УДК 61:75; 575

Original Article

WHOLE EXOME SEQUENCING OF A PATIENT WITH MORBID OBESITY: TRIO ANALYSIS, PRELIMINARY RESULTS

[*]Sikhayeva N.[1,*], Nakysh A.[2], Utupov T.[1], Zholdybayeva E.[1]

[1]*National Center for Biotechnology, 13/5, Korgalzhyn road, Astana, 010000, Kazakhstan*
[2]*Astana Medical University, 49a, Beybitshilik st., Astana, 010000, Kazakhstan*
[*]*ksnurgul@gmail.com*

ABSTRACT

Morbid obesity is a severe form of obesity that leads to numerous cardiovascular, metabolic diseases and cancers, as well as increased mortality. Whole-exome sequencing is the effective tool for studying more extreme forms of disease, such as morbid obesity. The aim of the study is to identify candidate genes involved in the development of morbid obesity using whole exome sequencing.

Here in this study, a family from Kazakh cohort having two siblings, one unaffected and one affected with morbid obesity was enrolled. Whole Exome Sequencing (WES) of trio with one affected and unaffected parents was done.

The trio analysis revealed a number of potential candidate genes predisposing to morbid obesity. Three genetic models were used: recessive, *de novo* and *compound heterozygote* models. In the recessive model, one variant (rs116253946) in the *KIAA1671* was identified as a candidate. In the *de novo* model, 15 polymorphisms located in 10 different genes were identified. In the *compound heterozygote* model, 6 polymorphisms, located in *FRG1* and *MST1L* genes, were identified.

This work presents the preliminary results of the study. Twelve genes identified by WES may be involved in the development of morbid obesity, and are grouped into three main pathophysiological pathways (immune response: *MST1L, HLA-DRB5, PRRC2A, KIR2DS4*; cell division and structure: *CEP170, TEKT4*; neuronal response: *OR2T34, ANKMY2*) or their function remains unknown to date (*KIAA1671, TMEM191C, PPP6R2, FRG1*).

Keywords: morbid obesity, gene, pathogenic mutation, whole exome sequencing.

INTRODUCTION

Morbid obesity is a severe form of obesity that leads to numerous cardiovascular, metabolic diseases and cancers, as well as increased mortality. At the population level, obesity is associated with environmental factors such as an abundance of unhealthy foods and lack of physical activity, but at the individual level, genetic factors explain most differences in body mass index (BMI). In general, the heredity of obesity is estimated at 40% to 70% [1].

The discovery of new genes associated with metabolic disorders makes it possible to identify biochemical and signaling pathways, as well as the mechanisms of their regulation that control the considered phenotypic traits. Such genes can be considered as candidate genes used to assess the risk of developing pathology, as well as potential targets for pharmacological interventions. Identification of genes associated with pathologies is carried out using various strategies, including family analysis, the study of candidate genes, as well as genome wide association studies. However, methodological approaches are not limited to these methods, and the era of next generation sequencing, including whole exome sequencing, should bring additional answers to the missing parts of the genetic underpinnings of metabolic diseases. It should be noted that next generation sequencing technologies have revolutionized the study of the molecular pathogenesis of many diseases. One such technology is whole-exome sequencing. Exome sequencing technology is already being used to study obesity. Thus, this technology has been validated for molecular diagnostics of 43 forms of diabetes and monogenic obesity [2]. That is, exome sequencing technology makes it possible to identify new variants in genes in obese patients. These variants may have a functional impact, thereby improving understanding of the pathophysiology of obesity.

The aim of the study is to identify candidate genes involved in the development of morbid obesity using whole exome sequencing.

MATERIALS AND METHODS

One family trio, the proband of which had morbid obesity (or BMI ≥ 35 kg/m^2) was included in this study. Written informed consent was obtained from all participants and the protocol was approved by the Ethics Committee of National Centre for Biotechnology, Astana, Kazakhstan (№ 10.14.03.2012).

Anthropometric indices including weight, height, waist circumference (WC), hip circumference (HC), and blood pressure were measured following standard protocols.

DNA was extracted from peripheral blood from the patient and parents for genetic analysis. DNA was extracted using the salting-out method [3]. The purity and concentration of the DNA obtained were determined using the NanoDrop spectrophotometer 1000 (Thermo Scientific).

Whole exome sequencing of 4 DNA samples from the same family was performed. Enrichment of exome target regions, including coding and untranslated regions regions, using SureSelect Targeted Enrichment Workflow (Agilent Technologies, USA) was realized from genomic DNA using DNA libraries. Sequencing was performed on the HiSeq 2500 platform (Illumina, USA) at Theragen Etex Bioinstitute (Republic of Korea), according to the protocol from the manufacturer. The quality control of primary sequences (reads) was done using the *FastQC program.* High quality sequences were then aligned to the reference human genome (UCSChg19, human genome assembly 19) using the BWA (Burrows-WheelerA-

ligner) program [4]. Duplicates as well as poorly aligned sequences ("Q-score" < 20) were removed. Aligned sequences subsequently recalibrated using the GATK software (genome analysis tool kit) [5] to improve the detection and nomination of variants. Finally, after quality control, polymorphisms were identified using the SAMtools program [6] and annotated using the SnpEff program [7].

RESULTS

Whole exome sequencing technology was used to identify new candidate genes linked to morbid obesity. In contrast to whole genome sequencing, which requires reading 3 billion bp of the human genome, exome sequencing means «capture» and targeted reading of only coding regions, which corresponds to 1-3% of the human genome. This study involved a family (trio: father, mother and child) of Kazakh nationality. The main characteristics of the study participants are shown in Table 1.

