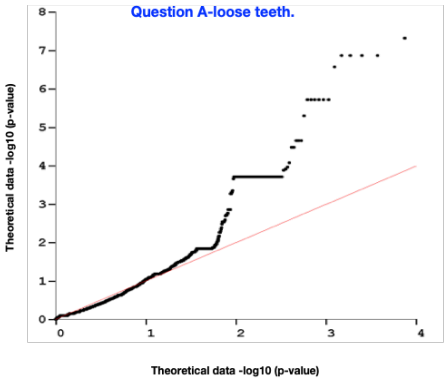


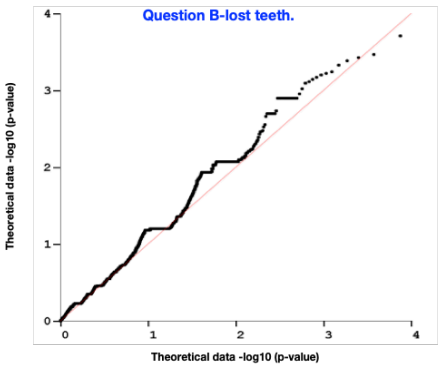
Figure 50. The vertical axis – is values -log10 p, and the horizontal axis – is the position of the base pair on chromosome 8. Logistic regression data reflect the ratio of SNP *DEFA1A3* in the population relative to the position of the base pair. A threshold for the SNP with significant p-values is shown as the red line. The spots above the red line demonstrate the most significant SNPs associated with each clinical feature.

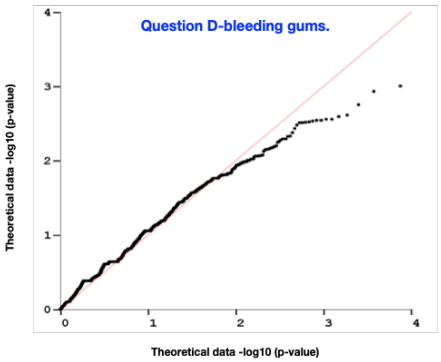
A Q-Q plot was used for GEMMA to compare the two probability distributions, like with PLINK. For each identified attribute, separate Q-Q plots were created to confirm the significance of the associations. The Q-Q plot allowed a visual check of the normality of the distribution for all observations. When comparing the GEMMA and PLINK plots, a significant deviation and excess of observed p-values (on the y-axis) are shown from the baseline in question A, which indicates a significant association of SNPs with the response to this question. However, the PLINK graph shows a sharper excess of p-values compared to GEMMA. GEMMA showed a more even and gradual increase in low p-values. Thus, the PLINK result showed a stronger association at low p-values than GEMMA. Question B showed even values on the baseline for all p-

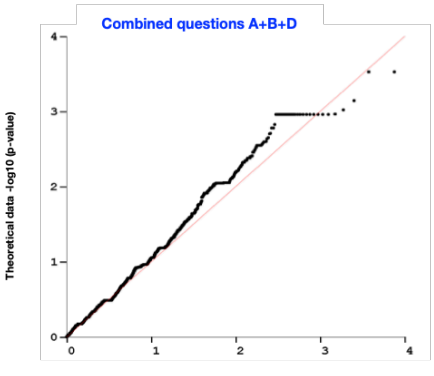
values the same in the GEMMA, in contrast to the uneven fluctuations in the PLINK. Question D shows the opposite picture between GEMMA and PLINK. In the last plot, PLINK shows a gradual increase of indicators compared to stable ones in GEMMA. This pattern indicates a bias in the PLINK analysis (Figure 51). According to the Q-Q plot, PLINK generally shows more abrupt deviations than GEMMA. The change begins to appear with an increase in the negative logarithm of p-values. However, the overall pattern is the same in both graphs.

PLINK

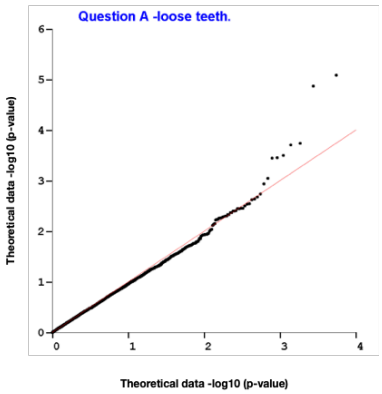

Question A-loose teeth.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)

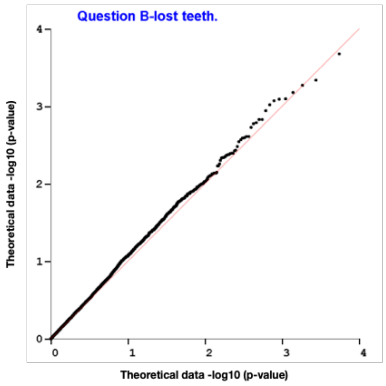

Question B-lost teeth.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)

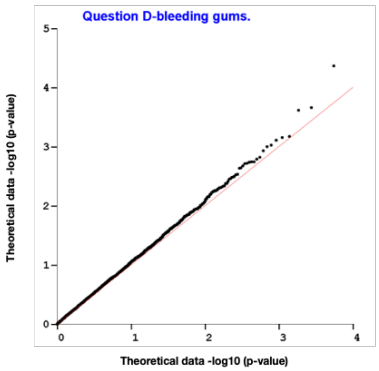

Question D-bleeding gums.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)


Combined questions A+B+D
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)

GEMMA


Question A -loose teeth.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)


Question B-lost teeth.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)


Question D-bleeding gums.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)

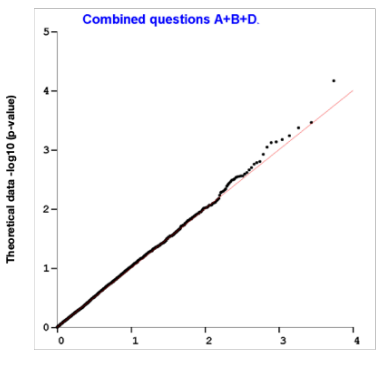

Combined questions A+B+D.
Theoretical data -log10 (p-value)
Theoretical data -log10 (p-value)

Figure 51. Q-Q plots of GEMMA and PLINK software. According to the Q-Q plot, the distribution of the observed p-values with question A deviates from the expected p-values for both software, which indicates the excess of the lowest p-values. Other plots correspond to Gaussian distribution. No changes were observed between the observed and expected data. The vertical axis is the observed data -log10 values of quantiles and the horizontal axis is the theoretical -log10 values of quantiles.

Statistical and comparative evaluation of algorithms for PLINK and GEMMA software

For statistical and comparative analysis of PLINK and GEMMA, all the data was used from question A, "loose teeth", since they showed a significant association with periodontitis compared to other questions. To evaluate SNPs, graphs were constructed with a combination of data from two software (Figures 52,53,54,55).

The first evaluation was made by the Manhattan plot (Figure 52). The results of association tests are shown in the Manhattan plot, in which the x-axis corresponds to the SNP positions, and the y-axis corresponds to the negative decimal logarithm of p-values. The higher the statistical significance of an individual attribute, the greater the -log10 p-value of the coordinate at the corresponding point on the graph. Further investigations in measuring the PLINK and GEMMA software results found no significant differences (Figure 53). The Q-Q plot, demonstrating the results of PLINK and GEMMA, also shows no significant differences. Q-Q plot displays the exact p-values (Figure 54). Overall, the results of PLINK and GEMMA complement each other. The correlation showed R^2=0.9870, which is a significant indicator since it is close to 1 (Figure 55).

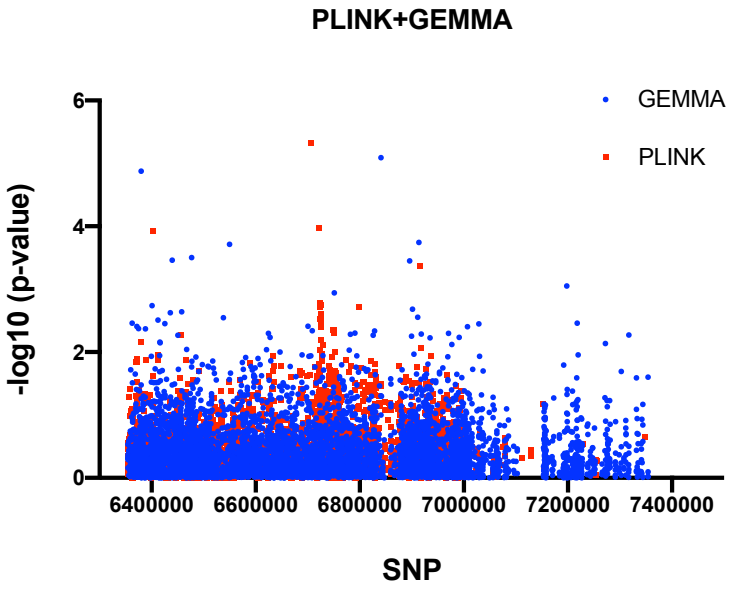


Figure 52. The Manhattan plot with GEMMA and PLINK. The Manhattan plot showed stronger signals of individual SNPs in both software. It should be noted that there are no apparent differences in the range of p-values and SNPs between the two software.

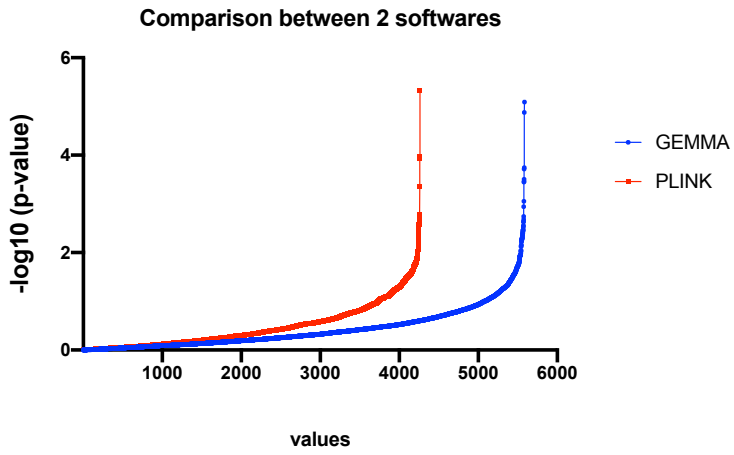


Figure 53. Comparison between the two software PLINK and GEMMA. When comparing the negative logarithm of the p-values with the number of samples of SNP, for individual question A, the number of SNPs in GEMMA was more than in PLINK. However, the overall distributions of p-values are almost the same in measuring the results and their values.

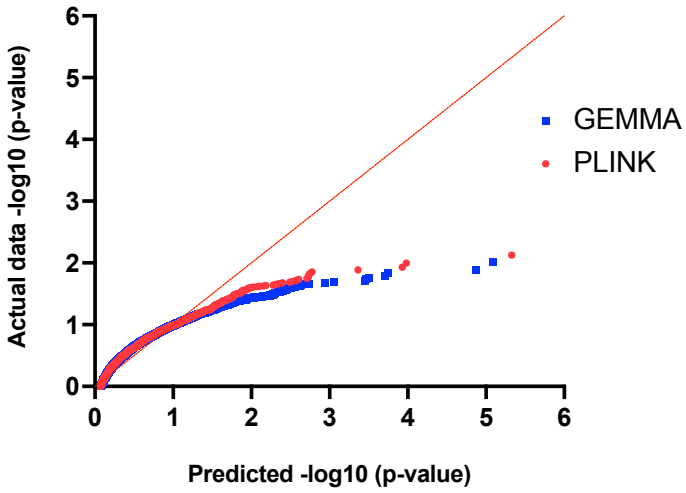


Figure 54. Statistical analysis of the algorithm implemented in the two software PLINK and GEMMA. This graph shows comparative estimates of two software that use the same twin data from minus 10 logarithms. The comparison was made in a Q-Q plot with the estimated and actual data in a multidimensional setting.

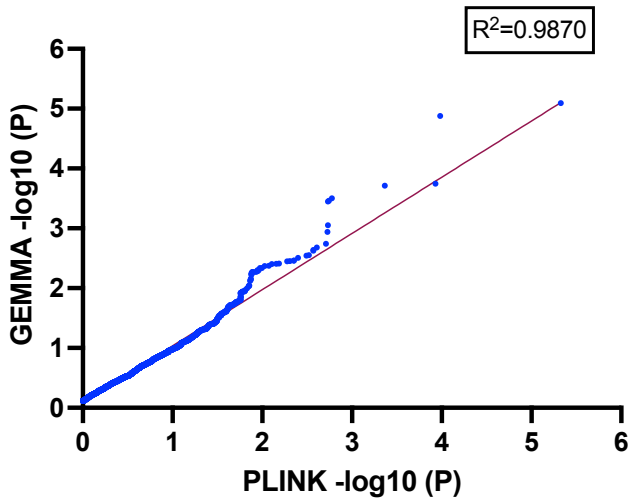


Figure 55. Comparison plot of compared data of the two software. When comparing two tools in accordance with the simulation, the p-values of the GEMMA increased starting from 3 logarithms; that is, an excess of low p-values (10^{-4}) are demonstrated. These indicators show a higher significance with a decrease in the p-value of GEMMA compared to PLINK for SNPs.

5.5 Discussion

One of the research targets was to determine and assess the degree of association between oral health parameters (which in this case are questionnaires) with the characteristics of genetic polymorphism and the role of the genotype (*DEFA* gene) in the development of susceptibility to relevant dental diseases, especially periodontitis. The approach for this analysis was calculating the significance of associations using statistical tests and software.

The methodological basis of the experimental part of the study was the comparison of the survey data with clinical indicators. Statistical methods were used to identify associations between the studied factors with clinical signs. The two software packages, PLINK and GEMMA, helped identify seven new SNPs for each clinical characteristic. These phenotypes were: loose teeth, tooth loss, bleeding gums, and a combination of all three signs. The identified four phenotypes are characterised by clinical features.

The study revealed that 51.4% of the participants were at risk for developing periodontitis based on the analysis of survey data. Two software approaches were used to validate the results. For comparison, four phenotypes that had previously shown the greatest association with periodontitis were reanalysed.

Analysis by PLINK software identified three significant SNPs in the *DEF* gene: rs56068450, rs2741680 and rs200409376, in all participants compared to the control population. The Bonferroni test showed that all SNPs had significant associations after correction. Taking into account the result, the SNPs and clinical symptoms can be the main and accurate for the diagnosis of periodontitis.

In a GEMMA cohort study, four new SNPs were identified: rs572069721, rs535649221, rs539794186 rs1448151183. The results with "loose teeth" demonstrated the best-associated SNP, compared to the rest of the questions, as well as the PLINK results. These SNPs have not previously been identified using PLINK for non-transformed phenotypes. GEMMA identified SNPs in *AGPAT5, DEFA1* and *MCPH1.* All identified SNPs are located in the 8p23.1 region of the chromosome.

Thus, the two (PLINK and GEMMA) software packages are considered complementary, not competing tools. Each of the software analyses data using its own personal computing techniques. And they

take into account different SNPs. By the results, it was found that GEMMA was as effective for prediction as PLINK with a complete set of SNP markers. Thus, the results obtained in the course of our work have revealed the important multifaceted role of *DEAFA1A3* in the development of chronic inflammatory diseases of the mouth, particularly periodontitis.

The results of our research are consistent with the data of some research groups. Thus, in their studies, Munz et al. identified an association between *DEFA1A3* flanking SNPs and aggressive periodontitis[469]. SNP identification at the *DEFA1A3* locus was also carried out in our research. Additionally, the genetic association of the *DEFA1A3* gene with the clinical symptoms of periodontitis was determined.

To date, a number of genes are known, the allelic state of which presumably affects the likelihood of developing periodontitis in an individual, as well as the rate of progression and severity of the disease. It should be noted that the studies of Taba et al. indicated that significant diagnostic or prognostic markers for identifying patients who are at risk of developing severe forms of periodontitis in terms of genetic predisposition to various forms of periodontitis[155].

In various studies, signs of periodontitis in populations are explained by hereditary factors (0.38 for all twins, 0.15 in family studies). In twin studies with disease prevalence between MZ twins and DZ twins, the score was 0.29 for all twin and family studies, with high heterogeneity[54]. In our study, the concordance rates ranged from 0.24-0.77 for MZ twins and 0.22-0.75 for DZ twins.

In GWAS studies, low heritability rates were determined. According to a meta-analysis, h^2 was 0.07[54]. Our study showed h^2 for the different parameters ranging from 3-17%, which was close to GWAS studies. Thus, for moderate chronic periodontitis[586] and severe chronic periodontitis[485], the range is from 0 to 0.14 and 0.23 to 0.24 for severe chronic periodontitis[486,586].

Some researchers, like Michalowicz et al., evaluated parameters associated with periodontitis (periodontal pocket depth, periodontal index, and plaque index) in 110 adult twins and concluded that between 38 and 82% of the population variability in these parameters is genetically determined. Observations of familial aggregation of aggressive forms of periodontitis and a high concordance of a decrease in alveolar bone density in pairs of MZ compared with DZ indicate a significant genetic component in the development of periodontal

diseases[479]. His other twin study has found that the risk of periodontitis in adults has a genetic basis, where the influence of genetic and environmental factors was taken into account. According to Michalowicz et al. study results, statistical significance with a genetic component was revealed. In MZ, all clinical symptoms were more similar than in DZ; this also concerned the degree and severity of periodontitis. The risk of inheritance and occurrence of periodontitis during the study was found to be approximately 50%[587]. In our study, the percentage of risk of periodontitis among questionnaire respondents was 51.4%. Thus, genetic changes may be an essential factor in determining the differentiated risk of periodontitis.

Defensins play an essential role in maintaining local immunity. In studies by Antunes et al., *DEFB1* gene polymorphisms are associated with developing persistent apical periodontitis. The study found the most significant association with periodontitis for the rs1047031 SNP of the *DEFB1* gene. The Japanese researchers also found an association of rs1800972 SNP with chronic periodontitis, which is presumably based on the low expression of *DEFB1* in gingival tissues in carriers of the rs1799946, rs1800972 and rs11362 genotypes, which in turn leads to a decrease in local resistance[588].

Some other genes have been found to be associated with periodontitis. Munz et al. in the study identified SNPs in the *SIGLEC5*[469], *MTND1P5* (chr8), *LOC107984137* (chr16)[589] genes to be connected with aggressive periodontitis. In the work of Suzuki et al. for patients from Japan, a relationship was established between the allelic state of several collagen genes and the development of periodontitis[590]. The work of Murakami discusses the molecular mechanisms of the pathogenesis of periodontal diseases. Long-term inflammation is characterised by continuous macrophage infiltration of periodontal tissues, which produces pro-inflammatory cytokines (TNF and IL-1). Being in the periodontal tissue, they intensify the differentiation and increase osteoclast activity, leading to the complete destruction of the alveolar bone. According to the author, inhibiting the production of pro-inflammatory cytokines will prevent or treat periodontal disease[591]. In Papadopoulos et al. studies, macrophages respond to *Porphyromonas gingivalis* by higher expression of TNF-α, interleukin-6, monocyte chemoattractant protein-1, and macrophage inflammatory protein-1a[592]. In the Yu et al. research, all macrophages perform a central part in the mechanism of periodontitis. The change in the ratio of macrophages of the M1/M2 phenotypes, which remains undetermined to date, is

considered to be the decisive mechanism in mediating damage to periodontal tissues[593]. Different periodontal phenotypes are examined as other parts of several similar factors and modulated as combinations of genetic risk loci. In this regard, various disease manifestations have common alleles and risk covariates.

