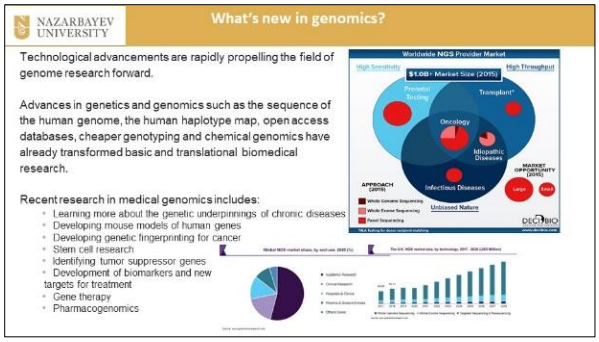
Dear Chairman, dear participants, dear colleagues and friends,
I would like to congratulate you with NASHIM 30th Anniversary !
Thank you very much for inviting me to present on this 30th Anniversary Symposium!

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 Unviersity - 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!"

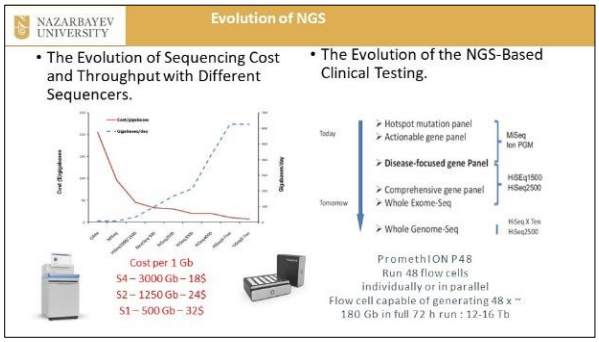


Technological advancements are rapidly propelling the field of genome research forward.
Advances in genetics and genomics such as the sequence of the human genome, the human haplotype map, open access databases, cheaper genotyping and chemical genomics have already transformed basic and translational 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.



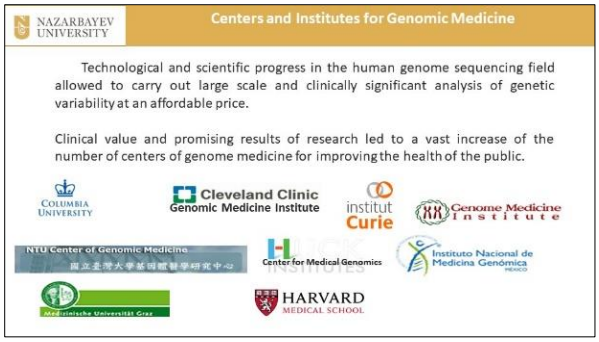
In the last decade, with development of sequencing platforms, cost per one gigabase decreases.

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.



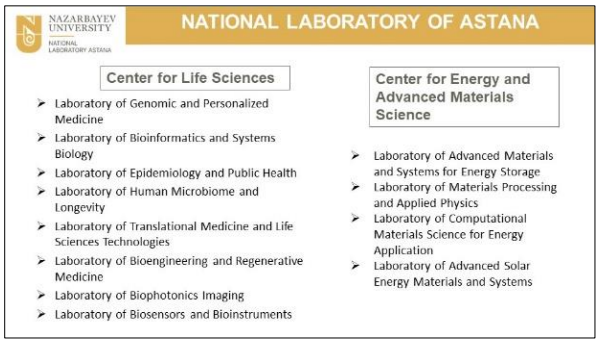
Technological and scientific progresss in the human genome sequencing field allowed to carry out large scale and clinically significant analysis of genetic variability at an affordable price.

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.

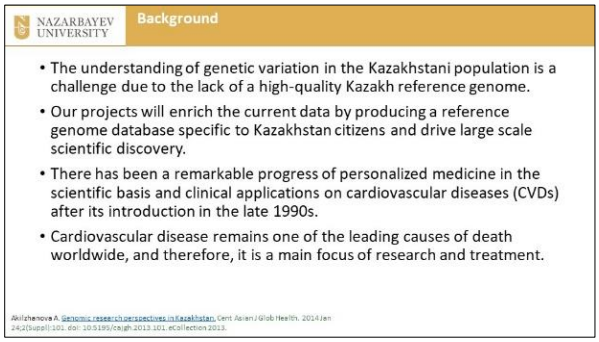
And this picture shows the main building of Medical School of our University and the University Medical Center.