As a clinical analysis result, data on clinically significant pathogenic polymorphisms present in study participants were obtained (Table 2). It should be noted that in all samples, missense mutation in the polymorphism rs1055138 (*CYP4V2* gene) was identified in the heterozygous state. Exome sequencing revealed a number of pathogenic polymorphisms associated with the risk of developing diastolic hypertension (rs11739136 in the *KCNMB1* gene), Bardet-Biedl syndrome (rs4784677 in the *BBS2* gene), obesity (rs2282440 in the *SDC3* gene), which lead to an increase in BMI. Pathogenic polymorphisms associated with the risk of developing certain orphan diseases have also been observed, such as maple syrup disease and Bietti's dystrophy, and the risk of developing neurological diseases, such as Alzheimer's disease (Table 2).

Table 1. Characteristics of study participants of morbid obesity

Sample		Age (years)	Weight (kg)	Height (sm)	BMI (kg/m^2)	WC (sm)	HC (sm)	SBP (mmHg)	DBP (mmHg)
Family	1.1 (Father)	55	100	170	34,60	116	111	120	80
	1.2 (Mother)	52	76	158	30,44	90	105	135	80
	1.3 (Daughter)	29	63	159	24,92	78	95	90	60
	1.4 (Son, proband)	30	132	177	42,13	132	126	140	85

Table 2. Pathogenic variants summary according to ClinVar

Reference sequence	Genotype	Exonic effect	Gene	Nucleotide change	Disease
Sample 1.1					
rs1055138	GG	missense	*CYP4V2*	c.64C>G	Bietti crystalline corneo retinal dystrophy
rs111033566	AT	missense	*PRSS1*	c.86A>T	Hereditary pancreatitis
rs114747203	GA	missense	*OR2J3*	c.677G>A	C3hex\x2c ability to smell
rs11558492	AG	missense	*GNPAT*	c.1556A>G	Rhizomelicchondro dysplasia Punctata type 2
rs11887534	GC	missense	*ABCG8*	c.55G>C	Gallbladder disease 4
rs12021720	TC	missense	*DBT*	c.1150A>G	Intermediate maples yrup Urine disease type 2
rs1208	GA	missense	*NAT2*	c.803G>A	Slow acetylator due to N-acetyltransferase enzyme variant
rs13306487	GA	missense	*ITGB3*	c.1544G>A	Ca/Tu alloantigen polymorphism, Thrombo cytopenia, neonatalallo immune
rs1726866	AA	missense	*TAS2R38*	c.785C>T	Phenylthiocarbamide tasting
rs1801253	GC	missense	*ADRB1*	c.1165G>C	Congestive heart failure and beta-blocker response
rs1801280	TC	missense	*NAT2*	c.341T>C	Slow acetylator due to N-acetyltransferase enzyme variant
rs1803274	CT	missense	*BCHE*	c.1699G>A	Butyryl cholinesterase activity
rs35383149	TC	missense	*ALG6*	c.391T>C	Congenital disorder of Glycosylation type1C

rs3733402	AA	missense	*KLKB1*	c.428G>A	Prekallikrein deficiency
rs3827760	AG	missense	*EDAR*	c.1109T>C	Hair morphology, hair thickness
rs3829241	AA	missense	*TPCN2*	c.2201G>A	Skin/hair/eye pigmentation
rs4148323	GA	missense	*UGT1A1*	c.211G>A	Gilbert's syndrome, Hyper bilirubinemia transient familial neonatal, Bilirubin
rs4244285	GA	samesense	*CYP2C19*	c.681G>A	Mephenytoin, poor metabolism of Proguanil and Clopidogrel response
rs429358	TC	missense	*APOE*	c.388T>C	Familial type 3 Hyper lipoproteinemia, Alzheimer disease due to APOE4 isoform
rs4784677	TT	missense	*BBS2*	c.209G>A	Bardet-biedl syndrome
rs5911	CC	missense	*ITGA2B*	c.2621T>G	Bakplatelet-specific antigen
rs61757294	AG	missense	*CYP11B2*	c.1157T>C	Corticosteronemethyl oxidase Type 2 deficiency
rs854560	AT	missense	*PON1*	c.163T>A	Coronary artery disease
Sample 1.2					
rs1055138	CG	missense	*CYP4V2*	c.64C>G	Bietti crystalline corneo retinal dystrophy
rs10246939	TC	missense	*TAS2R38*	c.886A>G	Phenylthiocarbamide tasting
rs1053878	GA	missense	*ABO*	c.464C>T	ABO blood group system, B (A) phenotype
rs111033566	AT	missense	*PRSS1*	c.86A>T	Hereditary pancreatitis
rs11739136	CT	missense	*KCNMB1*	c.193G>A	Hypertension diastolic
rs12021720	CC	missense	*DBT*	c.1150A>G	Intermediate maple syrup urine Disease type 2
rs1208	AA	missense	*NAT2*	c.803G>A	Slow acetylator due to N-acetyltransferase enzyme variant
rs1426654	AG	missense	*SLC24A5*	c.331A>G	Skin/hair/eye pigmentation
rs146973734	CT	missense	*GDF3*	c.584G>A	Microphthalmia\x2c isolated7
rs1726866	GA	missense	*TAS2R38*	c.785C>T	Phenylthiocarbamide tasting
rs1799853	CT	missense	*CYP2C9*	c.430C>T	Warfarin response
rs1799895	CG	missense	*SOD3*	c.691C>G	Superoxide dismutase, elevated extracellular
rs1799930	GA	missense	*NAT2*	c.590G>A	Slow acetylator due to N-acetyltransferase enzyme variant
rs1800497	GA	missense	*ANKK1*	c.2137G>A	Dopamine receptor d 2
rs2279343	AG	missense	*CYP2B6*	c.785A>G	Efavirenz response
rs2282440	GA	missense	*SDC3*	c.986C>T	Obesity
rs3733402	GA	missense	*KLKB1*	c.428G>A	Prekallikrein deficiency
rs3827760	GG	missense	*EDAR*	c.1109T>C	Hair morphology, hair thickness
rs3829241	GA	missense	*TPCN2*	c.2201G>A	Skin/hair/eye pigmentation
rs4244285	GA	samesense	*CYP2C19*	c.681G>A	Mephenytoin, poor metabolism of Proguanil and Clopidogrel response
rs4774518	CT	samesense	*DUOXA2*	c.738C>T	Thyroglobulin synthesis defect
rs4784677	TT	missense	*BBS2*	c.209G>A	Bardet-biedl syndrome