Analysis of information on the frequency of alleles in the test population allows us to quantify the effect associated with the disease of the allele on pathological processes, that is, the odds ratio, how many times the presence of this SNP increases the risk of developing the disease. In this work, the maximum odds ratios observed for significantly associated SNPs were in the range of 0.6783-6.271. An approach to detecting an allele associated with a disease is used, which is based on a standard case-control scheme.

5.6 Conclusion

Periodontitis, like many common diseases, is a complex multifactorial disease. A characteristic feature of this pathology is that it has a mild clinical picture at the beginning of its development, slowly progresses and eventually acquires a chronic course. Although the pathogenesis of periodontitis focuses on periodontopathogenic bacteria and external factors (such as alcohol, smoking, stress, etc.), the role of genetic factors is obvious[594]. The greatest severity of periodontitis is manifested by 40 years, so introducing a prognostic screening test for a younger population will help determine the risk of developing periodontitis at an early stage. Currently, the diagnosis of periodontal diseases in most countries is limited to making a diagnosis only on clinical and radiological data. Additionally, there is an invasive screening test called the Periodontal Screening Index (PSI). In the US, there is a periodontal sensitivity test (PST). This genetic test determines the risk of developing a severe form of oral disease. The procedure consists of taking a cheek swab and conducting a laboratory analysis. Based on the analysis results, an individual treatment plan can be created. This test helps to determine the risk, as early detection and intervention help prevent tooth loss and save teeth.

It should be noted that genomic areas of research, primarily related to the search for biomarkers of periodontal disease, taking into account

the genetic characteristics of patients, can help develop promising approaches to prevention and early diagnosis. Large (Biobank) scale association studies will likely be most productive for future progress in defining genetic factors.

CHAPTER VI. LIMITATIONS AND FUTURE PERSPECTIVES.

The main limitation of the study was the occurrence of COVID-19 and the lockdown for two years, which excluded the possibility of obtaining biological material (blood) from QMC patients and closed access to work in the laboratory.

The study of only one type of population limited the research possibilities in comprehensively analysing predisposition to MFDs. Obtaining patients with different populations would expand the analysis of association with diseases.

For a more precise definition of the structure and mechanism of CNV, using more than one research method would be more productive.

Questionnaire data collection was used in the study of periodontitis. For more accurate clinical characteristics, in addition to the questionnaire survey, it is necessary to collect data by examining the teeth and identifying a specific stage in the development of periodontitis.

Perspectives for the study:

1. Continuing research by increasing the number of samples to search for new gene markers and variants that may also be associated with these diseases.
2. Continue to study the pathogenetic mechanism of the MFD occurrence in more detail based on the expansion of data collection and analysis of their association, considering different populations.
3. Determination of more pronounced structural and functional features of rare pancreas and oral cavity diseases.
4. The exact genetic marker of periodontitis is currently unknown; searching for more candidate genes will improve understanding of the pathogenesis of inflammatory diseases of the oral cavity.
5. It is necessary to develop a prognostic criterion using the study results, which will allow to assess the risk of developing pancreatitis and periodontitis to the clinical and laboratory manifestations stage and justify the implementation of early pathogenetic measures.

6. Consider the possibility of the contribution of the obtained genetic information to scientific research in the clinical field. Conduct research in the field of implementation of the results obtained in clinical practice and compare. Analyse genetic findings and assess their impact and relationship to the preventive field of medicine.

CHAPTER VII. OVERALL CONCLUSION

Multifactorial diseases are the most common diseases, having an important role in the clinical practice of medicine. The problem of studying the genetic mechanisms of predisposition to diseases, the inheritance of which does not obey Mendelian rules, remains so far one of the least developed in human genetics. The study of the relationship between the genes predisposing to the development of multifactorial diseases will help solve the problems of modern clinical medicine and clarify the classification of the forms of diseases and their nosological affiliation. The problem of low effectiveness of therapeutic and preventive measures is associated with the lack of their aetiological orientation due to an insufficient understanding of the critical mechanisms for the formation of the vast majority of multifactorial diseases. Studies on causative genes in multifactorial diseases with marked pathological changes may help identify genes that contribute to the development of certain diseases.

The thesis explored mechanisms aimed at identifying the relationship of predisposition genes with multifactorial diseases using genetic methods. Namely, the responses of different people to the factors of the internal environment among the candidate genes for the role of predisposition. The genes *AMY1* and *AMY2*, *SPINK1* and *DEFA*, which were putative modifiers of diseases, were studied. Pancreatic diseases and periodontitis are multifactorial diseases with an expressed genetic predisposition. A diagnostic or prognostic marker is needed to determine diseases in patients that in a severe risk. Progress has been made in understanding the concept and the role of amylase and CNV in developing pancreatic diseases. The study determined the link between *AMY* CNV and pancreatic disease by using PRT. The study found that CNV *AMY1* and *AMY2* were higher in patients with pancreatic disease. The high risk of pancreatic disease is due to increased CNs, which may be one of the reasons for a greater predisposition to diseases in a combination of external factors than in isolation. According to the study, in cases from 2 to 16 CNV *AMY1* and up to 5 copies inclusive of *AMY2* were revealed. Duplications in *AMY2* have more association with pancreatic disease. The threshold risk for developing the pancreatic disease for *AMY1* (6 copies) was revealed. The presence of pancreatic diseases was characterised by a range from 6 to 9 CN *AMY1* among cases, and *AMY2* CNV was characterised by a wider range of copies.

Genetic studies of pancreatic diseases have an impact on the understanding of the aetiology of diseases. The study determined that patients with pancreatic diseases are characterised by normal and slightly increased body weight. A positive correlation was found between the development of pancreatic diseases at the age of 40 years and above; women are at more risk for the disease than men.

The simultaneous presence of several factors, such as age from 40 years and older, high CNs, and gender difference (females), significantly increases the risk of developing the pancreatic disease, compared with the influence of only one factor.

CNV is an integral part of disease pathogenicity analysis. Clinical practice shows the complexity of determining amylase in the blood as an enzyme since pancreatic amylase is limited in time. Therefore, testing the *AMY* gene as a diagnostic tool would allow more reliable determination of the clinical characteristics of pancreatic diseases. We expect our research to contribute to developing a more accurate diagnostic tool in the clinic in the future.

The relative CNs of amylase genes could be used as a biomarker to differentiate between different types of pancreatic diseases. May be helpful in the structure of healthcare, mainly at central service laboratories and remote locations with low laboratory equipment.

The presence of the N34S mutation of the *SPINK1* gene was detected in people with pancreatic diseases with a frequency of occurrence of 3.3%, which is higher than the population risk. The N34S mutation was detected only in a high CN in *AMY2*, while this feature was not observed in *AMY1*.

Studies on causative genes in multifactorial diseases with significant pathological changes in periodontal tissues can contribute to the identification of genes. Modern genetic studies of periodontal disease primarily include the identification of gene variants that play an essential role in the pathophysiology of the disease. Defensins are a barrier to infection, a robust regulatory system that harmonises the relationship of macro- and microorganisms, pro- and anti-inflammatory systems, and play a key role in developing the pathological process.

In our work, we investigated the genetic predisposition of the defensin gene to oral disease periodontitis. The data obtained fit into the general ideas about the development of periodontitis and show threshold indicators that diagnose periodontitis.

The analysis showed that 51.4% of the survey participants' responses had periodontitis risk indicators. Correlation analysis of the data showed a significant association between clinical factors and SNPs in the *DEFA1A3* gene region, indicating an association with an increased risk of periodontitis. The software identified a set of SNP markers (seven SNPs) related to specific clinical signs of periodontitis. The greatest severity of periodontitis is manifested by age 40, so introducing a prognostic screening test for a younger population will determine the risk of developing periodontitis at an early stage. Thus, the results obtained during our study revealed an important role of the *DEAFA1A3* gene in the development of chronic inflammatory diseases of the oral cavity, particularly periodontitis. It is necessary to interpret them in connection with the multifactorial nature of the pathogenesis of periodontitis with an aggressive course, a complex nature of inheritance and a large number of polymorphisms that determine the phenotype of this disease.

Identifying genetic predisposition will allow the development of new integrated approaches to differentiated, personalised diagnostics and a more accurate choice of tactics for treating and preventing the disease. This study can be used as one of the models explaining the genetic mechanisms of predisposition to multifactorial diseases.

Bibliography:

1 Duarte, C. W., Vaughan, L. K., Beasley, T. M. & Tiwari, H. K. in *Emery and Rimoin's Principles and Practice of Medical Genetics (Sixth Edition)* (eds David Rimoin, Reed Pyeritz, & Bruce Korf) 1-15 (Academic Press, 2013).

2 Fialkowski, A., Beasley, T. M. & Tiwari, H. K. in *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics (Seventh Edition)* (eds Reed E. Pyeritz, Bruce R. Korf, & Wayne W. Grody) 323-358 (Academic Press, 2019).

3 Martin, R. J. *Fanaroff and Martin's Neonatal-Perinatal Medicine*. 11 edn, Vol. 1 154-173 (Elsevier, Inc, 2020).

4 Baranov, V. S. Molecular medicine: Molecular diagnostics, preventive medicine, and gene therapy. *Molecular Biology* **34**, 590-600 (2000). https://doi.org:10.1007/BF02759567

5 Lango, H. & Weedon, M. N. What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice? *J Intern Med* **263**, 16-27 (2008). https://doi.org:10.1111/j.1365-2796.2007.01895.x

6 Biggio, J. R. in *Perinatal Genetics* (eds Mary E. Norton, Jeffrey A. Kuller, & Lorraine Dugoff) 1-10 (Elsevier, 2019).

7 Ingrid, L. Multifactorial Inheritance and Genetic Disease. *Nature Education* **1** (2008).

8 Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835-847 (2003). https://doi.org:10.1038/nature01626

9 Moffatt, M. F. Genes in asthma: new genes and new ways. *Current Opinion in Allergy and Clinical Immunology* **8**, 411-417 (2008). https://doi.org:10.1097/ACI.0b013e32830f1dc1

10 Waters, E. A., Muff, J. & Hamilton, J. G. Multifactorial beliefs about the role of genetics and behavior in common health conditions: prevalence and associations with participant characteristics and engagement in health behaviors. *Genetics in medicine : official journal of the American College of Medical Genetics* **16**, 913-921 (2014). https://doi.org:10.1038/gim.2014.49

11 Luyapan, J. *et al.* A new efficient method to detect genetic interactions for lung cancer GWAS. *BMC Med Genomics* **13**, 162 (2020). https://doi.org:10.1186/s12920-020-00807-9

12 Ropers, H. H. & van Karnebeek, C. D. Rare diseases: human genome research is coming home. *Cold Spring Harb Mol Case Stud* **8** (2022). https://doi.org:10.1101/mcs.a006210

13 Kere, J., Ripatti, S. & Perola, M. Genetics of multifactorial diseases. *Duodecim* **126**, 2305-2310 (2010).

14 Duarte, C. W., Vaughan, L. K., Beasley, M. T. & Tiwari, H. K. in *Reference Module in Biomedical Sciences* (Elsevier, 2014).

15 Kardia, S. L. & Wang, C. The role of health education and behavior in public health genetics. *Health Educ Behav* **32**, 583-588 (2005). https://doi.org:10.1177/1090198105278560

16 Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* **17**, 1520-1528 (2007). https://doi.org:10.1101/gr.6665407

17 Smerecnik, C. M. R., Mesters, I., de Vries, N. K. & de Vries, H. Educating the general public about multifactorial genetic disease: applying a theory-based framework to understand current public knowledge. *Genetics in Medicine* **10**, 251-258 (2008). https://doi.org:10.1097/GIM.0b013e31816b4ffd

18 Kliegman, R. M. *Nelson Textbook of Pediatrics*. 21 edn, 640-652 (Elsevier Inc, 2020).

19 Motulsky, A. G. *The genetic basis of common diseases*. (Oxford University Press, 2002).

20 Lesueur, F. *et al.* GEMO, a National Resource to Study Genetic Modifiers of Breast and Ovarian Cancer Risk in BRCA1 and BRCA2 Pathogenic Variant Carriers. *Frontiers in Oncology* **8** (2018). https://doi.org:10.3389/fonc.2018.00490

21 Klugman, S. & Rabin-Havt, S. S. in *Perinatal Genetics* (eds Mary E. Norton, Jeffrey A. Kuller, & Lorraine Dugoff) 11-19 (Elsevier, 2019).

22 Narimatsu, H. Gene–Environment Interactions in Preventive Medicine: Current Status and Expectations for the Future. *International Journal of Molecular Sciences* **18**, 302 (2017).

23 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009). https://doi.org:10.1038/nature08494

24 McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356-369 (2008). https://doi.org:10.1038/nrg2344

25 Igo, R. P., Jr., Kinzy, T. G. & Cooke Bailey, J. N. Genetic Risk Scores. *Curr Protoc Hum Genet* **104**, e95 (2019). https://doi.org:10.1002/cphg.95

26 Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biology* **18**, 77 (2017). https://doi.org:10.1186/s13059-017-1212-4

27 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* **12**, 745-755 (2011). https://doi.org:10.1038/nrg3031

28 Mahdavi, M. *et al.* Hereditary breast cancer; Genetic penetrance and current status with BRCA. *Journal of Cellular Physiology* **234**, 5741-5750 (2019). https://doi.org:https://doi.org/10.1002/jcp.27464

29 van der Groep, P., van der Wall, E. & van Diest, P. J. Pathology of hereditary breast cancer. *Cellular Oncology* **34**, 71-88 (2011). https://doi.org:10.1007/s13402-011-0010-3

30 Benes, L. B., Brandt, D. J., Brandt, E. J. & Davidson, M. H. How Genomics Is Personalizing the Management of Dyslipidemia and Cardiovascular Disease Prevention. *Current Cardiology Reports* **20**, 138 (2018). https://doi.org:10.1007/s11886-018-1079-3

31 Golfeyz, S., Lewis, S. & Weisberg, I. S. Hemochromatosis: pathophysiology, evaluation, and management of hepatic iron overload with a focus on MRI. *Expert Review of Gastroenterology & Hepatology* **12**, 767-778 (2018). https://doi.org:10.1080/17474124.2018.1496016

32 Alexander, J. & Kowdley, K. V. Hereditary Hemochromatosis: Genetics, Pathogenesis, and Clinical Management. *Annals of Hepatology* **4**, 240-247 (2005). https://doi.org:10.1016/S1665-2681(19)32046-0

33 Greulich, T. & Vogelmeier, C. F. Alpha-1-antitrypsin deficiency: increasing awareness and improving diagnosis. *Therapeutic Advances in Respiratory Disease* **10**, 72-84 (2015). https://doi.org:10.1177/1753465815602162

34 Steele, A. M. *et al.* Prevalence of vascular complications among patients with glucokinase mutations and prolonged, mild hyperglycemia. *Jama* **311**, 279-286 (2014). https://doi.org:10.1001/jama.2013.283980

35 Bishay, R. H. & Greenfield, J. R. A review of maturity onset diabetes of the young (MODY) and challenges in the management of glucokinase-MODY. *Medical Journal of Australia* **205**, 480-485 (2016). https://doi.org:https://doi.org/10.5694/mja16.00458

36 Amed, S. & Oram, R. Maturity-Onset Diabetes of the Young (MODY): Making the Right Diagnosis to Optimize Treatment. *Canadian Journal of Diabetes* **40**, 449-454 (2016). https://doi.org:10.1016/j.jcjd.2016.03.002

37 Kim, S.-H. Maturity-Onset Diabetes of the Young: What Do Clinicians Need to Know? *Diabetes Metab J.* , 468-477 (2015).

38 Steck, A. K. & Rewers, M. J. Genetics of type 1 diabetes. *Clin Chem* **57**, 176-185 (2011). https://doi.org:10.1373/clinchem.2010.148221

39 Noble, J. A. *et al.* The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *American journal of human genetics* **59**, 1134-1148 (1996).

40 Bell, G. I., Horita, S. & Karam, J. H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33**, 176-183 (1984). https://doi.org:10.2337/diab.33.2.176

41 Nisticò, L. *et al.* The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. *Hum Mol Genet* **5**, 1075-1080 (1996). https://doi.org:10.1093/hmg/5.7.1075

42 Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* **36**, 337-338 (2004). https://doi.org:10.1038/ng1323

43 Todd, J. A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* **39**, 857-864 (2007). https://doi.org:10.1038/ng2068
44 Ali, O. Genetics of type 2 diabetes. *World J Diabetes* **4**, 114-123 (2013). https://doi.org:10.4239/wjd.v4.i4.114
45 del Bosque-Plata, L., Martínez-Martínez, E., Espinoza-Camacho, M. Á. & Gragnoli, C. The Role of TCF7L2 in Type 2 Diabetes. *Diabetes* **70**, 1220-1228 (2021). https://doi.org:10.2337/db20-0573
46 Wolfe, C. M., Fitz, N. F., Nam, K. N., Lefterov, I. & Koldamova, R. The Role of APOE and TREM2 in Alzheimer's Disease—Current Understanding and Perspectives. *International Journal of Molecular Sciences* **20**, 81 (2019).
47 Tombelli, S., Mascini, M., Braccini, L., Anichini, M. & Turner, A. P. F. Coupling of a DNA piezoelectric biosensor and polymerase chain reaction to detect apolipoprotein E polymorphisms. *Biosensors and Bioelectronics* **15**, 363-370 (2000). https://doi.org:https://doi.org/10.1016/S0956-5663(00)00092-0
48 Martínez-Morillo, E. *et al.* Assessment of Peptide Chemical Modifications on the Development of an Accurate and Precise Multiplex Selected Reaction Monitoring Assay for Apolipoprotein E Isoforms. *Journal of Proteome Research* **13**, 1077-1087 (2014). https://doi.org:10.1021/pr401060x
49 Pagon, R. A. Genetic testing for disease susceptibilities: consequences for genetic counseling. *Trends Mol Med* **8**, 306-307 (2002). https://doi.org:10.1016/s1471-4914(02)02348-1
50 Ruderfer, D. M., Korn, J. & Purcell, S. M. Family-based genetic risk prediction of multifactorial disease. *Genome Med* **2**, 2 (2010). https://doi.org:10.1186/gm123
51 Wray, N. R. & Visscher, P. M. Estimating trait heritability. *Nature Education* **1** (2008).
52 Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255-266 (2008). https://doi.org:10.1038/nrg2322
53 Sadee, W. *et al.* Missing heritability of common diseases and treatments outside the protein-coding exome. *Hum Genet* **133**, 1199-1215 (2014). https://doi.org:10.1007/s00439-014-1476-7
54 Nibali, L. *et al.* What Is the Heritability of Periodontitis? A Systematic Review. *Journal of dental research* **98**, 632-641 (2019). https://doi.org:10.1177/0022034519842510
55 Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics*. 4 edn, (Longman, 1996).
56 Tenesa, A. & Haley, C. S. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet* **14**, 139-149 (2013). https://doi.org:10.1038/nrg3377
57 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010). https://doi.org:10.1038/ng.608