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.

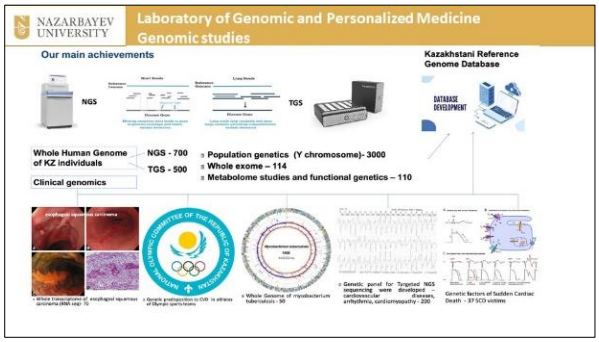


Understanding of genetic variation in the Kazakhstani population is a challenge, due to the lack of a high-quality Kazakh reference genome.

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.

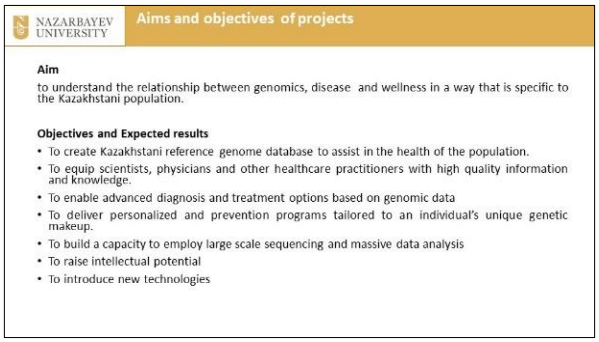


Our main achievements in laboratory and main directions for our projects are to generate comprehensive genomic data by using the most advanced sequencing technologies, both using short DNA fragments sequencing and long DNA fragments sequencing.

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 coronaviros epidemiology.



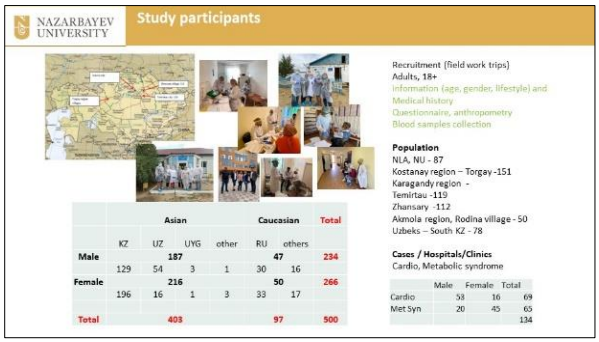
Overall aim of our project is to understand the relationship between genomics, disease and wellness in a way that is specific to the Kazakhstani population.

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.

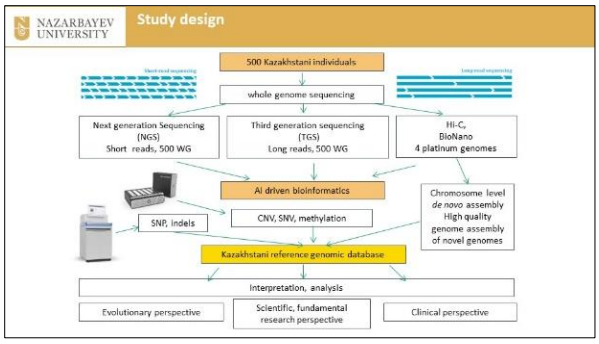
And of course, to build a capacity to employ large scale sequencing and bioinformatic analysis and raise intellectual potential, not only of our University but Kazakhstanian as well.

And of course to introduce new technologies.



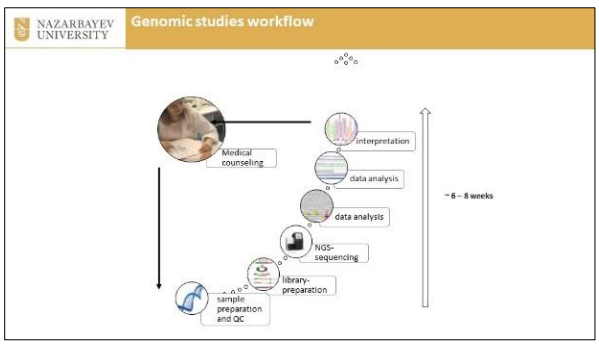
Study participants in our projects usually make field works and recruit participants from different parts of Kazakhstan.

