Scientific Appendix

Manipulating the Human Genome: Technology and Choice

As I was mid-way into writing the accompanying novel, *The Unedited*, an unexpected story hit the news. Dr. He Jiankui announced that he had enabled the birth of a pair of genetically edited twin girls by using a powerful new biotechnological tool called CRISPR-Cas9 to manipulate the girls' genomes at the early embryo stage. Although the report was not published in a peer-reviewed scientific journal, the claim seemed likely to be true. The experiment was quickly judged illegal and the scientific community largely condemned it (Cyranoski and Ledford, 2018). Unsurprisingly, the general press reacted as well. Whether plants or animals, genetically modified organisms, or GMOs, remain contentious for many. But producing genetically modified human beings? That is about as controversial as biotechnology can get. I will return to the specific experiment later, but the key message is this: Human genome editing is no longer just a matter of speculative fiction. It is here. And the issues this raises call for serious consideration from the scientific community and from society as a whole.

Does the advent of the means to edit the genome mean that "designer-babies" are right around the corner? I doubt it. Our biological know-how is not advanced enough for wide-ranging human genome editing, let alone "design", to make sense—even to the most eager. This may change, of course, but only slowly. In addition, public perception of editing is unlikely to be

P. Rørth, *The Unedited*, Science and Fiction, https://doi.org/10.1007/978-3-030-34624-9

favorable and societal regulations will most likely continue to outlaw it. This may change, as well, and potentially more quickly.

Public perception is likely to be guided by two important, but fundamentally different, concerns. The first, as with any new technology, is safety. Do we fully understand the consequences of editing the genome? Recombinant DNA technology and GMOs have been discussed in scientific and public forums over the past 50 years, resulting in policies and regulations controlling their use. Turn to human genomes and the stakes are much higher: genetic fixes may alleviate life-long suffering or early death, but mistakes could also be devastating. Is the risk worth the reward?

The second concern is reproductive choice. The rapidly advancing power of genome-based diagnostics is making choice a significant ethical issue facing us even today. This is a sensitive subject. We consider each individual human being to be unique and inviolate. From this perspective, it may feel wrong to choose or alter what someone *is supposed to be*. The technology to select embryos based on information about their genomes already exists and some prospective parents use it to ensure that their children do not carry specific disease mutations. In western society, views on embryo selection are likely to range from abhorrence of the very idea to an acceptance that gene-based selection has a place, given the right circumstances. Eliminating a known disease-causing mutation is a long way from designer-babies, but the newly available potential for at-will genome manipulation does bring a whole new dimension to these ethical issues.

In the novel *The Unedited*, I have assumed that the technological hurdles to human genome manipulation have been solved. There are number of factors implicit in this, but, as it stands, the underlying technologies are not far from the mark. Reproductive technologies, which generally depend on in vitro fertilization (IVF), are well established; for non-human mammals, this includes cloning and twinning. Genome sequencing is now fast and inexpensive, and the tools for individualized genome analysis are advancing rapidly. Precise genome editing is now also possible; there are some caveats, but technical improvements will continue. Less certain is to what extent we can predict the biological outcome of specific genome manipulations. The novel explores these areas in a fictional format. In the following, I will discuss the factual context: what is currently possible and what might be possible in the not-so-distant future.

Reproductive Technologies

IVF

The first baby conceived by IVF, Louise Brown, was born in 1978. Since then, the technique has gained wide acceptance and million of babies have been conceived this way. The original purpose of IVF was to enable people who, for one reason or another, have a difficult time conceiving to have children. Given that sperm normally swim around to find an egg and that the first one to make good contact fertilizes it, it is not surprising that the process can be recapitulated in vitro, in a petri dish. Sperm tend to be plentiful and can easily be frozen and stored for later use. While there are some ethical issues with using frozen sperm, these are fairly limited in scope: How much should prospective parents (and offspring) know about a donor? What if the donor dies? Etc. Possible use of surrogacy adds more ethical complexity and makes for popular storylines, but will not be discussed further here.

For IVF, the limiting factor is usually fertilizable eggs. The human ovary contains many immature germ cells, but just one or a few mature each month to be available for fertilization. Hormonal treatments to induce hyperovulation can provide more eggs, making the process more practical and improving the chances of success. But the numbers remain quite small. Technologies for cryopreservation and transplantation of human ovary tissue have been developed to help treat unwanted consequences of menopause. It is not hard to imagine that further developments along these lines might allow for organ culture, and possibly tissue expansion, to make fertilizable eggs both simpler to access and more plentiful. If neither sperm nor eggs are limiting, the issue of choice changes fundamentally as discussed in the end of this section.