58 Kinnersley, B. *et al.* Quantifying the heritability of glioma using genome-wide complex trait analysis. *Sci Rep* **5**, 17267 (2015). https://doi.org:10.1038/srep17267
59 Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* **2**, e41 (2006). https://doi.org:10.1371/journal.pgen.0020041
60 Valberg, M., Stensrud, M. J. & Aalen, O. O. The surprising implications of familial association in disease risk. *BMC Public Health* **18**, 135 (2018). https://doi.org:10.1186/s12889-018-5033-5
61 Aalen, O. O., Valberg, M., Grotmol, T. & Tretli, S. Understanding variation in disease risk: the elusive concept of frailty. *Int J Epidemiol* **44**, 1408-1421 (2015). https://doi.org:10.1093/ije/dyu192
62 Balmain, A., Gray, J. & Ponder, B. The genetics and genomics of cancer. *Nat Genet* **33 Suppl**, 238-244 (2003). https://doi.org:10.1038/ng1107
63 Riley, B. D. *et al.* Essential Elements of Genetic Cancer Risk Assessment, Counseling, and Testing: Updated Recommendations of the National Society of Genetic Counselors. *Journal of Genetic Counseling* **21**, 151-161 (2012). https://doi.org:https://doi.org/10.1007/s10897-011-9462-x
64 Frank, C., Fallah, M., Sundquist, J., Hemminki, A. & Hemminki, K. Population Landscape of Familial Cancer. *Sci Rep* **5**, 12891 (2015). https://doi.org:10.1038/srep12891
65 Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138**, 2044-2058 (2010). https://doi.org:10.1053/j.gastro.2010.01.054
66 Ripperger, T., Gadzicki, D., Meindl, A. & Schlegelberger, B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* **17**, 722-731 (2009). https://doi.org:10.1038/ejhg.2008.212
67 Johns, L. E. & Houlston, R. S. A systematic review and meta-analysis of familial prostate cancer risk. *BJU Int* **91**, 789-794 (2003). https://doi.org:10.1046/j.1464-410x.2003.04232.x
68 Hemminki, K., Li, X., Sundquist, J. & Sundquist, K. Familial risks for amyotrophic lateral sclerosis and autoimmune diseases. *Neurogenetics* **10**, 111-116 (2009). https://doi.org:10.1007/s10048-008-0164-y
69 Marder, K. *et al.* Risk of Parkinson's disease among first-degree relatives: A community-based study. *Neurology* **47**, 155-160 (1996). https://doi.org:10.1212/wnl.47.1.155
70 van den Oord, R. A. & Sheikh, A. Filaggrin gene defects and risk of developing allergic sensitisation and allergic disorders: systematic review and meta-analysis. *Bmj* **339**, b2433 (2009). https://doi.org:10.1136/bmj.b2433
71 Thomsen, S. F. The contribution of twin studies to the understanding of the aetiology of asthma and atopic diseases. *Eur Clin Respir J* **2**, 27803 (2015). https://doi.org:10.3402/ecrj.v2.27803

72 Thomsen, S. F., Duffy, D. L., Kyvik, K. O. & Backer, V. Genetic influence on the age at onset of asthma: a twin study. *J Allergy Clin Immunol* **126**, 626-630 (2010). https://doi.org:10.1016/j.jaci.2010.06.017
73 Thomsen, S. F., van der Sluis, S., Kyvik, K. O. & Backer, V. A study of asthma severity in adult twins. *Clin Respir J* **6**, 228-237 (2012). https://doi.org:10.1111/j.1752-699X.2011.00273.x
74 Koenig, L. B. in *The Wiley Encyclopedia of Personality and Individual Differences* 415-419 (2020).
75 Aulchenko U.S. Methodological approaches and strategies for mapping genes that control complex human traits. *Vestnik* **10**, 189-202 (2006).
76 Lee, D. S. *et al.* The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci U S A* **105**, 9880-9885 (2008). https://doi.org:10.1073/pnas.0802208105
77 Collister, J. A., Liu, X. & Clifton, L. Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists. *Frontiers in Genetics* **13** (2022). https://doi.org:10.3389/fgene.2022.818574
78 Klein, K. & Zanger, U. M. Pharmacogenomics of Cytochrome P450 3A4: Recent Progress Toward the "Missing Heritability" Problem. *Front Genet* **4**, 12 (2013). https://doi.org:10.3389/fgene.2013.00012
79 Sadee, W. The relevance of "missing heritability " in pharmacogenomics. *Clin Pharmacol Ther* **92**, 428-430 (2012). https://doi.org:10.1038/clpt.2012.116
80 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198 (2012). https://doi.org:10.1073/pnas.1119675109
81 Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* **20**, 467-484 (2019). https://doi.org:10.1038/s41576-019-0127-1
82 Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* **10**, 241-251 (2009). https://doi.org:10.1038/nrg2554
83 Aschard, H. *et al.* Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *American journal of human genetics* **90**, 962-972 (2012). https://doi.org:10.1016/j.ajhg.2012.04.017
84 Choquet, H. & Meyre, D. Genetics of Obesity: What have we Learned? *Curr Genomics* **12**, 169-179 (2011). https://doi.org:10.2174/138920211795677895
85 Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genetics* **36**, 949 (2004). https://doi.org:10.1038/ng1416
https://www.nature.com/articles/ng1416#supplementary-information
86 Sebat, J. *et al.* Large-Scale Copy Number Polymorphism in the Human Genome. *Science* **305**, 525-528 (2004). https://doi.org:10.1126/science.1098918

87 Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Reviews Genetics* **7**, 85-97 (2006). https://doi.org:10.1038/nrg1767

88 Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nature Reviews Genetics* **16**, 172 (2015). https://doi.org:10.1038/nrg3871 https://www.nature.com/articles/nrg3871#supplementary-information

89 McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics* **40**, 1166 (2008). https://doi.org:10.1038/ng.238 https://www.nature.com/articles/ng.238#supplementary-information

90 Pirooznia, M., Goes, F. S. & Zandi, P. P. Whole-genome CNV analysis: advances in computational approaches. *Front Genet* **6**, 138 (2015). https://doi.org:10.3389/fgene.2015.00138

91 Cooper, G. M., Nickerson, D. A. & Eichler, E. E. Mutational and selective effects on copy-number variants in the human genome. *Nature Genetics* **39**, S22 (2007). https://doi.org:10.1038/ng2054 https://www.nature.com/articles/ng2054#supplementary-information

92 Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature genetics* **39**, 1217-1224 (2007). https://doi.org:10.1038/ng2142

93 Reymond, A., Henrichsen, C. N. & Chaignat, E. Copy number variants, diseases and gene expression. *Human Molecular Genetics* **18**, R1-R8 (2009). https://doi.org:10.1093/hmg/ddp011

94 Schuster-Böckler, B., Conrad, D. & Bateman, A. Dosage sensitivity shapes the evolution of copy-number varied regions. *PloS one* **5**, e9474-e9474 (2010). https://doi.org:10.1371/journal.pone.0009474

95 Repping, S. *et al.* High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature Genetics* **38**, 463 (2006). https://doi.org:10.1038/ng1754 https://www.nature.com/articles/ng1754#supplementary-information

96 Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444-454 (2006). https://doi.org:10.1038/nature05329

97 Martin, C. L., Kirkpatrick, B. E. & Ledbetter, D. H. Copy number variants, aneuploidies, and human disease. *Clinics in perinatology* **42**, 227-vii (2015). https://doi.org:10.1016/j.clp.2015.03.001

98 Kashevarova, A. A. & Lebedev, I. N. Genomic architecture of human chromosomal diseases. *Russian Journal of Genetics* **52**, 447-462 (2016). https://doi.org:10.1134/S1022795416040062

99 Lupski, J. R. *et al.* DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**, 219-232 (1991). https://doi.org:https://doi.org/10.1016/0092-8674(91)90613-4

100 Lee, J. A. & Lupski, J. R. Genomic Rearrangements and Gene Copy-Number Alterations as a Cause of Nervous System Disorders. *Neuron* **52**, 103-121 (2006). https://doi.org:https://doi.org/10.1016/j.neuron.2006.09.027

101 Carter, G. T. *et al.* Neuropathic pain in Charcot-Marie-tooth disease. *Archives of Physical Medicine and Rehabilitation* **79**, 1560-1564 (1998). https://doi.org:10.1016/S0003-9993(98)90421-X
102 Baloh, R. H., Schmidt, R. E., Pestronk, A. & Milbrandt, J. Altered Axonal Mitochondrial Transport in the Pathogenesis of Charcot-Marie-Tooth Disease from Mitofusin 2 Mutations. *The Journal of Neuroscience* **27**, 422-430 (2007). https://doi.org:10.1523/jneurosci.4798-06.2007
103 Crespi, B., Summers, K. & Dorus, S. Genomic sister-disorders of neurodevelopment: an evolutionary approach. *Evolutionary applications* **2**, 81-100 (2009). https://doi.org:10.1111/j.1752-4571.2008.00056.x
104 Kim, Y. H., Kim, N. G., Lim, J. G., Park, C. & Kim, H. Chromosomal alterations in paired gastric adenomas and carcinomas. *The American journal of pathology* **158**, 655-662 (2001). https://doi.org:10.1016/S0002-9440(10)64007-2
105 Golzio, C. & Katsanis, N. Genetic architecture of reciprocal CNVs. *Current opinion in genetics & development* **23**, 240-248 (2013). https://doi.org:10.1016/j.gde.2013.04.013
106 Girirajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. *Human molecular genetics* **19**, R176-R187 (2010). https://doi.org:10.1093/hmg/ddq366
107 Girirajan, S. *et al.* Genotype-phenotype correlation in Smith-Magenis syndrome: evidence that multiple genes in 17p11.2 contribute to the clinical spectrum. *Genetics in medicine : official journal of the American College of Medical Genetics* **8**, 417-427 (2006). https://doi.org:10.109701.gim.0000228215.32110.89
108 Potocki, L. *et al.* Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *American journal of human genetics* **80**, 633-649 (2007). https://doi.org:10.1086/512864
109 Tassabehji, M. Williams–Beuren syndrome: a challenge for genotype–phenotype correlations. *Human Molecular Genetics* **12**, R229-R237 (2003). https://doi.org:10.1093/hmg/ddg299
110 Sharp, A. J. *et al.* Characterization of a recurrent 15q24 microdeletion syndrome. *Human Molecular Genetics* **16**, 567-572 (2007). https://doi.org:10.1093/hmg/ddm016
111 Li, L. *et al.* Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genetics* **16**, 243-251 (1997). https://doi.org:10.1038/ng0797-243
112 Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. & Elsea, S. H. Mutations in RAI1 associated with Smith–Magenis syndrome. *Nature Genetics* **33**, 466 (2003). https://doi.org:10.1038/ng1126
https://www.nature.com/articles/ng1126#supplementary-information
113 Cardoso, C. *et al.* Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of

17p13.3. *American journal of human genetics* **72**, 918-930 (2003). https://doi.org:10.1086/374320

114 Weiss, L. A. *et al.* Association between Microdeletion and Microduplication at 16p11.2 and Autism. *New England Journal of Medicine* **358**, 667-675 (2008). https://doi.org:10.1056/NEJMoa075974

115 Koolen, D. A. *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nature Genetics* **44**, 639 (2012). https://doi.org:10.1038/ng.2262
https://www.nature.com/articles/ng.2262#supplementary-information

116 Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *American journal of human genetics* **79**, 439-448 (2006). https://doi.org:10.1086/505915

117 Hollox, E. J. *et al.* Psoriasis is associated with increased beta-defensin genomic copy number. *Nature genetics* **40**, 23-25 (2008). https://doi.org:10.1038/ng.2007.48

118 Huse, K. *et al.* Genetic Variants of the Copy Number Polymorphic β-Defensin Locus Are Associated with Sporadic Prostate Cancer. *Tumor Biology* **29**, 83-92 (2008). https://doi.org:10.1159/000135688

119 Lupski, J. R. Genomic rearrangements and sporadic disease. *Nature Genetics* **39**, S43 (2007). https://doi.org:10.1038/ng2084

120 Forsberg, L. A., Absher, D. & Dumanski, J. P. Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. *Journal of medical genetics* **50**, 1-10 (2013). https://doi.org:10.1136/jmedgenet-2012-101322

121 Zhang, F., Gu, W., Hurles, M. E. & Lupski, J. R. Copy Number Variation in Human Health, Disease, and Evolution. *Annual Review of Genomics and Human Genetics* **10**, 451-481 (2009). https://doi.org:10.1146/annurev.genom.9.081307.164217

122 Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A* **107**, 961-968 (2010). https://doi.org:10.1073/pnas.0912629107

123 Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853 (2007). https://doi.org:10.1126/science.1136678

124 Hu, J. *et al.* CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *Journal of neurodevelopmental disorders* **7**, 26-26 (2015). https://doi.org:10.1186/s11689-015-9122-9

125 Lopatkina, M. E., Kashevarova, A. A. & Lebedev, I. N. Estimation of association of CNTN6 copy number variation with idiopathic intellectual disability. *Russian Journal of Genetics* **52**, 1004-1006 (2016). https://doi.org:10.1134/S102279541609009X

126 Conrad, D. F. *et al.* A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics* **38**, 1251 (2006). https://doi.org:10.1038/ng1911

https://www.nature.com/articles/ng1911#supplementary-information
127 Cooper, G. M., Nickerson, D. A. & Eichler, E. E. *Mutational and selective effects on copy-number variants in the human genome. Nat Genet 39: S22-S29*. Vol. 39 (2007).
128 Graubert, T. A. *et al.* A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS genetics* **3**, e3-e3 (2007). https://doi.org:10.1371/journal.pgen.0030003
129 Wong, K. K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *American journal of human genetics* **80**, 91-104 (2007). https://doi.org:10.1086/510560
130 Weerakkody, R. A. *et al.* Targeted next-generation sequencing makes new molecular diagnoses and expands genotype–phenotype relationship in Ehlers–Danlos syndrome. *Genetics In Medicine* **18**, 1119 (2016). https://doi.org:10.1038/gim.2016.14
https://www.nature.com/articles/gim201614#supplementary-information
131 Rebrikov D.V., Korostin D.O., Shubina E.S. & V.V., I. *NGS: High performance sequencing*. 2 edn, Vol. 232 (Binom. Laboratoriya znaniy, 2015).
132 LaDuca, H. *et al.* Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLOS ONE* **12**, e0170843 (2017). https://doi.org:10.1371/journal.pone.0170843
133 Baranov, V. S., Arsentyev, V. G., Serebryakova, E. A. & Shabalov, N. P. Genetics of Connective Tissue Diseases: State of the Art and Perspectives. *Russian Journal of Genetics* **54**, 788-795 (2018). https://doi.org:10.1134/S1022795418070025
134 Calus, M. P. L. & Vandenplas, J. SNPrune: an efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. *Genetics Selection Evolution* **50**, 34 (2018). https://doi.org:10.1186/s12711-018-0404-z
135 Balagué-Dobón, L., Cáceres, A. & González, J. R. Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure. *Briefings in Bioinformatics* **23**, bbac043 (2022). https://doi.org:10.1093/bib/bbac043
136 Mao, X., Young, B. D. & Lu, Y. J. The application of single nucleotide polymorphism microarrays in cancer research. *Curr Genomics* **8**, 219-228 (2007). https://doi.org:10.2174/138920207781386924
137 LaFramboise, T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* **37**, 4181-4193 (2009). https://doi.org:10.1093/nar/gkp552
138 Schaaf, C. P., Wiszniewska, J. & Beaudet, A. L. Copy Number and SNP Arrays in Clinical Diagnostics. *Annual Review of Genomics and Human Genetics* **12**, 25-51 (2011). https://doi.org:10.1146/annurev-genom-092010-110715
139 Skryabin, N. A., Vasilyev, S. A. & Lebedev, I. N. Epigenetic silencing of genomic structural variations. *Russian Journal of Genetics* **53**, 1072-1079 (2017). https://doi.org:10.1134/S1022795417100106

140 Goh, K.-I. & Choi, I.-G. Exploring the human diseasome: the human disease network. *Briefings in Functional Genomics* **11**, 533-542 (2012). https://doi.org:10.1093/bfgp/els032

141 Favorova O.O., Bashinskaya V.V., Kulakova O.G., Boyko A.N. & Favorov A.V. Genome-wide association study as a method to analyze the genome architecture in polygenic diseases, with the example of multiple sclerosis. *Molecular Biology* **48**, 496-507 (2014).

142 Chen, Y., Mei, Y. & Jiang, X. Universal and high-fidelity DNA single nucleotide polymorphism detection based on a CRISPR/Cas12a biochip. *Chemical Science* **12**, 4455-4462 (2021). https://doi.org:10.1039/D0SC05717G

143 Ahmad, T., Valentovic, M. A. & Rankin, G. O. Effects of cytochrome P450 single nucleotide polymorphisms on methadone metabolism and pharmacodynamics. *Biochemical Pharmacology* **153**, 196-204 (2018). https://doi.org:https://doi.org/10.1016/j.bcp.2018.02.020

144 van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418-426 (2014). https://doi.org:10.1016/j.tig.2014.07.001

145 Vallejos-Vidal, E. *et al.* Single-Nucleotide Polymorphisms (SNP) Mining and Their Effect on the Tridimensional Protein Structure Prediction in a Set of Immunity-Related Expressed Sequence Tags (EST) in Atlantic Salmon (Salmo salar). *Front Genet* **10**, 1406 (2019). https://doi.org:10.3389/fgene.2019.01406

146 Gill, P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* **114**, 204-210 (2001). https://doi.org:10.1007/s004149900117

147 Schaid, D. J. *et al.* Comparison of Microsatellites Versus Single-Nucleotide Polymorphisms in a Genome Linkage Screen for Prostate Cancer–Susceptibility Loci. *The American Journal of Human Genetics* **75**, 948-965 (2004). https://doi.org:https://doi.org/10.1086/425870

148 Middleton, F. A. *et al.* Genomewide Linkage Analysis of Bipolar Disorder by Use of a High-Density Single-Nucleotide–Polymorphism (SNP) Genotyping Assay: A Comparison with Microsatellite Marker Assays and Finding of Significant Linkage to Chromosome 6q22. *The American Journal of Human Genetics* **74**, 886-897 (2004). https://doi.org:https://doi.org/10.1086/420775

149 Daw, E. W., Heath, S. C. & Lu, Y. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. *BMC Genet* **6 Suppl 1**, S32 (2005). https://doi.org:10.1186/1471-2156-6-s1-s32

150 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108 (2005). https://doi.org:10.1038/nrg1521

151 Varzari, A., Deyneko, I. V., Tudor, E. & Turcan, S. Polymorphisms of glutathione S-transferase and methylenetetrahydrofolate reductase genes in Moldavian patients with ulcerative colitis: Genotype-

phenotype correlation. *Meta Gene* **7**, 76-82 (2016). https://doi.org:10.1016/j.mgene.2015.12.002

152 Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nature genetics* **41**, 885-890 (2009). https://doi.org:10.1038/ng.406

153 Yoo, S. S. *et al.* Putative functional variants of XRCC1 identified by Regulome DB were not associated with lung cancer risk in a Korean population. *Cancer Genetics* **208**, 19-24 (2015). https://doi.org:10.1016/j.cancergen.2014.11.004

154 Sobrino, B., Brión, M. & Carracedo, A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* **154**, 181-194 (2005). https://doi.org:10.1016/j.forsciint.2004.10.020

155 Taba Jr, M., Souza, Sergio Luis Scombatti de and Mariguela, Viviane Casagrande. Periodontal disease: a genetic perspective. *Brazilian Oral Research* **26**, 32-38 (2012). https://doi.org:https://doi.org/10.1590/S1806-83242012000700006

156 Kinane, D. F. & Hart, T. C. Genes and gene polymorphisms associated with periodontal disease. *Crit Rev Oral Biol Med* **14**, 430-449 (2003). https://doi.org:10.1177/154411130301400605

157 Baranov, V. S. Genome paths: a way to personalized and predictive medicine. *Acta Naturae* **1**, 70-80 (2009).

158 Underwood, P. C. & Read, C. Genetic association studies in nursing practice and scholarship. *Journal Of Nursing Scholarship* **40**, 212-218 (2008).

