BROAD INSTITUTE

415 Main Street
Cambridge, MA 02142
T 617-714-7000 F 617-714-8972
www.broadinstitute.org

Limits to our Understanding

Eric S. Lander
President and founding director, Broad Institute of MIT and Harvard

Remarks as delivered at the International Summit on Human Gene Editing
National Academy of Sciences, Washington, D.C.
December 1, 2015

My assignment from the organizing committee is to look at the genetic basis of human disease and to ask how does it inform our thinking about germline editing. The first step is to be clear about what diseases we are talking about. I'm going to divide things into two categories for the purpose of this discussion. First, there are rare Mendelian diseases, including relatively well-known conditions such as cystic fibrosis and Huntington's disease as well as more than 4,000 little-known conditions that have frequencies that might be 1 in 100,000 or 1 in a million. These conditions are caused by mutations in a single gene. Although genetically very simple, they can have devastating consequences.

Second, we have a large number of common diseases, which are, for the most part, polygenic. These include heart disease, Alzheimer's disease, and schizophrenia. We have identified genetic factors that play a role in these conditions, but each is only one of many factors that contribute to these conditions, and they are by no means determinative. There is a locus that has a significant effect, although by no means determinative, of Alzheimer's. Likewise, we know that there are many genes in which variations are linked to characteristics such as height or obesity, but we can't isolate or quantify their impact. The situation is more complex with things like 'intelligence' because we don't even have a clear definition of what we mean by intelligence and therefore no reliable way to measure it. Genes play some role, but we cannot specify what it is.

So what do people want to do with this evolving knowledge of genetics? There is a range of aspirations. I hear calls to eliminate all cases of genetic disease. Some people go so far as to say let us not just eliminate the cases of disease, but banish all disease alleles from the human population. That's with regard to these severe Mendelian diseases. With regard to the common genetic diseases, they ask why don't we eliminate or greatly reduce disease risk. And, some say, maybe let's enhance the population, give everybody all the protections that you might wish them to have against all possible diseases, infectious and otherwise. In other words, let's give everyone the best possible genome.

So that's what one hears in public discussions and in the popular media. But the scientific community knows that this speculation is unrealistic. I want to talk about what we actually know about human disease genes, and I will provide a brief overview of the progress of knowledge in the past 30 years.

Sometime in the 1980s, a bunch of very simple principles were laid out as to how we might find the genes for disease. For simple Mendelian diseases, a famous paper by David Botstein and others told us a recipe for mapping the genes for simple Mendelian diseases by studying meaningless spelling differences scattered around the genome and finding some that show the

same inheritance pattern as the disease gene.

If two things have a correlated inheritance pattern, they can't be too far away. For example, they might show genetic recombination only 1% of the time. Unfortunately, 1% recombination means that you are still about a million letters of DNA away. So you would still have to plod from a linked genetic marker to find the disease gene. In the case of cystic fibrosis that took about 5 years of work for a whole team, spending tens of millions of dollars to find that gene. When they did, it turned out to might appear to be a boring string of letters—though not boring at all to people with this genetic condition. Researchers found that three letters were deleted in the genetic code of the vast majority of patients who have cystic fibrosis, resulting in the deletion of one amino acid of that protein.

With this information, we could do a genetic test to see who carries the mutation. And, we could ask a computer if it had ever seen any proteins like that before; in fact, the computer came back and said the CF protein looks a lot like proteins that sit in the membrane and transport things: Congratulations, you probably discovered that the cystic fibrosis gene is a transporter!

That was the science, and it is fantastic. It was also very slow, but we will come to that in a second.

Also in the 1980s there were a lot of ideas about how to extend this concept beyond simple single-gene diseases. For example, do we need the families? Maybe we can get away without the families. And for simple diseases, new approaches were developed that don't require families, such as by looking just at inbred individuals who have a recessive genetic disease. These individuals will have a region of the genome that was transmitted on both sides of the family from a common ancestor, and we could recognize this by a bunch of genetic markers in a row that were all homozygous, that is the identical on both the maternal and paternal chromosomes.