Cloning and Twinning

The current uses of reproductive technologies in humans are quite restricted, but this is not the case for farm animals or animals used as model systems in research. Most people do not consider animals to have individual rights. Manipulation of animal genomes is therefore considered ethically unproblematic, unless it causes excessive suffering or somehow ends up posing a threat to human health. Dolly the sheep captured the imagination and raised public concern because of the implications for use of cloning technologies in humans, not because people lament sheep being genetic copies of one another rather than normal offspring.

Dolly became famous in the 1990s for being a cloned animal, but she was not the first. In 1970, John Gurdon cloned a frog from the nucleus of an adult cell. It only made it to the tadpole stage, but this was a key step in showing that a differentiated adult cell can be reprogrammed to become pluripotent and give rise to a whole animal. The information required, primarily genomic DNA, is present in every cell (with a few exceptions). Many years later, Shinya Yamanaka defined a specific set of proteins, or regulatory factors, that can do the reprogramming, thus generating so-called induced pluripotent stem cells (iPS cells). Gurdon and Yamanaka got a Nobel prize for this work in 2012 (Colman, 2013). After Dolly, many other animals have been cloned, including domestic pets.

The possibility of applying these tools to clone a human is not far-fetched. Reprogramming of adult cells is one possibility, but they may not achieve the required level of pluripotency without side effects. Immature embryonic cells are another option, in which case we may be dealing with twinning rather than cloning. For many mammals, including humans, identical twins occur naturally. The embryo-to-be starts out as a fertilized egg—the first cell with the new, complete genome—which divides to make more cells. In the beginning, the resulting cells are indistinguishable from one another. Normally, they are kept together and cooperate to form one individual. Occasionally, they split apart and, if two separate entities develop, make identical twins. In nature, this happens rarely, and seldom produces more than two identical offspring. In the laboratory, it is possible to deliberately separate cells at the early stage, which can then be used to produce identical twins. Deliberate twinning has been done for cattle, for example (Hashiyada, 2017). In the case of human embryos, cells may be recovered as in the twinning setup or by procedures related to the blastocyst biopsies currently used for genetic testing at IVF clinics. Blastocysts (early embryo cells) and many other cell types withstand freezing quite well, not just sperm. In other words, such cells may be used long after they were collected. If they were to be implanted to produce another child, this child would be an identical twin of the first embryo. In my novel, I call children produced in this way "*youngtwins*", and explore the reasons why parents, or indeed the first-born twin, might choose this route.

It may clarify the ethical debate if we simply think of a clone as an identical twin—a unique individual who has the same genetic makeup as another individual. The potentially more frightening aspect of cloning, namely making many copies rather than just one, will need separate consideration. Embryonic stem cells, which retain most of the properties of a fertilized egg but can be expanded to make a large number of genetically identical individuals, are routinely used to make genetically modified mice. Human embryonic stem cells

do exist, but are obviously not used in the same way and may not have exactly the same properties (Ilic and Ogilvie, 2017). Reprogramming of adult cells is theoretically also relevant in this context.

One possible rationale for cloning or twinning is to generate a person as closely resembling an already known person as possible. It could be a uniquely talented person or a much-loved person, possibly deceased. It is easy to imagine complications of "ownership" and decision-making power for frozen embryonic cells. If adult cells are used, this is simplified: Your cells belong to you, not your parents. Ownership is relevant for another possible rationale for cloning—a medical one: A clone could be a body copy of yourself, for spare parts. If we are talking about an actual person, or even a late embryo, this is clearly ethically unacceptable—though the idea has been considered in many speculative fictions. But what if only certain tissues or organs are made? Early embryonic cells can in principle be directed to form anything. There has also been considerable progress in growing tissues from adult-derived stem cells. From simple tissues to functional organs is a significant step. But it is not impossible. Current ideas include the use of human/nonhuman chimeras to host the growth of these cells into fully functional organs for transplantation (De Los Angeles et al., 2018). Chimeras, being partially human, open additional and complex ethical issues.