159 Flint, J. GWAS. *Current Biology* **23**, R265-R266 (2013).

160 The Wellcome Trust Case Control, C. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661 (2007). https://doi.org:10.1038/nature05911
https://www.nature.com/articles/nature05911#supplementary-information

161 The International HapMap, C., Altshuler, D. & Donnelly, P. A haplotype map of the human genome. *Nature* **437**, 1299 (2005). https://doi.org:10.1038/nature04226
https://www.nature.com/articles/nature04226#supplementary-information

162 Gudmundsson, S. *et al.* Variant interpretation using population databases: Lessons from gnomAD. *Hum Mutat* **43**, 1012-1030 (2022). https://doi.org:10.1002/humu.24309

163 Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012). https://doi.org:10.1038/nature11632

164 Gabriel, S. B. *et al.* The Structure of Haplotype Blocks in the Human Genome. *Science* **296**, 2225-2229 (2002). https://doi.org:10.1126/science.1069424

165 Gibbs, R. A. *et al.* The International HapMap Project. *Nature* **426**, 789 (2003). https://doi.org:10.1038/nature02168
https://www.nature.com/articles/nature02168#supplementary-information

166 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020). https://doi.org:10.1038/s41586-020-2308-7
167 Beck, T., Rowlands, T., Shorter, T. & Brookes, A. J. GWAS Central: an expanding resource for finding and visualising genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res* **51**, D986-d993 (2023). https://doi.org:10.1093/nar/gkac1017
168 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542 (2016). https://doi.org:10.1038/nature17671
169 Bonnefond, A. *et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet* **44**, 297-301 (2012). https://doi.org:10.1038/ng.1053
170 Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357-363 (2014). https://doi.org:10.1038/ng.2915
171 Majithia, A. R. *et al.* Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A* **111**, 13127-13132 (2014). https://doi.org:10.1073/pnas.1410428111
172 Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389 (2005). https://doi.org:10.1126/science.1109557
173 Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387-389 (2009). https://doi.org:10.1126/science.1167728
174 Garg, G. *et al.* Type 1 diabetes-associated IL2RA variation lowers IL-2 signaling and contributes to diminished CD4+CD25+ regulatory T cell function. *Journal of immunology (Baltimore, Md. : 1950)* **188**, 4644-4653 (2012). https://doi.org:10.4049/jimmunol.1100272
175 Alcina, A. *et al.* IL2RA/CD25 gene polymorphisms: uneven association with multiple sclerosis (MS) and type 1 diabetes (T1D). *PLoS One* **4**, e4137 (2009). https://doi.org:10.1371/journal.pone.0004137
176 Giza, S. *et al.* The role of PTPN22 C1858T gene polymorphism in diabetes mellitus type 1: first evaluation in Greek children and adolescents. *Biomed Res Int* **2013**, 721604 (2013). https://doi.org:10.1155/2013/721604
177 Chistiakov, D. A. & Turakulov, R. I. CTLA-4 and its role in autoimmune thyroid disease. *J Mol Endocrinol* **31**, 21-36 (2003). https://doi.org:10.1677/jme.0.0310021
178 Tomuschat, C. & Puri, P. RET gene is a major risk factor for Hirschsprung's disease: a meta-analysis. *Pediatr Surg Int* **31**, 701-710 (2015). https://doi.org:10.1007/s00383-015-3731-y
179 Yamamoto, S. & Ma, X. Role of Nod2 in the development of Crohn's disease. *Microbes Infect* **11**, 912-918 (2009). https://doi.org:10.1016/j.micinf.2009.06.005

180 McGovern, D. P., van Heel, D. A., Ahmad, T. & Jewell, D. P. NOD2 (CARD15), the first susceptibility gene for Crohn's disease. *Gut* **49**, 752-754 (2001). https://doi.org:10.1136/gut.49.6.752
181 Brand, A., Brand, H. & Schulte in den Bäumen, T. The impact of genetics and genomics on public health. *Eur J Hum Genet* **16**, 5-13 (2008). https://doi.org:10.1038/sj.ejhg.5201942
182 Doche, M. E. *et al.* Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest* **122**, 4732-4736 (2012). https://doi.org:10.1172/jci62696
183 Liu, R. *et al.* Rare Loss-of-Function Variants in NPC1 Predispose to Human Obesity. *Diabetes* **66**, 935-947 (2017). https://doi.org:10.2337/db16-0877
184 Saeed, S. *et al.* Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet* **50**, 175-179 (2018). https://doi.org:10.1038/s41588-017-0023-6
185 Grarup, N. *et al.* Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet* **50**, 172-174 (2018). https://doi.org:10.1038/s41588-017-0022-7
186 Loos, R. J. & Yeo, G. S. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nature reviews. Endocrinology* **10**, 51-61 (2014). https://doi.org:10.1038/nrendo.2013.227
187 Cardinale, C. J. *et al.* Targeted resequencing identifies defective variants of decoy receptor 3 in pediatric-onset inflammatory bowel disease. *Genes Immun* **14**, 447-452 (2013). https://doi.org:10.1038/gene.2013.43
188 Ellinghaus, D. *et al.* Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* **145**, 339-347 (2013). https://doi.org:10.1053/j.gastro.2013.04.040
189 Beaudoin, M. *et al.* Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet* **9**, e1003723 (2013). https://doi.org:10.1371/journal.pgen.1003723
190 Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881-888 (2008). https://doi.org:10.1126/science.1156409
191 Kronenberg, F. Genome-wide association studies in aging-related processes such as diabetes mellitus, atherosclerosis and cancer. *Exp Gerontol* **43**, 39-43 (2008). https://doi.org:10.1016/j.exger.2007.09.005
192 Ehret, G. B. *et al.* A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *American journal of human genetics* **91**, 863-871 (2012). https://doi.org:10.1016/j.ajhg.2012.09.013
193 Ruderfer, D. M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* **19**, 1017-1024 (2014). https://doi.org:10.1038/mp.2013.138

194 Baranov, V. *Gene polymorphism, ecogenetic diseases and predictive personalized medicine*. Vol. 9 (Scientific research centre of Obstetrics, Gynecology and Reproduction named by Otto D.O., 2011).
195 Goh, K.-I. *et al.* The human disease network. *Proceedings of the National Academy of Sciences* **104**, 8685-8690 (2007). https://doi.org:10.1073/pnas.0701361104
196 R.J., L. Linkage analysis and genome-wide association studies (GWAS). (2021). <https://med.libretexts.org/@go/page/38262>.
197 Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* **37**, 1243-1246 (2005). https://doi.org:10.1038/ng1653
198 Zondervan, K. T. & Cardon, L. R. The complex interplay among factors that influence allelic association. *Nat Rev Genet* **5**, 89-100 (2004). https://doi.org:10.1038/nrg1270
199 Cupples, L. A. *et al.* The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med Genet* **8 Suppl 1**, S1 (2007). https://doi.org:10.1186/1471-2350-8-s1-s1
200 Maouche, S. & Schunkert, H. Strategies Beyond Genome-Wide Association Studies for Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **32**, 170-181 (2012). https://doi.org:doi:10.1161/ATVBAHA.111.232652
201 Pulit, S. L., de With, S. A. & de Bakker, P. I. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol* **41**, 145-151 (2017). https://doi.org:10.1002/gepi.22032
202 Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-1120 (2015). https://doi.org:10.1038/ng.3390
203 Loh, P. R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-1392 (2015). https://doi.org:10.1038/ng.3431
204 Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**, 446-450 (2010). https://doi.org:10.1038/nrg2809
205 Feinberg, A. P. *et al.* Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Transl Med* **2**, 49ra67 (2010). https://doi.org:10.1126/scitranslmed.3001262
206 Costello, H. *et al.* A pilot study of potential brain donor satisfaction and attitudes towards telephone assessment. *Int J Geriatr Psychiatry* **32**, 1247-1256 (2017). https://doi.org:10.1002/gps.4605
207 Brookes, K. J. *et al.* Genotyping of the Alzheimer's disease Genome-Wide Association Study index single nucleotide polymorphisms in the Brains for Dementia research cohort. *Journal of Alzheimer's Disease* **64**, 355-362 (2018).

208 Craddock, N. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713-720 (2010). https://doi.org:10.1038/nature08979
209 Lluís-Ganella, C. *et al.* Additive Effect of Multiple Genetic Variants on the Risk of Coronary Artery Disease. *Revista Española de Cardiología (English Edition)* **63**, 925-933 (2010). https://doi.org:10.1016/S1885-5857(10)70186-9
210 Dupuis, J. & O'Donnell, C. J. Interpreting results of large-scale genetic association studies: separating gold from fool's gold. *Jama* **297**, 529-531 (2007). https://doi.org:10.1001/jama.297.5.529
211 Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704-712 (2010). https://doi.org:10.1038/nature08516
212 Murray, C. J. & Lopez, A. D. Evidence-based health policy--lessons from the Global Burden of Disease Study. *Science* **274**, 740-743 (1996). https://doi.org:10.1126/science.274.5288.740
213 Christensen, K. & Murray, J. C. What genome-wide association studies can do for medicine. *N Engl J Med* **356**, 1094-1097 (2007). https://doi.org:10.1056/NEJMp068126
214 Xiao, A. Y. *et al.* Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies. *Lancet Gastroenterol Hepatol* **1**, 45-55 (2016). https://doi.org:10.1016/s2468-1253(16)30004-8
215 Boxhoorn, L. *et al.* Acute pancreatitis. *Lancet (London, England)* **396** (2020). https://doi.org:10.1016/S0140-6736(20)31310-6
216 Karakayali, F. Y. Surgical and interventional management of complications caused by acute pancreatitis. *World journal of gastroenterology* **20**, 13412-13423 (2014). https://doi.org:10.3748/wjg.v20.i37.13412
217 Kokosis, G., Perez, A. & Pappas, T. N. Surgical management of necrotizing pancreatitis: an overview. *World journal of gastroenterology* **20**, 16106-16112 (2014). https://doi.org:10.3748/wjg.v20.i43.16106
218 Schepers, N. J. *et al.* Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut* **68**, 1044-1051 (2019). https://doi.org:10.1136/gutjnl-2017-314657
219 Hirota, M., Ohmuraya, M. & Baba, H. Genetic background of pancreatitis. *Postgraduate medical journal* **82**, 775-778 (2006). https://doi.org:10.1136/pgmj.2006.050591
220 Pongprasobchai, S., Vibhatavata, P. & Apisarnthanarak, P. Severity, Treatment, and Outcome of Acute Pancreatitis in Thailand: The First Comprehensive Review Using Revised Atlanta Classification. *Gastroenterol Res Pract* **2017**, 3525349-3525349 (2017). https://doi.org:10.1155/2017/3525349
221 Setiawan, V. W. *et al.* Prospective Study of Alcohol Drinking, Smoking, and Pancreatitis: The Multiethnic Cohort. *Pancreas* **45**, 819-825 (2016). https://doi.org:10.1097/MPA.0000000000000657

222 Vujasinovic, M. *et al.* Impact of a clinical pathway on treatment outcome in patients with acute pancreatitis. *World journal of gastroenterology* **21**, 9150-9155 (2015). https://doi.org:10.3748/wjg.v21.i30.9150
223 Mallédant, Y., Malbrain, M. L. N. G. & Reuter, D. A. What's new in the management of severe acute pancreatitis? *Intensive Care Med* **41**, 1957-1960 (2015). https://doi.org:10.1007/s00134-015-3903-1
224 Rawla, P., Bandaru, S. S. & Vellipuram, A. R. Review of Infectious Etiology of Acute Pancreatitis. *Gastroenterology Res* **10**, 153-158 (2017). https://doi.org:10.14740/gr858w
225 Cao, C. L. *et al.* Acute pancreatitis induced by etoposide-lobaplatin combination chemotherapy used for the treatment of lung cancer: A case report and literature review. *Medicine (Baltimore)* **96**, e7601 (2017). https://doi.org:10.1097/md.0000000000007601
226 Kopecky, K., Moreland, A., Hebert, C. & Colbert, G. B. Plasmapheresis for recurrent acute pancreatitis from hypertriglyceridemia. *Proc (Bayl Univ Med Cent)* **30**, 358-359 (2017). https://doi.org:10.1080/08998280.2017.11929648
227 Durie, P. R., Lerch, M. M., Lowenfels, A. B., Maisonneuve, P. & Ulrich, C. D. *Genetic disorders of the exocrine pancreas: an overview and update*. Vol. 1 (Karger Medical and Scientific Publishers, 2002).
228 Whitcomb, D. C. Pancreatitis: TIGAR-O Version 2 Risk/Etiology Checklist With Topic Reviews, Updates, and Use Primers. *Clin Transl Gastroenterol* **10**, e00027 (2019). https://doi.org:10.14309/ctg.0000000000000027
229 Whitcomb, D. C. *et al.* Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nature Genetics* **14**, 141-145 (1996). https://doi.org:10.1038/ng1096-141
230 Kavutharapu, S. *et al.* Role of proteases and antiprotease in the etiology of chronic pancreatitis. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association* **18**, 364-368 (2012). https://doi.org:10.4103/1319-3767.103427
231 Jansen, P. L. *et al.* Exocrine pancreatic insufficiency and idiopathic haemochromatosis. *Postgraduate medical journal* **60**, 675-678 (1984).
232 Sahin-Tóth, M. & Tóth, M. High-Affinity Ca2+ Binding Inhibits Autoactivation of Rat Trypsinogen. *Biochemical and Biophysical Research Communications* **275**, 668-671 (2000). https://doi.org:https://doi.org/10.1006/bbrc.2000.3355
233 Sadakari, Y. *et al.* Mutant KRAS and GNAS DNA Concentrations in Secretin-Stimulated Pancreatic Fluid Collected from the Pancreatic Duct and the Duodenal Lumen. *Clinical and translational gastroenterology* **5**, e62-e62 (2014). https://doi.org:10.1038/ctg.2014.14
234 Rosendahl, J., Bödeker, H., Mössner, J. & Teich, N. Hereditary chronic pancreatitis. *Orphanet journal of rare diseases* **2**, 1-1 (2007). https://doi.org:10.1186/1750-1172-2-1

235 Correlation between genotype and phenotype in patients with cystic fibrosis. *N Engl J Med* **329**, 1308-1313 (1993). https://doi.org:10.1056/nejm199310283291804
236 Witt, H. *et al.* Mutation in the SPINK1 trypsin inhibitor gene, alcohol use, and chronic pancreatitis. *Jama* **285**, 2716-2717 (2001).
237 Truninger, K. *et al.* Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. *The American journal of gastroenterology* **96**, 2657-2661 (2001). https://doi.org:10.1111/j.1572-0241.2001.04047.x
238 Howes, N. *et al.* Clinical and genetic characteristics of hereditary pancreatitis in Europe. *Clinical Gastroenterology and Hepatology* **2**, 252-261 (2004). https://doi.org:https://doi.org/10.1016/S1542-3565(04)00013-8
239 Witt, H. Chronic pancreatitis and cystic fibrosis. *Gut* **52 Suppl 2**, ii31-ii41 (2003). https://doi.org:10.1136/gut.52.suppl_2.ii31
240 Database, G. A. (2021).
241 Witt, H. *et al.* Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nature Genetics* **25**, 213 (2000). https://doi.org:10.1038/76088
242 Noone, P. G. *et al.* Cystic fibrosis gene mutations and pancreatitis risk: Relation to epithelial ion transport and trypsin inhibitor gene mutations. *Gastroenterology* **121**, 1310-1319 (2001). https://doi.org:https://doi.org/10.1053/gast.2001.29673
243 Le Maréchal, C. *et al.* Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nature Genetics* **38**, 1372 (2006). https://doi.org:10.1038/ng1904
https://www.nature.com/articles/ng1904#supplementary-information
244 Keim, V. Role of genetic disorders in acute recurrent pancreatitis. *World journal of gastroenterology* **14**, 1011-1015 (2008). https://doi.org:10.3748/wjg.14.1011
245 Correlation between Genotype and Phenotype in Patients with Cystic Fibrosis. *New England Journal of Medicine* **329**, 1308-1313 (1993). https://doi.org:10.1056/nejm199310283291804
246 Drenth, J. P. H., te Morsche, R. & Jansen, J. B. M. J. Mutations in serine protease inhibitor Kazal type 1 are strongly associated with chronic pancreatitis. *Gut* **50**, 687-692 (2002).
247 Banjar, H. H. *et al.* Genotype patterns for mutations of the cystic fibrosis transmembrane conductance regulator gene: a retrospective descriptive study from Saudi Arabia. *Ann Saudi Med* **40**, 15-24 (2020). https://doi.org:10.5144/0256-4947.2020.15
248 Ostedgaard, L. S. *et al.* CFTR with a partially deleted R domain corrects the cystic fibrosis chloride transport defect in human airway epithelia in vitro and in mouse nasal mucosa in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3093-3098 (2002). https://doi.org:10.1073/pnas.261714599

249 Whitcomb, D. C. Genetic risk factors for pancreatic disorders. *Gastroenterology* **144**, 1292-1302 (2013). https://doi.org:10.1053/j.gastro.2013.01.069
250 Ferrari, M. *et al.* Cystic fibrosis patients with liver disease are not genetically distinct. *American journal of human genetics* **48**, 815-816 (1991).
251 Dror, Y. *et al.* Draft consensus guidelines for diagnosis and treatment of Shwachman-Diamond syndrome. *Annals of the New York Academy of Sciences* **1242**, 40-55 (2011). https://doi.org:10.1111/j.1749-6632.2011.06349.x
252 Pearson, H. A. *et al.* A new syndrome of refractory sideroblastic anemia with vacuolization of marrow precursors and exocrine pancreatic dysfunction. *The Journal of Pediatrics* **95**, 976-984 (1979). https://doi.org:https://doi.org/10.1016/S0022-3476(79)80286-3
253 Clarke C. & Hadfield G. Congenital pancreatic disease with infantilism. *Quarterly Journal of Medicine* (1924).
254 Tarlow, M. J., Hadorn, B., Arthurton, M. W. & Lloyd, J. K. Intestinal enterokinase deficiency. A newly-recognized disorder of protein digestion. *Archives of disease in childhood* **45**, 651-655 (1970).
255 Attanasio, A., Castellaro, I., Rizzotti, P., Marini, M. & Vidal, M. «Macroamylasemia or pancreatitis? A diagnostic problem». *Minerva medica* **78**, 1347-1351 (1987).
256 Johanson, A. & Blizzard, R. A syndrome of congenital aplasia of the alae nasi, deafness, hypothyroidism, dwarfism, absent permanent teeth, and malabsorption. *The Journal of Pediatrics* **79**, 982-987 (1971). https://doi.org:https://doi.org/10.1016/S0022-3476(71)80194-4
257 Whitcomb, D. C. *et al.* Multicenter approach to recurrent acute and chronic pancreatitis in the United States: the North American Pancreatitis Study 2 (NAPS2). *Pancreatology : official journal of the International Association of Pancreatology (IAP) ... [et al.]* **8**, 520-531 (2008). https://doi.org:10.1159/000152001
258 Conwell, D. L. *et al.* Validation of Demographics, Etiology, and Risk Factors for Chronic Pancreatitis in the USA: A Report of the North American Pancreas Study (NAPS) Group. *Digestive diseases and sciences* **62**, 2133-2140 (2017). https://doi.org:10.1007/s10620-017-4621-z
259 Whitcomb, D. C. *et al.* Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. *Nature genetics* **44**, 1349-1354 (2012). https://doi.org:10.1038/ng.2466
260 Sakaguchi, T. *et al.* Cloning of the Human Claudin-2 5'-Flanking Region Revealed a TATA-less Promoter with Conserved Binding Sites in Mouse and Human for Caudal-related Homeodomain Proteins and Hepatocyte Nuclear Factor-1α. *Journal of Biological Chemistry* **277**, 21361-21370 (2002). https://doi.org:10.1074/jbc.M110261200

