DRAFT

ARMGENIA - HUMAN GENOME PROJECT IN ARMENIA

A Proposal by
ArmGenia Genetic Research Charitable Trust
for the Implementation of a Human Genome Project in Armenia, Creation of the Genetic Map of the Armenian People and the Establishment of a Genome Testing Laboratory and a Gene Bank

TABLE OF CONTENTS

1. BACKGROUND

After the completion of the Human Genome Project (HGP) in 2003 the field of Life Sciences and the scientific world as a whole entered into a new era which has already been rightly labeled as 'post-genomic'. The HGP, launched (and funded by) the US Government in 1990, proved to be the world's largest international collaborative initiative of its kind and yielded findings of paramount importance which can by no means be underestimated. This remarkable scientific achievement requires a totally new and advanced approach to the projects conducted in the area of molecular biology and molecular medicine. At present, virtually all activities in this field use the results of modern molecular genetic analysis, at least at some formative and hence vitally important stages of the work.

Through sequencing of the genomes of hundreds of thousands individuals and genome-wide association studies of the DNA differences (variations) between them, the acquired information helps to understand the genetic nature not only of hereditary disorders but also of different predisposition of individuals to common diseases and adverse side effects of many therapeutics. These data allow to understand the mechanisms of various diseases and, thus, facilitate developing and use of effective drugs.

The detection of disease-causing mutations has led to the identification of new biochemical pathways and has facilitated a greater understanding of the etiology of many diseases.

Furthermore, genome-wide association studies have provided information on how simple or complex genetic variability impacts on the risk for the development of various common and rare diseases.

Moreover, genomic information is used in historic, anthropological and human evolutionary studies, forensic applied sciences and in designing strategy for genetic, behavioral and environmental risk assessment. Another expected benefit of the genomics research is the commercial development of DNA-based therapeutic and molecular diagnostic products.

2. EXECUTIVE SUMMARY

Currently, many countries have launched their own national genome projects (e.g. Russia, Estonia, Iceland, Iran, etc.), which are closely linked with academic sciences, applied biomedical research and forensics. As described in details below, due to some unique historical and anthropological features of Armenians, unraveling the genomic make-up of the nation could provide unparalleled benefits for international community and, foremost, for the advancement of national science, healthcare and economy.

Unfortunately, the Armenian scientific community has not even discussed so far the necessity of national genome project, though some discrete patches are being conducted in Genetic Anthropology and medical genetics. This state of affairs hampers the advances in the national health care improvement and in both fundamental and applied biomedical research in Armenia. This, in turn, is seriously weakening our positions in the international scientific cooperation and impairing our abilities to develop national resources necessary for advancements of our health care based on the latest achievements of medical research and practice around the world.

The accomplishment of the project will enforce the development of advanced clinical practices and genomics research in Armenia, participation of Armenian researchers in internationally recognized human genome projects and creation of favorable conditions for implementing mutually beneficial collaborative research.

3. GOALS, AIMS AND OBJECTIVES OF THE ARMEGENIA PROJECT

The proposed Human Genome Project in Armenia - ArmGenia - is both an international and all-Armenian initiative one important target of which, among others as stated and described in this Proposal, is also to portray a genetic map of Historical Armenia as a mark of respect to the centenary of Armenian Genocide.

Unveiling human genome from a random sampling of Armenian population will make it possible to obtain in-depth information on individual's and population specific genetic markers.

Among the principle goals of the ArmGenia Project are therefore the establishment of a Genome Testing Laboratory and a National Gene Bank, a database consisting of phenotype and genotype data of the Armenian population aiming at conducting fundamental research, genetic and health studies.

The long-term goals of the project are:

1. Deepen our understanding of the genetic roots of the Armenian people;
2. Achieve a new level in Armenian healthcare through finding the optimal ways to improve population health and providing more efficient medical assistance;
3. Promote the development of biomedicine and modern medical and gene technologies in the Republic of Armenia thus increasing the competitiveness of medical services in the region;
4. Advance science and education ensuring the needed constant stream of qualified specialists in the high technology sector.

