



Dear Chairman, dear participants, dear colleagues and friends, I would like to congratulate you with NASHIM 30th Anniversary !

Thank you very much for inviting me to present on this 30th Anniversary Symposium!



I had a chance to participate in the NASHIM program in 2005, I got medical training in the field of Hematology and Molecular Epidemiology.

I am very grateful to NASHIM for this opportunity to acquaint with the records and guidance of treatment for atomic bomb victims and radiation disorder research results in the different departments Nagasaki University - Professor Yamashita, Professor Tomonaga, Professor Takamura and Sekine shared

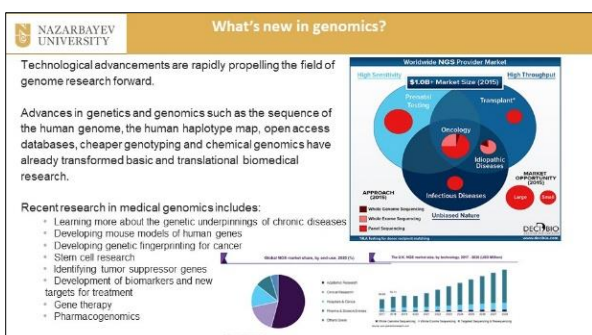
results from their departments.

Also in this photo you see that we met with different people, with different doctors in hospitals and clinics dedicated to healthcare system in Japan, in Nagasaki particularly.

Nagasaki people have a very special understanding with a word peace and apply many efforts not to repeat Hiroshima and Nagasaki like in 1945.

And NASHIM made a lot of great impact in this field.

Again, congratulations on the 30th anniversary!"



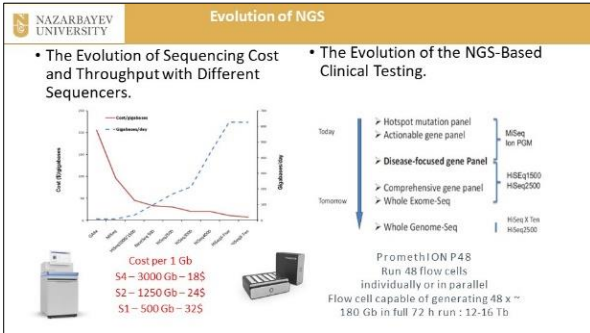
Technological advancements are rapidly propelling the field of genome research forward.

Advances in genetics and genomics such as the sequence of the human genome, the human haplotype map, open access databases, cheaper genotyping and chemical genomics have already transformed basic and translational research. Recent research in biomedical genomics includes almost all fields of

medicine.

And we can see increasing whole genome sequence in prenatal testing, oncology, transplantation, infectious diseases, etc.

And if only before it was all academic research, nowadays genomic research also increases in big pharma in private companies and in different practical purposes.



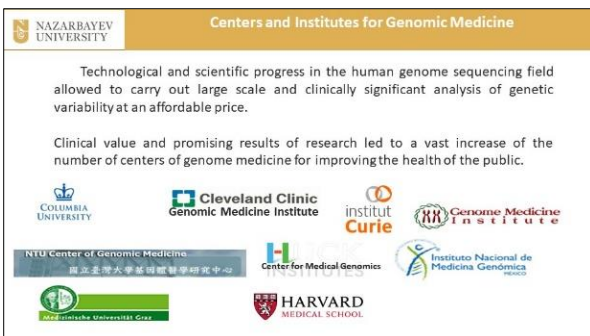
In the last decade, with development of sequencing platforms, cost per one gigabase decreases.

If whole genome completed in 2003 was almost 3 billion dollars, nowadays we can do whole genome sequencing in \$1,000 by using X-Ten and NovaSeq.

NovaSeq 6000 was launched in 2017, cost per gigabase using S4 flow cells, it will be almost \$18. And during 48 hours you can get 3,000 gigabases of information.

The Evolution of the NGS-Based Clinical Testing is also very promising. The hotspot mutation panels, actionable gene panels were used before by MiSeq and Ion Torrent machine, nowadays we can do whole exome-sequencing by using HiSeq series of platforms, and whole genome sequencing by using HiSeq X-Ten and NovaSeq.

Also PromethION 48, this is the highest throughput platform of Oxford Nanopore Technologies, which runs 48 flow cells individually or in parallel and gives the information up to 12 to 16 terabases in 72 hours, which generates along DNA fragments also can be used for large fragment DNA changes.



Technological and scientific progress in the human genome sequencing field allowed to carry out large scale and clinically significant analysis of genetic variability at an affordable price. NGS technologies are already an integral part of medical research in many countries with most major centers having access to genomic sequencing expertise.

In the slide you can see all the few centers and universities with very strong genomic platforms. Most of them are our collaborators in our current ongoing research project.



In this slide, I would like to show photos of Nazarbayev University campus, and the white arrows indicate where our Center for Life Sciences is located.

And this picture shows the main building of Medical School of our University and the University Medical Center.

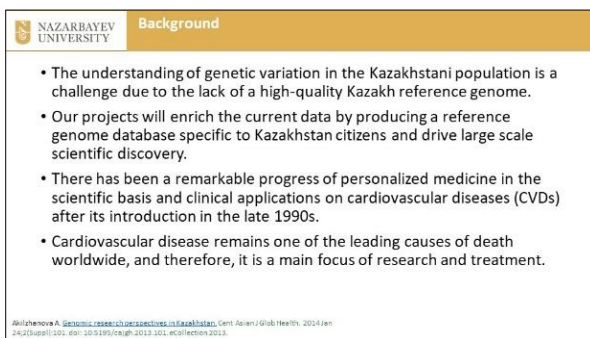


The Center for Life Sciences was organized in 2010, and in 2015 it was merged with Center for Energy and Advanced Materials Science to create the National Laboratory of Astana. Nowadays, we have eight laboratories in the Center for Life Sciences and four laboratories in the Center for Energy and Advanced Materials Science.