261 Schmidt, A. W. *et al.* Colocalization analysis of pancreas eQTLs with risk loci from alcoholic and novel non-alcoholic chronic pancreatitis GWAS suggests potential disease causing mechanisms. *Pancreatology : official journal of the International Association of Pancreatology (IAP) ... [et al.]* **22**, 449-456 (2022). https://doi.org:10.1016/j.pan.2022.03.007

262 Rosendahl, J. *et al.* Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut* **67**, 1855-1863 (2018). https://doi.org:10.1136/gutjnl-2017-314454

263 Lombardo, D. Bile salt-dependent lipase: its pathophysiological implications. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1533**, 1-28 (2001). https://doi.org:https://doi.org/10.1016/S1388-1981(01)00130-5

264 Nilsson, J. *et al.* cDNA cloning of human-milk bile-salt-stimulated lipase and evidence for its identity to pancreatic carboxylic ester hydrolase. *European Journal of Biochemistry* **192**, 543-550 (1990). https://doi.org:10.1111/j.1432-1033.1990.tb19259.x

265 Lidberg, U. *et al.* Genomic organization, sequence analysis, and chromosomal localization of the human carboxyl ester lipase (CEL) gene and a CEL-like (CELL) gene. *Genomics* **13**, 630-640 (1992). https://doi.org:https://doi.org/10.1016/0888-7543(92)90134-E

266 Madeyski, K., Lidberg, U., Bjursell, G. & Nilsson, J. Structure and organization of the human carboxyl ester lipase locus. **9**, 334-338 (1998). https://doi.org:https://doi-org.nottingham.idm.oclc.org/10.1007/s003359900762

267 Ræder, H. *et al.* Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nature Genetics* **38**, 54 (2005). https://doi.org:10.1038/ng1708
https://www.nature.com/articles/ng1708#supplementary-information

268 Ræder, H. *et al.* Carboxyl-ester lipase maturity-onset diabetes of the young is associated with development of pancreatic cysts and upregulated MAPK signaling in secretin-stimulated duodenal fluid. *Diabetes* **63**, 259-269 (2014). https://doi.org:10.2337/db13-1012

269 Fjeld, K. *et al.* A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. *Nature genetics* **47**, 518-522 (2015). https://doi.org:10.1038/ng.3249

270 Day, C. P. *et al.* Investigation of the role of polymorphisms at the alcohol and aldehyde dehydrogenase loci in genetic predisposition to alcohol-related end-organ damage. *Hepatology* **14**, 798-801 (1991). https://doi.org:https://doi.org/10.1002/hep.1840140509

271 Kimura, S. *et al.* Alcohol and aldehyde dehydrogenase polymorphisms in Japanese patients with alcohol-induced chronic pancreatitis. *Digestive diseases and sciences* **45**, 2013-2017 (2000). https://doi.org:10.1023/a:1005650902961

272 Chen, C. C. *et al.* Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism.

American journal of human genetics **65**, 795-807 (1999). https://doi.org:10.1086/302540

273 Goedde, H. W. *et al.* Population genetic studies on aldehyde dehydrogenase isozyme deficiency and alcohol sensitivity. *American journal of human genetics* **35**, 769-772 (1983).

274 Standop, J. *et al.* The pattern of xenobiotic-metabolizing enzymes in the human pancreas. *J Toxicol Environ Health A* **65**, 1379-1400 (2002). https://doi.org:10.1080/00984100290071603

275 van den Berg, F. F. *et al.* Meta-analysis and field synopsis of genetic variants associated with the risk and severity of acute pancreatitis. *BJS Open* **4**, 3-15 (2020). https://doi.org:10.1002/bjs5.50231

276 Liu, J. & Zhang, H.-x. A Comprehensive Study Indicates PRSS1 Gene Is Significantly Associated with Pancreatitis. *International Journal of Medical Sciences* **10**, 981-987 (2013). https://doi.org:10.7150/ijms.6164

277 Liu, J., Wang, Y. & Yu, Y. Meta-analysis reveals an association between acute pancreatitis and the risk of pancreatic cancer. *World J Clin Cases* **8**, 4416-4430 (2020). https://doi.org:10.12998/wjcc.v8.i19.4416

278 Zilio, M. B., Eyff, T. F., Azeredo-Da-Silva, A. L. F., Bersch, V. P. & Osvaldt, A. B. A systematic review and meta-analysis of the aetiology of acute pancreatitis. *HPB (Oxford)* **21**, 259-267 (2019). https://doi.org:10.1016/j.hpb.2018.08.003

279 Bhat, K. *et al.* Advances in biomarker research for pancreatic cancer. *Curr Pharm Des* **18**, 2439-2451 (2012). https://doi.org:10.2174/13816128112092439

280 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7-30 (2016). https://doi.org:10.3322/caac.21332

281 Klein, A. P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nature Communications* **9**, 556 (2018). https://doi.org:10.1038/s41467-018-02942-5

282 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018). https://doi.org:10.3322/caac.21492

283 Malvezzi, M., Bertuccio, P., Levi, F., La Vecchia, C. & Negri, E. European cancer mortality predictions for the year 2014. *Ann Oncol* **25**, 1650-1656 (2014). https://doi.org:10.1093/annonc/mdu138

284 Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* **74**, 2913-2921 (2014). https://doi.org:10.1158/0008-5472.Can-14-0155

285 Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114-1117 (2010). https://doi.org:10.1038/nature09515

286 Hruban, R. H. *et al.* An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms. *Am J Surg Pathol* **28**, 977-987 (2004). https://doi.org:10.1097/01.pas.0000126675.59108.80

287 Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60**, 277-300 (2010). https://doi.org:10.3322/caac.20073
288 Shi, C., Hruban, R. H. & Klein, A. P. Familial pancreatic cancer. *Arch Pathol Lab Med* **133**, 365-374 (2009). https://doi.org:10.5858/133.3.365
289 Klein, A. P. *et al.* Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer Res* **64**, 2634-2638 (2004). https://doi.org:10.1158/0008-5472.can-03-3823
290 Ottenhof, N. A., de Wilde, R. F., Maitra, A., Hruban, R. H. & Offerhaus, G. J. A. Molecular Characteristics of Pancreatic Ductal Adenocarcinoma. *Pathology Research International* **2011**, 620601 (2011). https://doi.org:10.4061/2011/620601
291 Ghiorzo, P. Genetic predisposition to pancreatic cancer. *World J Gastroenterol* **20**, 10778-10789 (2014). https://doi.org:10.3748/wjg.v20.i31.10778
292 Iwatate, Y. *et al.* Prognostic significance of p16 protein in pancreatic ductal adenocarcinoma. *Mol Clin Oncol* **13**, 83-91 (2020). https://doi.org:10.3892/mco.2020.2047
293 Hong, S. M., Park, J. Y., Hruban, R. H. & Goggins, M. Molecular signatures of pancreatic cancer. *Arch Pathol Lab Med* **135**, 716-727 (2011). https://doi.org:10.5858/2010-0566-ra.1
294 Collins, M. A. *et al.* Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. *J Clin Invest* **122**, 639-653 (2012). https://doi.org:10.1172/jci59227
295 Luo, J. KRAS mutation in pancreatic cancer. *Semin Oncol* **48**, 10-18 (2021). https://doi.org:10.1053/j.seminoncol.2021.02.003
296 Hartman, D. J., Davison, J. M., Foxwell, T. J., Nikiforova, M. N. & Chiosea, S. I. Mutant allele-specific imbalance modulates prognostic impact of KRAS mutations in colorectal adenocarcinoma and is associated with worse overall survival. *International Journal of Cancer* **131**, 1810-1817 (2012). https://doi.org:https://doi.org/10.1002/ijc.27461
297 Chiosea, S. I., Sherer, C. K., Jelic, T. & Dacic, S. KRAS mutant allele-specific imbalance in lung adenocarcinoma. *Modern Pathology* **24**, 1571-1577 (2011). https://doi.org:10.1038/modpathol.2011.109
298 Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405 (2012). https://doi.org:10.1038/nature11547
299 Krasinskas, A. M., Moser, A. J., Saka, B., Adsay, N. V. & Chiosea, S. I. KRAS mutant allele-specific imbalance is associated with worse prognosis in pancreatic cancer and progression to undifferentiated carcinoma of the pancreas. *Mod Pathol* **26**, 1346-1354 (2013). https://doi.org:10.1038/modpathol.2013.71
300 Takahashi, M., Hori, M., Mutoh, M., Wakabayashi, K. & Nakagama, H. Experimental animal models of pancreatic carcinogenesis for prevention studies and their relevance to human disease. *Cancers (Basel)* **3**, 582-602 (2011). https://doi.org:10.3390/cancers3010582

301 Hu, C. *et al.* Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. *Jama* **319**, 2401-2409 (2018). https://doi.org:10.1001/jama.2018.6228
302 Gnoni, A. *et al.* Carcinogenesis of pancreatic adenocarcinoma: precursor lesions. *Int J Mol Sci* **14**, 19731-19762 (2013). https://doi.org:10.3390/ijms141019731
303 Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e113 (2017). https://doi.org:10.1016/j.ccell.2017.07.007
304 AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov* **7**, 818-831 (2017). https://doi.org:10.1158/2159-8290.Cd-17-0151
305 Blackford, A. *et al.* SMAD4 gene mutations are associated with poor prognosis in pancreatic cancer. *Clin Cancer Res* **15**, 4674-4679 (2009). https://doi.org:10.1158/1078-0432.Ccr-09-0227
306 Balachandran, V. P. *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512-516 (2017). https://doi.org:10.1038/nature24462
307 Ryan, D. P., Hong, T. S. & Bardeesy, N. Pancreatic adenocarcinoma. *N Engl J Med* **371**, 1039-1049 (2014). https://doi.org:10.1056/NEJMra1404198
308 Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D. & Pavletich, N. P. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395**, 237-243 (1998). https://doi.org:10.1038/26155
309 Schlitter, A. M. *et al.* Molecular, morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes. *Sci Rep* **7**, 41064 (2017). https://doi.org:10.1038/srep41064
310 Ohtsubo, K. *et al.* Abnormalities of tumor suppressor gene p16 in pancreatic carcinoma: immunohistochemical and genetic findings compared with clinicopathological parameters. *J Gastroenterol* **38**, 663-671 (2003). https://doi.org:10.1007/s00535-003-1119-6
311 Caldas, C. *et al.* Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat Genet* **8**, 27-32 (1994). https://doi.org:10.1038/ng0994-27
312 Schutte, M. *et al.* Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Res* **57**, 3126-3130 (1997).
313 Ueki, T. *et al.* Hypermethylation of multiple genes in pancreatic adenocarcinoma. *Cancer Res* **60**, 1835-1839 (2000).
314 Lu, Y. *et al.* Identification of Recessively Inherited Genetic Variants Potentially Linked to Pancreatic Cancer Risk. *Frontiers in Oncology* **11** (2021). https://doi.org:10.3389/fonc.2021.771312
315 Zhang, M. *et al.* Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget* **7**, 66328-66343 (2016). https://doi.org:10.18632/oncotarget.11041

316 Streicher, S. A. *et al.* A pooled genome-wide association study identifies pancreatic cancer susceptibility loci on chromosome 19p12 and 19p13.3 in the full-Jewish population. *Hum Genet* **140**, 309-319 (2021). https://doi.org:10.1007/s00439-020-02205-8

317 Falchi, M. *et al.* Low copy number of the salivary amylase gene predisposes to obesity. *Nature Genetics* **46**, 492 (2014). https://doi.org:10.1038/ng.2939
https://www.nature.com/articles/ng.2939#supplementary-information

318 Choi, Y.-J. *et al.* Association between salivary amylase (AMY1) gene copy numbers and insulin resistance in asymptomatic Korean men. *Diabetic Medicine* **32**, 1588-1595 (2015). https://doi.org:10.1111/dme.12808

319 Mejía-Benítez, M. A. *et al.* Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children. *Diabetologia* **58**, 290-294 (2015). https://doi.org:10.1007/s00125-014-3441-3

320 Elder, P. J. D., Ramsden, D. B., Burnett, D., Weickert, M. O. & Barber, T. M. Human amylase gene copy number variation as a determinant of metabolic state. *Expert Rev Endocrinol Metab* **13**, 193-205 (2018). https://doi.org:10.1080/17446651.2018.1499466

321 Zhang, X. *et al.* 8p23 beta-defensin copy number determination by single-locus pseudogene-based paralog ratio tests risk bias due to low-frequency sequence variations. *BMC Genomics* **15**, 64 (2014). https://doi.org:10.1186/1471-2164-15-64

322 Butterworth, P. J., Warren, F. J. & Ellis, P. R. Human α-amylase and starch digestion: An interesting marriage. *Starch - Stärke* **63**, 395-405 (2011). https://doi.org:10.1002/star.201000150

323 Fernández, C. I. & Wiley, A. S. Rethinking the starch digestion hypothesis for AMY1 copy number variation in humans. *American Journal of Physical Anthropology* **163**, 645-657 (2017). https://doi.org:10.1002/ajpa.23237

324 Rohleder, N., Nater, U. M., Wolf, J. M., Ehlert, U. & Kirschbaum, C. Psychosocial Stress-Induced Activation of Salivary Alpha-Amylase: An Indicator of Sympathetic Activity? *Annals of the New York Academy of Sciences* **1032**, 258-263 (2004). https://doi.org:10.1196/annals.1314.033

325 Bompard-Gilles, C., Rousseau, P., Rougé, P. & Payan, F. Substrate mimicry in the active center of a mammalian α amylase: structural analysis of an enzyme–inhibitor complex. *Structure* **4**, 1441-1452 (1996). https://doi.org:https://doi.org/10.1016/S0969-2126(96)00151-7

326 Svensson, B., Fukuda, K., Nielsen, P. K. & Bønsager, B. C. Proteinaceous α-amylase inhibitors. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1696**, 145-156 (2004). https://doi.org:https://doi.org/10.1016/j.bbapap.2003.07.004

327 Tsujita, T. Persimmon-Tannin, an α-Amylase Inhibitor, Retards Carbohydrate Absorption in Rats. *Journal of Nutritional Science and Vitaminology* **62**, 192-197 (2016). https://doi.org:10.3177/jnsv.62.192

328 Matsushita, H., Takenaka, M. & Ogawa, H. Porcine pancreatic alpha-amylase shows binding activity toward N-linked oligosaccharides of glycoproteins. *The Journal of biological chemistry* **277**, 4680-4686 (2002). https://doi.org:10.1074/jbc.M105877200

329 Little, T. J., Gupta, N., Case, R. M., Thompson, D. G. & McLaughlin, J. T. Sweetness and bitterness taste of meals per se does not mediate gastric emptying in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **297**, R632-R639 (2009). https://doi.org:10.1152/ajpregu.00090.2009

330 Whitcomb, D. C. & Lowe, M. E. Human Pancreatic Digestive Enzymes. *Digestive diseases and sciences* **52**, 1-17 (2007). https://doi.org:10.1007/s10620-006-9589-z

331 Ferey-Roux, G. *et al.* The human pancreatic α-amylase isoforms: isolation, structural studies and kinetics of inhibition by acarbose. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1388**, 10-20 (1998). https://doi.org:https://doi.org/10.1016/S0167-4838(98)00147-2

332 Brannon, P. M. Adaptation of the Exocrine Pancreas to Diet. *Annual Review of Nutrition* **10**, 85-105 (1990). https://doi.org:10.1146/annurev.nu.10.070190.000505

333 Giorgi, D., Bernard, J. P., Lapointe, R. & Dagorn, J. C. Regulation of amylase messenger RNA concentration in rat pancreas by food content. *EMBO J* **3**, 1521-1524 (1984).

334 Snook, J. Dietary regulation of pancreatic enzymes in the rat with emphasis on carbohydrate. *American Journal of Physiology-Legacy Content* **221**, 1383-1387 (1971). https://doi.org:10.1152/ajplegacy.1971.221.5.1383

335 Carpenter, D. *et al.* Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. *Human molecular genetics* **24**, 3472-3480 (2015). https://doi.org:10.1093/hmg/ddv098

336 Groot, P. C. *et al.* The human α-amylase multigene family consists of haplotypes with variable numbers of genes. *Genomics* **5**, 29-42 (1989). https://doi.org:https://doi.org/10.1016/0888-7543(89)90083-9

337 Groot, P. C., Mager, W. H. & Frants, R. R. Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family. *Genomics* **10**, 779-785 (1991).