rs5911	AC	missense	*ITGA2B*	c.2621T>G	Bakplatelet-specific antigen
rs6256	GT	samesense	*PTH*	c.247C>A	Primary hyper parathyroidism
rs662	TC	missense	*PON1*	c.575A>G	Coronary artery disease
rs713598	CG	missense	*TAS2R38*	c.145G>C	Phenylthiocarbamide tasting
rs77010315	CA	missense	*SLC36A2*	c.260G>T	Iminoglycinuria, Hyperglycinuria
rs1055138	GG	missense	*CYP4V2*	c.64C>G	Bietti crystalline corneo retinal dystrophy
rs854560	AT	missense	*PON1*	c.163T>A	Coronary artery disease Microvascular complications of diabetes
Sample 1.3					
rs111033566	AT	missense	*PRSS1*	c.86A>T	Hereditary pancreatitis
rs114747203	GA	missense	*OR2J3*	c.677G>A	C3hex\x2c ability to smell
rs11558492	AG	missense	*GNPAT*	c.1556A>G	Rhizomelic chondro dysplasia Punctata type 2
rs11887534	GC	missense	*ABCG8*	c.55G>C	Gallbladder disease 4
rs12021720	CC	missense	*DBT*	c.1150A>G	Intermediate maple syrup urine Disease type 2
rs1208	AA	missense	*NAT2*	c.803G>A	Slow acetylator due to N-acetyltransferase enzyme variant
rs1726866	AA	missense	*TAS2R38*	c.785C>T	Phenylthiocarbamide tasting
rs1801280	TC	missense	*NAT2*	c.341T>C	Slow acetylator due to N-acetyltransferase enzyme variant
rs2282440	GA	missense	*SDC3*	c.986C>T	Obesity
rs35383149	TC	missense	*ALG6*	c.391T>C	Congenital disorder of Glycosylation type 1C
rs3733402	AA	missense	*KLKB1*	c.428G>A	Prekallikrein deficiency
rs3827760	AG	missense	*EDAR*	c.1109T>C	Hair morphology, hair thickness
rs3829241	AA	missense	*TPCN2*	c.2201G>A	Skin/hair/eye pigmentation
rs4148323	GA	missense	*UGT1A1*	c.211G>A	Gilbert's syndrome, Hyper bilirubinemia transient Familial neonatal, Bilirubin
rs4244285	GA	samesense	*CYP2C19*	c.681G>A	Mephenytoin, poor metabolism of Proguanil and Clopidogrel response
rs429358	TC	missense	*APOE*	c.388T>C	Familial type 3 Hyper lipoproteinemia, Alzheimer disease due to APOE4 isoform
rs4784677	TT	missense	*BBS2*	c.209G>A	Bardet-biedl syndrome
rs5911	CC	missense	*ITGA2B*	c.2621T>G	Bakplatelet-specific antigen
rs854560	AT	missense	*PON1*	c.163T>A	Coronary artery disease
Sample 1.4					
rs10246939	TC	missense	*TAS2R38*	c.886A>G	Phenylthiocarbamide tasting
rs1053878	GA	missense	*ABO*	c.464C>T	ABO blood group system, B (A) phenotype
rs1055138	CG	missense	*CYP4V2*	c.64C>G	Bietti crystalline corneo retinal dystrophy

rs12021720	TC	missense	*DBT*	c.1150A>G	Intermediate maple syrup Urine disease type2
rs1208	AA	missense	*NAT2*	c.803G>A	Slow acetylator due to N-acetyltransferase enzyme variant
rs1726866	GA	missense	*TAS2R38*	c.785C>T	Phenylthiocarbamide tasting
rs1799895	CG	missense	*SOD3*	c.691C>G	Superoxide dismutase, elevated extracellular
rs1801253	GC	missense	*ADRB1*	c.1165G>C	Congestive heart failure and beta-blocker response
rs1801280	TC	missense	*NAT2*	c.341T>C	Slow acetylator due to N-acetyltransferase enzyme variant
rs2282440	GA	missense	*SDC3*	c.986C>T	Obesity
rs35383149	TC	missense	*ALG6*	c.391T>C	Congenital disorder of glycosylation type 1C
rs3733402	GA	missense	*KLKB1*	c.428G>A	Prekallikrein deficiency
rs3827760	AG	missense	*EDAR*	c.1109T>C	Hair morphology, hair thickness
rs3829241	GA	missense	*TPCN2*	c.2201G>A	Skin/hair/eye pigmentation
rs4244285	GA	samesense	*CYP2C19*	c.681G>A	Mephenytoin, poor metabolism of Proguanil and Clopidogrel response
rs429358	TC	missense	*APOE*	c.388T>C	Familial type 3 Hyper lipoproteinemia, Alzheimer disease due to APOE4 isoform
rs4784677	TT	missense	*BBS2*	c.209G>A	Bardet-biedl syndrome
rs5911	AC	missense	*ITGA2B*	c.2621T>G	Bakplatelet-specific antigen
rs61757294	AG	missense	*CYP11B2*	c.1157T>C	Corticosteronemethyl oxidase Type 2 deficiency Corticosteronemethyl oxidase type1deficiency
rs6256	GT	samesense	*PTH*	c.247C>A	Primary hyperpara thyroidism
rs713598	CG	missense	*TAS2R38*	c.145G>C	Phenylthiocarbamide tasting
rs854560	AT	missense	*PON1*	c.163T>A	Coronary artery disease, Microvascular complications of diabetes

A trio analysis (father, mother, son-proband) was performed to identify candidate genes responsible for morbid obesity. The information obtained after sequencing was filtered. First, all DNA changes previously described in databases of apparently healthy people, as well as synonymous nucleotide substitutions in DNA, were excluded. Then, mutations in the genes most likely to have functional and clinical significance were identified.

