THE ROCKEFELLER UNIVERSITY HOSPITAL

CENTER FOR CLINICAL AND TRANSLATIONAL SCIENCE

Explaining Next Generation Sequencing (NGS) Studies to Research Volunteers in Plain Language

Human Genomics Core Working Group
Center for Clinical and Translational Science, Rockefeller University

This brochure is intended for investigators who design and conduct research involving Next Generation Sequencing (NGS). The goal is to provide both context and language to aid investigators in explaining research involving NGS to research volunteers. The IRB-approved Informed Consent Form (ICF) for a given study captures the required elements to be conveyed about a specific research study. However, research participants come from a broad range of backgrounds which presents various challenges to obtaining truly informed consent. Barriers to comprehension include limited medical literacy, limited scientific understanding, limited research literacy, and therapeutic misconception. Investigators may need additional information and strategies to overcome these challenges. The content below reflects a consensus of a Working Group of the Rockefeller Center for Clinical and Translational Science consisting of IRB members, professionals specializing in human subject protections and research conduct, faculty, genetic counselors, and IRB and CCTS leadership. In addition to the content below, the Working Group composed a participant tri-fold pamphlet by modifying an existing excellent pamphlet developed by the New England Research Subject Advocacy Group and contributing partners (http://catalyst.harvard.edu/services/rsa; http://catalyst.harvard.edu/regulatory/language.pdf) that you may want to provide to prospective participants. It includes a series of suggested questions that the participant may ask you as the investigator. This document is designed to assist you in developing the answers that fit your study and convey complex information clearly. Studies by Appelbaum et al. (Genet Med. 2014 May;16(5):367-73), Jamal et al. (Am J Med Genet A. 2013 May;161A(5):935-50), and Yu et al. (Am J Hum Genet. 2014 Jul 3;95(1):77-84) provide information on the views of genetic professionals on what information should be included in obtaining consent and what information should be shared with participants. A video designed to educate potential participants is available in YouTube (https://www.youtube.com/watch?v=IXamRS85hXU).

The text below in bold is targeted to you as an investigator. Text in regular font is potential wording to initiate discussion with the research volunteer who is considering entering your study. The questions demarcated with an * are ones that are included in the Rockefeller research participant pamphlet.

---

*What is the purpose of the study?

Is the study designed purely around a basic research question? Is the study designed to answer a medical diagnosis or care question? Is the study designed to address both basic science aims and questions with clinical implications? Participants may have misconceptions about whether NGS research can answer a specific medical question. It is important to be clear about whether the design will generate any results other than the basic research results, and if so, whether the results will be certified by a laboratory licensed by New York State (i.e., CLEP-licensed) so they can be shared with participants and their health care providers. It is important to explain whether research results involving known disease-predictive variants will undergo confirmatory sequencing (usually Sanger) as part of the protocol or whether findings will trigger referral for definitive testing elsewhere. If no known disease-predictive variants are to be studied, and results cannot be shared, it is important that the volunteer understands that the results will not answer a medical care question.

*Why perform a genetic analysis?

Some participants have little or no understanding of genetics and so to obtain informed consent it is important that the basics are conveyed in simple, clear terms. Comprehension should be assessed, just as with consent, by the teach-back method in which the volunteer explains back to you what you have communicated. A summary for volunteers might include the following basic facts:

- The study of "genetics" is the study of how traits such as hair color, height, and risk for disease are passed from parents to their children. A person's genetic information is called their genetic code or "genome." Genetic code that is passed on to children is said to be inherited by the child. Genetic code can also change before or after you are born, through mutations, which may result from random chance, or from environmental factors.
- We are studying your genome to better understand how the human body works and/or what causes it to not work well, as when someone has a disease.

*How will you collect my genetic sample?

It is important to explain what will be collected -- blood, saliva, urine, skin biopsy, tumor sample, other – and the mechanics of how the sample will be obtained, e.g. blood draw, skin biopsy, cheek swab, other. It is important to explain the implications of taking samples of the genes out of the body; some populations may believe that removing some of their genes means they have lost them.