338 Hamid, A. K., Andersson-Assarsson, J., Ericson, U. & Sonestedt, E. Interaction Effect Between Copy Number Variation in Salivary Amylase Locus (AMY1) and Starch Intake on Glucose Homeostasis in the Malmö Diet and Cancer Cohort. *Front Nutr* **7**, 598850-598850 (2021). https://doi.org:10.3389/fnut.2020.598850

339 Cantsilieris, S. & White, S. J. Correlating Multiallelic Copy Number Polymorphisms with Disease Susceptibility. *Human Mutation* **34**, 1-13 (2013). https://doi.org:10.1002/humu.22172
340 Barnes, C. *et al.* A robust statistical method for case-control association testing with copy number variation. *Nature Genetics* **40**, 1245 (2008). https://doi.org:10.1038/ng.206
https://www.nature.com/articles/ng.206#supplementary-information
341 Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature Genetics* **37**, 1243-1246 (2005). https://doi.org:10.1038/ng1653
342 Zanda, M. *et al.* A genome-wide assessment of the role of untagged copy number variants in type 1 diabetes. *PLoS genetics* **10**, e1004367-e1004367 (2014). https://doi.org:10.1371/journal.pgen.1004367
343 Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* **39**, 1256 (2007). https://doi.org:10.1038/ng2123
https://www.nature.com/articles/ng2123#supplementary-information
344 Yang, Z.-M. *et al.* The roles of AMY1 copies and protein expression in human salivary α-amylase activity. *Physiology & Behavior* **138**, 173-178 (2015). https://doi.org:https://doi.org/10.1016/j.physbeh.2014.10.037
345 Mandel, A. L., Peyrot des Gachons, C., Plank, K. L., Alarcon, S. & Breslin, P. A. S. Individual differences in AMY1 gene copy number, salivary α-amylase levels, and the perception of oral starch. *PloS one* **5**, e13352-e13352 (2010). https://doi.org:10.1371/journal.pone.0013352
346 Usher, C. L. *et al.* Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nature Genetics* **47**, 921 (2015). https://doi.org:10.1038/ng.3340
https://www.nature.com/articles/ng.3340#supplementary-information
347 Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genetics* **36**, 949-951 (2004). https://doi.org:10.1038/ng1416
348 Gibbs, R. A. *et al.* The International HapMap Project. *Nature* **426**, 789-796 (2003). https://doi.org:10.1038/nature02168
349 Hasegawa, T. *et al.* Impact of salivary and pancreatic amylase gene copy numbers on diabetes, obesity, and functional profiles of microbiome in Northern Japanese population. *Sci Rep* **12**, 7628 (2022). https://doi.org:10.1038/s41598-022-11730-7
350 Falchi, M. *et al.* Low copy number of the salivary amylase gene predisposes to obesity. *Nature genetics* **46**, 492-497 (2014). https://doi.org:10.1038/ng.2939
351 Santos, J. L. *et al.* Copy Number Polymorphism of the Salivary Amylase Gene: Implications in Human Nutrition Research. *Lifestyle Genomics* **5**, 117-131 (2012). https://doi.org:10.1159/000339951

352 León-Mimila, P. *et al.* Low Salivary Amylase Gene (AMY1) Copy Number Is Associated with Obesity and Gut Prevotella Abundance in Mexican Children and Adults. *Nutrients* **10**, 1607 (2018). https://doi.org:10.3390/nu10111607
353 Marcovecchio, M. L. *et al.* Low AMY1 Gene Copy Number Is Associated with Increased Body Mass Index in Prepubertal Boys. *PLoS One* **11**, e0154961 (2016). https://doi.org:10.1371/journal.pone.0154961
354 Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science (New York, N.Y.)* **330**, 641-646 (2010). https://doi.org:10.1126/science.1197005
355 Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science (New York, N.Y.)* **334**, 105-108 (2011). https://doi.org:10.1126/science.1208344
356 Kovatcheva-Datchary, P. *et al.* Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell Metabolism* **22**, 971-982 (2015). https://doi.org:https://doi.org/10.1016/j.cmet.2015.10.001
357 Choi, Y. J. *et al.* Association between salivary amylase (AMY1) gene copy numbers and insulin resistance in asymptomatic Korean men. *Diabetic medicine : a journal of the British Diabetic Association* **32**, 1588-1595 (2015). https://doi.org:10.1111/dme.12808
358 Mandel, A. L. & Breslin, P. A. S. High endogenous salivary amylase activity is associated with improved glycemic homeostasis following starch ingestion in adults. *J Nutr* **142**, 853-858 (2012). https://doi.org:10.3945/jn.111.156984
359 Zhao, Y., Zhang, J., Zhang, J., Wu, J. & Chen, Y. Metabolic syndrome and diabetes are associated with low serum amylase in a Chinese asymptomatic population. *Scandinavian Journal of Clinical and Laboratory Investigation* **74**, 235-239 (2014). https://doi.org:10.3109/00365513.2013.878469
360 Nakajima, K. *et al.* Low serum amylase in association with metabolic syndrome and diabetes: A community-based study. *Cardiovasc Diabetol* **10**, 34-34 (2011). https://doi.org:10.1186/1475-2840-10-34
361 Atkinson, F. S., Hancock, D., Petocz, P. & Brand-Miller, J. C. The physiologic and phenotypic significance of variation in human amylase gene copy number. *The American journal of clinical nutrition* **108**, 737-748 (2018). https://doi.org:10.1093/ajcn/nqy164
362 Hancock, A. M. *et al.* Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci U S A* **107 Suppl 2**, 8924-8930 (2010). https://doi.org:10.1073/pnas.0914625107
363 Inchley, C. E. *et al.* Selective sweep on human amylase genes postdates the split with Neanderthals. *Scientific reports* **6**, 37198-37198 (2016). https://doi.org:10.1038/srep37198

364 Marcadier, J. L. *et al.* Congenital sucrase-isomaltase deficiency: identification of a common Inuit founder mutation. *CMAJ* **187**, 102-107 (2015). https://doi.org:10.1503/cmaj.140657

365 Pedersen, C. T. *et al.* The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. *Genetics* **205**, 787-801 (2017). https://doi.org:10.1534/genetics.116.193821

366 Noto, Y., Sato, T., Kudo, M., Kurata, K. & Hirota, K. The Relationship Between Salivary Biomarkers and State-Trait Anxiety Inventory Score Under Mental Arithmetic Stress: A Pilot Study. *Anesthesia & Analgesia* **101**, 1873-1876 (2005). https://doi.org:10.1213/01.Ane.0000184196.60838.8d

367 Guglielminotti, J. *et al.* Assessment of salivary amylase as a stress biomarker in pregnant patients. *International journal of obstetric anesthesia* **21**, 35-39 (2012). https://doi.org:10.1016/j.ijoa.2011.09.008

368 Clavien, P.-A., Burgan, S. & Moossa, A. R. Serum enzymes and other laboratory tests in acute pancreatitis. *BJS* **76**, 1234-1243 (1989). https://doi.org:10.1002/bjs.1800761205

369 Pieper-Bigelow, C., Strocchi, A. & Levitt, M. D. Where does serum amylase come from and where does it go? *Gastroenterology clinics of North America* **19**, 793-810 (1990).

370 Vissers, R. J., Abu-Laban, R. B. & McHugh, D. F. Amylase and lipase in the emergency department evaluation of acute pancreatitis11Clinical Laboratory in Emergency Medicine is coordinated by Jonathan S. Olshaker, MD, of the University of Maryland Medical Center, Baltimore, Maryland. *The Journal of Emergency Medicine* **17**, 1027-1037 (1999). https://doi.org:https://doi.org/10.1016/S0736-4679(99)00136-5

371 Nadhem, O. & Salh, O. Acute Pancreatitis: An Atypical Presentation. *Case Rep Gastroenterol* **11**, 359-363 (2017). https://doi.org:10.1159/000475920

372 Galassi, E. *et al.* A 5-Year Experience of Benign Pancreatic Hyperenzymemia. **43** (2014). https://doi.org:10.1097/MPA.0000000000000138

373 Amodio, A. *et al.* Prospective evaluation of subjects with chronic asymptomatic pancreatic hyperenzymemia. *The American journal of gastroenterology* **107**, 1089-1095 (2012). https://doi.org:10.1038/ajg.2012.125

374 Layer, P., Zinsmeister, A. R. & DiMagno, E. P. Effects of decreasing intraluminal amylase activity on starch digestion and postprandial gastrointestinal function in humans. *Gastroenterology* **91**, 41-48 (1986). https://doi.org:https://doi.org/10.1016/0016-5085(86)90436-1

375 Williams, J. A. Amylase. *Pancreapedia: The Exocrine Pancreas Knowledge Base* (2019).

376 Mehta, D. I., Wang, H. H., Akins, R. E., Wang, L. & Proujansky, R. Isolated pancreatic amylase deficiency: Probable error in maturation. *The Journal of Pediatrics* **136**, 844-846 (2000). https://doi.org:https://doi.org/10.1016/S0022-3476(00)96936-1

377 Sjölund, K. *et al.* Selective deficiency of pancreatic amylase. *Gut* **32**, 546-548 (1991). https://doi.org:10.1136/gut.32.5.546

378 Feinberg, A. P. Genome-scale approaches to the epigenetics of common human disease. *Virchows Arch* **456**, 13-21 (2010). https://doi.org:10.1007/s00428-009-0847-2

379 McKay, J. A. & Mathers, J. C. Diet induced epigenetic changes and their implications for health. *Acta Physiologica* **202**, 103-118 (2011). https://doi.org:10.1111/j.1748-1716.2011.02278.x

380 Christensen, B. C. *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *PLOS Genetics* **5**, e1000602 (2009). https://doi.org:10.1371/journal.pgen.1000602

381 Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci* **13**, 3245-3276 (2012). https://doi.org:10.3390/ijms13033245

382 Os, P., Eijk Van, G. C. & Schouten, J. P. in *PCR Mutation Detection Protocols* (eds Bimal D. M. Theophilus & Ralph Rapley) 97-126 (Humana Press, 2011).

383 Groth, M. *et al.* High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. *Human Mutation* **29**, 1247-1254 (2008). https://doi.org:https://doi.org/10.1002/humu.20751

384 Field, S. F. *et al.* Experimental aspects of copy number variant assays at CCL3L1. *Nat Med* **15**, 1115-1117 (2009). https://doi.org:10.1038/nm1009-1115

385 Aldhous, M. C. *et al.* Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Human Molecular Genetics* **19**, 4930-4938 (2010). https://doi.org:10.1093/hmg/ddq411

386 Hollox, E. J. in *Genotyping: Methods and Protocols* (eds Stefan J. White & Stuart Cantsilieris) 127-146 (Springer New York, 2017).

387 Armour, J. A. *et al.* Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res* **35**, e19 (2007). https://doi.org:10.1093/nar/gkl1089

388 Algady, W. *et al.* Genotyping complex structural variation at the malaria-associated human glycophorin locus using a PCR-based strategy. *Annals of Human Genetics* **85**, 7-17 (2021). https://doi.org:https://doi.org/10.1111/ahg.12405

389 Cantsilieris, S., Baird, P. N. & White, S. J. Molecular methods for genotyping complex copy number polymorphisms. *Genomics* **101**, 86-93 (2013). https://doi.org:https://doi.org/10.1016/j.ygeno.2012.10.004

390 Carpenter, D., Mitchell, L. M. & Armour, J. A. L. Copy number variation of human AMY1 is a minor contributor to variation in salivary amylase expression and activity. *Hum Genomics* **11**, 2-2 (2017). https://doi.org:10.1186/s40246-017-0097-3
391 Carpenter, D., Walker, S., Prescott, N., Schalkwijk, J. & Armour, J. A. L. Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders. *BMC Genomics* **12**, 418 (2011). https://doi.org:10.1186/1471-2164-12-418
392 Shwan, N. A. A., Louzada, S., Yang, F. & Armour, J. A. L. Recurrent Rearrangements of Human Amylase Genes Create Multiple Independent CNV Series. *Human Mutation* **38**, 532-539 (2017). https://doi.org:10.1002/humu.23182
393 Khan, F. F. *et al.* Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. *BMC Genomics* **14**, 719 (2013). https://doi.org:10.1186/1471-2164-14-719
394 Ouyang, G. *et al.* The global, regional, and national burden of pancreatitis in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *BMC Med* **18**, 388 (2020). https://doi.org:10.1186/s12916-020-01859-5
395 Qu, Y. J. *et al.* Association of copy numbers of survival motor neuron gene 2 and neuronal apoptosis inhibitory protein gene with the natural history in a Chinese spinal muscular atrophy cohort. *J Child Neurol* **30**, 429-436 (2015). https://doi.org:10.1177/0883073814553271
396 Usher, C. L. *et al.* Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nature genetics* **47**, 921-925 (2015). https://doi.org:10.1038/ng.3340
397 Andersen, M. K. & Sandholt, C. H. Recent Progress in the Understanding of Obesity: Contributions of Genome-Wide Association Studies. *Current Obesity Reports* **4**, 401-410 (2015). https://doi.org:10.1007/s13679-015-0173-8
398 Nayema, Z. *et al.* Genetic factors associated with serum amylase in a Japanese population: combined analysis of copy-number and single-nucleotide variants. *J Hum Genet* **68**, 313-319 (2023). https://doi.org:10.1038/s10038-022-01111-3
399 Yong, R. Y. *et al.* Complex Copy Number Variation of AMY1 does not Associate with Obesity in two East Asian Cohorts. *Hum Mutat* **37**, 669-678 (2016). https://doi.org:10.1002/humu.22996
400 Viljakainen, H. *et al.* Low Copy Number of the AMY1 Locus Is Associated with Early-Onset Female Obesity in Finland. *PloS one* **10**, e0131883-e0131883 (2015). https://doi.org:10.1371/journal.pone.0131883
401 Elks, C. E. *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)* **3**, 29-29 (2012). https://doi.org:10.3389/fendo.2012.00029
402 Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.

Science (New York, N.Y.) **316**, 889-894 (2007). https://doi.org:10.1126/science.1141634

403 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* **42**, 937 (2010). https://doi.org:10.1038/ng.686
https://www.nature.com/articles/ng.686#supplementary-information

404 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197 (2015). https://doi.org:10.1038/nature14177
https://www.nature.com/articles/nature14177#supplementary-information

405 Bonnefond, A. *et al.* Relationship between salivary/pancreatic amylase and body mass index: a systems biology approach. *BMC Med* **15**, 37-37 (2017). https://doi.org:10.1186/s12916-017-0784-x

406 Al-Akl, N., Thompson, R. I. & Arredouani, A. High plasma salivary α-amylase, but not high AMY1 copy number, associated with low obesity rate in Qatari adults: cross-sectional study. *Scientific Reports* **10**, 17918 (2020). https://doi.org:10.1038/s41598-020-74864-6

407 Higuchi, R., Iwane, T., Iida, A. & Nakajima, K. Copy Number Variation of the Salivary Amylase Gene and Glucose Metabolism in Healthy Young Japanese Women. *Journal of Clinical Medicine Research* **12**, 184-189 (2020). https://doi.org:10.14740/jocmr4082

408 Venkatapoorna, C. M. K. *et al.* Association of Salivary Amylase (AMY1) Gene Copy Number with Obesity in Alabama Elementary School Children. *Nutrients* **11**, 1379 (2019). https://doi.org:10.3390/nu11061379

409 Kang, J. U. *et al.* Genetic alterations in primary gastric carcinomas correlated with clinicopathological variables by array comparative genomic hybridization. *J Korean Med Sci* **21**, 656-665 (2006). https://doi.org:10.3346/jkms.2006.21.4.656

410 Byman, E. *et al.* Alpha-amylase 1A copy number variants and the association with memory performance and Alzheimer's dementia. *Alzheimers Res Ther* **12**, 158 (2020). https://doi.org:10.1186/s13195-020-00726-y

411 Bonnefond, A. *et al.* Relationship between salivary/pancreatic amylase and body mass index: a systems biology approach. *BMC Med* **15**, 37 (2017). https://doi.org:10.1186/s12916-017-0784-x

412 Iskow, R. C., Gokcumen, O. & Lee, C. Exploring the role of copy number variants in human adaptation. *Trends Genet* **28**, 245-257 (2012). https://doi.org:10.1016/j.tig.2012.03.002

413 Stankiewicz, P. & Lupski, J. R. Structural Variation in the Human Genome and its Role in Disease. *Annual Review of Medicine* **61**, 437-455 (2010). https://doi.org:10.1146/annurev-med-100708-204735

414 Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Human Mutation* **21**, 12-27 (2003). https://doi.org:https://doi.org/10.1002/humu.10147

415 Hegyi, E. & Sahin-Tóth, M. Genetic Risk in Chronic Pancreatitis: The Trypsin-Dependent Pathway. *Digestive diseases and sciences* **62**, 1692-1701 (2017). https://doi.org:10.1007/s10620-017-4601-3
416 Kereszturi, É., Király, O. & Sahin-Tóth, M. Minigene analysis of intronic variants in common SPINK1 haplotypes associated with chronic pancreatitis. *Gut* **58**, 545-549 (2009). https://doi.org:10.1136/gut.2008.164947
417 Király, O., Wartmann, T. & Sahin-Tóth, M. Missense mutations in pancreatic secretory trypsin inhibitor (SPINK1) cause intracellular retention and degradation. *Gut* **56**, 1433-1438 (2007). https://doi.org:10.1136/gut.2006.115725
418 Zou, W.-B. *et al.* SPINK1,PRSS1,CTRC, andCFTRGenotypes Influence Disease Onset and Clinical Outcomes in Chronic Pancreatitis. *Clinical and Translational Gastroenterology* **9** (2018).
419 Pfützer, R. H. *et al.* SPINK1/PSTI polymorphisms act as disease modifiers in familial and idiopathic chronic pancreatitis. *Gastroenterology* **119**, 615-623 (2000). https://doi.org:https://doi.org/10.1053/gast.2000.18017
420 Schneider, A. *et al.* SPINK1/PSTI mutations are associated with tropical pancreatitis and type II diabetes mellitus in Bangladesh. *Gastroenterology* **123**, 1026-1030 (2002). https://doi.org:https://doi.org/10.1053/gast.2002.36059
421 Truninger, K., Ammann, R. W., Blum, H. E. & Witt, H. Genetic aspects of chronic pancreatitis: insights into aetiopathogenesis and clinical implications. *Swiss Med Wkly* **131**, 565-574 (2001).
422 Lempinen, M. *et al.* Mutations N34S and P55S of the SPINK1 gene in patients with chronic pancreatitis or pancreatic cancer and in healthy subjects: A report from Finland. *Scandinavian Journal of Gastroenterology* **40**, 225-230 (2005). https://doi.org:10.1080/00365520510011560
423 Jalaly, N. Y. *et al.* An Evaluation of Factors Associated With Pathogenic PRSS1, SPINK1, CTFR, and/or CTRC Genetic Variants in Patients With Idiopathic Pancreatitis. *The American journal of gastroenterology* **112**, 1320-1329 (2017). https://doi.org:10.1038/ajg.2017.106
424 Sahin-Tóth, M. Genetic risk in chronic pancreatitis: the misfolding-dependent pathway. *Curr Opin Gastroenterol* **33**, 390-395 (2017). https://doi.org:10.1097/mog.0000000000000380
425 Sun, Z., Kolossváry, I., Kozakov, D., Sahin-Tóth, M. & Vajda, S. The N34S mutation of SPINK1 may impact the kinetics of trypsinogen activation to cause early trypsin release in the pancreas. *bioRxiv*, 2020.2008.2021.262162 (2020). https://doi.org:10.1101/2020.08.21.262162
426 Teich, N. & Mössner, J. Genetic aspects of chronic pancreatitis. *Med Sci Monit* **10**, Ra325-328 (2004).
427 Lerch, M. M. *et al.* Advances in the etiology of chronic pancreatitis. *Dig Dis* **28**, 324-329 (2010). https://doi.org:10.1159/000319408

428 Laskowski Jr, M. & Kato, I. Protein inhibitors of proteinases. *Annual review of biochemistry* **49**, 593-626 (1980).
429 Rinderknecht, H. *Pancreatic secretory enzymes.* 2 edn, (The pancreas: biology, pathobiology and disease., 1993).
430 Perri, F. *et al.* Mutation analysis of the cystic fibrosis transmembrane conductance regulator (CFTR) gene, the cationic trypsinogen (PRSS1) gene, and the serine protease inhibitor, Kazal type 1 (SPINK1) gene in patients with alcoholic chronic pancreatitis. *Eur J Hum Genet* **11**, 687-692 (2003). https://doi.org:10.1038/sj.ejhg.5201035
431 Bernardino, A. L. *et al.* CFTR, PRSS1 and SPINK1 mutations in the development of pancreatitis in Brazilian patients. *Jop* **4**, 169-177 (2003).
432 Aoun, E. *et al.* Pathways to injury in chronic pancreatitis: decoding the role of the high-risk SPINK1 N34S haplotype using meta-analysis. *PLoS One* **3**, e2003 (2008). https://doi.org:10.1371/journal.pone.0002003
433 Chandak, G. R. *et al.* Absence of PRSS1 mutations and association of SPINK1 trypsin inhibitor mutations in hereditary and non-hereditary chronic pancreatitis. *Gut* **53**, 723-728 (2004). https://doi.org:10.1136/gut.2003.026526
434 Sengupta, D., Choudhury, A., Basu, A. & Ramsay, M. Population Stratification and Underrepresentation of Indian Subcontinent Genetic Diversity in the 1000 Genomes Project Dataset. *Genome Biol Evol* **8**, 3460-3470 (2016). https://doi.org:10.1093/gbe/evw244
435 Lee, K. H. *et al.* Mutation Analysis of SPINK1 and CFTR Gene in Korean Patients with Alcoholic Chronic Pancreatitis. *Digestive diseases and sciences* **50**, 1852-1856 (2005). https://doi.org:10.1007/s10620-005-2950-9
436 Keim, V. *et al.* The Course of Genetically Determined Chronic Pancreatitis. *JOP : Journal of the pancreas* **4**, 146-154 (2003).
437 Szabó, A. *et al.* Defective binding of SPINK1 variants is an uncommon mechanism for impaired trypsin inhibition in chronic pancreatitis. *The Journal of biological chemistry* **296**, 100343 (2021). https://doi.org:10.1016/j.jbc.2021.100343
438 CHEN, J.-M., MERCIER, B., AUDREZET, M.-P. & FEREC, C. Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene in hereditary and sporadic chronic pancreatitis. *Journal of Medical Genetics* **37**, 67-69 (2000). https://doi.org:10.1136/jmg.37.1.67
439 Witt, H. *et al.* Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet* **25**, 213-216 (2000). https://doi.org:10.1038/76088
440 Kereszturi, É. & Sahin-Tóth, M. Pancreatic Cancer Cell Lines Heterozygous for the SPINK1 p.N34S Haplotype Exhibit Diminished Expression of the Variant Allele. *Pancreas* **46**, e54-e55 (2017). https://doi.org:10.1097/mpa.0000000000000817
441 Chen, J. M., Mercier, B., Audrezet, M. P. & Ferec, C. Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI)

gene in hereditary and sporadic chronic pancreatitis. *J Med Genet* **37**, 67-69 (2000). https://doi.org:10.1136/jmg.37.1.67