For data analysis, taking into account the types of all known obesity mutations, an initial selection of polymorphisms was performed, including only those variants that are functional, that is: 1) «stoploss/stopgain» variants, resulting in the loss or addition of a stop codon, respectively; 2) insertions and deletions in the reading frame; 3) non-synonymous variants; and finally 4) splice site variants. The second filtering stage included only those variants having at least one «D score» according to PolyPhen2, SIFT. The PolyPhen2 and SIFT programs are used to predict the pathogenicity of variants. SIFT predicts whether an amino acid substitution affects the function of a protein, its score ranges from 0 to 1. The lower the score, the more likely it is that the SNP has a damaging effect, that is, SIFT assumes that substitutions with scores less than 0.05 are damaging. PolyPhen predicts the possible effect of amino acid substitution on protein structure and function, and is used in the evaluation of rare alleles potentially involved in complex phenotypes: a score between 0.957 and 1 indicates a possible damaging effect (probably damaging (D)), a score between 0.453 and 0.956 indicates a likely damaging effect (possibly damaging (P)), and a score from 0 to 0.452 indicates a moderate or favorable effect (benign (B)). The third filtering stage included only variants with a score greater than 0.2 according to phastCons100way. This indicator (phastConservation score) is calculated based on multiple alignments of 100 vertebrate (including human) genomes. The higher the index, the more conservative the studied site is. Variants with a frequency of less than 1%, according to the 1000 Genome Project database, passed the fourth filtering stage (Table 3).

Table 3. Candidate gene variants for three genetic models according to filtration stages

Filtration	Recessive model	*De novo*	*Compound heterozygote*
Raw data	1591	169	3312
Filter 1	347	38	360
Filter 2	108	23	118
Filter 3	67	17	45
Filter 4	1	15	6

Three genetic models were used: recessive, *de novo* and *compound heterozygote* models. In the recessive model, variants found in the homozygous state in the patient and in the heterozygous state in the parents (not having the disease under study) were considered as candidates. In the *de novo* model, a *variant* was *considered* as candidate if it was *present* only in the patient, but not in the parents. And in *compound heterozygote* models, all pairs of variants composed of polymorphisms located in the same gene in the heterozygous state in the patient were considered as candidates, while in the parents the heterozygous state was observed for only one variant.

As can be seen from Table 3, as a result of the analysis, after passing through four stages of filtration, one variant was identified according to the recessive model, fifteen variants were identified according to the *de novo* model, and six variants were identified according to the *compound heterozygote* model.

In the recessive model, one variant in the *KIAA1671* gene is identified as a candidate (Table 4).

Table 4. Candidate gene variants selected by recessive model

Chromosome	Reference sequence/ position	Alleles	Function	Gene	DNA change	Protein change	SIFT	Polyphen
22	rs116253946/ 25425282	A/G	missense	*KIAA1671*	c.1316A>G	p.Lys 439Arg	T	D
Notes: 1 T-tolerated; 2 D-damaging								

As shown in Table 4, the identified polymorphism rs116253946 leads to a missense mutation that results in a lysine to arginine substitution at codon 439. The amino acid substitution likely has a damaging effect on protein structure and function according to Polyphen. In the *de novo* model, 15 polymorphisms located in 10 different genes were identified (Table 5).

Table 5. Candidate gene variants selected by de novo model

Chromosome	Reference sequence/ position	Alleles	Function	Gene	DNA change	Protein change	SIFT	Polyphen
1	rs11260920/ 17083816	G/A	nonsense	*MST1L*	c.1981 C>T	p.Arg 661*	T	-
1	rs199844027/ 243328110	G/A	missense	*CEP170*	c.3152 C>T	p.Pro 1051Leu	T	P,B
1	rs147489167/ 248737664	C/T	missense	*OR2T34*	c.395 G>A	p.Cys132 Tyr	D	D
2	-/95542482	C/A	missense	*TEKT4*	c.1276 C>A	p.Pro426 Thr	T	D
6	-/31597469	A/C	missense	*PRRC2A*	c.2101 A>C	p.Lys701 Gln	D	D
6	rs139583918/ 32497910	T/C	missense	*HLA-DRB5*	c.92A >G	p.Asp31 Gly	D	D
6	rs146966122/ 32497913	C/A	missense	*HLA-DRB5*	c.89 G>T	p.Gly30 Val	D	P
7	rs77708061/ 16676082	T/A	splicing	*ANKMY2*	c.68-2 A>T	-	-	-