   Specific objectives
5. Anthropological Genetics: Reconstruction of the genetic history and origins of Armenians.

6. Medical Genetics: Structure of the Armenian gene pool according to the frequencies of deleterious mutations.
7. Pharmacogenetics: Structure of the Armenian gene pool according to the polymorphisms of Drug Metabolism Enzymes (DME) controlling genes.
8. Personalized medicine: Set up a healthcare database of individual health and genetic data of Armenians based on medical genetics and pharmacogenetics studies to ensure access to their own data to benefit from the personalized medicine.
9. Education: Support the education in the field of molecular genetics, bioinformatics, social sciences and biomedicine as an important component of the Project.
10. Public awareness: Increase awareness of the public of developments in the field of gene technology and biotechnology in general to raise popular interest.

Comprehensive genomic information obtained as an outcome of this project would be used in historic, anthropological and national evolutionary studies, forensic applied sciences and in designing strategy for genetic, behavioral and environmental risk assessment that can be used also for personalized medical care for our future generations and a wider range of other purposes.

4. PHASES OF THE PROJECT

The first phase of ArmGenia project will start with collecting information and data about the genetic peculiarities of 17 different geographic/dialect groups of Armenians for the purpose of creation of the Genetic Map of Armenia.

Although not covering the whole area of historical Armenia, this will be fairly enough to reconstruct the basic picture of the spatial diversity of the Armenians' genetic legacy lost after the Genocide.

This work will be continued to achieve a whole coverage of the area of historical Armenia, which used to include about 30 provinces. The samples will be typed for paternally transmitted Y-chromosomal and maternally inherited mitochondrial DNA markers.

It is planned to collect blood samples from donors whose ancestors both from paternal and maternal lines originated from the same geographic province of historical Armenia. Currently, in cooperation with Armenia's Ministry of Labor and Social Affairs around 100 individuals (age range from 90 to 100) have been identified as potential DNA sample donors. On average, 10-15 DNA samples will be collected in each of 30 territorial groups (in total of 300-450 individuals). The DNA samples will be subjected to whole genome sequencing, this will allow us to draw Genetic Map of Armenia in verified form.

It is the ultimate goal of the ArmGenia project to create DNA database/Gene Bank and a Laboratory that will serve as a source of valuable data on fundamental and applied sciences (bio-archeology, anthropology, human genomics evolution and clinical

medicine).

The second phase of the project will start with collecting blood samples in nuclear families (two parents and one offspring) in the population of the Republic of Armenia. Ten families in each of 30 administrative regions of modern Armenia will be studied. The sampling includes also the collection of detailed demographic data concerning health status, genealogy, lifestyle, and drug response.

From technical point of view, this project entails several organizational, technological and technical components of high importance.

5. ESTABLISHMENT OF A GENOME TESTING LABORATORY

In order to ensure the efficiency of the scientific aspect of the Project it is fundamentally important to establish a Genome Testing Laboratory (GTL) main functions of which shall be:

- the sequencing of DNA from blood samples and buccal swabs;
- the assemblage of these sequences to create a representation of the original genomes;
- the annotation and analysis of that representation to evaluate the multiplicity of the individual's single nucleotide polymorphisms (SNP).

The GTL will serve as a source of valuable data on fundamental and applied sciences (bio-archeology, anthropology, human genomics evolution and clinical medicine)

The activities of the Laboratory would proceed through three consecutive phases:

- Phase 1 (pre-analytic): selection and specimen collection from individuals and families chosen as project subjects, after obtaining of an informed consent;
- Phase 2 (analytic): extensive testing, establishing of a database to document genetic commonalities and range of variations (diversities), monitoring and verification of tests performed;
- Phase 3 (post-analytic): archiving records, release data for anthropological, medical and pharmacogenetics, as well as for personalized care and proficiency testing and training.

6. ESTABLISHMENT OF A GENE BANK

The core constituent of the national genome projects is nationwide DNA database, i.e. creation of a Gene Bank database including health and genetic data of the population.