Selection and Choice

Choice remains a major—and contentious—issue in human reproduction. Thanks to modern contraception, having sex no longer has to mean the risk of producing a child. Contraception is ethically uncomplicated for many people. This is fortunate on a personal level but also for our planet, challenged as it is by the hugely expanded human population. Abortion is a more complex issue. Our society currently accepts decisions to abort a fetus for one of two reasons: either the woman does not want the pregnancy *and* the fetus has not developed very far—or the fetus has a severe defect, such as a chromosome abnormality or a mutant allele of a gene known to cause disease. With IVF, it is possible to identify embryos with genetic defects beforehand and choose not to implant them. This is accepted and done. We do not currently accept the idea of positive selection of embryos, for example based on sex. In the future, couples wishing to become parents may no longer accept such strict constraints on their ability to choose. Imagine if current bottlenecks of reproductive technology are overcome and a woman could have twenty, fifty or a hundred well-developed eggs available for fertilization. The sperm do their job

and there are now many zygotes to choose from to initiate one pregnancy. The future parents might want to avoid alleles linked to high cancer risk or early onset dementia. More controversially, there might be traits that they want to select for. If full genome information is available, it may be tempting to use it to make active choices.

Reading the Genome

The human genome is large—about 3.2 billion base pairs divided into the 23 chromosomes of a haploid cell. The genome is written in a four-letter code with the bases G, A, T and C. This code is interpreted in many different ways, simultaneously, by the cell. Scientists are very good at reading the parts that encode proteins. This is roughly 1% of the total base pairs, making up a total of 20–25 thousand protein-coding genes. These regions of the genome and the proteins they encode are close to identical in all humans. There are, however, some naturally occurring variants in the population, which may be disease-related or not. In addition, a small number of proteins are hypervariable, such as those encoded by the major histocompatibility complex (MHC). Proteins are generally very similar in related animals, as well, with a significant fraction being identical in humans and chimps, for example. Other features of the genome—conserved, non-coding genes and certain types of regulatory sequences—are also quite well understood. But there is a lot of DNA in the genome, much of it transcribed into RNA, for which we do not know the function. Some of it may have no function whatsoever: remnants of viruses or transposons or other passive "junk". Much of it, however, is likely to have some regulatory effects, determining how much of a certain gene product is produced in which cells, at which time and in response to which stimuli. In this way, such genomic regions are also part of the code that determines how a complex organ like the brain develops and works. Although regulatory functions are important, they are difficult to decipher directly from the DNA sequence, as they are not read by the cell in as rigid a manner as the protein coding parts are. What *is* clear is that much of the variation between humans and other species, as well as between individual humans, can be found in these less-well-understood areas of the genome. So, while sequencing the first complete human genome—The Human Genome Project—and identifying all the protein-coding genes was an important landmark, sequencing many more individual genomes will continue to be useful for a whole different set of reasons.

One of the practical byproducts of the human genome project is the development of ever faster and cheaper DNA sequencing technology. Until fairly

recently, state of the art DNA sequencing meant using the enzyme DNA polymerase coupled with a neat trick of modified building blocks to read longish strings of DNA—typically 500–1000 base pairs. New technologies, called next-generation sequencing (NGS) or massive parallel sequencing, produces a very large number of short sequence-reads at once, which are then assembled and often aligned to a reference sequence by computer (van Dijk et al., 2014). Acquiring the first complete genome sequence of a species is by far the hardest, as it requires fitting together many bits of DNA sequence from scratch and orienting them correctly. Subsequent sequencing of other individuals is much easier because the reference sequence can serve as a scaffold. But, even so, it is remarkable that while sequencing and assembling the first human genome cost about 500 million US dollars and took several years, it now costs about $600 to have your genome sequenced by a commercial sequencing provider and, in principle, takes just one day. A gradual decrease in sequencing costs had been anticipated, but the magnitude and speed of the decrease might not have been. We are now suddenly in the "individual genome era", with all the possibilities and problems that this entails.

So, it is now possible to read the full genome code of every single person on a routine basis—and to store this information. How do we make use of that? One practical application is unambiguous identification, for example for forensics. Given the size of the genome and the presence of variable, even hyper-variable regions, this "fingerprint" can in principle tell each and every one of us apart. Another use of genome-based identification is to assess biological family relationships. This includes straightforward issues like paternity as well as more distant ancestry. The widespread availability of human genome data is having interesting consequences. Some long-sought criminals have been identified through DNA relationships with living relatives. Family trees—in particular the male line—can be reevaluated and unexpected contributions uncovered.