- In general, you have the same copies of your genes in all of your cells. One exception might be cells from a tumor that contains mutations not found in the rest of your body. Removing the DNA and genes contained in your blood sample/saliva/urine/tumor/skin biopsy will not change your genetics or affect your health related to your genes.

*What will you look for in my genetic information?

If the volunteer does not understand the science of genes and variants, you may suggest viewing the video produced by investigators at Mount Sinai that provides some of the basics about genetics and whole genome sequencing (https://www.youtube.com/watch?v=IXamRS85hXU) before the conversation starts. By asking the volunteer questions, you can assess whether she or he understands at a basic level what is DNA, what is a gene, what is a variant, and what is a mutation.

There are several good sources of information for patients on the basics of genetics and DNA-based clinical genetic testing, but few resources that explain genetics in the context of basic and translational research. Below is an example of basic information provided in simple language.

- Your genetic code is contained in the make-up of your DNA. This code consists of four elements, or building blocks, which are also called bases. The code consists of these bases in the same way that words consist of letters. In fact, each of these four bases is named by a letter, which is an abbreviation for the molecule it signifies [adenine (A), thymine (T), guanine (G), cytosine (C)].

- Although there are four types of bases, each "word" of DNA contains only three of them, so the resulting DNA depends on which three of the four bases it contains, and the order (or sequence) in which the bases appear.

- Once the DNA is formed, it provides the code that determines the selection of a larger building block, called an amino acid. These amino acids, in turn, form still larger building blocks called proteins, and the order of the amino acids in a given protein needs to be just right for the protein to work properly.

- Finally, these proteins combine in various special ways to make your body function, for example by building tissues like muscle, bone, and skin; by triggering responses to various injuries; by functioning as signaling molecules; such as hormones and cytokines; or by acting as enzymes, which speed up chemical reactions in the body.

- **DNA bases** (A,T,G,C) build → **genetic code** (GCT-ACT-GTC-) that selects → **amino acids** (Ala, Thr, Cys) that build → **proteins** (ala-cys-thr-val-leu-ala-pro) that build → **tissues** (skin, muscle, blood, heart).

- In order to understand your genetic code, in the research, we will analyze the sequence of bases [adenine (A), thymine (T), guanine (G), cytosine (C)] that is present in your DNA and then compare it to the most common sequence of DNA bases that have already been recorded for thousands and thousands of other people. This type of analysis is called DNA "sequencing." We will note if there are any differences, for example, substituting a G for a C, an A for a T, or a T for a G; these differences are termed "variants" because they vary from the most common sequence.

- We know that in addition to the small proportion of the genetic code that contains the blueprints for the amino acids of proteins, there are large sections your DNA that are important in controlling or "regulating" whether other genes are turned on and off in each cell. Variations in these regions of the genetic code that affect gene regulation may also be important in causing bodily dysfunction or disease. We may analyze the sequence of the ~3 billion DNA bases that make up your entire genome ("whole genome sequencing," abbreviated WGS) or just the ~1-2% of your genome that directs the sequence of amino acids that make up the proteins in your body. The latter approach, which is limited to analysis of the genes that contain the blueprints for proteins, is termed "whole exome sequencing" or WES.

- Although DNA is primarily contained in the nucleus of each cell, and this nuclear DNA is inherited from both of your parents, there is also a small amount of DNA present in mitochondria, which are small structures within cells that help generate energy. Some diseases are associated with abnormalities in mitochondrial DNA. WGS analysis, but not WES, will include analysis of mitochondrial DNA. Mitochondrial DNA, unlike nuclear DNA, is inherited only from the mother.