Laboratory of Genomic and Personalized Medicine and Laboratory of Bioinformatics and Systems Biology, we work

closely together in the frame of our joint project.

As well we study and collaborate with other laboratories in our big projects, interdisciplinary projects.



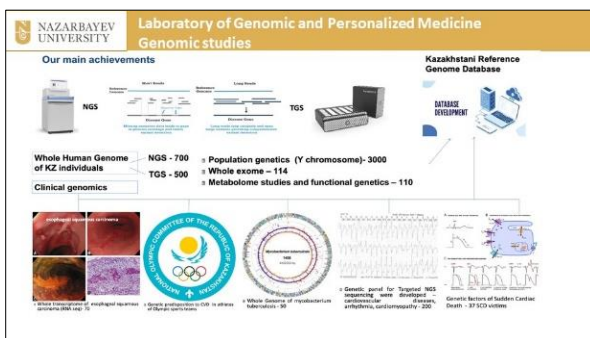
Understanding of genetic variation in the Kazakhstani population is a challenge, due to the lack of a high-quality Kazakh reference genome.

Our projects will enrich the current data by producing a reference genome database specific to Kazakhstan citizens and drive large scale scientific discovery.

The objective of the whole genome project is to promote health management and position of Nazarbayev University's omics-

driven research and innovation hub.

There has been a remarkable progress in personalized medicine in the field of clinical application on cardiovascular disorders, because its introduction was in the late 1990s and nowadays the researches are still increasing. Cardiovascular disease remains as one of the leading causes of mortality and morbidity in Kazakhstan, that's why it's one of our priority research fields in our projects.



Our main achievements in laboratory and main directions for our projects are to generate comprehensive genomic data by using the most advanced sequencing technologies, both using short DNA fragments sequencing and long DNA fragments sequencing.

Utilization of these advanced bio-informatics tools to detect and analyze genomic data in creation of good bioinformatics class and storage systems, which promotes our large-scale whole

genome projects.

Implement semantic technology to extract and convert genome data into useful information for clinics.

And in parallel, we are making clinical genomic projects including oncological projects, cardiovascular disorders in athletes of Olympic games in Kazakhstan, different predisposition to sudden cardiac death in different diseases, and also we study whole genomes of bacteria and viruses which predispose and have a high impact now during the coronavirus epidemiology.

NAZARBAYEV UNIVERSITY Aims and objectives of projects

Aim
to understand the relationship between genomics, disease and wellness in a way that is specific to the Kazakhstani population.

Objectives and Expected results

- To create Kazakhstani reference genome database to assist in the health of the population.
- To equip scientists, physicians and other healthcare practitioners with high quality information and knowledge.
- To enable advanced diagnosis and treatment options based on genomic data
- To deliver personalized and prevention programs tailored to an individual's unique genetic makeup.
- To build a capacity to employ large scale sequencing and massive data analysis
- To raise intellectual potential
- To introduce new technologies

Overall aim of our project is to understand the relationship between genomics, disease and wellness in a way that is specific to the Kazakhstani population.

We expect to create Kazakhstani reference genome database to assist in the health care of the population of Kazakhstan.

We would like to equip scientists, physicians, and other healthcare professionals with high quality information and knowledge, and enable advanced diagnosis and treatment

options based on genomic data, to deliver personalized and prevention programs tailored to an individual's unique genetic makeup.

And of course, to build a capacity to employ large scale sequencing and bioinformatic analysis and raise intellectual potential, not only of our University but Kazakhstani as well.

And of course to introduce new technologies.

NAZARBAYEV UNIVERSITY Study participants

Recruitment (Field work trips)
Adults, 18+
Information (age, gender, lifestyle) and Medical history
Questionnaire, anthropometry
Blood samples collection

Population
NIA, NU - 87
Kostanay region - Torgay -151
Karagandy region - Temirtau -119
Zhamansay -112
Aimola region, Rodina village -50
Uzbeks - South KZ -78

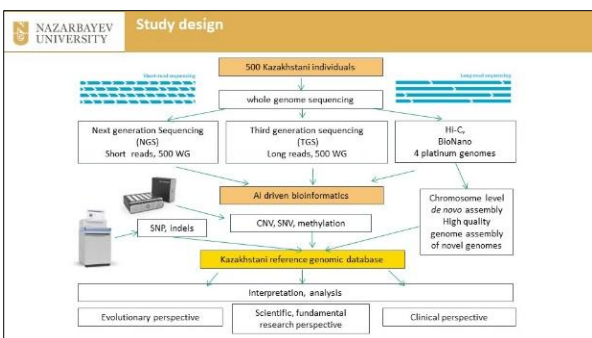
	Asian				Caucasian			Total
	KZ	UZ	UWG	other	RU	others		
Male	129	54	3	1	30	16	234	
Female	196	16	1	3	33	17	266	
Total		403			97		500	

Cases / Hospitals/Clinics
Cardio, Metabolic syndrome

	Male	Female	Total
Cardio	53	16	69
Met syn	20	45	65
			134

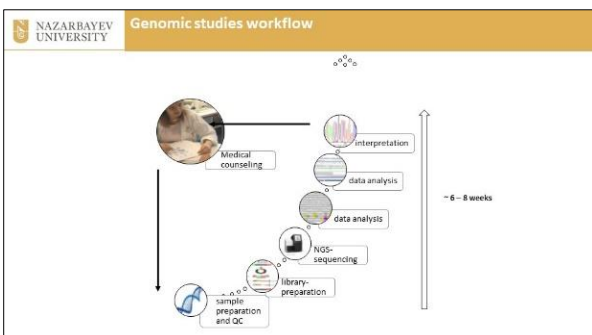
Study participants in our projects usually make field works and recruit participants from different parts of Kazakhstan.

