BIOSCIENTIST
The Salford Biomedicine Society Magazine

An insight into genomics advancements, the future of whole exome sequencing: Dr Arijit Mukhopadhyay

Mary Yuhanna

Novel genomic research comes with very promising and revolutionising prospects in disease monitoring and treatment. The newest findings of whole exome sequencing and gene editing CRISPR technologies are major contributing factors to a newly found DNA profile-targeted approach in the growing field of personalised medicine. In this interview, Mary Yuhanna spoke with Dr. Arijit Mukhopadhyay, a reader in human genetics at the University of Salford about his passion for human genetics, dedicated interest in glaucoma neurodegeneration research, and the future of personalised medicine-driven National Health Service.

Mary Yuhanna: How and when did you become so passionate about genetics?

Arijit Mukhopadhyay: I do not think there was one specific day when I suddenly decided, it is complex. Interestingly, I always liked biology, but I also liked pure sciences: physics, chemistry, maths. When I had to choose my undergraduate subjects, I chose chemistry with physics and maths, and I went on to do a master's in pure chemistry as well. Then, my career started with working in a chemical industry as a research technician. After 18 months I decided to jump in the unknown and started a PhD in human genetics. My father is an anthropologist, I grew up listening to him, talking to his students about human pedigrees and human evolution, I am sure that played a role.

I grew up in India and was privileged to grow up in an academic family, which played a role in shaping my views. I was fortunate enough that I did not have to think about earning money to support my family. I want the young readers to know that it is completely random luck that I had those privileges. This made it easier for me to take risks without thinking about job security, otherwise it would not be possible.

MY: Most people in the biological field studying genetics or genomics tend to move toward RNA as it is a little bit more complex. What shifted your mindset from DNA to RNA?

AM: I cannot pinpoint a particular reason or a particular time or day. As a scientist when you do your PhD, which is a training in science, you understand how to ask questions, how to design experiments, and how to interpret it objectively. When you start to think about bigger questions, you realise that you are a small player in a big canvas. Then you start to think about bigger questions and in the biomedical space, you think about either understanding nature or solving problems, diseases in this case. Technically speaking, you start your journey with DNA, then you realise that DNA does not give you all the answers.

Measurable phenotypes are very complex outcomes of many factors. DNA is an important contributor, but it is just one contributor. After my PhD, I spent about six years part as a postdoctoral fellow, part as a principal investigator on DNA, trying to answer bigger questions based on DNA. With every project I increasingly realised that DNA would only answer so much and not beyond. I moved to the next level of complexity in biological system, which is RNA. If you think about the rest of your career as an academic scientist, you understand the limitations of your field, and then you try to incorporate more complex things

DOI: https://doi.org/10.57898/biosci.68

to get a clearer answer. When you realise DNA can do so much, you incorporate RNA. It is not one replacing the other, it is adding the complexity because your ability to handle complexities improves, and that is the natural transition, nothing special about me.

MY: Recently you published an article about using whole exome sequencing, which revealed novel candidate genes in familial forms of glaucomatous neurodegeneration. What is whole exome sequencing and how far can we take this in genetic, this genetic advancement in scientific research?

AM: First, glaucoma is a disease of the eye. It is a wide variety of diseases, not one particular phenotype, all these put together cause neurodegeneration. That is why the technical term for glaucoma is glaucomatous neurodegeneration. The paper is in glaucoma genetics. Exome means all the exons of all the genes put together. In biomedical science field these days, you hear a lot of words, which as an -ome at the end. Proteome, exome, lipidome genome. -Ome means all.

Coming from to the central dogma molecular biology, the DNA makes RNA and RNA makes proteins. For human genome, for our cells, that central dogma DNA to RNA to protein is the linear part of central dogma, there are non-linear parts as well. Only 2% of our genome makes proteins. When you deal with human diseases, the first place to start looking for problems are the proteins, which do not answer all the questions. That is why scientists figured out a way, back in around 2010, to capture only this 2% from the entire genome, and remove the 98% that is not coding for protein. Because of this technology, exons (which make mRNA and proteins) are all grouped together, that is why it is called an exome.

The human genome is 3 billion base pairs. 2% of that, about 62 megabases, that is what exome is. It has been around for more than 10 years now, and it is quite widely adopted. If you have a reason to hypothesise that the disease or the phenotype, you are trying to figure out genetically has a protein as a key molecule, then it makes sense to not look at the rest. That is why exome sequencing has been powerful. For the experimental part, it is not necessarily cheaper than whole genome sequencing because capturing this 2% is cost intensive, but the analysis is easier because you are only looking at 62 million base pairs instead of 3 billion base pairs.