And then we turned to populations such as that of Finland, which started 2,000 years ago with a relatively small founding group. Almost anybody in Finland who has a particular rare genetic disease inherited it from a common ancestor, and we can recognize the chunk of the ancestral chromosome inherited from the common ancestor in almost all the people today who have the disease. So we can map those genes just by looking at the population.

Well, that gave rise to the idea that maybe we could go further. Maybe we could take this concept of mapping disease genes in populations and extend it from places like Finland to larger populations. Finland goes back 2,000 years, but Europe goes back 8,000 years, and other populations go back 10,000 years. With a denser and denser genetic map, we could find the regions of the genome where people who have the disease on average more often share the same spelling differences in a region.

So there were methods developed for mapping diseases based on what we would call common variant association studies. That is, the people who have the disease share certain particular common spellings more often than the people who don't.

More recently, there is the idea that you don't need to look only at common variants. We could just observe that the people who have a disease have a higher frequency of rare mutations in a particular gene than the people who don't have the disease.

Now, to use any of these approaches, we needed a tremendous amount of data and technology. Just for simple Mendelian diseases we needed to have genetic signposts up and

down the genome. Then we needed to be able to march along the chromosome and find the disease gene and sequence it.

And well, if we wanted to then apply the common variant methods, we would need to study hundreds of thousands of variants in thousands of people. And, if we wanted to study the rare genetic variants, we couldn't look just at signposts. We'd have to study the entire genome sequence in thousands of people.

So that is exactly what the scientific community did. It got its act together and developed the tools and methods. First was the Human Genome Project. The goal of the Human Genome Project was to find those genetic signposts, connect them together, read out all the sequence, make a list of all the genes, and ensure that all that information is freely and immediately available to anybody wherever they are in academia or industry or in a developed country or developing country.

I will completely skate over the 13-year process of completing the human genome. Suffice it to say: it got done.

So then, of course, having a sequence of one person was not enough. We needed to know virtually all the genetic variation in the entire human population.

So, that happened as well. There were international projects to collect all the genetic variants, and today something like 99.8% of the genetic variants in this audience are already in the databases. So we essentially have all human genetic variation, and we can genotype it really quickly: there are little genotyping chips where we take a little bit of your DNA, wash it over the chip, and read out 2.5 million or 10 million genetic variants.

Moreover, if we actually wanted to do full genomic sequencing, not just the 2 million signposts, to look for even the rarest genetic variants, the costs have dropped so dramatically that this is now feasible. It's about 2 million times cheaper today to do that than when we first did the human genome. That's pretty good for a relatively short time. I don't know anything that has ever dropped 2 million-fold in cost over about a decade.

So it means that now for about $1,500 you can get a genome sequenced. We can actually do all these things that we dreamed about and thought about and developed methods for in the 1980s, using the tools and information that were created in the 1990s and the 2000s.

So what about mapping disease genes? There has been a lot of that that has been going on. The mapping of disease genes—those simple Mendelian disease genes such as cystic fibrosis that used to take so much work—are now almost a piece of cake. It's almost a rotation project for a student to map the location of the disease gene, if you happen to have the samples in hand.

The genetics community knew about 70 disease genes in 1990. By the time the Human Genome Project was completed, it was about 1,700. Today it's about 4,000, and the limitation is just having enough samples to study. Interestingly, though, the family-based mapping for common genetic diseases failed miserably. There was a theory that for common diseases, such as heart disease and Alzheimer's, we map them like Mendelian diseases within families. That approach failed completely. Almost nothing was found that way, because in fact common disease is not just the whole collection of individual rare Mendelian diseases. It has a different genetic architecture.

The handful of exceptions that were found were things like ApoE4 in Alzheimer's and NOD2

in Crohn's disease. These were cases where you had things that were fairly common in the population, not like rare mutations, and they had effect sizes that were modest--two, three, four, fivefold. That suggested that instead of mapping in families, we should look at whole populations—by looking at the frequencies of the common genetic variants and pileups of rare genetic variants.

So folks did that. They took these collections of common genetic variants and they started mapping. The world found very little in 2001, 2002, 2003, 2004. Then as the better technology and richer information became available, we saw a big uptick in 2005 and larger increases after that. So, today, there are something like 4,000 or 5,000 such discoveries across hundreds of different diseases and traits. We have huge piles of common genetic variants associated with different diseases.