- A DNA base variant may or may not affect the way your body functions. For example, a variation in a DNA base that determines the sequence of amino acids in a protein may or may not change the amino acid, and a change in an amino acid may or may not affect whether the protein functions properly or is properly made. This could happen for several reasons. First, there is more than one code for each amino acid, so the change in base sequence might not lead to a change in the amino acid. Second, if the code change does insert a new amino acid into the protein in place of the amino acid originally in the code, its impact depends on how different the two amino acids are, and where this change occurs in the protein. If the change occurs in a critical part of the protein, it may have a huge impact. If the change occurs in a less important part of the protein, it may not have any effect on health. Some changes to the genetic code prematurely stop a protein from being made. Depending on whether that stop occurs early or late in the production

of the protein, the impact could range from minimal to profound. The impact of variants that occur in the genetic code of regulatory regions of DNA may require additional research to understand.

- Finally, if the change in the amino acid sequence has not previously been reported in a patient, it may be difficult or impossible to predict whether the variant will alter the protein's production or function, and thus whether the variant is likely to produce a disorder in bodily function or disease.

- If you have a variant in your genetic code, you may have inherited the variant from one of your parents – or the variant may have entered into your DNA at the time you became an embryo or at another stage. Changes to DNA may occur when sperm and egg join at conception, or later when other influences cause DNA to change. Changes to the DNA can occur due to the insertion of DNA bases, the deletion of DNA bases, and the breaking of a DNA strand, followed by its joining another DNA strand in a way that brings together DNA sequences that don't belong next to each other. All of these changes or base variants affect the sequence of the bases and thus the genetic code.

Since each research study is different, it is important to explain specifically what will be done in the study. A few examples are given below.

- In this research study we will be looking for potential variants in one (or a series of) genes that we think may be altered involving **(*discuss disease or biological process here*).**
- We are not sure which genes are likely to be involved and so we want to analyze your entire exome (WES) (OR, your entire genome (WGS) to look for variants that…..
- In studies of tumors, we commonly are trying to identify variants present in the tumor that are not present in your other tissues. In this approach, we are testing whether the variants in the tumor may be responsible for the abnormal growth and behavior of the tumor or the response to therapy. Understanding the biology of the tumor may thus help guide the choice of therapy.

*What type of genetic testing will be performed?

From a regulatory standpoint, "genetic testing" means using a test with known predictive and diagnostic value. However, volunteers may use this term more loosely to mean any sequence analysis. It is important, therefore, to clarify:

- **that not all genetic research tests are the equivalent to doing a "genetic test" in a clinical lab**
- **whether you will be conducting a genetic test that has diagnostic or predictive value**
- **whether the results of the genetic test can be shared with the participant/volunteer and/or their health care provider**

*Will you also want to test members of my family? If yes, why and what is the potential impact on me?

If family members will be tested, explain why the genetics of other family members is needed for study. Multiple issues arise related to NGS testing of family members. It is important to explain the issues of privacy, autonomy, and confidentiality that are raised by family testing. It is also important to explain that when participants share information about the identification of their significant variants, this may affect the privacy of other family members. As a result, it is important for participants to understand this and decide which family members will receive which data. They also need to understand that family members' results will not be returned to the research volunteer unless the family member agrees. In

the case of studies involving children, for variants without disease-predictive significance, genetic counseling best practices suggest disclosing results to children only when they reach majority; this approach may be at odds with parental expectations.

*How reliable is the genetic testing?

- While genetic testing using the latest techniques is very reliable, it is not perfect. To improve the reliability, researchers program the machines so that they perform the same analysis many times, and this improves the reliability since a single error will be obvious when compared with a large number of identical results. For some regions of the genome there are greater technical challenges and thus in these regions variants may be harder to detect and confirm.

- For studies that include testing to help make a diagnosis or to guide therapy, laboratories usually confirm the results found with the new method using a technique that is known to be very reliable. If a variant is found that may be important for you to know about for your health, we may confirm the presence of the variant by asking a laboratory that specializes in identifying and confirming DNA variants to repeat the test. Such laboratories are regulated and inspected by New York State or the federal government. The federal program falls under the Clinical Laboratory Improvement Act (CLIA) and so these labs are commonly termed CLIA-approved laboratories. In New York State, laboratories are approved under the Clinical Laboratory Evaluation Program (CLEP) and so such laboratories are designated as CLEP-certified.