The Gene Bank is based on the systematic data and tissue samples collection,

which leads to the development of a unique database allowing large scale association studies, revealing new information about genes that cause and influence common diseases. This information could predetermine more precise and efficient use of medicaments, introduction of new diagnostic tests, individualized treatment opportunities and disease-related risk assessment. The database enables different genetic study designs. It can also be used for the benefit of the persons included in the Gene Bank through family physicians, disease registers and other joint projects (eligible to have access to the information) in order to assess health risks or determine treatment.

The Gene Bank may be used only for scientific and public health research, statistical and educational purposes. The Gene Bank database should be of great interest among different research institutions, bioinformatics, biotechnology and pharmaceutical companies since it enables to analyze genotype and phenotype relationship based on population level.

It is planned to select about 2000 individuals of Armenian descent, genomic information of which is expected to provide the necessary clues from anthropological and medical predisposition prospective. This cohort will include individuals solicited through the nation-wide campaign, across the Republic and Diaspora, who can trace their origin from different regions of present and historic Armenia. Data shall be collected from both healthy and sick donors to establish relevant control groups inside the database. The data collection process consists of the collection of data concerning health status (health data), genealogy (genealogical data), lifestyle, environmental factors and drug response. The selected individuals will provide either blood or buccal swab specimens. The specimen collection will be done exclusively on the voluntary basis, with a written consent and full protection of the participants' privacy and ownership of the individual's data.

A special IT infrastructure of the database will be created with the aim of enabling application of a digital questionnaire and collection, processing, preservation and protection of data. Data enabling identification of gene donors shall not be available through the external computer network of the National Gene Bank. Data can be issued from the Gene Bank only in coded form.

The DNA samples will be sequenced in collaboration with the world leading institutions (e.g. Illumina Inc, San Diego, CA). The information obtained will be securely stored and analyzed for all genetic variations, including those expected to be markers for human genealogy and clinical predispositions. Initial genealogical history and frequency of clinical markers will be estimated already at this point.

Based on the genomic information obtained from 1,000 individuals, a full panel of the identified variations (SNP-Chip) will be provided by the sequencing vendor (i.e. Illumina, Inc) that will include all variations found in Armenians on a single test platform. This panel will be used for a cost-effective detection of the variations presence across much larger cohort of participants, around 10,000 individuals.

The combined information on the genetic variations will be pooled together in a single database and available for further in-depth analyses.

7. APPLICATION IN ANTHROPOLOGICAL GENETICS

The history of Armenia is a difficult one to reconstruct due to several reasons. The vast majority of valuable historical evidences have been destroyed or lost during numerous invasions and earthquakes in the Armenian homeland. Besides, the modern political division of Historical Armenia among neighboring states have made archival and archaeological research a sensitive, and often difficult, task.

Armenia has been little studied genetically, but the origins of its people may have important implications for theories of ancient Middle Eastern population expansion and the spread of Indo-European languages. Located at the intersection of Europe and the Near East, Armenia has acted as a focal point for multicultural fusion. However, the precise identification of the groups, which were the most significant contributors to the formation of the Armenian ethnotype has been contentious and remains unresolved. Linguistic analyses indicate that the Armenian language shares greatest similarity with the Indo-European languages of Greece and the Balkans. Furthermore, of the living Indo-European languages, Armenian appears to be the oldest, and therefore is potentially the closest living relative of the ancestral Proto-Indo-European dialects. Such relationships seem to indicate that Armenians are primarily descendants of the Indo- European tribes of the Armenian Highland and incorporated elements of Hurro-Urartian culture as they spread through the region.

Complicating the genetic structure of Armenians is a history characterized by invasions and occupations from various foreign empires. Armenia's complex history has likely had a substantial effect upon the genetic architecture of the ethnicity as invading empires introduced foreign genetic components into the region. The majority of genetic examinations of the Armenian population demonstrate their genetic lineage as intermediate to Europe and the Middle East, lending support for archaeological evidence indicating that Armenia served as a conduit of Neolithic migrations from the Middle East to the Balkan Peninsula. However, the genetic distance separating Armenia from Europe and the Middle East varies between marker systems, suggesting gender-based population events also played an important role in establishing the regions current genetic structure.