Recently, human geneticists are finding genes with rare genetic variants associated with these diseases. Now I'm not going to go into all of the biology, but I will tell you very quickly one particular story, because it will be relevant to what I want to say about human editing. I am going to relate a story of work by three particular people, Mark Daly, Steve McCarroll, Beth Stevens, all at the Broad, and their collaborators. They are Harvard professors who work on schizophrenia, and their latest research will be coming out in late January. I'm not going to go into the biological details, but I want to give you a sense of what things look like.

Some years ago, they and a bunch of colleagues around the world did a genetic study with 6,000 people with schizophrenia; they looked at 2 million genetic variants in 6,000 people and found absolutely nothing. At least nothing that was statistically significant. That was very depressing.

But they knew that there was something there. They could tell statistically there was some signal there, and so they expanded it to 20,000 people and found five genes. They expanded it to 50,000 people and they found 62 genes, and they expanded it to 110,000 people, and there are now 108 genes associated with schizophrenia. They published these results about a year and a half ago.

Interestingly, they found the genetic signal on chromosome 6 in a region called the HLA complex, which is related to immune reactions to infectious disease. They wanted to figure out what it was. Many people assumed this meant schizophrenia must have something to do with infectious disease.

But you can't tell until you actually look. People had a lot of theories about which infectious disease it was. The *Atlantic* magazine had an article about how you get mental illness from your cat. That was *Toxoplasma gondii*.

So what was the gene? I will just make a very long story, which will be told very thoroughly in the paper, very short to say they found the gene and it's called C4. I will tell you one detail about it. The gene C4 has a ‘day job’ functioning in the immune system. It is involved in marking microbes for destruction by the immune system.

But that is not the job relevant to schizophrenia. It turns out it also has a ‘night job’. The same gene is involved in the brain in marking synapses, connections between nerves, for destruction. It is a gene involved in the pruning of synaptic connections, and the paper shows genetic variants associated with higher expression of C4 confer higher risk of schizophrenia. This strongly suggests that schizophrenia is a disease related, at least in part, to the level of

synaptic pruning. That actually fits with some really old neuroanatomical observations, namely that the brains of patients with schizophrenia appear to have fewer synaptic connections, although nobody ever knew whether that was a cause or an effect of having schizophrenia, and also that a tremendous amount of synaptic pruning goes on in late adolescence, which is the time of onset of schizophrenia.

So it is possible, and I think the evidence is now piling up for other reasons, that schizophrenia may turn out to be at least in part a disease of excess synaptic pruning. That presents a therapeutic hypothesis that you can't act on today but might be able to act on at some point in the future by modulating synaptic pruning in some way.

I raise this not because I want to discuss schizophrenia but to say this is why we do genetic mapping. It points us to biological processes that might lead to treatments. I am going to come back to C4 in a moment.

So let's now turn to the implications for gene editing. Let's start with the rare Mendelian diseases. Let's start with a dominant disease. The vast majority of the time the affected person is a heterozygote. They have one copy, not two copies of the disease-causing variant.

Well, this means that 50 percent of their offspring will inherit that and 50 percent will not. So half of their embryos will be free of that disease. You could use preimplantation genetic diagnostics (PGD) to identify which half is not going to inherit the genetic disease and implant those embryos. It's not immediately clear why you'd want to use gene editing as a fix - selecting those that did get the genetic disease and trying to fix them by gene editing is much more difficult.

Now there is the possibility that we might not have enough embryos that have inherited the healthy gene, but half is pretty big fraction. Still, it is possible to have a situation in which an insufficient number of unaffected embryos are produced. Of course, if the dream of being able to turn somatic cells into germ cells and to culture them, then we wouldn't face this problem because we'd have an unlimited supply.