*What happens if the tests reveal a medical issue?

The answer to this question needs to be worked out with the IRB in advance of the informed consent discussion. For example, if the study involves genes with known disease-predictive variants, the IRB will likely have required a plan for verifying and/or returning those results. The plan might include confirmation of the variant through Sanger sequencing at a CLEP-certified lab and referral for genetic counseling and appropriate follow-up care. The informed consent form should contain information as to whether the participant will incur additional costs for confirmatory sequencing, genetic counseling, and/or clinical care related to the diagnosis.

*What if you find something that you did not expect?

Such variants are commonly termed "incidental" findings. It is important to explain the difference between the incidental detection of recognized variants with known significance in relation to a specific disease, and incidental detection of variants of unknown significance. Of recognized variants, some are "actionable" variants – that is, there is a therapeutic or preventative intervention known to make a difference in the outcome for persons with that variant. In contrast, for variants that are not actionable, there is no known intervention that has a positive effect on disease course/outcome. The American College of Medical Genetics (ACMG) periodically publishes a consensus list of actionable mutations and this list is considered the most authoritative assessment currently available. Recent research suggests that the likelihood of finding such an "actionable" variant is probably approximately 1-2% (Amendola et al., Genome Res. 25:305, 2015).

- There are several categories of unexpected variants, which are given the name "incidental findings" because they were not the primary purpose of the research study. One category consists of variants that have a high likelihood of indicating a disease or a predisposition to a condition or disease for which there are known effective preventive measures or treatments. These are called

"actionable" variants because one can take action to limit or prevent their effects. All such "incidental" findings need to be confirmed in a CLIA/CLEP-certified laboratory. There is a strong consensus among researchers and bioethicists that it is important to provide information to participants about "actionable" variants since this knowledge may be important for their health. Some people, however, feel strongly that research participants should have the option of declining to receive such information.

One way, but not the only way, to address the question of how to deal with research volunteers who refuse to receive data on actionable incidental findings is to exclude such individuals from the study using a statement such as:

- As a researcher who cares about your health, I am uncomfortable about possessing knowledge about you that I think may be important for your health and not be able to provide that information to you because you declined to receive it. As a result, I do not want to include you in this study if you are unwilling to receive this information.

If the investigator does not want to exclude research volunteers who do not want actionable results reported back to them and if the known "actionable" genes are not likely to be relevant to the research study, an alternative approach may be to exclude analysis of those genes. In this case the investigator will eliminate the possibility of being in the position of knowing that the participant has an actionable variant.

To address the issue of variants becoming actionable in the future, one can consider providing participants with their DNA sequences and encourage them to enter it into a program such as My46 available on the internet that provides medical interpretations of DNA sequence. The participant can then periodically update the interpretation.

- It is possible that over time some variants that are "not actionable" today may become "actionable" in the future as new treatments and preventive measures become available. We are not able to provide future tracking of all potentially actionable variants. If you want to know if one or more of your variants become actionable in the future, we will provide you with your sequence data and direct you to web sites that offer confidential ongoing updated analysis of your DNA variants. You may, however, have to pay for these repeat analyses.

- A second category of incidental findings includes variants that are not expected to alter the sequence or production of a protein and therefore are not expected to lead to a disease or predisposition to disease.

- A third category of incidental findings are ones in which it is difficult or impossible to predict whether the variant will affect protein production and/or function in a way that would lead to a disorder or a predisposition to a disorder. These are commonly called variants of unknown significance. Because of the uncertainty about the effects of these variants, we will not usually report them to you. If you like, we can discuss the types of incidental findings you would like to know about.

* Will I, or anyone else, receive results from this study?