Analysis wise, it is quicker. In the UK, the NHS regularly uses it as part of clinical service, not for all diseases though. In diseases where the phenotype is not obvious, where we do not know the genes, but it looks like a rare and penetrant disease, then doctors in the NHS can prescribe for exome sequencing. This would be done in one of the genomic services of NHS to get the data, so that that is not even futuristic; it is widely used across the world.

MY: Do you know, per patient, how much that costs?

AM: The way I do it as an academic researcher in the lab, and the way NHS will do it would be different. What I mean by that is for any omics type of work and the very nature of these technologies, are that if you do a smaller number of samples, per sample, cost is higher compared to the greater number of samples. When NHS does it, they do work in volume. When academics do it, they work on one family at set a time. For the paper you mentioned, we worked on total 30 individuals. That experiment happened three to four years ago. That experiment was done in India on Illumina platform. All these parameters are important because when you change the technology and the country, the costs change. This was done in India for a smaller number of samples, and that costed us approximately £3000 per sample.

MY: How far can we take this genetic advancement in scientific research currently or in the future?

AM: If you think that all these technologies, these are all at the DNA level. First, any genetic condition that is not on DNA, can be an epigenetics, RNA level problem, an environmental problem and would not be answered by exome sequencing. If the answer is in DNA, then it may or may not be in that 2% of the genome.
Diseases like cystic fibrosis, Duchenne muscular dystrophy, Huntington disease, are quite aggressive in genetic terms. We call it highly penetrant. If you come up with a patient with those phenotypes, genetically well-known phenotypes, and you know which gene to look at, then you do not have to go through the

DOI: https://doi.org/10.57898/biosci.68

process of capturing these 2% of the genome and sequencing them. You can use exome sequencing when the disease looks penetrant enough, but the gene is not known, so that it is rational to hypothesise that the mutation would be disrupting a protein function. When you are dealing with a condition with more uncertainty, the whole genome sequencing is a better approach, as it gives a wider search space to get the answer to your question.

MY: Some may state that due to the current genetic advancements such as CRISPR/CAS9 technology and understanding of the epigenome, we are heading towards a disease-free future by 2050. Do you agree with this statement? If yes or no, why?

AM: It is not going to happen. I think it should be mandated that everyone reads the book called 'The Code Breaker.' 'The Code Breaker' book came out in 2021. It is on the CRISPR discovery, from the point of view of Jennifer Doudna. If you said 100 years, I'd probably have said yes or 200 years, I would probably have said yes. By 2050, which is 25 years, probably not. We now have a technology where we can change any nucleotide on the DNA. We learnt only to rewrite the code post 2012. The technology is getting perfected, you might change something else, inadvertently. Unless it is perfected to that level, we would not be able to cure all diseases.

Known inherited diseases mentioned before will probably get cured through CRISPR in 25 years. We understand how to read those pages of the genome so well that we feel comfortable rewriting them. CRISPR, as it stands now, is now being applied to RNA. In addition to DNA, CRISPR is now starting to come out that they can edit epigenome as well, but these are much earlier in the process of a perfection. They will take longer; most diseases will have an epigenetic component.

For sickle cell anaemia, thalassemia, it is quite advanced. It might happen in next five years that it is quite advanced across the world, not only in the UK or US. Many places in the world are doing that, so that will happen sooner for HIV. There were efforts where this has been successfully done in the stem cell. If you think of applying CRISPR to cure diseases, it depends on at what level you detect the disease. If you are dealing with an adult who has trillions of cells in their body, and if they have an inherited mutation, which is presenting all cells of their body, and if they suffer from a disease that affect multiple organs, then where do you correct using CRISPR? Conceptually speaking, the best outcome of a gene editing technology for curing the disease would come if you can cure at the stem cell stage because then all the cells will be cured, the corrected genome. That will also play a crucial role in which disease this can be cured and in which it cannot.

We are not far from a time where you can buy a gene editing kit on a super supermarket shelf, people will be able to do it at home. It will be as simple as at administering insulin onto your body. It will be as simple as doing a pregnancy test or doing Covid test at home. These are also biology experiments if you think about it, but we have made it so streamlined that people can do it at home. As this famous line goes, that with greater power comes greater responsibility. Now, where do you draw the line? Who decides that? Does a parents get to choose and change their foetus? Designer baby? Where do we draw the line so that we do not fall in the trap of eugenics again? You are trying to see that before CRISPR is widely used to cure all the problems we must first define what is a problem and what is not a problem. The ethics and the legality of it must catch up before these can be utilised for greater good so that we do not create monsters.

MY: The NHS has completed the a 'Hundred Thousand Genomes Project.' Their aim was to be able to have genomic sequencing as part of their healthcare routine. What do you think about the concept of personalised medicine and what impact will it have in the future?
AM: What I think about personalised medicine I think that should happen. Consider siblings, we have all seen you both take a tablet of medicine, and you develop an allergy, and your sibling does not. The idea is that you would react differently to an exposure, or you react differently to a drug, which is also an exposure is quite common. I do not think it is practical to think that there will be medicine designed for you as a person.