Here you can see one of our field works where we make cohort of Central Asian Kazakhstan regions up to 500 people including Asian and Caucasian nationalities living in our regions, here you can see photos, as well we are working with hospitals and clinics to create our clinical cohorts of patients - cardio syndromes, metabolic syndromes, diabetes, and etc.



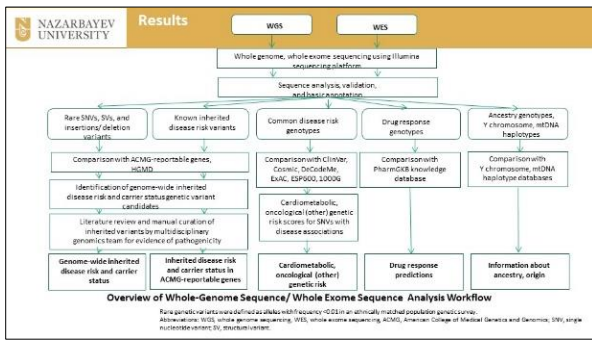
Overall study design includes whole genome sequencing, using next generation sequencing and third generation sequencing, and all data generated by these technologies will be analyzed and form Kazakhstani reference genomic database.

After interpretation of analysis, we can do evolutionary perspective research, scientific, fundamental research perspective, and clinical perspective.



Genomic studies workflow includes several steps, in here in the slide you can see sample preparation, library preparation, sequencing itself and data analysis, it also includes medical counseling, which is the most important in our studies, where we give all interpreted data to our physicians and they talk with the case patient and their families about diagnosis, treatment, and further management.

Usually it takes up to 6 or 8 weeks depending on diseases.



Overall workflow of analysis of whole genome data and whole exome sequencing data includes also analysis of rare SNVs, structural variations, insertion/deletion variants, known inherited disease risk variants, common disease risk genotypes, drug response genotypes, ancestry genotypes, y chromosome, mitochondrial DNA haplotypes.

While doing analysis of all of these parts, we use different databases that exist in the international book like Human Gene

Mutation Databases, the ACMG guidelines, comparison with ClinVar, Cosmic, DecodeMe and other databases, comparison with PharmGKB knowledge database where all pharmacogenetics results of the pharmacogenetics studies are included.

As a result, we can say genome-wide inherited disease risks or carrier status, inherited disease risk and carrier status in ACMG-reportable genes, cardiometabolic, oncological genetic risk, or maybe in patient with other complication disorders risk, drug response predictions, and ancestry genotypes of Y chromosome and mitochondrial DNA.

SNP summary		WE	WG	Total
Annotated with ANNOVAR: ENCODE gene	frameshift deletion	583	388	635
	frameshift insertion	338	209	362
	nonframeshift deletion	622	777	642
	nonframeshift insertion	322	91	326
	missense mutation	27,562	12,255	26,559
	stopgain	311	237	338
	stoploss	49	47	56
	synonymous SNV	27,521	15,455	26,459
	unknown	387	368	426
	Intronic	139,467	2,883,847	2,904,811
Splicing	241	238	306	
UTR	53,960	60,966	72,584	
UTR3	8,552	6,452	12,260	
UTR5	63	65	101	
UTR5/UTR3	385	2,745	3,622	
Upstream/Downstream	3,857	53,723	53,750	
Upstream	5,078	68,146	66,371	
Downstream	84,155	4,181,867	4,188,454	
Intergenic	5,897	43,370	45,956	
ncRNA	13,137	840,351	843,869	
ncRNA_exonic	28	249	265	
ncRNA_intronic	4,800	43,706	53,642	
ncRNA_UTR5	504	383	701	
ncRNA_UTR3	7	7	8	

As a result of our preliminary data of whole genome and whole exomes, we found many thousands of exonic variations and intronic variations.

And total is more than 635 frameshift deletions and frameshift insertions, etc., which will be divided by categories.