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.



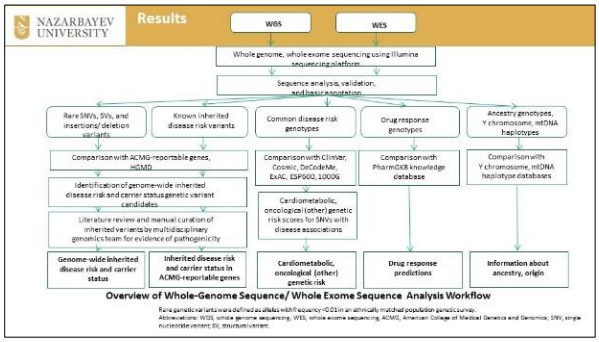
Overall study design includes whole genome sequencing, using next generation sequencing and third generation sequencing, and all data generated by these technologies will be analyzed and form Kazakhstani reference genomic database.

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

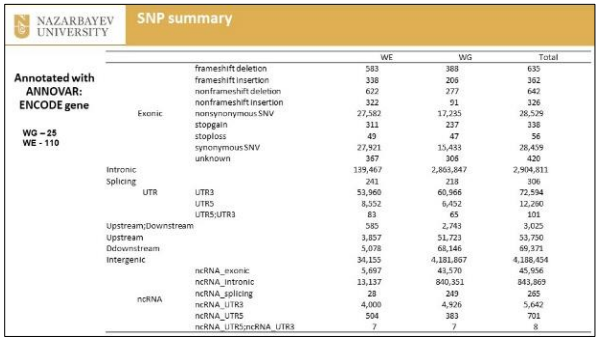
Usually it takes up to 6 or 8 weeks depending on diseases.



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, mitochondrian DNA haplotypes.

While doing analysis of all of these parts, we use different databases that exist in the international abook 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.

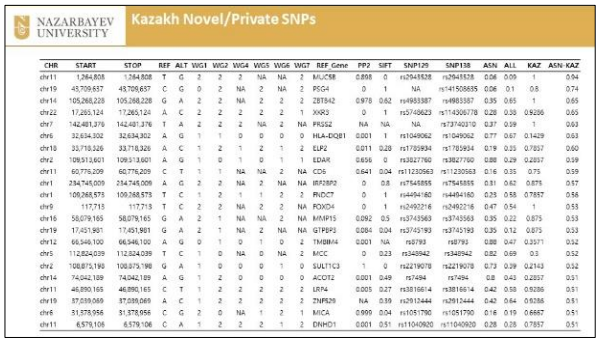
As a result, we can say genome-wide inherited disease risks or carrier status, inherited disease risk and carrier status in ACMG-reportable genes, cardiometabolic, oncological genetic risk, or maybe in patient with other complication disorders risk, drug response predictions, and ancestry genotypes of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



			WE	WG	Total
Exonic		frameshift deletion	583	388	635
		frameshift insertion	338	206	362
		nonframeshift deletion	622	277	642
		nonframeshift insertion	322	91	326
		nonsynonymous SNV	27,582	17,235	28,529
		stopgain	311	237	338
		stoploss	49	47	56
		synonymous SNV	27,973	15,488	28,459
		unknown	367	308	420
Intronic			139,467	2,863,847	2,904,811
Splicing			241	218	306
UTR		UTR3	53,960	60,966	72,594
		UTR5	8,552	6,452	12,260
		UTR5;UTR3	83	65	101
Upstream;Downstream			585	2,743	3,025
Upstream			3,857	51,723	53,750
Ddownstream			5,078	68,146	69,371
Intergenic			34,155	4,181,897	4,188,454
ncRNA		ncRNA_exonic	5,697	43,570	45,956
		ncRNA_intronic	13,137	840,351	843,869
		ncRNA_splicing	28	249	265
		ncRNA_UTR3	4,000	4,926	5,642
		ncRNA_UTR5	504	383	701
		ncRNA_UTR5;ncRNA_UTR3	7	7	8