The challenging case is when an individual is *homozygous* for a dominant disorder, having two copies of the gene. In that case, every one of the offspring will inherit the genetic disease. This is a serious case in which genome editing would be useful. It's also an extremely rare case. Both of those statements are true, and I make no value judgment about them. But I want to emphasize just how rare this is. In the entire scientific literature, the number of instances of Huntington's disease is measured in the dozens worldwide, and for many other diseases it is likely to be lower. So it is very unlikely—although still possible—that an individual would be homozygous for a dominant disease, and this particular individual would really benefit right now from germline editing.

What about recessive diseases? Well, here you have two parents who typically are unaffected. They are heterozygous carriers. They usually don't even *know* that they carry the gene, yet they can have a child who has the genetic disease. Of course, if they do know they carry the disease, they could use PGD: 75% of the embryos would not have the genetic disease. The numbers game is even more in their favor, although it is always possible that they continue to always draw embryos that are homozygous for the disease alleles. However, it is pretty unlikely that that will be the case.

The truth is—and this should be said here—if we really care about helping parents avoid

cases of genetic disease, germline editing is not the first, second, third, or fourth thing that we should be thinking about. What we should be thinking about is that the vast majority of people who have children with a recessive disease were never aware they were carriers. Most such recessive disease arises unexpected in a family. The most important thing is to make genetic diagnostics available so they could know they are carriers and be able to avail themselves of PGD. This would be the most effective option both from a societal basis and from helping the largest number of parents. It is important that we think about how we want to deploy our resources here.

In addition, if we wanted to *eliminate* all genetic disease, we would have to do more than simply eliminate the production of these homozygous embryos. We would have to eliminate the production of heterozygous embryos as well.

I should note that means all of you should not be engaging in natural reproduction, because all of you carry multiple, probably about a dozen, genetic disease genes in the heterozygous state. So if we want to get serious about eliminating all these genetic variants in that might cause disease, it would require everybody's effort to stop reproducing naturally and instead use PGD. I don't think that is very likely to happen.

The case that is most plausible for germline editing is where you have two parents who are homozygotes. Then all the kids are going to inherit the genetic disease. It happens. But, it just isn't very common, because most genetic diseases are exceedingly rare to start with. Many of them are very debilitating. And situations where two parents with the same genetic disease marry and have kids are exceedingly rare. The most likely case I can think of is two parents with congenital deafness due to the same gene. There are actually multiple genes that could cause deafness; if the parents have deafness due to different genes, it's not a problem. But there can be cases when they carry the same gene, although again, we're talking about relatively small numbers here. No value judgments in that, but we should understand the size of the compelling case there.

So now let's turn to common polygenic diseases? Well, the suggestion is that we could decrease disease risk and thereby enhance the human population. But I have to let you in on this secret. All those genes that have been linked to for common disease have exceedingly modest effects on risk. There are a handful that have effect sizes like three- or four- or five-fold. But 99-plus-plus percent have effect sizes less than a 1.2-fold increase.

Why do they tend to have weak effects? Well, the reason is that evolution beats down the frequency of alleles that have very strong effects. Evolution selects against them, and they become less frequent. In addition, it turns out disease processes are complicated and they are buffered. A missing gene will make a difference only when it occurs along with a number of other factors.

Let me put this in concrete terms by going back to C4. Schizophrenia risk in the population is 1%. If you happen to inherit the C4 allele associated with higher risk of schizophrenia, you now are facing a risk of 1.1%. That's it. It's a really important discovery from the point of view of uncovering a disease mechanism that is likely at work in many patients and may lead to real treatments over time, But, the risk conferred by the common genetic variants in this gene is not a risk you would even notice.

Now what if we were to take the top 108 genes that I told you about and create a polygenic

risk score for each? We will note which allele is more risk-causing and which less risk-causing, We will weighted each locus by how much it contributes. If look at the top decile of the population for gene scores associated with the top 100 loci, these individuals have about a 3% risk of schizophrenia.

What if we include the top 10,000 loci in the human genome? That means we are willing to include many things that are not statistically significant (because we can't tell what is statistically significant). If we use a gene score based on the top 10,000 scores, the top decile of the population now has a 10% risk of schizophrenia. So if you were interested in 'fixing' 10,000 loci in the human genome, the risk for the top decile could be brought back down.

But is that a free lunch? Or will things happen when we mess with 100 or 1,000 or 10,000 loci in the genome?