Kazakh Novel/Private SNPs		REF	ALT	WG1	WG2	WG3	WG4	WG5	WG6	WG7	REF	Gene	PP3	SIFT	SNP126	SNP138	ASN	ALL	KAZ	ASN-KAZ
chr11	1,264,808	T	G	2	2	2	NA	NA	2	NA	2	MUC6	0.88	0	rs941838	rs4341838	0.06	0.09	1	0.94
chr19	43,709,637	C	G	0	2	NA	2	NA	2	NA	2	PCGA	0	1	NA	rs14508885	0.06	0.11	0.8	0.74
chr14	105,260,228	G	A	2	2	NA	2	2	2	2	2	TRPA2	0.97	0.62	rs1983387	rs4985397	0.36	0.65	1	0.65
chr22	17,285,124	A	C	2	2	2	2	2	2	2	2	XORR	0	1	rs7579823	rs114556178	0.28	0.30	0.9266	0.85
chr7	142,481,376	A	A	2	2	2	NA	2	NA	PRSS2	NA	NA	NA	NA	rs73740210	0.37	0.39	1	0.60	
chr6	32,634,302	A	G	1	1	0	0	0	0	0	0	HLA-DQB1	0.021	1	rs1049082	rs1049082	0.77	0.67	0.1429	0.63
chr18	23,713,228	A	T	1	2	1	2	2	2	2	2	ELF2	0.11	0.29	rs1702954	rs1763934	0.19	0.33	0.7827	0.60
chr2	108,513,601	A	G	1	0	1	0	1	1	1	1	EDAR	0.656	0	rs3227780	rs3227780	0.88	0.29	0.2937	0.59
chr11	60,776,209	C	T	1	1	NA	NA	2	NA	CDS	0.641	0.04	rs11235963	rs11235963	0.16	0.35	0.75	0.59		
chr1	234,745,029	A	G	2	2	NA	2	NA	NA	IRF3BP2	0	0.8	rs7543893	rs7543893	0.91	0.62	0.875	0.57		
chr11	106,268,878	T	C	1	1	1	2	2	2	PRKCT	0	1	rs4446146	rs4446146	0.28	0.18	0.7887	0.56		
chr9	117,715	T	C	2	2	NA	2	2	NA	FOXO4	0	1	rs2492216	rs2492216	0.47	0.54	1	0.53		
chr16	58,078,165	G	A	2	1	NA	NA	2	NA	MMP15	0.092	0.5	rs3743583	rs3743583	0.95	0.22	0.875	0.53		
chr19	17,421,901	G	A	2	1	NA	2	NA	NA	GTF2F5	0.084	0.04	rs3743789	rs3743789	0.20	0.12	0.875	0.53		
chr12	66,346,100	A	G	0	1	0	1	0	1	TRIM84	0.021	NA	rs7878	rs7878	0.88	0.67	0.3371	0.52		
chr5	112,824,039	T	C	1	0	NA	0	NA	2	MCC	0	0.23	rs348942	rs348942	0.82	0.69	0.3	0.52		
chr2	108,878,190	G	A	1	0	0	1	0	1	SULT1C3	1	0	rs2219078	rs2219078	0.73	0.29	0.2143	0.52		
chr14	74,042,189	A	G	1	2	0	0	0	0	AC0212	0.021	0.69	rs1504	rs1504	0.8	0.42	0.2027	0.51		
chr11	46,390,165	A	T	1	2	2	2	2	2	LRPA	0.005	0.27	rs3316614	rs3316614	0.42	0.19	0.5038	0.51		
chr19	87,038,069	A	C	1	2	2	2	2	2	ZNF229	NA	0.59	rs10912444	rs10912444	0.42	0.64	0.5038	0.51		
chr6	51,878,956	C	G	2	0	NA	1	2	1	MICA	0.999	0.04	rs1051790	rs1051790	0.16	0.19	0.6667	0.51		
chr11	6,576,106	C	A	1	2	2	2	2	2	DNAH21	0.021	0.51	rs11016202	rs11016202	0.28	0.18	0.7397	0.51		

Also we found after analysis and comparison with different databases, Kazakh Novel or Private SNPs, which is the first SNPs functionally studied and confirmed in different, other studies.

Processed Databases

COSMIC
Catalogue of somatic mutations in cancer

Genes: 25660
Samples: 98172
Mutations: 1627878
Unlike variants: 1292597
Fusions: 10251
Genomic rearrangements: 7584

SNPedia

52 440 known SNPs
33 989 genotypes
1635 genes
291 medical conditions
182 medicines

ClinVar
An archive of medically relevant variants
<http://www.ncbi.nlm.nih.gov/clinvar/>

Total genes represented: **18827**
Total variations represented: **104237**

IGSR: The International Genome Sample Resource
Providing ongoing support for the 1000 Genomes Project data

Genome Asia 100K

ExAC Browser (Beta) | Exome Aggregation Consortium 60706 individuals exome data

We process different databases to analyze our genome data like Cosmic, SNPedia, ClinVar, Genome Asia, etc. for the analysis of Genome Aggregation Consortium.

Kazakh WG: Summary info						
rs #	SNP	REF. Gene	PP2	SFT	MAG	W01
r307325	r307325(G>T)	GLTPD1	NA	NA	2	25% decrease in autoresponsivity extra tasting ability?
r307327	r307327(C>T)	TAS2R38	0	0.35	4	1.3x risk
r4664028	r4664028(C>T)	SPPL5	NA	NA	1.5	with out for higher in diet
r1811218	r1811218(C>G)	FRS3	0	0.02	3	0.3x decreased risk for bladder cancer
r1112186	r1112186(T>C)	IRF102/RODD4	NA	NA	2	Higher risk cancer
r2910364	r2910364(C>G)	MIR1464	NA	NA	2.5	59 lighter hair
r2491292	r2491292(T>C)	G13	NA	NA	2	risk quiescence
r360399	r360399(G>C)	MTH9	0	1	1.1	causative. Ribavirin-induced anemia. Ribavirin-induced anemia during anti-hepatitis C virus therapy
r1121954	r1121954(A>G)	ITPA	0.792	0.32	3	

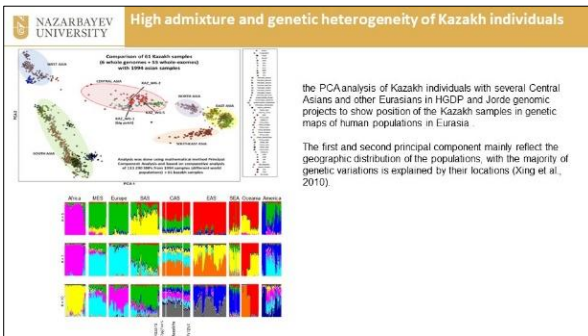
W02						
rs #	SNP	REF. Gene	PP2	SFT	MAG	Summary
r10754339	r10754339(A>G)	VTG1	NA	NA	1.3	1.3x increased risk of breast cancer
r4661461	r4661461(A>G)	HU100492700/TM021	NA	NA	1.5	1.5x increased risk for open angle glaucoma
r4663058	r4663058(A>C)	BAT2B	NA	NA	2	2x increased risk among Europeans for sudden cardiac death
r1919122	r1919122(C>T)	IRF102/RODD4	NA	NA	2	2-3x higher prostate cancer risk if routinely exposed to the pesticide tolfenox
r466306	r466306(G>C)	DNF3	NA	NA	2.5	Associated with Alcohol Dependence, associated with risk of dependence in the brain
r1818830	r1818830(C>C)	MET	NA	NA	2	2x risk of autism
r2237373	r2237373(T>C)	MET	NA	NA	3	reduced abilities needed for navigation and ability to recognize faces
r4491677	r4491677(G>T)	IRF102/RODD4	NA	NA	1.1	For type-2 diabetes, 1.3x increased nephropathy risk
r1688889	r1688889(A>G)	IRF102/RODD4	NA	NA	1.4	1.4x higher risk for colorectal cancer
r2272697	r2272697(A>G)	SNCG2	0.008	0.15	1	Adverse reaction more likely to be idiosyncratic/epileptic patients
r2412297	r2412297(C>T)	HU100492700/TM021	NA	NA	1.2	Associated with (slight) increase in the consumption
r1690908	r1690908(A>G)	COMT	0.043	0.38	2.5	slightly higher risk for nicotine dependence, lower risk for cocaine dependence
r2068915	r2068915(C>G)	NOO2	1	0.02	3	3x higher risk for Crohn's disease