Another use of personal genomes and one that is most relevant for this discussion is identifying gene variants associated with diseases and other traits. Monogenic diseases caused by mutations in a single gene, such as sickle cell anemia or cystic fibrosis, are numerous but relatively rare. Many common diseases, including diabetes, neurodegenerative diseases and various types of cancer, do however have a clear genetic component. Gene variants can be identified that confer high or low risk of that particular disease; they may also correlate with specific treatments being more or less likely to be effective. Disease genes, or more correctly, disease alleles, come with variable degrees of knowledge attached. Sometimes the exact way in which a mutation affects a gene product and a cellular process is known. There may also be direct evidence

of causality—for example if the mutation has been studied in an animal model system and correcting the genetic defect was shown to eliminate the disease. At the other end of the spectrum, the link between a DNA sequence and a disease may be based solely on statistics. A sequence variant can be associated with high disease risk without being in any way responsible for the disease—for example by being close to something else in the genome. In that case, fixing the variant would obviously be pointless. In principle, extensive whole genome sequencing allows simultaneous and unbiased correlation of all gene variants with all diseases and traits. It does, however, depend on the appropriate medical diagnostics or trait measurements being performed. As more and more individual genomes are sequenced and analyzed in this manner, even complex associations involving multiple interdependent genes can be uncovered.

Personal genome data give a lot of information about personal risk. As with other rich sources of personal information, this brings forth ethical issues related to privacy, in particular in countries with insurance-based systems for healthcare. From a commercial point of view, statistically demonstrated high risk of disease is just like any other liability: something an insurer will want to avoid taking on if at all possible. Iceland and a Reykjavik-based company, deCODE genetics, has become famous for collecting knowledge about the genomes and genetics of the whole population (Halldorsson et al., 2019). Fortunately, Iceland has a universal healthcare system. Aside from healthcare, there is the question of how much individuals may want to know about disease risks implied by their personal genome, in particular if there is nothing they can do about it.

Finally, returning to the issue of reproduction and choice, genome data for an embryo gives just as much information as for an adult. With sensitive sequencing techniques and prior knowledge of the two contributing genomes (the parents), it should be possible to get this information from one or a few cells. This means that it will be possible to have the complete genome read prior to implantation if using IVF. In other words, even without considering active genome manipulation, there will soon be scope for highly informed and directed choice about which of the potential children to have, should that be desired by parents and allowed by society.

Manipulating the Genome

As with many technological advances in biological sciences, methods for directed genome manipulation have arisen from scientists studying the basic biology of what cells and organisms normally do and how they do it.

Our genomes are used as operational code in every single cell of our body. This includes very long-lived cells like neurons of the brain as well as short-lived blood cells that patrol the body to catch invaders. Despite the passage of years for long-lived cells, and many rounds of genome duplication in rapidly dividing cells, the genome remains (essentially) the same. Epigenetic changes may affect its readability but the DNA sequence remains. This is not because DNA, as such, is unchangeable. It is easily damaged by UV light and by reactive chemical agents, including endogenous metabolites generated when cells "breathe". Errors also occasionally occur when DNA is copied to prepare for cell division. In normal cells, a dedicated DNA surveillance system detects damage and errors; it then promotes the necessary repairs (Jackson and Bartek, 2009). The genomes of many cancer cells accumulate an excess of mutations over time because this system is defective.

DNA strands can also break, and rejoin. In a controlled manner, this process is responsible for homologous recombination, or exchange, which is used to make the unique haploid genomes of germ cells. It is also required to make the diverse antibodies of the adaptive immune system. Reactive chemical agents can lead to unintentional DNA strand breaks. The cell can repair such breaks using matched DNA from the intact strand or from a homologous chromosome as a template. In addition to the risk of breaking up a gene, a double-stranded break, where both strands of the genomic DNA are broken at once, poses a particular problem because of the way chromosomes are partitioned in dividing cells. For this reason, active mechanisms exist to repair these breaks by rejoining the free ends. This may or may not involve using matched DNA as a template.