In the frames of the ArmGenia Project it will be possible to perform at an ultimate level of details the genetic mapping of historical Armenia and to identify the location of Armenians on the genetic landscape of the Middle East and beyond focusing on the following main goals:

1. To reveal the spatial stratification of the Armenian genetic pool;
2. To study the role of the Armenian Highland in the spread of different migration waves of Neolithic agriculturalists from the Near East to Europe;

3. To provide genetic evidences for the indigenous nature of Armenians in the Armenian Highland;
4. To test the presence of Arabic and Central Asian genetic traces in the Armenian genetic pool;
5. To reconstruct ancient genetic contacts of Armenians with other Middle Eastern autochthonous populations.

While most of the historic Armenian territory is now not populated by Armenians, direct descendants of people from different regions, towns and even villages can still be traced. ArmGenia project will allow to reconstruct the genetic make-up of people lived in the historic regions, i.e. creating a Living Map of Armenia. This will be a great tribute to our ancestors and will ensure that even a single village will not be lost from our historic memories.

8. TARGET GROUPS INVOLVED IN THE PROJECT

During 2014-2015 it is planned to collect DNA samples (buccal swabs and saliva) in 10 geographically distinct Armenian populations whose ancestors originated from different regions of historical Armenia. Currently these are the areas from eastern Turkey to western Azerbaijan. On average 200 DNA samples will be collected in each group. The samples will be typed for Y chromosome and mitochondrial DNA markers to reconstruct the patrilineal and matrilineal histories of Armenians. DNA extraction and genetic typing will be performed in the laboratories of our colleagues in Russia, Estonia and Lebanon with which we are collaborating within the frames of anthropogenetic study of the populations of South Caucasus and Near East.

9. ETHICAL, SOCIAL AND LEGAL ASPECTS OF THE PROJECT

As more genes are identified, the field of behavioral genetics becomes more complicated. If genes that indicate susceptibility for criminality, intelligence, or homosexuality are discovered, how should our society respond?

Should the government prevent insurance companies from demanding this genetic information from their customers? If the government does not intervene, would the insurance companies simply deny coverage to healthy people that may simply have the genetic potential for developing a particular disease?

The subject of the debate is intricate, sometimes controversial and fire at us serious questions from diverse angels.

The various specific socio-ethical decisions that our society must make are based on a few fundamental points; who should have access to your genetic information? Your

doctor? Your insurance company? Your employer? The court of justice? The police department? Who owns this information? How should this information be used during reproduction? Should couples test their unborn baby for non-disease traits (such as intelligence) and make reproductive decisions based on those tests? Who will standardize genetic tests and ensure their reliability?

Will people who desire to be tested make accurate decisions while considering the possible disease complications linked to gene-environment interactions? (e.g. heart disease, type 2 diabetes, various cancers). Can people overcome their genetically chosen behaviors? How much does free-will affect behavior, and how many of our actions are genetically predetermined? How should courts weigh these factors?

Who owns the genes and gene sequences? How will we keep this knowledge from being used by the "wrong hands?"

We need to address all ethical issues and make the issue public before the second phase of the project is launched.

10. THE IMPACT AND BENEFITS OF THE PROJECT

Application in Medical Genetics

We already know or can easily evaluate the current occurrence of any type of disorders (e.g. cardiovascular diseases, diabetes, cancers, etc.) or failure of their treatment among Armenians in general, in the Republic or different regions, with some of them being relatively high in comparison to other regions or nationalities. The ArmGenia project will help understand to what extent the higher frequency of disease might be or, alternatively, might not be explained by a different level of genetic predispositions. If there is an increased genetic predisposition, efforts should be focused on identification of the specific risk groups in the population, for prophylactic, diagnostic or treatment purposes, by implementing molecular tests targeting specific type of variations that were found to be overrepresented. If the increased occurrence cannot be explained by genetics, then efforts need to be focused instead on the search for underlying environmental, social or medical risk factors.