After interpretation and comparison with these databases, we can form to each person for whom we meet for genome analysis like summary, which genes and which variants in genes will produce certain conditions.

Summary: predispositions

1. Watch out for high fat in diet (PPARG)
2. No smoking (FAS3G)
3. Predisposition to Hypertension: 1.6x increased risk for high blood pressure (ADD1)
4. High probability of dark hair and skin. If European ancestry – dark hair (SLC45A2)
5. Rheumatoid arthritis: 1.9x increased risk (TGAF, STAT4)
6. Coronary Artery Disease: 1.9x increased risk (rs1333049)
7. Prostate cancer (RFK6, rs6983267, ELAC2)
8. Macular degeneration (ARMS2, HTRA1)
9. Heart diseases (rs2943654, rs10757276, rs2383206)
10. Increased risk of type 2 diabetes (CDKAL1, SLC30A8)
11. Increased risk of male pattern baldness (rs2180439)
12. Ribavirin-induced anemia during anti-hepatitis C virus therapy (ITPA)
13. Lower risk of dementia and Alzheimer's disease (CETP)
14. Dry earwax – likely Asian ancestry (ABCC11)
15. No body odour (ABCC11)
16. Extra tasting ability (TAS1R3, TAS2R38)
17. May be better long-term memory (PRNP)
18. Advantage in memory and attention tasks (COMT)

And after this summary, we can create the conclusion to each patient with a different recommendation about the modification of their health and lifestyle.



Our genomic analysis also gives us a chance to understand our genetic background. High admixture and genetic heterogeneity were found in Kazakh individuals.

We performed the PCA comparison analysis of Kazakh individuals with several Central Asian population and Eurasian population in Jorde genomic projects and Human Gene Database Project to show position of the Kazakh samples in genetic maps of the human populations in Eurasia. First and

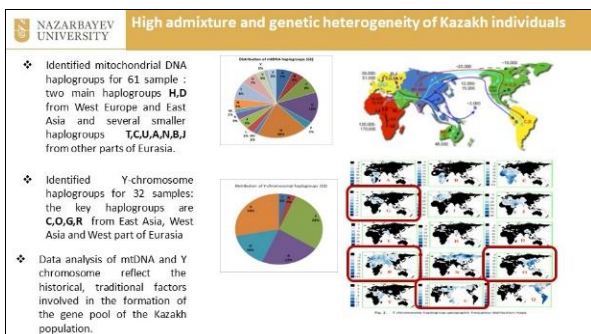
second principle components mainly reflect geographical distribution of the population. And you can see Kazakh population is clustered in Central Asia, which reflects geographic map of the world.

Also, ADMIXTURE is a common tool in genomics to analyze populational SNP data.

Here we performed ADMIXTURE analysis using more than 3000 whole genomes from different datasets, and you can see that from 5 to 10 ancestor groups forming our population, which shows the genetic heterogeneity of Kazakh population.

And you can see it is very heterogeneous in Kazakh population and very close to Hazara, Uyghurus, and Kyrgyz.

And as a comparison, African population, East Asian population are more homogeneous compared to ours.



One of the important findings is the mitochondrial DNA haplogroups in our samples.

Two main haplogroups are H and D from West Europe and East Asia, and several smaller haplogroups are T, C, U, A, N, B, J from other parts of Eurasia.

Y-chromosome analysis showed prevalence of C, O, G, R haplogroups from East Asia, West Asia, and West part of Eurasia.

Analysis of mitochondrial DNA and Y-chromosome reflect the historical, traditional factors involved in the formation of our nation genetic pool, also nomadic lifestyle, and development of silk road trade.

Targeted sequencing

Implementation of „Cardio-gene-panel“ sequencing by NGS technologies –

One Design to cover the various clinical phenotypes

Genome size: 3,200,000,000 bp (3.2 billion bp)
 ~ 1% of the genome encodes an extract (genes encoding a protein) = 30,000 genes
 1-15 genes are usually sufficient for routine diagnostic testing covering > 65% of cases

Whole genome sequencing itself is still a little bit costly, for wide range.

That's why targeted sequencing of certain genes is very important.

If genome size is more than 3.2 billion base pairs, only 1% of the whole genome encodes genes encoding proteins, around 25,000-30,000 genes.

And up to 15 genes are usually sufficient for routine diagnostic testing covering more than 65% of cases.

That's why our effort to create cardiogenetic panel, using NGS technologies - it's one design to cover the various clinical phenotypes in cardiology.

That's why our effort to create cardiogenetic panel, using NGS technologies - it's one design to cover the various clinical phenotypes in cardiology.

Cardiopanel development

HALOPLEX Cardiomyopathy (34 gene) and Arrhythmia Research Panels (21 gene)

Designed cardiopanel 96 genes for targeted exome sequencing
 Haloplex technology, Agilent Technologies

We included in our gene panel existed 2 panels, like HALOPLEX Cardiomyopathy (34 genes) and Arrhythmia panels (21 genes).