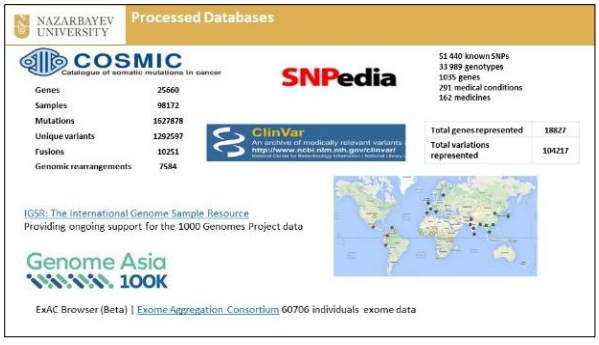
As a result of our preliminary data of whole genome and whole exomes, we found many thousands of exonic variations and intronic variations.

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

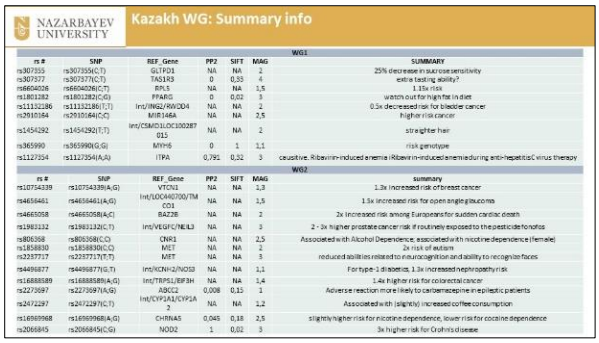


CHR	START	STOP	REF	ALT	WG1	WG2	WG4	WG5	WG6	WG7	REF_Gene	PP2	SIFT	SNP129	SNP138	ASN	ALL	KAZ	ASN/KAZ
chr11	1,264,808	1,264,808	T	G	2	2	2	NA	NA	2	MUC5B	0.898	0	rs2643528	rs2643528	0.06	0.09	1	0.94
chr19	43,709,657	43,709,657	C	G	0	2	NA	2	NA	2	PSG4	0	1	NA	rs141508655	0.06	0.1	0.8	0.74
chr14	105,268,228	105,268,228	G	A	2	2	NA	2	2	2	ZBTB42	0.978	0.62	rs4983387	rs4983387	0.35	0.65	1	0.65
chr22	17,265,124	17,265,124	A	C	2	2	2	2	2	1	XKR3	0	1	rs5746623	rs114306778	0.28	0.38	0.9286	0.65
chr7	142,481,376	142,481,376	T	A	2	2	2	NA	2	NA	PRSS2	NA	NA	NA	rs73740310	0.37	0.39	1	0.63
chr6	32,634,302	32,634,302	A	G	1	1	0	0	0	0	HLA-DQB1	0.001	1	rs1049062	rs1049062	0.77	0.67	0.1429	0.63
chr10	33,718,326	33,718,326	A	C	1	2	1	2	1	2	ELP2	0.011	0.28	rs1785934	rs1785934	0.19	0.35	0.7857	0.60
chr2	109,513,601	109,513,601	A	G	1	0	1	0	1	1	EDAR	0.656	0	rs3827760	rs3827760	0.88	0.29	0.2857	0.59
chr11	60,776,209	60,776,209	C	T	1	1	1	NA	NA	2	CD6	0.641	0.04	rs11230563	rs11230563	0.16	0.35	0.75	0.59
chr1	234,745,009	234,745,009	A	G	2	2	NA	2	NA	NA	IRF2BP2	0	0.8	rs7545885	rs7545885	0.31	0.62	0.875	0.57
chr1	109,268,573	109,268,573	T	C	1	2	1	1	2	2	PADC7	0	1	rs4494160	rs4494160	0.25	0.58	0.7857	0.56
chr9	117,715	117,715	T	C	2	2	NA	2	2	NA	FOXD4	0	1	rs2492216	rs2492216	0.47	0.54	1	0.53
chr16	58,079,165	58,079,165	G	A	2	1	NA	NA	2	NA	MMP15	0.092	0.5	rs3743563	rs3743563	0.35	0.22	0.875	0.53
chr19	17,451,981	17,451,981	G	A	2	1	NA	2	NA	NA	GTPBP3	0.084	0.04	rs3745193	rs3745193	0.35	0.12	0.875	0.53
chr12	66,546,100	66,546,100	A	G	0	1	0	1	0	2	TMBIM4	0.001	NA	rs8793	rs8793	0.88	0.47	0.3571	0.52
chr5	112,824,039	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	108,875,198	108,875,198	G	A	1	0	0	0	1	0	SULT1C3	1	0	rs2219078	rs2219078	0.73	0.39	0.2143	0.52
chr14	74,042,189	74,042,189	A	G	1	2	0	0	0	0	ACOT2	0.001	0.49	rs7494	rs7494	0.8	0.43	0.2857	0.51
chr11	46,890,165	46,890,165	C	T	1	2	2	2	2	2	LRP4	0.005	0.27	rs3816814	rs3816814	0.42	0.58	0.9286	0.51
chr19	37,039,069	37,039,069	A	C	1	2	2	2	2	2	ZNF529	NA	0.39	rs2912444	rs2912444	0.42	0.64	0.9286	0.51
chr6	31,378,956	31,378,956	C	G	2	0	NA	1	2	1	MICA	0.999	0.04	rs1051790	rs1051790	0.16	0.19	0.6667	0.51
chr11	6,579,106	6,579,106	C	A	1	2	2	2	1	2	DNHD1	0.001	0.51	rs11040920	rs11040920	0.28	0.28	0.7857	0.51