In the lab, scientists have found ways to make use of the cellular DNA recombination and repair machinery to deliberately modify the genome. The first generation of this technology involved simply providing cells with a modified exogenous template DNA, which could be used for homologous recombination with the host chromosome and at some low frequency replace the host sequence. The process is quite inefficient, in part due its dependence on a random DNA break to initiate it. Despite this, the existence of mouse embryonic stem cells and optimization of their care have allowed scientists to engineer a large number of genetic changes in mice. This includes inactivating,

or knocking-out, individual genes in order to determine their biological role (see for example White et al., 2013). While more challenging, it is also possible to use this method to precisely engineer DNA sequence changes, in other words, to edit the genome. This has often been done to produce mutations in mice that match disease-associated mutations in the human genome and thereby make the most accurate mouse model for that particular disease.

So, directed genome editing is not new; it has been done in mice and other organisms for quite some time. What has changed is the ease with which it can be done. CRISPR/Cas is a bacterial immune system that protects the bacterial genome from invading viruses and plasmids by recognizing their DNA as foreign and cleaving it (Garneau et al., 2010). Genome editing using CRISPR/Cas is powerful and efficient because the CRISPR/Cas enzyme can be targeted to any predefined spot in a genome by the enzyme's RNA component (Jinek et al., 2012). Once there, it will make a break in the DNA, which can then be repaired by the cell. If the goal is just to inactivate the target gene, scientists can take advantage of the naturally occurring non-homologous end joining to create mutations. If a precise editing event is required, exogenous DNA that carries the desired change may be introduced to allow repair via homologous recombination. CRISPR/Cas technology is now being widely used to modify the genomes of cells and model organisms. While it is a major technological advance, there are still some issues with accuracy and unintended consequences that I will discuss in a moment.

Before getting into the safety issues, I would like to point out that there are two different ways in which genome editing technology can be used in human cells. The ethical issues for the two differ substantially. The first and most consequential use would be to edit a fertilized egg or very early embryo. This would, in principle, change every cell in the body of the resulting child and be passed on into future generations. A second use—somatic editing—would be to edit cells from a consenting adult with the intent of introducing them back into to the individual they came from. This could include adult-derived stem cells to correct genetic deficiencies, possibly via in vitro-derived organs, or modified immune cells directed to kill specific cancer cells. If such manipulations cover a significant medical need, one that is currently untreatable or not effectively treated, many people would probably be in favor of it, despite it being a form of human genetic engineering.

The safety issues associated with genome editing must be considered carefully before proceeding with even the most benign-looking application of it. There are two main concerns. The first is accuracy. Basic CRISPR/Cas-mediated DNA editing involves DNA cleavage as well as homologous recombination with exogenous donor DNA. It is known from model systems that

either of these processes can lead to unintended changes elsewhere in the genome. More recent method developments, such as base editing and prime editing (Anzalone et al., 2019), allow genome editing without double-stranded breaks or donor DNA and appear to have a lower frequency of unintended changes. In either case, genome sequencing can, and should, be used to check that editing has occurred as planned and that no other changes have been introduced into the genome. Sequencing of the whole genome is essential if the aim is to edit an embryo, but may also be needed for somatic applications.

The second safety concern is unintended consequences of a more subtle kind. This is the tricky bit. Our knowledge of how the genome is used to make a healthy, thinking, feeling person, with all the complexity this implies, is rudimentary. We can make educated guesses about the consequences of altering a gene in humans from studying animal models. But even in well-studied animal models, it is often impossible to predict with certainly what will happen (this will be explored in more detail in the next section). If, however, an allele or variant of a gene is already common in the human population and is not associated with a disease or defect, then it is probably safe to introduce that allele into a specific cell. In practice, this means introducing a normal allele in place of a mutant one, such as CF or BRCA1 mutations, or shifting from one common variant to another. If an allele preexists in the population, we know whether it is basically functional or not. If it is a common allele, we also know how it behaves in combination with many different variants of other genes, which adds confidence. Every other change to the genome is a gamble and it will come down to risk versus reward. For a consenting adult suffering from a devastating disease it may well be worth taking the chance of editing his or her own stem cells and reintroducing them. For reproductive editing, it is much harder to disregard the potential risk, including for subsequent generations. There would have to be a very compelling reason to do it. In the fictional world of *The Unedited*, genome editing is actively pursued because it is the only way to ensure that a particularly vicious virus does not infect people. This scenario is not completely unrealistic, but let us hope it never comes to that.