Plus we included another gene associated with different cardiological syndromes and create our own panel.

Selected cardiac disorders and associated genes.

List of 96 target genes

Arrhythmogenic syndromes	Associated genes
Long QT syndrome	KCNQ1, KCNH2, SCN5A, KCNE1, KCNE2, ANK2, KCND2, CACNA1c, Cw3, SCN4B, AKAP9, KCN15, SNTA1
Shortened QT syndrome	KCNQ1, KCNH2, KCN2
Brugada syndrome (BrS)	SCN5A, CACNB2, GPD1L, SCN1b, KCNE3, SCN3b, CACNA1c, MOGL1, KCNE5, KCND3, HCN4
Idiopathic ventricular arrhythmia	KCNAS, KCNE2, KCNQ1, NPPA, NUP155, LMNA, SCN5A, KCN8, ABCG2, GJA5, KCN2
Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT)	RyR2, CASQ2, KCN2
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2, DSG2, DSC2, DSP, JUP, TMEM43, TGFB3, RYR2
Dilated cardiomyopathy	LMNA, LDB3, TNNT2, PLN, MYH7, MYBPC3, SCN5A, DES, SOD3, CSRP3, TCAP, ACTC, TNNC1, TNNT3, TTR, ILK, EMD, CRYAB, BAG3, CHRM2, SGC8, DSP, DSP, TPST1, NEBL, DSG2, TTN, EYA4, ABCG3, TMPO, PSEN1, PSEN2, ACTN2, TAZ, VCL, ANKRD1, FRTN, LAMP2, NEKN, TBK20, DTNA, MYFN, LAMA4, FHL2, LAMA2, DMD, RBM20, SERCA2A, MYH6, CRYAB, BAG3, CHRM2, SGC8, NEBL, DSG2
Hypertrophic Cardiomyopathies	MYH7, TNNT2, TPM1, MYBPC3, TNNT3, PRKAG2, MYL3, TNNT1, MYL2, ACTC, LAMP2, SIA, Cw3, TTR, FHL1, TTN, CSRP3, MYOZ2, MYH6, MYL2, MYO6, LDB1, TCAP, VCL, ACTN2, PLN, JPH2, CRYAB, NDUFB2, GAA, CALR3, CTF1, NEKN
Restrictive Cardiomyopathies	MYH7, TNNT3, MYBPC3, TPM1, TNNT2, TNNT, ACTC, MYL2, MYL3, TCAP, LMNA, DES, CSRP3, TAZ, LDB3, MYOZ2, PLN, GLA

In this table, we can see arrhythmogenic syndromes like long QT syndrome, shortened QT syndrome, Brugada syndrome, dilated cardiomyopathy, etc.

And all of these genes which are encoding to be shown as associated with these diseases, included in our (study).

Cardiopanel development

Design of study

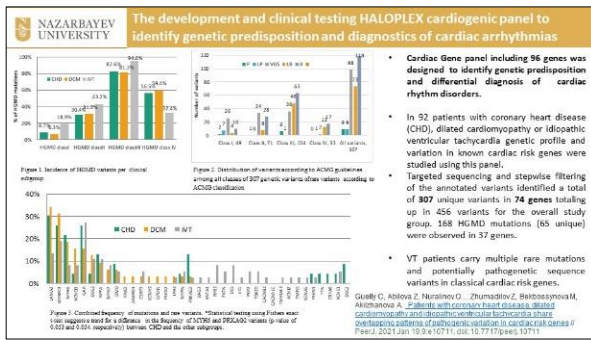
Ventricular tachycardia

- CHD, n=23
- DCM, n=32
- iVT, n=40

Material and Methods

- DNA isolation
- Library prep
- HaloPlex Target Enrichment System Protocol (version D.5, May 2013, Agilent Technologies, CA, US) using the standard HaloPlex 96 indexing primer cassette.
- Quality check - 2100 BioAnalyzer (Agilent Technologies, CA, US)
- Sequencing - HiSeq2000 platform using 2x150bp paired-end standard sequencing conditions.

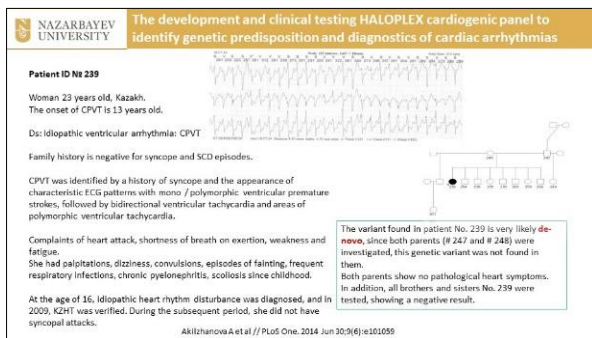
We will validate our developed panel in patients with ventricular tachycardia including patients with coronary heart disease, dilated cardiomyopathy, and idiopathic forms to find different genetic background of these patients using our panel.



In this group, so we found 307 unique variants in 74 genes, and among them 168 was presented also in Human Gene Mutation Database, and 65 genetic variants were not found in the existing databases, which was unique.

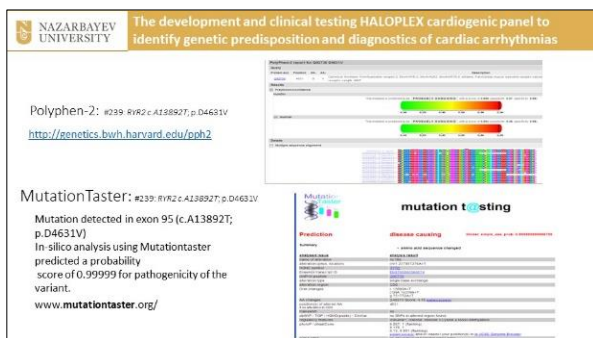
In our patients, we found also according to ACMG guidelines, up to 10 pathological mutations and likely pathological mutations.