442 Chandak, G. R. *et al.* Mutations in the pancreatic secretory trypsin inhibitor gene (PSTI/SPINK1) rather than the cationic trypsinogen gene (PRSS1) are significantly associated with tropical calcific pancreatitis. *Journal of Medical Genetics* **39**, 347-351 (2002). https://doi.org:10.1136/jmg.39.5.347

443 Di Leo, M. *et al.* Meta-analysis of the impact of SPINK1 p.N34S gene variation in Caucasic patients with chronic pancreatitis. An update. *Dig Liver Dis* **49**, 847-853 (2017). https://doi.org:10.1016/j.dld.2017.04.023

444 Hollemans, R. A. *et al.* Pancreatic exocrine insufficiency following acute pancreatitis: Systematic review and study level meta-analysis. *Pancreatology : official journal of the International Association of Pancreatology (IAP) ... [et al.]* **18**, 253-262 (2018). https://doi.org:10.1016/j.pan.2018.02.009

445 Koziel, D., Gluszek, S., Kowalik, A., Chlopek, M. & Pieciak, L. Genetic mutations in SPINK1, CFTR, CTRC genes in acute pancreatitis. *BMC Gastroenterology* **15**, 70 (2015). https://doi.org:10.1186/s12876-015-0302-6

446 O'Reilly, D. A. *et al.* The SPINK1 N34S variant is associated with acute pancreatitis. *Eur J Gastroenterol Hepatol* **20**, 726-731 (2008). https://doi.org:10.1097/MEG.0b013e3282f5728c

447 Rai, P., Sharma, A., Gupta, A. & Aggarwal, R. Frequency of SPINK1 N34S mutation in acute and recurrent acute pancreatitis. *Journal of Hepato-Biliary-Pancreatic Sciences* **21**, 663-668 (2014). https://doi.org:https://doi.org/10.1002/jhbp.111

448 Threadgold, J. *et al.* The N34S mutation of SPINK1 (PSTI) is associated with a familial pattern of idiopathic chronic pancreatitis but does not cause the disease. *Gut* **50**, 675-681 (2002). https://doi.org:10.1136/gut.50.5.675

449 Teich, N., Mössner, J. & Keim, V. Screening for mutations of the cationic trypsinogen gene: are they of relevance in chronic alcoholic pancreatitis? *Gut* **44**, 413-416 (1999). https://doi.org:10.1136/gut.44.3.413

450 Gomez-Lira, M. *et al.* Mutations in the SPINK1 gene in idiopathic pancreatitis Italian patients. **11**, 543-546 (2003). https://doi.org:10.1038/sj.ejhg.5200989

451 He, C., Holme, J. & Anthony, J. SNP genotyping: the KASP assay. *Methods Mol Biol* **1145**, 75-86 (2014). https://doi.org:10.1007/978-1-4939-0446-4_7

452 Alvarez-Fernandez, A. *et al.* KASP: a genotyping method to rapid identification of resistance in Plasmodium falciparum. *Malaria Journal* **20**, 16 (2021). https://doi.org:10.1186/s12936-020-03544-7

453 Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *The*

Journal of cell biology **160**, 629-633 (2003). https://doi.org:10.1083/jcb.200210140

454 Clegg, R. Förster resonance energy transfer—FRET: what is it, why do it, and how it's done. *«FRET and FLIM Techniques. Laboratory Techniques in Biochemistry and Molecular Biology»* **33**, 1-57 (2009). https://doi.org:10.1016/S0075-7535(08)00001-6. ISBN 978-0-08-054958-3

455 *KASP™ Genotyping Technology*, <http://info.biosearchtech.com/> (

456 Suo, W., Shi, X., Xu, S., Li, X. & Lin, Y. Towards low cost, multiplex clinical genotyping: 4-fluorescent Kompetitive Allele-Specific PCR and its application on pharmacogenetics. *PLOS ONE* **15**, e0230445 (2020). https://doi.org:10.1371/journal.pone.0230445

457 Dombrowski, J. G. *et al.* G6PD deficiency alleles in a malaria-endemic region in the Western Brazilian Amazon. *Malaria Journal* **16**, 253 (2017). https://doi.org:10.1186/s12936-017-1889-6

458 Semagn, K., Babu, R., Hearne, S. & Olsen, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* **33**, 1-14 (2014). https://doi.org:10.1007/s11032-013-9917-x

459 Ciosi, M. *et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**, 568-580 (2019). https://doi.org:10.1016/j.ebiom.2019.09.020

460 Pihlstrom, B. L., Michalowicz, B. S. & Johnson, N. W. Periodontal diseases. *The Lancet* **366**, 1809-1820 (2005). https://doi.org:https://doi.org/10.1016/S0140-6736(05)67728-8

461 *Global Burden of Disease Study*, <http://ghdx.healthdata.org/gbd-results-tool> (2020).

462 Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1211-1259 (2017). https://doi.org:https://doi.org/10.1016/S0140-6736(17)32154-2

463 West, N. in *Highlights of Perio Workshop 2016 on the Boundaries Between Dental Caries and Periodontal Diseases* (ed Søren Jepsen Mariano Sanz, Iain Chapple, Maurizio Tonetti) (European Federation of Periodontology, British Society of Periodontology, 2016).

464 Hodge, P. & Michalowicz, B. Genetic predisposition to periodontitis in children and young adults. *Periodontology 2000* **26**, 113-134 (2001). https://doi.org:doi:10.1034/j.1600-0757.2001.2260106.x

465 Holt, S. C. & Ebersole, J. L. Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia: the 'red complex', a prototype polybacterial pathogenic consortium in periodontitis. *Periodontology 2000* **38**, 72-122 (2005). https://doi.org:https://doi.org/10.1111/j.1600-0757.2005.00113.x

466 Tonetti, M. S., Greenwell, H. & Kornman, K. S. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *Journal of Periodontology* **89**, S159-S172 (2018). https://doi.org:https://doi.org/10.1002/JPER.18-0006

467 Balta, M. G., Loos, B. G. & Nicu, E. A. Emerging Concepts in the Resolution of Periodontal Inflammation: A Role for Resolvin E1. *Frontiers in Immunology* **8** (2017). https://doi.org:10.3389/fimmu.2017.01682

468 Eke, P. I., Dye, B. A., Wei, L., Thornton-Evans, G. O. & Genco, R. J. Prevalence of Periodontitis in Adults in the United States: 2009 and 2010. *Journal of dental research* **91**, 914-920 (2012). https://doi.org:10.1177/0022034512457373

469 Munz, M. *et al.* A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Human Molecular Genetics* **26**, 2577-2588 (2017). https://doi.org:10.1093/hmg/ddx151

470 Preshaw, P. M. *et al.* Periodontitis and diabetes: a two-way relationship. *Diabetologia* **55**, 21-31 (2012). https://doi.org:10.1007/s00125-011-2342-y

471 Scapoli, C. *et al.* Gene–gene interaction among cytokine polymorphisms influence susceptibility to aggressive periodontitis. *Genes And Immunity* **12**, 473 (2011). https://doi.org:10.1038/gene.2011.28

472 Sahingur, S. E. & Cohen, R. E. Analysis of host responses and risk for disease progression. *Periodontology 2000* **34**, 57-83 (2004). https://doi.org:doi:10.1046/j.0906-6713.2002.003425.x

473 Nazir, M. A. Prevalence of periodontal disease, its association with systemic diseases and prevention. *International journal of health sciences* **11**, 72-80 (2017).

474 Indriolo, A., Greco, S., Ravelli, P. & Fagiuoli, S. What can we learn about biofilm/host interactions from the study of inflammatory bowel disease. *Journal of clinical periodontology* **38**, 36-43 (2011). https://doi.org:https://doi.org/10.1111/j.1600-051X.2010.01680.x

475 Toy, V. & Uslu, M. Do genetic polymorphisms affect susceptibility to periodontal disease? A literature review. *Nigerian Journal of Clinical Practice* **22**, 445-453 (2019). https://doi.org:10.4103/njcp.njcp_462_18

476 Offenbacher, S. *et al.* Genome-wide association study of biologically informed periodontal complex traits offers novel insights into the genetic basis of periodontal disease. *Human Molecular Genetics* **25**, 2113-2129 (2016). https://doi.org:10.1093/hmg/ddw069

477 Miley, D. D. *et al.* Heritability of alveolar bone loss from periodontal disease in a baboon population: a pilot study. *J Periodontol* **82**, 575-580 (2011). https://doi.org:10.1902/jop.2010.100189

478 Hiyari, S. *et al.* Heritability of periodontal bone loss in mice. *Journal of periodontal research* **50**, 730-736 (2015). https://doi.org:10.1111/jre.12258

479 Michalowicz, B. S. *et al.* Periodontal findings in adult twins. *J Periodontol* **62**, 293-299 (1991). https://doi.org:10.1902/jop.1991.62.5.293
480 Mucci, L. A., Björkman, L., Douglass, C. W. & Pedersen, N. L. Environmental and heritable factors in the etiology of oral diseases--a population-based study of Swedish twins. *Journal of dental research* **84**, 800-805 (2005). https://doi.org:10.1177/154405910508400904
481 Corey, L. A., Nance, W. E., Hofstede, P. & Schenkein, H. A. Self-reported periodontal disease in a Virginia twin population. *J Periodontol* **64**, 1205-1208 (1993). https://doi.org:10.1902/jop.1993.64.12.1205
482 Michalowicz, B. S. *et al.* Periodontal Findings in Adult Twins. *Journal of Periodontology* **62**, 293-299 (1991). https://doi.org:https://doi.org/10.1902/jop.1991.62.5.293
483 Dowsett, S. A. *et al.* The effect of shared genetic and environmental factors on periodontal disease parameters in untreated adult siblings in Guatemala. *J Periodontol* **73**, 1160-1168 (2002). https://doi.org:10.1902/jop.2002.73.10.1160
484 Divaris, K. *et al.* Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum Mol Genet* **22**, 2312-2324 (2013). https://doi.org:10.1093/hmg/ddt065
485 Teumer, A. *et al.* Genome-wide association study of chronic periodontitis in a general German population. *Journal of clinical periodontology* **40**, 977-985 (2013). https://doi.org:10.1111/jcpe.12154
486 Feng, P. *et al.* Genome wide association scan for chronic periodontitis implicates novel locus. *BMC Oral Health* **14**, 84 (2014). https://doi.org:10.1186/1472-6831-14-84
487 Diomedi, M., Leone, G. & Renna, A. The role of chronic infection and inflammation in the pathogenesis of cardiovascular and cerebrovascular disease. *Drugs of today (Barcelona, Spain : 1998)* **41**, 745-753 (2005). https://doi.org:10.1358/dot.2005.41.11.917342
488 Thorstensson, H., Kuylenstierna, J. & Hugoson, A. Medical status and complications in relation to periodontal disease experience in insulin-dependent diabetics. *Journal of clinical periodontology* **23**, 194-202 (1996).
489 Lalla, E. & Papapanou, P. N. Diabetes mellitus and periodontitis: a tale of two common interrelated diseases. *Nature reviews. Endocrinology* **7**, 738-748 (2011). https://doi.org:10.1038/nrendo.2011.106
490 McDevitt, M. J., Russell, C. M., Schmid, M. J. & Reinhardt, R. A. Impact of increased occlusal contact, interleukin-1 genotype, and periodontitis severity on gingival crevicular fluid IL-1beta levels. *J Periodontol* **74**, 1302-1307 (2003). https://doi.org:10.1902/jop.2003.74.9.1302
491 Pizzo, G., Guiglia, R., Lo Russo, L. & Campisi, G. Dentistry and internal medicine: from the focal infection theory to the periodontal medicine

concept. *European journal of internal medicine* **21**, 496-502 (2010). https://doi.org:10.1016/j.ejim.2010.07.011

492 Newman, K. L. & Kamada, N. Pathogenic associations between oral and gastrointestinal diseases. *Trends in Molecular Medicine* **28**, 1030-1039 (2022). https://doi.org:https://doi.org/10.1016/j.molmed.2022.05.006

493 El-Shinnawi, U. & Soory, M. Associations between periodontitis and systemic inflammatory diseases: response to treatment. *Recent patents on endocrine, metabolic & immune drug discovery* **7**, 169-188 (2013).

494 Bobkova, I. L. & Zinovenko, O. G. Dysbacteriosis as an Aspect of the Pathogenesis of Periodontal Disease. *Dentistry. Estetics. Inovation*, 158-164 (2021). https://doi.org:10.34883/PI.2021.5.2.005

495 Amato, M. *et al.* Probiotics in Periodontal and Peri-Implant Health Management: Biofilm Control, Dysbiosis Reversal, and Host Modulation. *Microorganisms* **10**, 2289 (2022).

496 Prevention methods and programmes for oral diseases. Report of a WHO Expert Committee. *World Health Organization technical report series* **713**, 1-46 (1984).

497 Nath, S. G. & Raveendran, R. Microbial dysbiosis in periodontitis. *J Indian Soc Periodontol* **17**, 543-545 (2013). https://doi.org:10.4103/0972-124x.118334

498 Van Dyke, T. E., Bartold, P. M. & Reynolds, E. C. The Nexus Between Periodontal Inflammation and Dysbiosis. *Front Immunol* **11**, 511 (2020). https://doi.org:10.3389/fimmu.2020.00511

499 Fowler, E. B., Breault, L. G. & Cuenin, M. F. Periodontal disease and its association with systemic disease. *Mil Med* **166**, 85-89 (2001).

500 Kinane, D. F. & Marshall, G. J. Periodontal manifestations of systemic disease. *Aust Dent J* **46**, 2-12 (2001). https://doi.org:10.1111/j.1834-7819.2001.tb00267.x

501 Banoczy, J., Rugg-Gunn, A. & Woodward, M. Milk fluoridation for the prevention of dental caries. *Acta medica academica* **42**, 156-167 (2013). https://doi.org:10.5644/ama2006-124.83

502 Rosenstein, E. D., Greenwald, R. A., Kushner, L. J. & Weissmann, G. Hypothesis: the humoral immune response to oral bacteria provides a stimulus for the development of rheumatoid arthritis. *Inflammation* **28**, 311-318 (2004). https://doi.org:10.1007/s10753-004-6641-z

503 Kostrigina, E., Zyul'kina, L. & Ivanov, P. A MODERN VIEW OF THE ETIOPATHOGENESIS OF PERIODONTITIS (A REVIEW OF LITERATURE). *University proceedings. Volga region. Medical sciences* (2017). https://doi.org:10.21685/2072-3032-2017-3-13

504 Campisi, G. *et al.* Saliva variations in gastro-oesophageal reflux disease. *Journal of dentistry* **36**, 268-271 (2008). https://doi.org:10.1016/j.jdent.2008.01.003

505 Miskiewicz, A. *et al.* The correlation between pancreatic dysfunction markers and selected indices of periodontitis. *Adv Clin Exp Med* **27**, 313-319 (2018). https://doi.org:10.17219/acem/64937

506 Chang, J. S., Tsai, C. R., Chen, L. T. & Shan, Y. S. Investigating the Association Between Periodontal Disease and Risk of Pancreatic Cancer. *Pancreas* **45**, 134-141 (2016). https://doi.org:10.1097/mpa.0000000000000419

507 Jagelavičienė, E., Vaitkevičienė, I., Šilingaitė, D., Šinkūnaitė, E. & Daugėlaitė, G. The Relationship between Vitamin D and Periodontal Pathology. *Medicina (Kaunas)* **54** (2018). https://doi.org:10.3390/medicina54030045

508 Kvaratskhelia, S., Gogiberidze, M., Orjonikidze, M., Puturidze, S. & Jorbenadze, T. The Prevalence of Malocclusion and Periodontal Diseases and Their Correlation in Samegrelo Region, Georgia. *Iran J Public Health* **51**, 207-209 (2022). https://doi.org:10.18502/ijph.v51i1.8314

509 Ababneh, K. T., Taha, A. H., Abbadi, M. S., Karasneh, J. A. & Khader, Y. S. The association of aggressive and chronic periodontitis with systemic manifestations and dental anomalies in a jordanian population: a case control study. *Head & Face Medicine* **6**, 30 (2010). https://doi.org:10.1186/1746-160X-6-30

510 Adler, I. *et al.* Helicobacter pylori associated with glossitis and halitosis. *Helicobacter* **10**, 312-317 (2005). https://doi.org:10.1111/j.1523-5378.2005.00322.x

511 Elahi, M., Telkabadi, M., Samadi, V. & Vakili, H. Association of oral manifestations with ulcerative colitis. *Gastroenterology and hepatology from bed to bench* **5**, 155-160 (2012).

512 Pillet, S., Pozzetto, B. & Roblin, X. Cytomegalovirus and ulcerative colitis: Place of antiviral therapy. *World journal of gastroenterology* **22**, 2030-2045 (2016). https://doi.org:10.3748/wjg.v22.i6.2030

513 Rajkarnikar, J., Thomas, B. S. & Rao, S. K. Inter- relationship between rheumatoid arthritis and periodontitis. *Kathmandu University medical journal (KUMJ)* **11**, 22-26 (2013).

514 Margiotta, V., Campisi, G., Mancuso, S., Accurso, V. & Abbadessa, V. HIV infection: oral lesions, CD4+ cell count and viral load in an Italian study population. *J Oral Pathol Med* **28**, 173-177 (1999). https://doi.org:10.1111/j.1600-0714.1999.tb02019.x

515 Bravo, I. M. *et al.* Prevalence of oral lesions in HIV patients related to CD4 cell count and viral load in a Venezuelan population. *Medicina oral, patologia oral y cirugia bucal* **11**, E33-39 (2006).