To close this section, I would like to return to the case of the Chinese twins and the rationale behind that experiment. The twins had one or both of their CCR5 genes inactivated by CRISPR/Cas-9 activity. The stated aim was to make the twins resistant to infection by HIV, which their father carries. CCR5 does encode part of the landing pad for HIV, and people with deletions in their CCR5 genes already exist in the population at about 1% frequency, so the risk of unintended consequences—at this site—could be argued to be

limited. Inactivating a gene also requires less accuracy than editing it to make an altered but functional gene product. Thus the approach per se may have technical validity, assuming the twins do not acquire other, unintended mutations as a consequence of editing. By sequencing the twins and their parents, this is in principle knowable (if a bit late). Technical issues aside, other measures could have been used to alleviate the risk of HIV infection. So, not only was the manipulation illegal, it was also non-essential and for that reason not worth the risk of unintended consequences for those two children. Whatever our views on the specifics of this case may be, the fact of it did push human genome editing back into the public discourse—and rightfully so. The possibilities enabled by the new technology may create demand. We need to think carefully about the ethical issues at all levels to allow rational decisions about where society should go with it.

Predicting and Designing

One of the more emotionally weighted phrases used in discussions about reproductive choice is "designer-babies". Does it reflect anything real? In principle, we could eliminate known disease mutations and switch some other gene variants around. But can we actually design anything, given that we are working with an enormously complex genome that we barely understand? Evolution has created an amazing diversity of plant and animal species displaying a vast array of traits, abilities and peculiarities (Gould, 1980). But evolution is messy; it works with what happens by chance. The genome of any species, including our own, was not designed with every element serving a precise, well-defined function, like in man-made objects. Subtle variations in the genome lead to subtle variations in traits of individuals, but again, not in a predictable, structured way. How do we untangle this?

Leaving aside the issue of design for now, let us start by considering how much we can expect to predict in terms of traits or phenotypes from a given personal genome. By traits I mean both negative traits such as risk of specific diseases and positive traits like strength, beauty, intelligence, musicality etc. The first question is how large the genetic component of a trait of interest actually is, as opposed to the contributions of environment and pure chance. The gold standard for determining this is, or used to be, identical twins reared apart—same genome, different (post-natal) environment. If they share a rare or very specific trait, it probably has a strong genetic component; if they do not, it probably does not. Given how rare identical twins are, and even rarer that they are raised apart, using this comparison as a standard has its limitations. But it is obvious from this approach that certain traits, even if complex,

have a very strong genetic component—facial features, for example. Other family studies as well as genome-wide association studies (GWAS) can also give information about the strength of genetic components. My impression is that most traits, including personality and abilities, have *some* genetic component. Just ask a person who is adopted. This does not mean the trait is predetermined, merely that it is influenced by the genetic background. Speaking about disease-risk in terms of genetics is not contentious. Viewing other traits, such as IQ or creativity, in the same manner is not to everyone's liking. But in order to have an informed discussion about future uses of whole genome data and possible genome manipulations, it seems wrong *not* to consider this aspect.

Assuming that there is a genetic component to a trait, how easy is it to identify the contributing genes? This depends on the magnitude and the complexity of the genetic contribution—the larger it is and the fewer genes involved, the higher the predictive power. The two are not necessarily linked: Eye color is genetically simple, facial features are not, but both have a strong genetic component. It also depends on the number of individuals who have had their DNA analyzed, ideally their genomes sequenced, *and* this trait reliably measured. Genome sequencing is expanding and the data are accumulating rapidly. Trait measurements are more variable. Physical characteristics are the most straightforward, along with well-defined disease states. But even these tend to vary over time and require an objective observer or an objective test. Certain tests, like IQ measurements that give a number, seem straightforward but are not. People can improve their scores by training; that does not mean they become smarter. Finally, measurements and genomes need to be reliably paired. For diseases, doctors, clinics and lab-tests are responsible for objective diagnoses and patients generally cooperate because it is in their best interest. But wider use of this information requires good recordkeeping and external access to the data, which has ethical complications as discussed previously. Keeping genetic tests and diagnoses connected but dissociated from patients' names or ID numbers can in principle help maintain anonymity. But as we shift toward using complete genome sequences as DNA data, this may break down. The personal genome is, after all, the ultimate identifier. In conclusion, if a trait has a high genetic component *and* it is properly recorded, genome sequencing may allow reliable prediction of that trait in the future, even if it cannot do so now. It will, however, remain a matter of statistics and rarely a certainty.