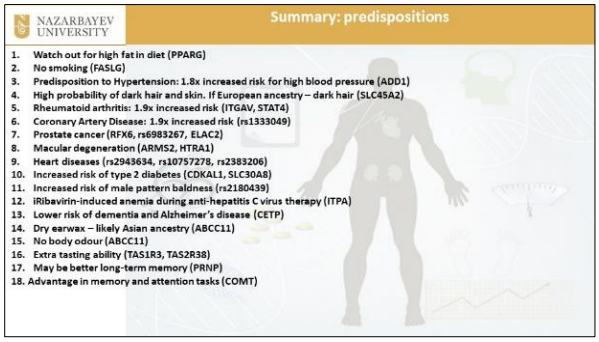
Also we found after analysis and comparison with different databases, Kazakh Novel or Private SNPs, which is the first SNPs functionally studied and confirmed in different, other studies.



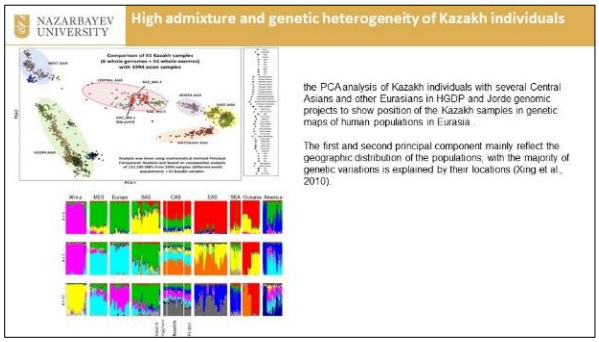
We process different databases to analyze our genome data like Cosmic, SNPedia, ClinVar, Genome Asia, etc. for the analysis of Genome Aggregation Consortium.



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.



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.

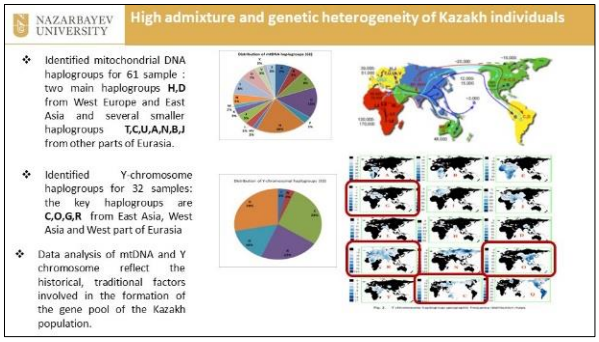
We performed the PCA comparison analysis of Kazakh individuals with several Central Asian population and Eurasian population in Jorde genomic projects and Human Gene Database Project to show position of the Kazakh samples in genetic maps of the human populations in Eurasia. First and second principle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

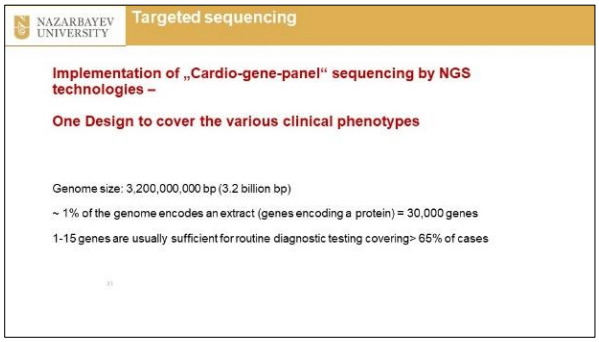


One of the important findings is the mitochondrial DNA haprogroups in our samples.