If research results involve known disease-predictive variants, results certified by a CLEP-certified laboratory can be shared with participants and their health care providers. Research that identifies variants without established disease-predictive value, or which have not undergone confirmatory sequencing in a CLEP-certified laboratory, cannot be shared. As part of research confidentiality and privacy, participants should be informed that in the event results can be shared, the participant has control over whether anyone else will receive their results. Any data shared for research-only purposes would be coded and not identifiable as belonging to the participant. Information about who will receive the results should be included in the IRB-approved consent form.

*Will the test results become part of my medical record?

Research tests are not entered into the medical record at Rockefeller University. Details of the specific protections afforded participants, for example, a Certificate of Confidentiality from NIH, should be in the informed consent form.

*How do you protect the confidentiality and security of the information in the genetic material?

Assigning a code number to the DNA sample rather than the participant's name and keeping the table that links the code numbers to the names in a locked cabinet or in a password-protected computer – file is the standard method to protect research results from being disclosed to unauthorized individuals. Computers and USB flash drives that contain research information should be encrypted.

- To protect the security of the results of your genetic test, your sample will be given a code number and the information linking your name to the code will only be known by those investigators authorized by the ethics committee overseeing this research to have that information. Thus, the results of the genetic testing will only be identified by the code number and only the authorized investigators who have access to the linking table will know which DNA sample came from which participant. The computers and the computer programs that contain the DNA data meet high standards of data security, although it is not possible to guarantee that unauthorized individuals who are very sophisticated in breaking into computer systems won't be able to view the data.

*Will test results impact my health insurance coverage in the future?

U.S. law prohibits an employer or health insurance company from discriminating against individuals based on genetic testing. The law does not, however, extend to life insurance or disability insurance companies. Researchers will not disclose genetic information to any company except as may be required by law.

*Will my DNA sequencing data be shared with other researchers?

If you plan to share the data with a public database such as dbGAP, you should explain to the research volunteer that you are be required by the research funding agency (NIH) to share the coded data, but emphasize that no obvious identifiers will be shared, only coded sequence data.

- There is a strong consensus among researchers and the governmental agency that funds most of the biomedical research in the United States, the National Institutes of Health (NIH), that the results of research funded by the public should be made publicly available to authorized investigators so that it can be used to further advance biomedical science and improve human health. As a result, the NIH has created national databases for investigators to deposit their coded research data. These databases are designed for maximum security against unauthorized entry, but despite this very high level of security, it is impossible to guarantee that there will not be some unauthorized entry. As with the data kept by the researcher, however, the data are stripped of any information that would make it easy for someone to connect the results to a specific research participant. Nonetheless, it is not possible to guarantee that someone could not connect the results to the participant.

*Will my samples be used for future research? If so, will I need to give my consent?

This question is covered in the Informed Consent form as required by New York State law.

*What impact might my participation in the study have on my family planning and on members of my family?

If you anticipate your research may touch on these issues, you may want to include a genetic counselor as part of your team to discuss this complex issue. Only data obtained from a CLEP-certified lab can be shared for this purpose.

- Information you receive from participating in a research study that analyzes your genetic information may be of potential interest to you and your family members in family planning. For example, some genetic disorders occur only when both copies of the gene (one inherited from the mother and the other from the father) have variants that alter the function or production of a protein. In this case, individuals who have a variant in only one gene copy feel entirely well but are capable of passing the variant on to their children. These people are called “silent carriers.” Participating in a genetic research study may result in your finding out that you are a silent carrier of a variant that causes or predisposes a person to a particular disorder. As a result, you may want to consider testing a potential mate for a variant in the same gene, or other techniques such as pre-implantation or prenatal diagnosis. Similarly, if you learn that you have a variant, it is possible that other members of your family may have inherited the same variant, with potential implications for their family planning. It is advisable, therefore, before joining study, to think about how you would feel about having such information, and how you would feel about sharing that information with other family members. We will provide a genetic counselor to you to discuss the nature of the variant you have and how you might like to convey this information to family members who may be able to benefit from this information.