Patients with coronary heart disease, ARVC, idiopathic ventricular tachycardia are showing overlapping a developing pattern of genetic mutations in different cardiac diseases.

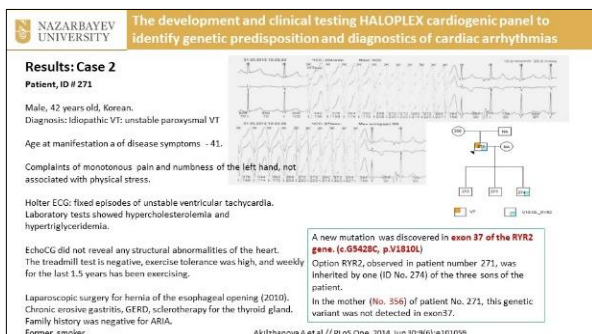


Here is a case report of a patient with cardiovascular diseases, with ventricular tachycardia catecholaminergic polymorphic form.

And we found in this patient de novo mutation in human ryanodine receptor gene, which was likely to know because we didn't confirm such mutation in her siblings and parents.



And this mutation was in pathogenic predicted by Polyphen and by MutationTaster and also showing that the disease caused mutation.



Another sample is a man with arrhythmias with paroxysmal VT, also found mutation in the same gene, the ryanodine (RYR2) receptor gene in exon 37.

We also found that in his son. And after genetic counseling, it was recommended to do more medical evaluation by a cardiologist for his son and for him to proper management of this patient.

NAZARBAYEV UNIVERSITY The development and clinical testing HALOPEX cardiogenic panel to identify genetic predisposition and diagnostics of cardiac arrhythmias

Polyphen-2: #271: RVR2 c.65428C; p.V181D.

MutationTaster: #271: RVR2 c.65428C; p.V181D.

mutation testing

Prediction: Disease causing

His mutation was also found in the damaging zone, it was also shown by MutationTaster that the disease caused it.

NAZARBAYEV UNIVERSITY Genome-associated personalized antithrombotic therapy in patients with high risk of thrombosis and bleeding

- Genome guided treatment approach in anti-thrombotic therapy for patients with high risk of thrombosis and bleeding were implemented into practice of the National Scientific Cardiac Surgery Center, Astana, Kazakhstan.
- August 2012 - first heart transplantation in Kazakhstan, LVAD implantations >100
- August 2018 -implantation of a fully artificial heart into 60-year-old patient, who was suffering terminal stage heart failure. A group of specialists from CARMAT a medtech company, and Airbus created the one million euro (US\$ 1.16 million) artificial heart.
- 100 patients with implanted left ventricular assist device (LVAD) operated at National Scientific Cardiac Surgery Center, Astana, Kazakhstan were included into study.
- Polymorphisms in genes involved in the metabolism of antithrombotic drugs (VKORC1, CYP2C9, CYP2C19, CYP3A4, UGT1A1, P450, etc.) were genotyped. According to the genotyping results optimal dose of Warfarin were calculated and recommended.
- Genetic screening for markers of sudden cardiac death in patients with cardiac arrhythmias and personalized approach for antithrombotic treatment were implemented into practice of the National Scientific Cardiac Surgery Center, Astana, Kazakhstan.
- Genetic counseling for patients

Another clinical usage of our genomic data is genome guided treatment approach in anti-thrombotic therapy for patients with high risk of thrombosis and bleeding.

And this methodology was implemented into practice in National Center for Cardiac Surgery in Kazakhstan.

This center is very famous in our first heart transplantation in Kazakhstan, LVAD implantations, and implantation of a fully artificial heart in our patients, in collaboration with a

technological company Carmat and Airbus.

And we implemented panel of genes including genes involved in the metabolism of antithrombotic drugs, and according to genotyping results, optimal doses of Warfarin, aspirin, as anticoagulant, will be calculated and recommended to such patients.

NAZARBAYEV UNIVERSITY Genomic and transcriptomic profile of esophageal cancer

Esophageal cancer is the sixth common cancer in Kazakhstan, and usually not detected until it has progressed to an advanced incurable stage.

Aim: To identify genetic basis of esophageal squamous cancer by performing whole human genome/exome and large scale RNA (transcriptome) sequencing study.

- total RNA sequencing of Kazakhstani patients with Esophageal cancer using HiSeq2000 Illumina was performed.
- genes involved in carcinogenesis of esophageal cancer - potential targets for the development of diagnostic markers and cancer therapy

54 tissue samples were sequenced for complete transcriptome (27 normal and 27 cancer samples) of 27 Kazakhstani patients with esophageal squamous cell carcinoma (ESCC) using next-generation sequencing technologies (Illumina HiSeq 2000).

1072 down-regulated and 1969 up-regulated genes were identified.

Functional analysis of up-regulated genes revealed the most significant enrichment for genes encoding products in the category of "cell cycle", "DNA replication" and "lysosome".

down-regulated genes in the category "metabolism of lipids and lipoproteins", "amine, leucine and isoleucine degradation" and "propanoate metabolism".

Principal component analysis

Biological process network

Biological signaling

Esophageal Cancer in Kazakhstan: Multi-omic Research Challenge. Rakhimova S, Akhbarova A, Zhukov Y, Omarov M, Zhumadillov Z. Cent Asian J Biol Med. 2014 Dec; 12(3(Suppl)):170. doi: 10.5195

Also we use genomic research and other omics research in oncology. One of the examples is the esophageal cancer which is the sixth common cancer in Kazakhstan, and usually not detected until an advanced, incurable stage is already developed. We performed total RNA sequencing, genome sequencing of such patient, and found down-regulated and up-regulated genes in different categories of "cell cycle," "DNA replication," and "Lysosome", and trying to find targets for treatment and

prognosis for this patient. And now we continue with this study and increasing the scale and the number of p(atients.)