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

Y-chromosome analysis showed prevalence of C, O, G, R haprorgroups 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.



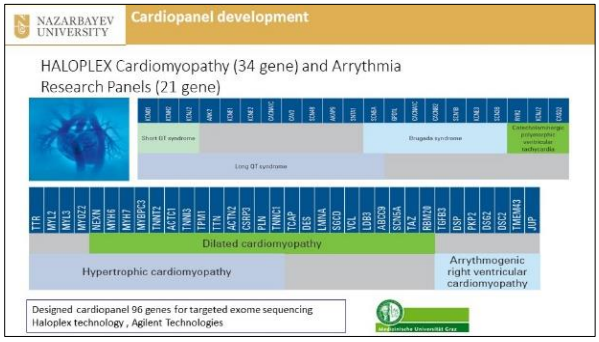
Whole genome sequencing itself is still a little bit costly, for wide range.

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 enconding proteins, around 25,000-30,000 genes.

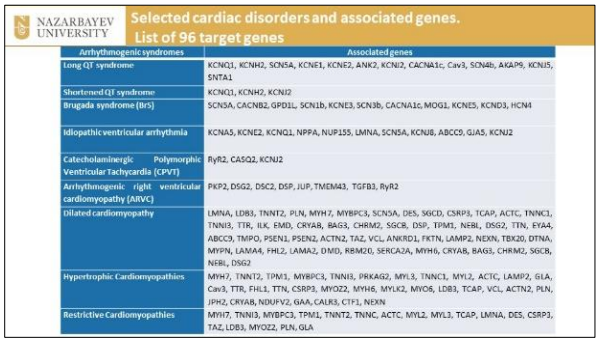
And up to 15 genes are usually sufficient for routine diagnostic testing covering more than 65% of cases.

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



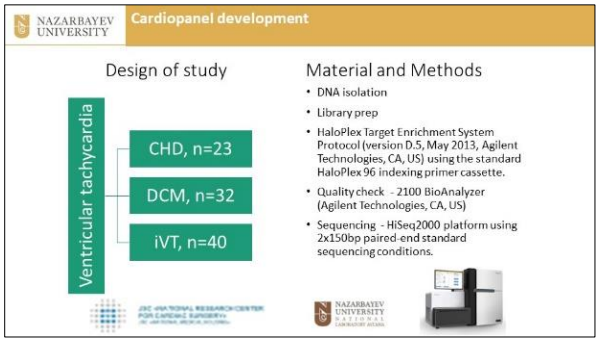
We included in our gene panel existed 2 panels, like HALOPLEX Cardiomyopathy (34 genes) and Arrythmia panels (21 genes).

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

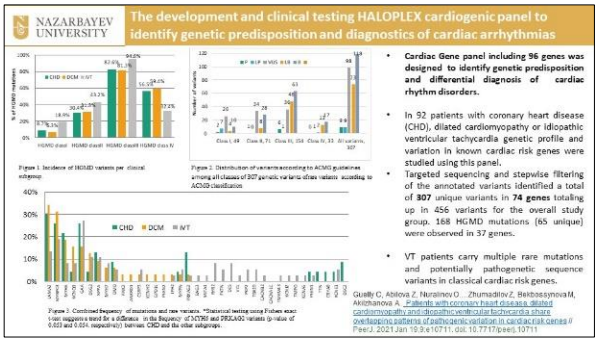


In this table, we can see arrhythmogenic syndromes like long QT syndrome, shortened QT syndrome, Brugada syndrome, dialated cardiomyopathy, etc.