19	-/55351014	G/C	missense	*KIR2DS4*	c.502 G>C	p.Ala168 Pro	-	-
19	-/55351098	G/A	missense	*KIR2DS4*	c.586 G>A	p.Val196 Met	-	-
19	-/55351105	T/A	missense	*KIR2DS4*	c.593 T>A	p.Leu198 His	-	-
19	-/55351111	C/T	missense	*KIR2DS4*	c.599 C>T	p.Thr200 Met	-	-
19	-/55351132	T/C	missense	*KIR2DS4*	c.620 T>C	p.Ile207 Thr	-	-
22	rs113026558/ 21822665	T/C	missense	*TMEM191C*	c.494 T>C	p.Leu165 Pro	-	-
22	rs199679407/ 50878419	C/T	missense	*PPP6R2*	c.2315 C>T	p.Pro772 Leu	D	D,D,D
Notes: 1 T-tolerated; 2 B-benign; 3 D-damaging; 4 P-possibly damaging; 5 – no data								

According to the Table 5, all identified polymorphisms appear as missense mutations, except for the rs77708061 polymorphism, where the nucleotide base is replaced in the splicing region. All polymorphisms have a damaging effect on protein function, except for rs199844027, which has a moderate/favorable effect. For 7 out of 15 polymorphisms, there is no data on the digital reference sequence. In the *compound heterozygote* model, 6 polymorphisms, located in *FRG1* and *MST1L* genes, were identified (Table 6).

Table 6. Candidate gene variants selected by *compound heterozygote* model

Chro mo some	Reference sequence/ position	Alleles	Function	Gene	DNA change	Protein change	SIFT	Poly phen
4	rs114792270/ 190862176	C/A	nonsense	*FRG1*	c.12C>A	p.Tyr4*	T	-
4	rs199625825/ 190883075	C/T	missense	*FRG1*	c.728C>T	p.Thr243 Met	T	P
4	rs200322789/ 190876192	A/C	missense/ splicing	*FRG1*	c.318A>C	p.Arg106 Ser	-	P
1	-/17083767	A/G	missense	*MST1L*	c.2030T>C	p.Val677 Ala	T	P, B
1	rs201784242/ 17085006	C/T	missense	*MST1L*	c.1469G>A	p.Gly490 Glu	T	D, D
1	rs2297532/ 17085791	G/A	missense	*MST1L*	c.1030C>T	p.Arg344 Cys	D	D
Notes: 1 T-tolerated; B-benign; D-damaging; 4 P-possibly damaging; 5 – no data								

Polymorphisms in the *FRG1* and *MST1L* genes appear as missense mutations, except for the rs114792270 polymorphism, for which a nonsense mutation is observed. As a result of the replacement of the nucleotide base at position c.12C>A, a stop codon appears. However, according to SIFT, amino acid substitution has a tolerated effect on protein function.

DISCUSSION

In all samples, the polymorphism rs1055138 was detected. The polymorphism rs1055138 is located in the *CYP4V2* gene. This gene encodes a member of the cytochrome P450 hemi-thiolate protein superfamily, which are involved in the oxidation of various substrates in metabolic pathways. It is involved in the metabolism of fatty acid precursors. Mutations in this gene result in Bietti crystalline chorioretinal dystrophy. Crystalline dystrophy is a rare autosomal recessive disorder and seems to be more common in people of Asian origin. It should also be noted that some identified pathogenic polymorphisms are associated with diseases such as diastolic hypertension (rs11739136 in the *KCNMB1* gene), Bardet-Biedl syndrome (rs4784677 in the *BBS2* gene), obesity (rs2282440 in the *SDC3* gene), which lead to obesity and increased BMI [8].

The trio analysis revealed a number of potential candidate genes predisposing to morbid obesity. In the recessive model, one variant (rs116253946) in the *KIAA1671* gene was identified as a candidate. To date, no specific function has been identified for the *KIAA1671* gene, which is located on chromosome 22q11.23. According to the NCBI database, the *KIAA1671* gene has an effect on endothelial cells of blood vessels [9, 10]. Interestingly, polymorphism (rs6004419) in the *KIAA1671* gene associated with the risk of developing type 2 diabetes with cataracts in the Taiwanese population [11].

In the *de novo* model, 15 polymorphisms located in 10 different genes were identified. The *MST1L* gene encodes a macrophage stimulating 1-like protein. A recent study showed that the *MST1L* gene highly upregulated in the diabetic beta cell and not only induces its death, but also directly impairs

insulin secretion by stimulating proteasomal degradation of the beta-cell transcription factor, pancreatic and duodenal homeobox 1 (PDX1). PDX1 is critical in insulin secretion [12]. The *CEP170* (centrosomal protein 170) gene encodes a protein that is a component of the centrosome in animal cells. The centrosome performs various functions in the cell. Two centrosomes organize the bipolar spindle of division necessary for splitting the two sister chromosomes and breeding them to opposite poles of the cell. Centrosomal proteins participate in the microtubule growth [13].

The *OR2T34* (olfactory receptor family 2 subfamily T member 34) gene encodes functional proteins of olfactory receptors. Olfactory receptors interact with odor molecules in the nose to initiate a neuronal response that triggers the perception of smell. Olfactory receptor proteins belong to the family of G-protein-coupled receptors. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitters and hormone receptors and are responsible for recognition and G protein-mediated transduction of odorous signals. The family of olfactory receptor genes is the largest in the genome [10]. The TEKT4 gene encodes tecton 4, a constitutive microtubule protein in cilia, flagella, basal bodies, and centrioles [14].

The *PRRCA2A* gene encodes for a proline rich protein (proline rich coiled-coil 2A) also known as *BAT2* gene. The BAT1-BAT5 gene cluster is localized near the TNF alpha and TNF beta genes. These genes are located in the human major histocompatibility complex class III region (major histocompatibility complex, MHC). *MHC* is a group of genes and the cell surface antigens they encode that play a critical role in foreign recognition and the development of an immune response. This gene has microsatellite repeats that are associated with insulin-dependent diabetes mellitus and may be associated with the inflammatory process of pancreatic beta cells in the development of insulin-dependent diabetes mellitus. The *PRRCA2A* gene also a candidate gene for the development of rheumatoid arthritis. The *HLA-DRB5* gene belongs to class II genes of the major histocompatibility complex. The human major histocompatibility complex is called Human Leukocyte Antigens (HLA). *HLA-DRB5* belongs to the HLA class II beta chain paralogues. This class II molecule is a heterodimer consisting of an alpha (DRA) and a beta (DRB) chains anchored in the membrane [10]. According to a study by Zhao et al., the *HLA-DRB5* gene affects the risk of developing type 1 diabetes and pancreatic islet autoantibodies [15].