Can we predict the consequences of deliberately altering the genome? That is, can we predict the impact of introducing a specific change in the context of a specific, personal genome? Can we re-design a person? That depends on what is meant by design. Correcting mutations by reintroducing wild type

alleles or substituting variable alleles may soon be possible with existing technology, has limited risks and fairly predictable outcomes. But this stays within the range of naturally occurring variation between humans and hardly qualifies as design. What about *real* changes? By this I mean introducing changes to the genome that are *not* already present in the population. We can call these genetic novelties. Such changes might be tempting if they seem to be extremely favorable. Imagine inactivating a gene that is directly involved in the development of Alzheimer's or other dementia; imagine a simple mutation that is likely to increase active lifespan. Making more drastic changes to our biology, what we might call *de novo* design, seems unrealistic on the pure biology side. We simply do not understand genome logic and its translation into real-life biology well enough to design reliably that way. One type of "design" merits a separate mention: copying. In the case of a delayed identical twin or a clone, we have a pretty good idea what the outcome would be. Twinning or cloning humans is likely to become technically feasible, in one form or the other, in the near future. The question is whether there will ever be public interest in making it legal. Another route that limits possible unforeseen consequences is what I previously termed somatic manipulations. Any change is restricted to the particular cells being manipulated and the ethical concerns are minimized as only the person agreeing to the treatment is affected. Such genome editing approaches seem likely to move forward and may eventually become quite sophisticated.

Any genetic novelty that one considers introducing into the genome of a human embryo should be checked in a model system first. This seems reasonable. One of the reasons we study genetics of model systems such as mice, fruit flies and roundworms is to understand the contribution of specific genes and gene products to the function of the organism as a whole. No cell line, tissue or even whole organ grown in culture can address this adequately. The genome did, after all, evolve as a blueprint for the whole animal, not for isolated cells. Genetics continuously throws up surprises, however, even in the best-studied model systems. This is true for forward genetics, where scientists look for a gene responsible for a trait/phenotype of interest, as well as for reverse genetics, which looks at the consequences of manipulating a predefined gene. An example of the first kind is the discovery of a whole new class of non-coding genes, micro-RNAs, from changes in the timing of differentiation events in the roundworm (Lee et al., 1993). An example of the second kind is inactivating the mouse gene encoding uPA (urokinase-type plasminogen activator). Interest in uPA comes largely from its association with cancer progression and tissue remodeling. The mutant mouse, however, is timid: it shows reduced exploratory activity (Rantala et al., 2015). In fact, many mouse

mutants have unexpected phenotypes (White et al., 2013). Also, the same mutation introduced into different inbred mouse strains can have surprisingly different effects (Threadgill et al., 1995). In humans, this corresponds to the same variant of a gene having different effects depending on the genetic background, that is, which regulatory landscape and which variants of other genes are present in a person's genome. Such interdependencies may turn out to be the rule rather than the exception for naturally occurring mutations and the traits that we care about.

In conclusion, the effects of genetic perturbations are often not what we would expect them to be. Add to this that any change made to the genome of a zygote will be present in each and every cell of the body, including the brain, and we face what may be an insurmountable challenge. No matter what model system is used for testing, it will never be possible to ensure that a genetic novelty will not trigger unintentional neurological or psychological effects—psychosis, depression, inability to learn syntax of a language or some other quirk of the brain. We cannot ask mice or flies—or even monkeys—how they feel. Nor could we trust their answers to reflect how we would feel.

Concluding Remarks: Choice Versus Chance

Technological advances give us possibilities. Fast, inexpensive whole genome sequencing is one such advance. Simply having this information, in particular for that little ball of cells that could produce a whole new person, presents any number of challenges from an ethical perspective. It is all about choice. The possibility of manipulating—or editing—the genome brings even more possibilities and more choices. We, as individuals or as a society, will be responsible for the consequences whether we act on the choices or not.