Furthermore, ArmGenia will identify most common type of health-relevant variations in the Armenian population. This will allow developing refined diagnostics tests optimally tailored to Armenian genotypes that will function significantly better than tests developed without the knowledge about Armenians-specific variations distribution.

Application in Pharmacogenetics

It is well recognized that different patients respond in different ways to the same medication, often requiring empirical strategies to find the appropriate drug therapy for

each patient. The existence of large population differences with small intra-patient variability is consistent with inheritance as a determinant of drug response; it is estimated that genetics can account for 20 to 95 percent of variability in drug disposition and effects. Although many non-genetic factors influence the effects of medications, including age, organ function, concomitant therapy, drug interactions, and the nature of the disease, there are now numerous examples of cases in which inter-individual differences in drug response are due to sequence variants in genes encoding drug-metabolizing enzymes, drug transporters, or drug targets. Unlike other factors influencing drug response, inherited determinants generally remain stable throughout a person's lifetime.

Unfortunately, no studies have been conducted so far to reveal the pattern of distribution of different variants in genes encoding Drug Metabolizing Enzymes (DMEs) in the Armenian population.

In the frames of the ArmGenia Project it will be possible to evaluate molecular genetic variation in selected DMEs affecting drug efficacy and safety in the Armenian population focusing on the following main goals:

1. To identify a group of diseases common in the Armenian population (e.g. essential hypertension, coronary artery disease, diabetes mellitus and other metabolic disorders);
2. To identify drugs commonly used to treat the diseases identified;
3. To identify DMEs likely to be involved in the metabolism of the drugs identified;
4. To identify known variation in the DMEs as recorded in the literature and available databases, as well as to define methods to categorize genomically detailed variants;
5. To establish variation in DMEs using DNA samples collected in the Armenian population;
6. To evaluate implications for drug efficacy and safety in the Armenian population.

Personalized medicine

Thanks to immense data accumulated and technological improvements in the last 10 years, human genetic testing is already leading to implementation of so-called personalized medicine, which is currently the most advanced trend in healthcare and pharmaceutical industry. Under the personalized medicine approach, diagnosis and treatment choice is done not only based on the patient's symptoms, demographics and biochemical test results, but also his/her genetic make-up. As a result, implementing human genome studies in general can be useful and effective in: 1) preventing disease and clinical conditions; 2) monitoring disease progression; 3) finding most effective treatment, and 4) reducing the cost of treatment through its optimization and preventive measures. Thus, genomic testing becomes a critical tool for predictive and precise (personalized) medicine.

The analysis of risk factors of disease development or accurate disease diagnostics is important, with a great example being the detection of BRCA 1/2 mutations pre-disposing to the breast cancer. However, the most immediate promise for human genomics testing is in the safer and more efficacious use of drugs in the patient i.e. the improvement of cure. Indeed, adverse effect of drugs, their inefficiency and clinical complications are currently among the main factors that are defining the effectiveness and cost of patient's management.

Furthermore, the genomic testing is found to be useful in the transplant surgery, where it helps to reduce the risk of graft-versus-host disease reactions, as well as to defining tissue-specific mutations directly responsible for tumor growth in the body or diagnosis of pre-malignant conditions. One of the first application of genomics in oncology is the analysis of SNP in various sub-types of human leukemia that allowed to find personalized treatment options for the disease with similar clinical manifestations, but genetically different.

Development of economy and generating revenues

The human skills and technological/computational infrastructure generated by ArmGenia will stimulate the economy of Armenia by creating canters for genetic testing and counseling that could serve not only Armenia but also generate revenues by providing services to other countries in the region. Also, the developed expertise and identified needs will create a favorable environment for biotech and drug companies to be established in Armenia for manufacturing and distribution the commercial test kits, instruments and/or pharmaceutical products.