The function of the *ANKMY2* gene (ankyrin repeat and MYND domain containing 2) is poorly characterized. Reduced or overexpression of *ANKMY2* gene in mouse embryonic fibroblasts resulted in down- and up-regulation of Shh (Sonic hedgehog) signaling, respectively [16]. The sonic hedgehog protein is considered to be the most important determinant of the neurogenic pathway of germinal ectoderm cell development. Shh begins to be expressed by chordal cells even before the onset of active growth of the neural folds, and is involved in the primary determination of neural plate cells [17].

The *KIR2DS4* (killer cell immunoglobulin like receptor, two Ig domains and short cytoplasmic tail 4) gene belongs to the *KIR* family of genes encoding the killer cell immunoglobulin-like receptor (KIR). They play an important role in the regulation of the functional activity of natural killers. KIR receptors are transmembrane glycoproteins with two or three extracellular immunoglobulin-like domains (KIR2D and KIR3D, respectively) and a long (L) or short (S) cytoplasmic region. Receptors with L-regions conduct an inhibitory signal, and those with S-regions carry out an activating one. KIR receptors, interacting with HLA class I antigens and transmitting an activating or inhibitory signal, are involved in the regulation of the functional activity of natural killers, the most important component of innate immunity [18]. Thus, according to a study by Sanjeevi et al., the interaction of KIR receptors with HLA-C ligands is significant and some combinations contribute to susceptibility or protection against type 1 diabetes [19]. The *TMEM191C* encodes the 191c transmembrane protein; the role of this gene is not fully understood. The *PPP6R2* encodes protein phosphatase regulatory subunits that modulate catalytic activity, substrate recognition, and determination of intracellular localization of the holoenzyme [10].

In the *compound heterozygote* model, six polymorphisms were identified located in the two *FRG1* and *MST1L* genes. The human *FRG1* gene is located on chromosome 4q35 and is considered as a candidate locus in humeroscapular-facial myopathy, but its function has not been elucidated. The FRG1 protein has a conserved structure and contains a lipocalin-like motif, which indicates its possible transport functions [20].

CONCLUSION

Thus, the results of the studies showed that 12 genes identified by whole exome sequencing may be involved in the development of morbid obesity, and are grouped into three main pathophysiological pathways (immune response: *MST1L, HLA-DRB5, PRRC2A, KIR2DS4*; cell division and structure : *CEP170, TEKT4*; neuronal response: *OR2T34, ANKMY2*) or their function remains unknown to date (*KIAA1671, TMEM191C, PPP6R2, FRG1*). This work presents the preliminary results of the study. In the future, it is planned to validate the identified mutations, as well as an additional study of new families in the proband of which there is morbid obesity.

FUNDING

This study was supported by the Ministry of Education and Science of the Republic of Kazakhstan, grant number AP09057948

LITERATURE

1. Obesity and overweight / World Health Organization // http://www.who.int/mediacentre/ru/. 12.08.2022.

2. Bonnefond, A., Philippe, J., Durand, E., Muller, J., Saeed, S. et al. Highly sensitive diagnosis of 43 monogenic forms of diabetes or obesity through one-step PCR-based enrichment in combination with next-generation sequencing // Diabetes Care. – 2014. – Vol. 37, №2. - P. 460-467.

3. Miller, S., Dykes, D., Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells // Nucleic Acids Res. – 1988. – Vol. 16. – P. 1215.

4. Li, H., Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform // Bioinformatics. – 2010. – Vol. 26. – P. 589-595.

5. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data // Genome Res. – 2010. – Vol. 20. – P. 1297-1303.

6. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. et al. 1000 Genome project data processing subgroup. The Sequence Alignment/Map format and SAMtools // Bioinformatics. – 2009. – Vol. 25. – P. 2078-2087.

7. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 // Fly (Austin). – 2012. – Vol. 6, №2. – P. 80-92.

8. SNPedia // www.snpedia.com/index.php/SNPedia. 28.08.2022.

9. Zhou, J., Chen, C., Liu, S., Zhou, W., Du, J., Jiang, Y., Dai, J., Jin, G., Ma, H., Hu, Z., Chen, J., Shen, H. Potential functional variants of KIAA genes are associated with breast cancer risk in a case control study // Ann. Transl. Med. – 2021. – Vol. 9(7). – P.549.

10. National Center for Biotechnology Information // www.ncbi.nlm.nih.gov/. 20.07.2022.

11. Lin, H., Huang, Y., Lin, J., Liao, W., Wu, J. et al. Novel susceptibility genes associated with diabetic cataract in a Taiwanese population // Ophthalmic Genet. – 2013. – Vol. 34, №1-2. – P. 35-42.

12. Ardestani, A., Maedler, K. *MST1*: a promising therapeutic target to restore functional beta cell mass in diabetes // Diabetologia. – 2016. - Vol. 59, №9. – P. 1843-1849.

13. Stcherbic, V.V., Buchatsky, L.P. Centrosome structure and Clifford algebra // Problems of ecological and medical genetics and clinical immunology. – 2013. – №. 6. – С. 44-60.