What might happen is, or what is already happening, and that will get better with time, is that rather than giving a generic drug for a generic condition, will be able to subgroup people into their predispositions based on their genetic makeup. We would be able to tell, this group of people is allergic to this drug, so we

DOI: https://doi.org/10.57898/biosci.68

should be able to check from the genetic data and then give an alternative drug. It is personalised, but as a group, rather than a single person. In another scenario, there might be a future where you go into a clinic, they investigate your DNA profile and your other biological parameters, and then maybe there is a drug 3D printed for you. 3D printed medicine, 3D printed organs are happening now. These are possible, but I think it is more likely that we will be able to form smaller groups of people who would behave similarly given an external exposure either to a drug or to an environmental exposure such as Covid, for example. In the NHS led 100K Genome project, the summary of that vision is that NHS foresees a future where genetic information would be the first layer of health management.

The project has not only sequenced whole genome, but they have also done whole exome for a part of the samples. There are different layers of data, and all this data is part of UK Biobank, which researchers can access for a fee. Now, there is a newborn screening programme announced as part of Hundred Thousand Genome Project. I foresee a future where every child soon after they are born, the parents will their genome information in a card, like a credit card with a microchip. I dream of a future (albeit with my personal bias) where when a person goes to a doctor, a swipe of that card will tell the doctor of all the disease predispositions encoded in their genome. The major hurdle is the ethics of consent, confidentiality, and ownership of the data. We are entering an exciting era in healthcare where it will be data driven, from multiple sources, and artificial intelligence will integrate it together to get a better understanding of health. My fantasy is that you will have a kitchen top device next to your toaster and the microwave, where you can occasionally check your DNA and RNA for epigenetic changes.

MY: Do you think that it will happen? Are you saying that it should happen?

AM: I think it will happen. The companies like Illumina, who holds more than 90% share in the genomics market, has a vision to make genomics a consumer item like any other daily consumables such as toothpaste. It means they want to get it to every household so that public does not have to deal with this complexity of talking to researchers and people like me who talks in language that most people do not get, so that barrier is removed. I am not saying it should happen, I am saying it will most likely happen.

The ethics will become increasingly more important with increasing access to genomic technologies. What do you do with the data? We need more genetic counsellors who will be that bridge between lab scientists or clinicians and the members of public who can alleviate the anxieties, who can tell them what to do, what not to do, and how not to extrapolate, how not to sensitise themselves with the information. And even before that, empower the public to choose if they even want to know what their genome holds? Those things need to be figured out before it reaches a consumer level. Technically, we are there now, but ethically we are far away. We have sequencing machines that are smaller than a smart phone.
If a baby has a genome sequenced, who should know that information?' Should the future employers know? Should the insurance companies know? The reason I emphasise on the babies is that people who cannot decide for themselves, they are the most vulnerable. The same applies to the people who are dead, who should decide?

British Society of Genomic Medicine is the formal organisation in this country who informs parliament and the ministries about the policies, guidelines, and changes. Back in 2019, they published a new guideline for ethics of genomics data. I was fortunate enough to part of the sub-committee who drafted that guideline. I went through the process of the debate and discussions what should be there or should not be there. It is not that we reached a definitive answer, but the point I am trying to make is that debate needs to be continued across all stakeholders, clinicians, scientists, students of science, public, everybody should have a say on these. Until we have figured it out a situation or everybody is reasonably satisfied. Until then, we are not ready to use that technology.

DOI: https://doi.org/10.57898/biosci.68

MY: You spoke about obviously personalised medicine and how much data we are getting from the NHS now. Do you think currently we have enough people to look through and analyse this data? And if not, why is that?

AM: The answer is no, we do not. Currently for 64 million people, we have less than 500 genetic counsellors only in this country, most of the other countries are even worse. For your question about data analysis, it is even less promising now. What we know now is that genomics is more data science than biology. The reason I see it that way is because most of the data generation part, which is the biological experimental side part, has now been automated.

Now, you do not need a human being to operate a genomic instrument, a robot can do the analysis, most of it; the basic analysis is also automated. It happens in the cloud with AI but when you are trying to explain a phenotype, which is still a very human driven side of the process. We do not have enough people who understands data and can deal with complex data parameters from a computational point of view and understands biology. Biologists cannot do it. Data scientists cannot do it. We need people skilled in interdisciplinary areas such as genetic counsellors and genome informaticians.

MY: Finally, do you have a thought-provoking question for the audience?

AM: Just because we can, should we sequence everyone's genomes?

DOI: https://doi.org/10.57898/biosci.68