The ArmGenia databank by itself will be a very valuable information resource for commercial companies interested in research and development of novel tests and/or drugs. They could pay appropriate license fees for a controlled access to the anonymous data, thus enabling ArmGenia to become a self-sustained entity and source of revenues.

Other Immediate and Long-term Benefits

The basic outcome of the project is the reconstruction of the genetic history of the Armenians population and their ancient genetic contacts with other populations. We anticipate identifying, in broad historical context, the role the Armenians played in the history of the Middle East considering the area of their habitation, the Armenian Highland, as a main transition corridor for anatomically modern humans' migration to Europe both in Paleolithic and Neolithic. We also expect to get strong genetic evidence for the indigenous nature of eastern Armenians. The role of the Armenian plateau in the spread of different migration waves of the Near Eastern agriculturalists to Europe (through Anatolia and the North Caucasus) will be elucidated. For the contemporary Armenian gene pool, we will estimate the genetic consequences of medieval invasions of

Arabic and Turkic-speaking tribes in Armenia.

Another important and, potentially, long-lasting benefit of the ArmGenia project is the significant advancement in the diagnostics and treatment of a vast variety of human diseases, both hereditary and sporadic, rare and common. The ArmGenia project will make the most of the wealth of information on genetic variations associated with disease and treatment failure that has already been obtained by massive number of genome projects around the world. Moreover, the databank of information that will be established could be used continuously in the future, i.e. for new association markers that undoubtedly will be constantly emerging from the global genomic studies. Besides implementing the knowledge obtained by international science, ArmGenia by itself will contribute into the search for novel markers, especially for diseases that might be common among Armenians, like Familial Mediterranean Fever (FMF).

The ArmGenia project will result in the evaluation of the frequencies of different genetic predispositions specifically among Armenians. Many of them are likely to be either significantly lower or higher than general estimates across different human populations. Even more important (and to a great extent unique for Armenians) is the potential that much more refined associations of genetic predispositions could be done by mapping the frequencies of specific genetic variations along the genealogical branches of the Armenian tree.

On the whole, the ArmGenia project will bring the national healthcare on par with one of the most advanced aspects of modern medicine and will allow using the latest achievements in the diagnostic and treatment of common diseases. This project will be the foundation for developing and applying personalized medicine approach in Armenia. This will lead to a significant reduction of mis- or underdiagnosed diseases and adverse treatment effects.

11. EDUCATION AND PUBLIC AWARENESS

The initiation of the genome project will put Armenia to the forefront of world-level clinical, biomedical, social as well as computational science. The project itself will be a collaborative effort with the leading research institutions in US and other countries that will enable to recruit or train Armenian scientists that will participate in the project. The ArmGenia implementation will develop technological and analytical resources (skills and infrastructure) that will dramatically increase Armenian scientific potential in medicine, anthropology as well as bioinformatics and computational sciences. It will allow Armenian scientists to establish close collaborations and exchange with international academic and applied institutions promoting the integration into international scientific cooperation.

In addition, ArmGenia will popularize the anthropological and medical science in general population, increasing their awareness about national history as well as increase

health conscious of the nation in general. It will decrease or even eliminate the bias against genomics that is still common among lay people and educate them how genetic tests can improve their personal health prospects or determine their family genealogy.

Finally, ArmGenia project will represent an investment in the future generations of Armenia by improving the education curriculums on biomedical, social and computational sciences.

12. FUNDING AND BUDGET

ArmGenia is a unique, multifaceted and innovative project. To set the clock on initiation and, afterwards, to guarantee a safely continued long-term operation, it needs nonstop multi-stream funding from different sources - the Government of Armenia, international and Armenian donor and charitable organizations, individual benefactors, international grants, etc.

ArmGenia requires a lot of organizational effort, experienced personnel, a sizeable set of state-of-the-art technology and sophisticated equipment, involvement of experts from respective Government and other agencies, universities and scientific institutions, and a lot more, including foreign operations and maintenance costs.