14. Zheng, Z., Zhou, X., Cai, Y., Chen, E., Zhang, X., Wang, O., Wang, Q., Liu, H. TEKT4 promotes papillary thyroid cancer cell proliferation, colony formation, and metastasis through activating PI3K/Akt pathway// Endocr. Pathol. – 2018. – Vol. 29 (4). – P. 310-316.

15. Zhao, L., Alshiekh, S., Zhao, M., Carlsson, A., Larsson, H. et al. DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes // Diabetes. – 2016. - Vol. 65, №3. – P. 710-718.

16. Saita, S., Shirane, M., Ishitani, T., Shimizu, N., Nakayama, K. Role of the *ANKMY2-FKBP38* Axis in regulation of the sonic hedgehog (Shh) signaling pathway // J. Biol. Chem. – 2014. – Vol. 298, №37 – P. 25639-25654.

17. Somatilaka, B., Hwang, S., Palicharla, V., White, K., Badgandi, H., Shelton, J., Mukhopadhyay, S. Ankmy2 prevents smoothened-independent hyperactivation of the hedgehog pathway via cilia-regulated adenylyl cyclase signaling // Dev. Cell. – 2020. – Vol. 54(6). – P. 710-726.

18. Castrillon, M., Marin, N., Karduss-Urueta, A., Velasquez, S., Alvarez, C. Killer-cell immunoglobulin-like receptor diversity in an admixed south American population // Cells. – 2022. – Vol. 11(18). – P. 2776.

19. Sanjeevi, S., Sun, C., Kanungo, A., Sanjeevi, C. Killer immunoglobulin receptor genes and their HLA-C ligand are associated with type 1 diabetes in an Eastern Indian population // Diabet. Med. – 2016. - Vol. 33, №1. – P. 91-96.

20. Grewal, P., Todd, L., van der Maarel, S., Frantz, R., Hewitt, E. *FRG1*, a gene in the FSH muscular dystrophy region of human chromosome 4q35, is highly conserved in vertebrates and invertebrates // Gene N. – 1998. - Vol. 216. – P. 13-19.

REFERENCES

1. Obesity and overweight / World Health Organization // http://www.who.int/mediacentre/ru/. 12.08.2022.

2. Bonnefond, A., Philippe, J., Durand, E., Muller, J., Saeed, S. et al. Highly sensitive diagnosis of 43 monogenic forms of diabetes or obesity through one-step PCR-based enrichment in combination with next-generation sequencing // Diabetes Care. – 2014. – Vol. 37, №2. - P. 460-467.

3. Miller, S., Dykes, D., Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells // Nucleic Acids Res. – 1988. – Vol. 16. – P. 1215.

4. Li, H., Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform // Bioinformatics. – 2010. – Vol. 26. – P. 589-595.

5. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data // Genome Res. – 2010. – Vol. 20. – P. 1297-1303.

6. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. et al. 1000 Genome project data processing subgroup. The Sequence Alignment/Map format and SAMtools // Bioinformatics. – 2009. – Vol. 25. – P. 2078-2087.

7. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 // Fly (Austin). – 2012. – Vol. 6, №2. – P. 80-92.

8. SNPedia // www.snpedia.com/index.php/SNPedia. 28.08.2022.

9. Zhou, J., Chen, C., Liu, S., Zhou, W., Du, J., Jiang, Y., Dai, J., Jin, G., Ma, H., Hu, Z., Chen, J., Shen, H. Potential functional variants of KIAA genes are associated with breast cancer risk in a case control study // Ann. Transl. Med. – 2021. – Vol. 9(7). – P.549.

10. National Center for Biotechnology Information // www.ncbi.nlm.nih.gov/. 20.07.2022.

11. Lin, H., Huang, Y., Lin, J., Liao, W., Wu, J. et al. Novel susceptibility genes associated with diabetic cataract in a Taiwanese population // Ophthalmic Genet. – 2013. – Vol. 34, №1-2. – P. 35-42.

12. Ardestani, A., Maedler, K. *MST1*: a promising therapeutic target to restore functional beta cell mass in diabetes // Diabetologia. – 2016. - Vol. 59, №9. – P. 1843-1849.

13. Stcherbic, V.V., Buchatsky, L.P. Centrosome structure and Clifford algebra // Problems of ecological and medical genetics and clinical immunology. – 2013. – №. 6. – С. 44-60.

14. Zheng, Z., Zhou, X., Cai, Y., Chen, E., Zhang, X.,

Wang, O., Wang, Q., Liu, H. TEKT4 promotes papillary thyroid cancer cell proliferation, colony formation, and metastasis through activating PI3K/Akt pathway// Endocr. Pathol. – 2018. – Vol. 29 (4). – P. 310-316.

15. Zhao, L., Alshiekh, S., Zhao, M., Carlsson, A., Larsson, H. et al. DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes // Diabetes. – 2016. - Vol. 65, №3. – P. 710-718.

16. Saita, S., Shirane, M., Ishitani, T., Shimizu, N., Nakayama, K. Role of the *ANKMY2-FKBP38* Axis in regulation of the sonic hedgehog (Shh) signaling pathway // J. Biol. Chem. – 2014. – Vol. 298, №37 – P. 25639-25654.

17. Somatilaka, B., Hwang, S., Palicharla, V., White, K., Badgandi, H., Shelton, J., Mukhopadhyay, S. Ankmy2 prevents smoothened-independent hyperactivation of the hedgehog pathway via cilia-regulated adenylyl cyclase signaling // Dev. Cell. – 2020. – Vol. 54(6). – P. 710-726.

18. Castrillon, M., Marin, N., Karduss-Urueta, A., Velasquez, S., Alvarez, C. Killer-cell immunoglobulin-like receptor diversity in an admixed south American population // Cells. – 2022. – Vol. 11(18). – P. 2776.