A fictional example illustrates how challenging this can be: If, based on a large-scale IQ test, it is found that individuals with the allele combination a4a4, b2b2, c1c4 have a 40% of having an IQ > 145 (exceptionally gifted) whereas for a4a4, b2b2, c1c1 it is 0%, would you be tempted to implant a c1c4 embryo over a c1c1 embryo if given the choice? Or would you decline to know and effectively toss a coin? What if the choice is about 0% versus 40% chance of having an IQ < 35, that is, severely impaired and likely to need lifetime care? Would you then choose the c1c4 embryo? Would society always pay for care if you didn't? In either case, would you choose to edit a c1 allele to c4 (assuming it is technically safe) for an embryo? What if we substitute IQ for creativity or freedom from depression or low cancer risk? What if you choose not to know or not to edit and the affected child realizes this later on?

In the novel *The Unedited,* I explore questions like these via the fictional lives of five young characters. This is not just future and fiction, however: fertility clinics are already offering the choice to select against certain genetic variants. Currently these choices are limited, but then sequencing the whole genome has also only recently become easy and inexpensive.

Choice versus chance—that is the question. In most areas of life we think choice is best. In reproduction, having children, it gets complicated. A new child is an unknown, a person created by two people plus a hefty dose of chance; this seems natural and comfortable to us. Some parents-to-be even prefer *not* to know the sex of their child before birth. So, it is possible that there will never be any public pressure to make complete genome information available and allow informed choices—beyond what is considered critical for a child's wellbeing. Nazi eugenics casts a long shadow in our culture. But attitudes may change as genome-based predictions become more sophisticated and as new diseases or new viruses arise. We are already close to this dilemma: Is it ethical to bring a child into this world if she has 70% chance of breast cancer or early-onset Alzheimer's or severe depressions? Also, people differ. Some parents are likely to want as much information and control as possible. IVF clinics offer some DNA screening of pre-implantation embryos; sending the DNA to a whole genome sequencing facility is a small step. These questions and issues are unlikely to ever be easy, for the individual or for society. But they will, hopefully, be at the center of a sensible debate moving forward.

Bibliography

Anzalone, A. V. et al: Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. doi:https://doi.org/10.1038/s41586-019-1711-4 (2019)

Colman, A.: Profile of John Gurdon and Shinya Yamanaka, 2012 Nobel Laureates in Medicine or Physiology. Proc Natl Acad Sci U S A. **110**(15), 5740–5741 (2013)

Cyranoski, D. and Ledford, H.: Genome-edited baby claim provokes international outcry. Nature. **563**, 607–608 (2018)

De Los Angeles, A., Pho, N. and Redmond, D. E. Jr.: Generating Human Organs via Interspecies Chimera Formation: Advances and Barriers. Yale J Biol Med. **91** (3), 333–342 (2018)

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath P., Magadán, A.H. and Moineau, S.: The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. **468** (7320), 67–71 (2010)

Gould, S. J.: The Panda's Thumb. Publ.: New York: W. W. Norton. (1980)

Halldorsson, B.V. et al.: Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. **363**, (6425), eaau1043 (2019)

Hashiyada, Y.: The contribution of efficient production of monozygotic twins to beef cattle breeding. J Reprod Dev. **63** (6), 527–538 (2017)

Ilic, D. and Ogilvie, C.: Human Embryonic Stem Cells-What Have We Done? What Are We Doing? Where Are We Going? Stem Cells. **35** (1), 17–25 (2017)

Jackson, S.P. and Bartek, J.: The DNA-damage response in human biology and disease. Nature. **461** (7267), 1071–8. (2009)

P. Rørth, *The Unedited*, Science and Fiction, https://doi.org/10.1007/978-3-030-34624-9

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A, and Charpentier, E.: A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. **337**(6096), 816–821 (2012)

Lee, R. C., Feinbaum, R. L. and Ambros, V.: The *C. Elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. **75** (5), 843–854 (1993).

Rantala, J., Kemppainen, S., Ndode-Ekane, X.E., Lahtinen, L., Bolkvadze, T., Gurevicius, K., Tanila, H. and Pitkänen, A. Urokinase-type plasminogen activator deficiency has little effect on seizure susceptibility and acquired epilepsy phenotype but reduces spontaneous exploration in mice. Epilepsy & Behavior. **42**, 117–28. (2015)

Threadgill, D. W. et al.: Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. Science. **269** (5221), 230–234 (1995)

van Dijk, E.L., Auger, H., Jaszczyszyn, Y. and Thermes, C.: Ten years of next-generation sequencing technology. Trends in Genetics. **30** (9), 418–26 (2014)

White, J.K. et al.: Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes. Cell. **154** (2), 452–464 (2013)