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



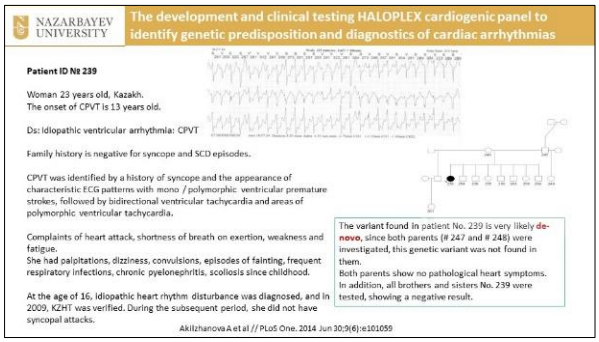
We will validate our developed panel in patients with venticular tachycardia including patients with coronary heart disease, dialated cardiomyopathy, and idiopathic forms to find different genetic background of these patients using our panel.



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.

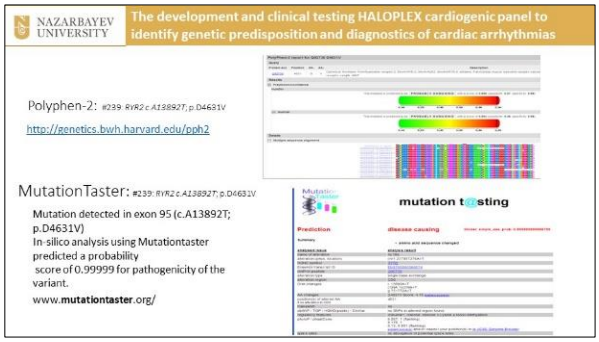
In our patients, we found also according to ACMG guidelines, up to 10 pathological mutations and likely pathological mutations.

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

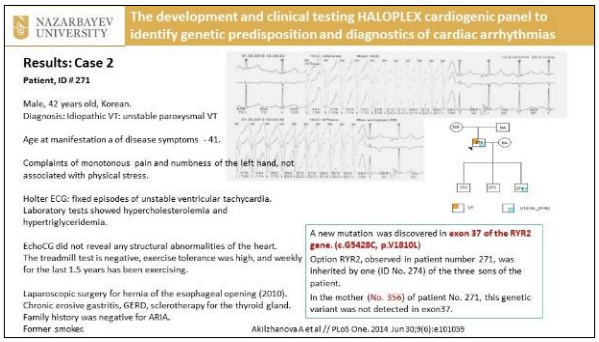


Here is a case report of a patient with cardiovascular diseases, with venticular tachycardia catecholaminergic polymorphic form.

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.

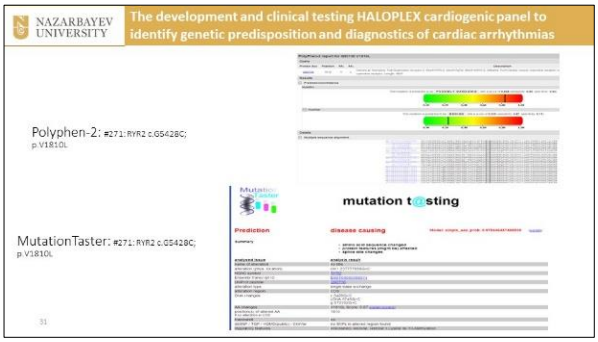


And this mutation was in pathogenic predicted by Polyphen and by MutationTaster and also showing that the disease caused mutation.

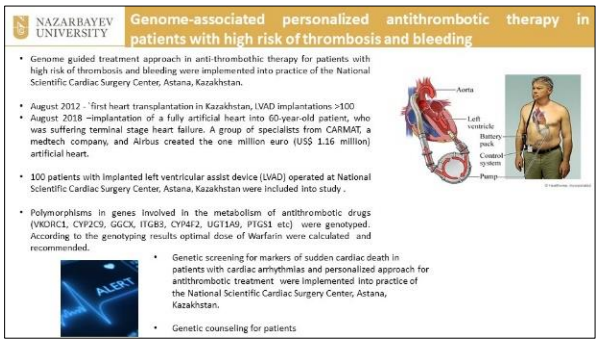


Another sample is a man with arrhythmias with paroxysmall VT, also found mutation in the same gene, the ryanodine (RYR2) receptor gene in exon 37.

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.