Well, there is no free lunch, or at least rarely a free lunch. Genetic variants tend to have what we call pleiotropic effects. They affect many different things, and they also interact with environment. For example, we know that there are variants that lower risk of one thing but increase the risk of something else. For example, eliminating the CCR5 gene greatly lowers your risk of getting HIV, but it also increases your risk 13-fold for acquiring a fatal case of West Nile. We could edit the whole human population to eliminate the CCR5 gene and we will be protected against HIV, but now we will be much more at risk for fatal cases of West Nile.

There tends not to be free lunches, and in any case we have a pretty poor idea for most things about what their overall effects are. We have relatively little data. We discover that a variant has an effect on some disease. We don't have an index of all the other things it may be doing with respect to risk of other diseases, especially in combination with other genes in different environments.

So is it plausible to think about avoiding all the deleterious variants in the genome? No. Most have very small effects, and there is a small number with large effects. Should we go around and bestow upon everybody protective variants with large effects?

Well, there are a handful of cases of genetic variants that offer three- or fourfold protection against a disease, as with CCR5 and HIV. But to know that it would be safe, they'd have to be pretty common, so that we could observe them in the homozygous state. We don't want to give people a variant that turns out to be good in the one dose but bad when you inherit two doses. And, we'd ideally want them to have no downsides—no undesired pleiotropic effects, which isn't that easy to know at this point. With electronic medical records and fill genotypes on millions of people, we might be able to be able to get to the point that we could know, but today we simply don't know enough.

The best two candidates I can think of are ApoE and PCSK9.

ApoE2 and 3 look like they are much better to have than ApoE4, which increases your risk of Alzheimer's disease and some forms of heart disease. However, I can't swear that there isn't some problem here, because after all, ApoE4 has remaining around 3% allele frequency in essentially every human population. If ApoE4 had no benefits, nature should probably have gotten rid of it by now. But maybe not. Perhaps it is just at 3% frequency in all populations just by chance. Still we ought to keep our minds open It's not to say we shouldn't do something about it, but we have to recognize the limits of our knowledge here.

Individuals who do not carry functional copies of PCSK9 have much lower LDL levels and much lower risk for heart attack. There are some homozygotes in the population, and the few that have been looked at seem pretty healthy. The most famous case is an aerobics instructor that is homozygous null for PCSK9. Still, we have quite incomplete knowledge. We have very few cases like this, and we have pretty incomplete knowledge on most of them.

What's the upshot? The decisions to be made about whether and where to use human germline editing involve many value judgments, but they have to be informed by the facts of the genetic architecture of disease.

With rare Mendelian diseases, the vast majority of situations can currently be addressed by in vitro fertilization and PGD. There are some compelling cases, although they are rare, and if we wish to avoid devastating genetic diseases, the best thing to do from a population point of view, from a public health point of view, from a caring about parents point of view, is to make sure that parents have easy access to genetic testing so that they know that they are at risk and can use conventional PGD.

With common polygenic diseases, it is great to hear people talk about how we are going to give intelligence to the human population and athletic ability and all sorts of things like that, but the truth is these are very complex traits—often influenced by hundreds of genes. Height, for example, has at least 400 contributors, and together they explain only a small part of it. These are really important scientifically because they reveal the actual underlying biology and the processes and the point of therapy. But for the vast majority of individual variants, the risk is small, and even if you want to gang together and CRISPR-ize hundreds at a time, the effect will be relatively small.

Currently I can only think of a relative handful of things that would be plausible cases to try to do, and even then I don't know for sure that they are a good idea, because, again, if it is such a good idea, I want to scratch my head and ask why didn't evolution think of doing increasing its frequency in the population.

Bottom line: My prescription is humility. It is always good to remind ourselves, especially when we have in our hand an amazingly powerful tool like CRISPR gene editing, that we exist in a state of very limited knowledge, and human genetic disease is complex. We still have a lot to learn, and it might, might, might be a good idea that—before we make permanent changes to the human gene pool—we should exercise considerable caution.

For a video of this talk, please visit:

https://vimeo.com/album/3703972/video/149192705