516 Campo, J. *et al.* Oral candidiasis as a clinical marker related to viral load, CD4 lymphocyte count and CD4 lymphocyte percentage in HIV-infected patients. *J Oral Pathol Med* **31**, 5-10 (2002). https://doi.org:10.1034/j.1600-0714.2002.310102.x

517 Greenspan, D. *et al.* Incidence of oral lesions in HIV-1-infected women: reduction with HAART. *Journal of dental research* **83**, 145-150 (2004). https://doi.org:10.1177/154405910408300212

518 Leao, J. C., Ribeiro, C. M. B., Carvalho, A. A. T., Frezzini, C. & Porter, S. Oral complications of HIV disease. *Clinics (Sao Paulo, Brazil)* **64**, 459-470 (2009). https://doi.org:10.1590/S1807-59322009000500014

519 Levy, J. A. HIV pathogenesis: 25 years of progress and persistent challenges. *Aids* **23**, 147-160 (2009). https://doi.org:10.1097/QAD.0b013e3283217f9f

520 Forbi, J. C. & Agwale, S. M. Inverted CD4+/CD8+ ratio associated with AIDS event and death in HIV-1 infected individuals in Nasarawa State, Nigeria. *Tanzan J Health Res* **11**, 144-148 (2009). https://doi.org:10.4314/thrb.v11i3.47701

521 Klein, R. S. *et al.* Oral candidiasis in high-risk patients as the initial manifestation of the acquired immunodeficiency syndrome. *N Engl J Med* **311**, 354-358 (1984). https://doi.org:10.1056/nejm198408093110602

522 Reichart, P. A. *et al.* Oral hairy leukoplakia: observations in 95 cases and review of the literature. *J Oral Pathol Med* **18**, 410-415 (1989). https://doi.org:10.1111/j.1600-0714.1989.tb01573.x

523 Hilton, J. F. *et al.* Development of oral lesions in human immunodeficiency virus-infected transfusion recipients and hemophiliacs. *Am J Epidemiol* **145**, 164-174 (1997). https://doi.org:10.1093/oxfordjournals.aje.a009087

524 Steidley, K. E. *et al.* A comparison of T4:T8 lymphocyte ratio in the periodontal lesion of healthy and HIV-positive patients. *J Periodontol* **63**, 753-756 (1992). https://doi.org:10.1902/jop.1992.63.9.753

525 Wahid, A., Chaudhry, S., Ehsan, A., Butt, S. & Ali Khan, A. Bidirectional Relationship between Chronic Kidney Disease & Periodontal Disease. *Pak J Med Sci* **29**, 211-215 (2013). https://doi.org:10.12669/pjms.291.2926

526 A. S. Simbirtsev, A. Y. G., A. V. Rydlovskaya. The role of cytokine gene polymorphism in regulation inflammation and immunity. *Medical Academic Journal* **6** (2006).

527 Rawlinson, A., Grummitt, J. M., Walsh, T. F. & Ian Douglas, C. W. Interleukin 1 and receptor antagonist levels in gingival crevicular fluid in heavy smokers versus non-smokers. *Journal of clinical periodontology* **30**, 42-48 (2003). https://doi.org:10.1034/j.1600-051X.2003.300107.x

528 Kishimoto, T. Interleukin-6: discovery of a pleiotropic cytokine. *Arthritis Research & Therapy* **8**, S2 (2006). https://doi.org:10.1186/ar1916

529 Singh, P., Gupta, N. D., Bey, A. & Khan, S. Salivary TNF-alpha: A potential marker of periodontal destruction. *J Indian Soc Periodontol* **18**, 306-310 (2014). https://doi.org:10.4103/0972-124x.134566

530 Pradeep, A. R., Roopa, Y. & Swati, P. P. Interleukin-4, a T-helper 2 cell cytokine, is associated with the remission of periodontal disease. *Journal of periodontal research* **43**, 712-716 (2008). https://doi.org:10.1111/j.1600-0765.2007.01079.x

531 Rawlinson, A., Dalati, M. H., Rahman, S., Walsh, T. F. & Fairclough, A. L. Interleukin-1 and IL-1 receptor antagonist in gingival crevicular fluid. *Journal of clinical periodontology* **27**, 738-743 (2000). https://doi.org:10.1034/j.1600-051x.2000.027010738.x

532 Herr, D. R. in *International Review of Cell and Molecular Biology* Vol. 297 (ed Kwang W. Jeon) 45-81 (Academic Press, 2012).
533 Kornman, K. S. & di Giovine, F. S. Genetic variations in cytokine expression: a risk factor for severity of adult periodontitis. *Annals of periodontology* **3**, 327-338 (1998). https://doi.org:10.1902/annals.1998.3.1.327
534 Brett, P. M. *et al.* Functional Gene Polymorphisms in Aggressive and Chronic Periodontitis. *Journal of dental research* **84**, 1149-1153 (2005). https://doi.org:10.1177/154405910508401211
535 McDevitt, M. J. *et al.* Interleukin-1 Genetic Association With Periodontitis in Clinical Practice. *Journal of Periodontology* **71**, 156-163 (2000). https://doi.org:10.1902/jop.2000.71.2.156
536 Kornman, K. S. Interleukin 1 genetics, inflammatory mechanisms, and nutrigenetic opportunities to modulate diseases of aging. *The American journal of clinical nutrition* **83**, 475s-483s (2006). https://doi.org:10.1093/ajcn/83.2.475S
537 McLoughlin, R. M. *et al.* Differential regulation of neutrophil-activating chemokines by IL-6 and its soluble receptor isoforms. *Journal of immunology (Baltimore, Md. : 1950)* **172**, 5676-5683 (2004).
538 Sugiyama, A. *et al.* Activation of human gingival epithelial cells by cell-surface components of black-pigmented bacteria: augmentation of production of interleukin-8, granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor and expression of intercellular adhesion molecule 1. *Journal of medical microbiology* **51**, 27-33 (2002). https://doi.org:10.1099/0022-1317-51-1-27
539 Biedermann, T. *et al.* Mast cells control neutrophil recruitment during T cell-mediated delayed-type hypersensitivity reactions through tumor necrosis factor and macrophage inflammatory protein 2. *The Journal of experimental medicine* **192**, 1441-1452 (2000).
540 Parameswaran, N. & Patial, S. Tumor necrosis factor-alpha signaling in macrophages. *Critical reviews in eukaryotic gene expression* **20**, 87-103 (2010).
541 Telepnev, M., Golovliov, I., Grundström, T., Tärnvik, A. & Sjöstedt, A. Francisella tularensis inhibits Toll-like receptor-mediated activation of intracellular signalling and secretion of TNF-α and IL-1 from murine macrophages. *Cellular Microbiology* **5**, 41-51 (2003). https://doi.org:10.1046/j.1462-5822.2003.00251.x
542 Takeichi, O. *et al.* Cytokine Profiles of T-lymphocytes from Gingival Tissues with Pathological Pocketing. *Journal of dental research* **79**, 1548-1555 (2000). https://doi.org:10.1177/00220345000790080401
543 Kurtiş, B. *et al.* Gingival Crevicular Fluid Levels of Monocyte Chemoattractant Protein-1 and Tumor Necrosis Factor-Alpha in Patients With Chronic and Aggressive Periodontitis. *Journal of Periodontology* **76**, 1849-1855 (2005). https://doi.org:10.1902/jop.2005.76.11.1849

544 Liu, X. & Li, H. A Systematic Review and Meta-Analysis on Multiple Cytokine Gene Polymorphisms in the Pathogenesis of Periodontitis. *Frontiers in Immunology* **12** (2022). https://doi.org:10.3389/fimmu.2021.713198
545 Norman, A. W. A vitamin D nutritional cornucopia: new insights concerning the serum 25-hydroxyvitamin D status of the US population. *The American journal of clinical nutrition* **88**, 1455-1456 (2008). https://doi.org:10.3945/ajcn.2008.27049
546 Inagaki, K., Krall, E. A., Fleet, J. C. & Garcia, R. I. Vitamin D receptor alleles, periodontal disease progression, and tooth loss in the VA dental longitudinal study. *J Periodontol* **74**, 161-167 (2003). https://doi.org:10.1902/jop.2003.74.2.161
547 Provvedini, D. M., Tsoukas, C. D., Deftos, L. J. & Manolagas, S. C. 1,25-dihydroxyvitamin D3 receptors in human leukocytes. *Science* **221**, 1181-1183 (1983).
548 R, W. (Wageningen Academic Publishers, The Netherlands, 2013).
549 Mathieu, C. Vitamin D and the immune system: Getting it right. **8**, 178-186 (2011). https://doi.org:10.1138/20110505
550 Amano, Y., Komiyama, K. & Makishima, M. Vitamin D and periodontal disease. *Journal of oral science* **51**, 11-20 (2009).
551 Umasuthan, N., Mothishri, M. S., Thulasitha, W. S., Nam, B. H. & Lee, J. Molecular, genomic, and expressional delineation of a piscidin from rock bream (Oplegnathus fasciatus) with evidence for the potent antimicrobial activities of Of-Pis1 peptide. *Fish Shellfish Immunol* **48**, 154-168 (2016). https://doi.org:10.1016/j.fsi.2015.11.005
552 Steinstraesser, L., Kraneburg, U., Jacobsen, F. & Al-Benna, S. Host defense peptides and their antimicrobial-immunomodulatory duality. *Immunobiology* **216**, 322-333 (2011). https://doi.org:10.1016/j.imbio.2010.07.003
553 Bahri, R., Saidane-Mosbahi, D. & Rouabhia, M. Candida famata modulates toll-like receptor, beta-defensin, and proinflammatory cytokine expression by normal human epithelial cells. *J Cell Physiol* **222**, 209-218 (2010). https://doi.org:10.1002/jcp.21939
554 Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* **3**, 710-720 (2003). https://doi.org:10.1038/nri1180
555 Linzmeier, R. M. & Ganz, T. Human defensin gene copy number polymorphisms: Comprehensive analysis of independent variation in α- and β-defensin regions at 8p22–p23. *Genomics* **86**, 423-430 (2005). https://doi.org:https://doi.org/10.1016/j.ygeno.2005.06.003
556 Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nature Reviews Immunology* **3**, 710-720 (2003). https://doi.org:10.1038/nri1180
557 Guani-Guerra, E., Santos-Mendoza, T., Lugo-Reyes, S. O. & Teran, L. M. Antimicrobial peptides: general overview and clinical implications in human health and disease. *Clinical immunology (Orlando, Fla.)* **135**, 1-11 (2010). https://doi.org:10.1016/j.clim.2009.12.004

558 Fu, L. B., Yu, J. L. & Liu, W. H. [Biological characteristics of defensin and its disease-resistance genetic engineering]. *Yi chuan = Hereditas* **33**, 512-519 (2011).
559 Kim, C. *et al.* Human alpha-defensins neutralize anthrax lethal toxin and protect against its fatal consequences. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4830-4835 (2005). https://doi.org:10.1073/pnas.0500508102
560 Feng, Z., Dubyak, G. R., Lederman, M. M. & Weinberg, A. Cutting edge: human beta defensin 3--a novel antagonist of the HIV-1 coreceptor CXCR4. *Journal of immunology (Baltimore, Md. : 1950)* **177**, 782-786 (2006). https://doi.org:10.4049/jimmunol.177.2.782
561 Kim, J. Y. *et al.* Protease inhibitors from plants with antimicrobial activity. *Int J Mol Sci* **10**, 2860-2872 (2009). https://doi.org:10.3390/ijms10062860
562 Li, X. *et al.* The expression of human β-defensins (hBD-1, hBD-2, hBD-3, hBD-4) in gingival epithelia. *Archives of oral biology* **66**, 15-21 (2016). https://doi.org:10.1016/j.archoralbio.2016.01.012
563 Salzman, N. H., Ghosh, D., Huttner, K. M., Paterson, Y. & Bevins, C. L. Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin. *Nature* **422**, 522-526 (2003). https://doi.org:10.1038/nature01520
564 Zilbauer, M. *et al.* Expression of human beta-defensins in children with chronic inflammatory bowel disease. *PLoS One* **5**, e15389 (2010). https://doi.org:10.1371/journal.pone.0015389
565 Kohlgraf, K. G., Pingel, L. C., Dietrich, D. E. & Brogden, K. A. Defensins as anti-inflammatory compounds and mucosal adjuvants. *Future microbiology* **5**, 99-113 (2010). https://doi.org:10.2217/fmb.09.104
566 Bals, R. Epithelial antimicrobial peptides in host defense against infection. *Respir Res* **1**, 141-150 (2000). https://doi.org:10.1186/rr25
567 Lai, Y. & Gallo, R. L. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol* **30**, 131-141 (2009). https://doi.org:10.1016/j.it.2008.12.003
568 Sahl, H. G. *et al.* Mammalian defensins: structures and mechanism of antibiotic activity. *J Leukoc Biol* **77**, 466-475 (2005). https://doi.org:10.1189/jlb.0804452
569 Zhao, H. & Kinnunen, P. K. Modulation of the activity of secretory phospholipase A2 by antimicrobial peptides. *Antimicrob Agents Chemother* **47**, 965-971 (2003). https://doi.org:10.1128/aac.47.3.965-971.2003
570 Peters, B. M., Shirtliff, M. E. & Jabra-Rizk, M. A. Antimicrobial peptides: primeval molecules or future drugs? *PLoS Pathog* **6**, e1001067 (2010). https://doi.org:10.1371/journal.ppat.1001067
571 Bedi, T., Mahendra, J. & Ambalavanan, N. Defensins in periodontal health. *Indian Journal of Dental Research* **26**, 340-344 (2015). https://doi.org:10.4103/0970-9290.167627

572 Gursoy, U. K., Kononen, E., Luukkonen, N. & Uitto, V. J. Human neutrophil defensins and their effect on epithelial cells. *J Periodontol* **84**, 126-133 (2013). https://doi.org:10.1902/jop.2012.120017

573 Derradjia, A. *et al.* alpha-tocopherol decreases interleukin-1beta and -6 and increases human beta-defensin-1 and -2 secretion in human gingival fibroblasts stimulated with Porphyromonas gingivalis lipopolysaccharide. *Journal of periodontal research* **51**, 295-303 (2016). https://doi.org:10.1111/jre.12308

574 Fruitwala, S., El-Naccache, D. W. & Chang, T. L. Multifaceted immune functions of human defensins and underlying mechanisms. *Semin Cell Dev Biol* **88**, 163-172 (2019). https://doi.org:10.1016/j.semcdb.2018.02.023

575 Ganz, T. *et al.* Defensins. Natural peptide antibiotics of human neutrophils. *The Journal of clinical investigation* **76**, 1427-1435 (1985). https://doi.org:10.1172/JCI112120

576 Kaiser, V. & Diamond, G. Expression of mammalian defensin genes. *Journal of Leukocyte Biology* **68**, 779-784 (2000). https://doi.org:https://doi.org/10.1189/jlb.68.6.779

577 Ganz, T. Antimicrobial polypeptides. *Journal of Leukocyte Biology* **75**, 34-38 (2004). https://doi.org:https://doi.org/10.1189/jlb.0403150

578 Ganz, T., Metcalf, J. A., Gallin, J. I., Boxer, L. A. & Lehrer, R. I. Microbicidal/cytotoxic proteins of neutrophils are deficient in two disorders: Chediak-Higashi syndrome and "specific" granule deficiency. *The Journal of Clinical Investigation* **82**, 552-556 (1988). https://doi.org:10.1172/JCI113631

579 Young, A. N. *et al.* Beta defensin-1, parvalbumin, and vimentin: a panel of diagnostic immunohistochemical markers for renal tumors derived from gene expression profiling studies using cDNA microarrays. *Am J Surg Pathol* **27**, 199-205 (2003). https://doi.org:10.1097/00000478-200302000-00008

580 Bauer, B., Wex, T., Kuester, D., Meyer, T. & Malfertheiner, P. Differential expression of human beta defensin 2 and 3 in gastric mucosa of Helicobacter pylori-infected individuals. *Helicobacter* **18**, 6-12 (2013). https://doi.org:10.1111/hel.12000

581 Sun, C. Q. *et al.* Human beta-defensin-1, a potential chromosome 8p tumor suppressor: control of transcription and induction of apoptosis in renal cell carcinoma. *Cancer Res* **66**, 8542-8549 (2006). https://doi.org:10.1158/0008-5472.Can-06-0294

582 Cox, D. P. & Weathers, D. R. Leukocyte adhesion deficiency type 1: an important consideration in the clinical differential diagnosis of prepubertal periodontitis. A case report and review of the literature. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **105**, 86-90 (2008). https://doi.org:10.1016/j.tripleo.2007.02.026

583 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**, 559-575 (2007).

584 Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods* **11**, 407-409 (2014). https://doi.org:10.1038/nmeth.2848
585 Lu, B. *et al.* Cloning and characterization of murine 1-acyl-sn-glycerol 3-phosphate acyltransferases and their regulation by PPARalpha in murine heart. *Biochem J* **385**, 469-477 (2005). https://doi.org:10.1042/bj20041348
586 Divaris, K. Fundamentals of Precision Medicine. *Compend Contin Educ Dent* **38**, 30-32 (2017).
587 Michalowicz, B. S. *et al.* Evidence of a substantial genetic basis for risk of adult periodontitis. *J Periodontol* **71**, 1699-1707 (2000). https://doi.org:10.1902/jop.2000.71.11.1699
588 Ikuta, T. *et al.* Gene polymorphism of β-defensin-1 is associated with susceptibility to periodontitis in Japanese. *Odontology* **103**, 66-74 (2015). https://doi.org:10.1007/s10266-013-0139-9
589 Munz, M. *et al.* Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. *Eur J Hum Genet* **27**, 102-113 (2019). https://doi.org:10.1038/s41431-018-0265-5
590 Suzuki, A. *et al.* Single nucleotide polymorphisms associated with aggressive periodontitis and severe chronic periodontitis in Japanese. *Biochem Biophys Res Commun* **317**, 887-892 (2004). https://doi.org:10.1016/j.bbrc.2004.03.126
591 Murakami, T. [Understanding and treatment strategy of the pathogenesis of periodontal disease based on chronic inflammation]. *Clin Calcium* **26**, 766-772 (2016).
592 Papadopoulos, G. *et al.* Immunologic environment influences macrophage response to Porphyromonas gingivalis. *Molecular oral microbiology* **32**, 250-261 (2017). https://doi.org:10.1111/omi.12168
593 Yu, T. *et al.* Enhanced Activity of the Macrophage M1/M2 Phenotypes and Phenotypic Switch to M1 in Periodontal Infection. *J Periodontol* **87**, 1092-1102 (2016). https://doi.org:10.1902/jop.2016.160081
594 Delima, A. J. *et al.* Soluble antagonists to interleukin-1 (IL-1) and tumor necrosis factor (TNF) inhibits loss of tissue attachment in experimental periodontitis. *Journal of clinical periodontology* **28**, 233-240 (2001). https://doi.org:10.1034/j.1600-051x.2001.028003233.x

Supplementary material

Table 1.

Diagnosis	Quantity	Gender		Age	BMI
		female	male		
Gallstone	56	40	16	22-87	20-49
Alcoholic Pancreatitis	15	5	10	24-78	20-46
Idiopathic Pancreatitis	30	16	14	20-82	31-42
Malignancy	7	6	1	43-81	21-31
CBD stones	3	1	2	62,65,76	26,31,38
Congenital malformation	1	1	-	69	18
Medication	1	-	1	33	23
Necrosis	3	1	2	66-2, 71	22,25,29
Polyp	1	-	1	41	26
Post ERCP	5	4	1	34-68	24-38
Post operation	1	1	-	36	33
Pseudocyst	1	-	1	43	24