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

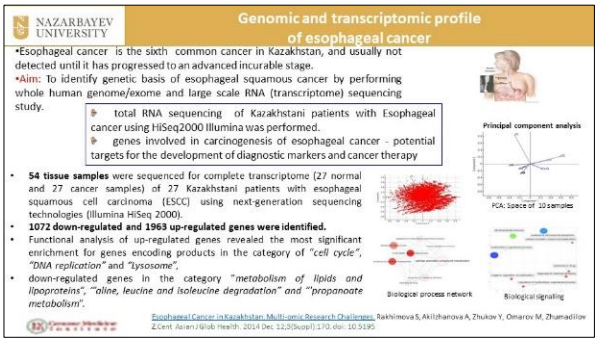


Another clinical usage of our genomic data is genome guided treatment approach in anti-thrombothic therapy for patients with high risk of thrombosis and bleeding.

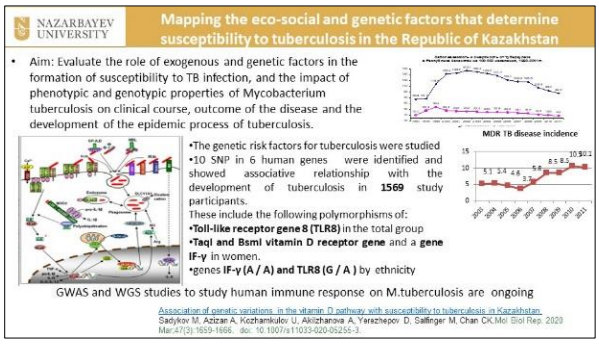
And this methodology was implemented into practice in National Center for Cardiac Surgery in Kazakhstan.

This center is very famous in our first heart transplantation in Kazakhstan, LVAT 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.

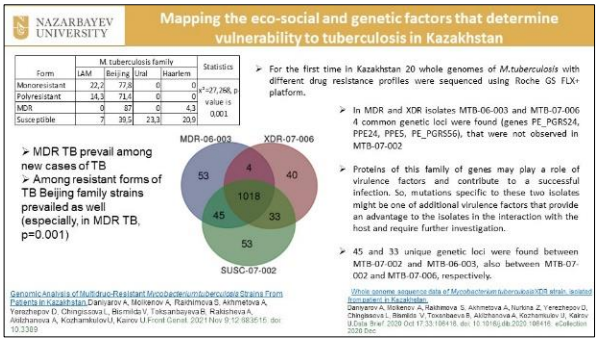


Also we use genomic research and other omics research in oncology. One of the examples is the esophageal cancer which is the sixth common cancer in Kazakhsan, and usually not detected until an advanced, incurable stage is already developed. We performed total RNA sequencing, genome sequencing of such 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.)



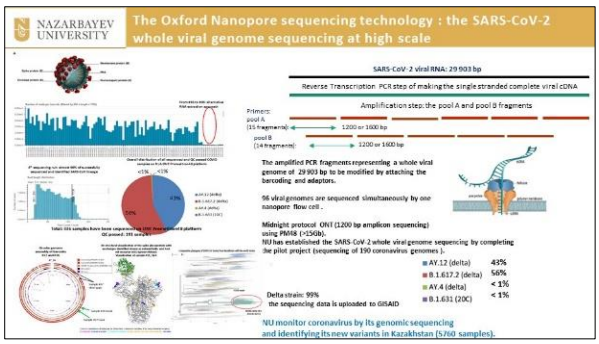
We perform genomic project in the field of tubeculosis which is still hard burden for healthcare in Kazakhstan.

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.

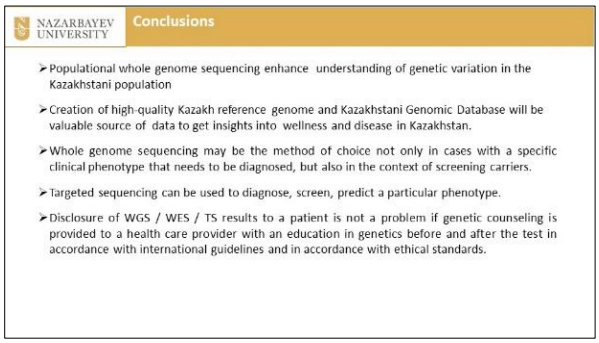
Among resistance forms of tuberculosis, the Beijing family is predominant. This study is still continued to increase scale and increase the number of patients and number of microbacteria which will be whole genome sequenced.



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 straints 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.

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.