NAZARBAYEV UNIVERSITY Mapping the eco-social and genetic factors that determine susceptibility to tuberculosis in the Republic of Kazakhstan

Aim: Evaluate the role of exogenous and genetic factors in the formation of susceptibility to TB infection, and the impact of phenotypic and genotypic properties of Mycobacterium tuberculosis on clinical course, outcome of the disease and the development of the epidemic process of tuberculosis.

The genetic risk factors for tuberculosis were studied

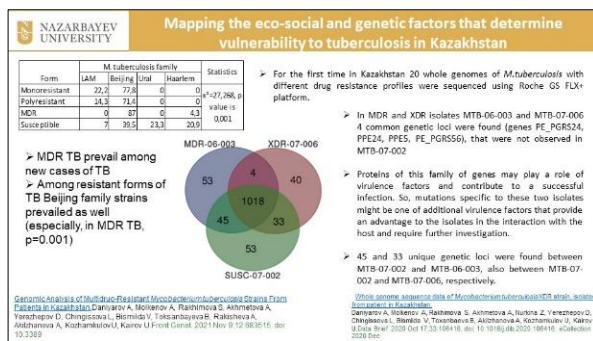
- 10 SNP in 6 human genes were identified and showed associative relationship with the development of tuberculosis in 1569 study participants.
- These include the following polymorphisms of:
 - Toll-like receptor gene B (TLR8) in the total group
 - Toll and Bsm1 vitamin D receptor gene and a gene IF-γ in women.
 - genes IF-γ (A / A) and TLR8 (G / A) by ethnicity

GWAS and WGS studies to study human immune response on M.tuberculosis are ongoing

Association of genetic variations in the vitamin D pathway with susceptibility to tuberculosis in Kazakhstan. Sapelkov M, Azzaz A, Kichanbulatov I, Alqabbarova A, Yarszapper D, Gallegos M, Chen CK. Mol Biol Rep. 2019 Mar;47(3):1659-1666. doi: 10.1007/s11033-020-05255-3

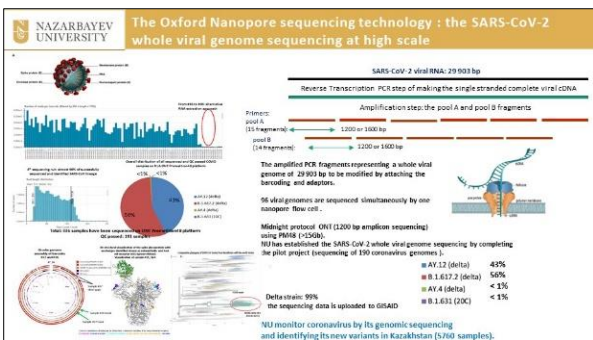
We perform genomic project in the field of tuberculosis which is still hard burden for healthcare in Kazakhstan.

We estimate host-pathogen interaction by studying of genomic immune response, engaged Mycobacterium tuberculosis in patient who is affected by this disease.



And by doing whole genome sequencing of Mycobacterium tuberculosis circulating in Kazakhstan, and found that among all primary cases of tuberculosis, 87% belongs to MDR.

Among resistance forms of tuberculosis, the Beijing family is predominant. This study is still continued to increase scale and increase the number of patients and number of microbacteria which will be whole genome sequenced.

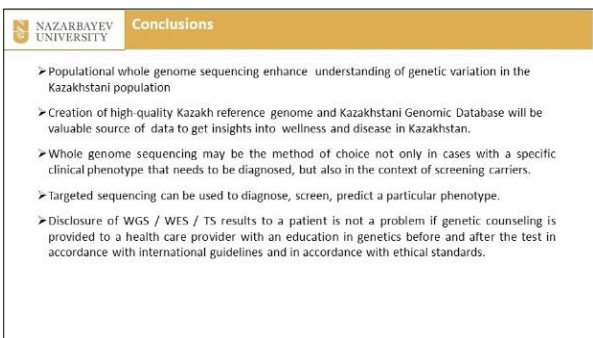


One of the good examples of using sequencing technologies is the Oxford Nanopore technology sequencing in the case of SARS-CoV-2 epidemics.

We use this technology by establishing the Midnight protocol to monitor coronavirus pervading by its whole genome sequencing to find new variants.

In our pilot study we performed in November, we found 99% all strains sequenced by that time it was Delta strain, and

nowadays appear more Omicron strains.



And we uploaded the whole genome sequencing in GISAID platform which is internationally available and all data for whole genomes of coronavirus is integrated in this database.

Let me make conclusions.

Populational whole genome sequencing enhance understanding of genetic variation in the Kazakhstani population.

Creation of high-quality Kazakh reference genome and Kazakhstani Genomic Database will be valuable source of data

to get insights into wellness and disease in Kazakhstan.

Results of whole genome sequencing will increase the current data by producing reference genome database specific to Kazakhstani citizens and drive large-scale scientific discovery.

Whole genome sequencing may be the method of choice not only in cases with a specific clinical phenotype that needs to be diagnosed, but also in the context of screening carriers.

Targeted sequencing can be used to diagnose, to screen, and to predict a particular phenotype.

Disclosure of whole genome sequencing, whole exome sequencing, and targeted sequencing results to a patient is not a problem if genetic counseling is provided to a health care provider with an education in genetics before and after the sequencing in accordance with international guidelines and in accordance with ethical standards.

Congratulations to NASHIM's 30th Anniversary!



I would like to acknowledge our laboratory staff, it's the Laboratory of Genomic Personalized Medicine and the Laboratory of Bioinformatics and System Biology and all of our National Laboratory staff.

From Nur-Sultan, I would like to congratulate to NASHIM's 30th anniversary!



Thank you very much for your attention.