19. Sanjeevi, S., Sun, C., Kanungo, A., Sanjeevi, C. Killer immunoglobulin receptor genes and their HLA-C ligand are associated with type 1 diabetes in an Eastern Indian population // Diabet. Med. – 2016. - Vol. 33, №1. – P. 91-96.

20. Grewal, P., Todd, L., van der Maarel, S., Frantz, R., Hewitt, E. *FRG1*, a gene in the FSH muscular dystrophy region of human chromosome 4q35, is highly conserved in vertebrates and invertebrates // Gene N. – 1998. - Vol. 216. – P. 13-19.

ПОЛНОЭКЗОМНОЕ СЕКВЕНИРОВАНИЕ ПАЦИЕНТА С МОРБИДНЫМ ОЖИРЕНИЕМ: ТРИО АНАЛИЗ, ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Сихаева Н.[1,*], Накыш А.[2], Утупов Т.[1], Жолдыбаева Е.[1]

[1]Национальный центр биотехнологии, Коргалджинское шоссе 13/5, Астана, 010000, Казахстан
[2]Медицинский университет Астана, ул. Бейбитшилик 49а, Астана, 010000, Казахстан
[*]ksnurgul@gmail.com

АБСТРАКТ

Морбидное ожирение – тяжелая форма ожирения, приводящая к многочисленным сердечно-сосудистым, метаболическим и онкологическим заболеваниям, а также к повышенной смертности. Полноэкзомное секвенирование является эффективным инструментом для изучения более экстремальных форм заболеваний, таких как морбидное ожирение. Цель исследования — выявить гены-кандидаты, участвующие в развитии морбидного ожирения с помощью полноэкзомного секвенирования (WES).

В исследовании участвовала семья, в пробанде которой наблюдалось морбидное ожирение. Проведено полноэкзомное секвенирование членов семьи (отец, мать, сын).

Трио анализ выявил ряд потенциальных генов-кандидатов участвующие в развитии морбидного ожирения. Три генетических моделей были использованы: рецессивная , *de novo* и *compound heterozygote* модели. В рецессивной модели, один вариант (rs116253946) в гене *KIAA1671* был выявлен как кандидат. В *de novo* модели, были выявлены 15 полиморфизмов расположенные в 10 различных генах. И в *compound heterozygote* модели были выявлены 6 полиморфизмов расположенные в двух *FRG1* и *MST1L* генах.

Данная работа представляет предварительные результаты исследования. 12 генов, выявленные с помощью WES, могут быть вовлечены в развитие морбидного ожирения, и могут быть группированы в три основные патофизиологические пути (иммунный ответ: *MST1L, HLA-DRB5, PRRC2A, KIR2DS4*; структура и деление клетки: *CEP170, TEKT4*; нейрональный ответ: *OR2T34, ANKMY2*) или их функция еще неизвестна на сегодняшний день (*KIAA1671, TMEM191C, PPP6R2, FRG1*).

Ключевые слова: морбидное ожирение, ген, патогенная мутация, полноэкзомное секвенирование

МОРБИДТІК СЕМІЗДІГІ БАР ПАЦИЕНТТІ ТОЛЫҚ ЭКЗОМДЫҚ СЕКВЕНИРЛЕУ: ТРИО АНАЛИЗ, БАСТАПҚЫ НӘТИЖЕЛЕР

Сихаева Н.[1,*], Накыш А.[2], Утупов Т.[1], Жолдыбаева Е.[1]

[1]Ұлттық биотехнология орталығы, Қорғалжын тасжолы13/5, Астана, 010000, Қазақстан
[2]Астана Медициналық Университеті, Бейбітшілік к. 49а, Астана, 010000, Қазақстан
[*]ksnurgul@gmail.com

ТҮЙІН

Морбидтік семіздік - көптеген жүрек-қан тамырлары, зат алмасу және онкологиялық ауруларға, сондай-ақ өлімнің артуына әкелетін семіздіктің ауыр түрі. Толық экзомдық секвенирлеу морбидтік семіздік сияқты аурудың экстремалды түрлерін зерттеудің тиімді құралы болып табылады. Зерттеудің мақсаты - толық экзомдық секвенирлеу (WES) көмегімен морбидтік семіздіктің дамуына қатысатын кандидат-гендерді анықтау.

Зерттеуге пробандасы морбидтің семіздікпен ауыратын отбасы қатысты. Отбасы мүшелеріне (әкесі, анасы, баласы) толық экзомдық секвенирлеу жүргізілді.

Трио анализ морбидтік семіздіктің дамуына қатысатын бірқатар ықтимал кандидат-гендерді анықтады. Үш генетикалық модель қолданылды: рецессивті, *de novo* және *compound heterozygote* модельдер. Рецессивті моделінде, *KIAA1671* генінде бір вариант (rs116253946) анықталды. *De novo* моделінде 10 түрлі генде орналасқан 15 полиморфизм анықталды. *Compound heterozygote* моделінде екі *FRG1* және *MST1L* генінде орналасқан 6 полиморфизм анықталды.

Бұл мақалада зерттеудің алғашқы нәтижелері берілген. WES анықтаған 12 ген морбидтің семіздіктің дамуына қатысуы мүмкін және үш негізгі патофизиологиялық жолға топтастырылуы мүмкін (иммундық жауап: MST1L, HLA-DRB5, PRRC2A, KIR2DS4; жасуша құрылымы мен бөлінуі: CEP170, TEKT4; нейрондық жауап : OR2T34, ANKMY2) немесе олардың қызметі әлі белгісіз (KIAA1671, TMEM191C, PPP6R2, FRG1).

Негізгі сөздер: морбидтік семіздік, ген, патогендік мутация, толық экзомдық секвенирлеу