Transcript: What is a bioinformatician?

**Naimah:** Here at Genomics England, we work with a lot of bioinformaticians, but what exactly do they do? I'm Naimah Callachand, and today I'm going to be speaking to Jamie Ellingford, who is a Lead Genomic Data Scientist for rare diseases here at Genomics England.

**Naimah:** So Jamie, first of all, can you tell me what is a bioinformatician?

**Jamie:** Bioinformaticians are a mix between biologists and informaticians. It smashed those two words together, but in reality, it uses many more skill sets and has people from very different backgrounds than just those two disciplines. We are a discipline of people that uses maths, statistics, computer science, software programming, logic, and principles of scientific thought, as well as biology, to try and solve particular problems or questions. So, here at Genomics England, we are specialists in genomics bioinformatics - and bioinformatics can be applied across lots of different disciplines - but here we apply that to the study of the genome. Within our company, there are a number of different roles and very diverse roles, all of whom would call themselves a Bioinformatician. Just to give you some examples of those, we have people that have the job roles of Clinical Bioinformaticians, Software Developers, Data Architects, Bioinformatic Engineers, and Genome Data Scientists. Within each of those specialisms, there are different hierarchies – people who are very much the doers, the core of the business that drive some of the analysis we do, through to managers and people that help to direct and provide strategic direction of the workload.

**Naimah:** Do you specialise in any area?

**Jamie:** I'm a Genomic Data Scientist and I'm part of a really mixed team. Our squad is composed of both bioinformatics, engineers and genomic data scientists, as well as specialists in machine learning. We also interact with much wider people in the business. Like product managers and service owners, who ensure that the things that we do day to day can be translated appropriately to our users. So, in our case, the users of the rare disease service here, and we can innovate in an appropriate way that the users really need. And so, whilst I'm a Genomic Data Scientist, I work day to day with lots of people, different people with different expertise.

**Naimah:** It sounds like bioinformaticians have a really varied skillset. Are there any common skillsets that you all share?

**Jamie:** I think that that's a really difficult question to answer because we all do slightly different things. I guess the common theme in modern-day genomics is we all work with bigger data and so we have to have the skillsets to be able to work with data sets that are really large scale.

**Naimah:** Can you just tell me what big data actually means?

**Jamie:** It means a lot of storage space and a lot of computer power to be able to process it. What we mean in our context is the generation of as much data as possible for the whole human genome. So, our human genome is comprised of individual letters or Nucleotides, and we have just under three and a half billion of those letters. The specific sequencer and the way that they're arranged is what makes our genome our own and so it looks slightly different in everyone, but there's a huge amount of data that we need to generate to be able to survey that appropriately. To put that into context, please don't quote me on the accuracy of this, but the Lord of the Rings trilogy has about two and a half million letters in, we're talking at a completely different scale, almost three and a half billion letters that comprise each of our unique genomes.

**Naimah:** Then to go back to the question about the other common skill sets that bioinformaticians have, can you give me a bit more information about those?

**Jamie:** Being able to work with data at the scale requires that we can work in really tailored environments and so can access things like high-performance computing clusters. So instead of running analysis and software programmes on our everyday laptops, we access really powerful computing systems to be able to process and analyse that data. I guess that's also a common theme that we all develop work with help to benchmark computer programmes, and decision support systems that help us to look at the genome in a specific way.

**Naimah:** Can yougo into a little bit more detail about how or what the process is of making sense of genomic data?

**Jamie:** There are lots of distinct parts of trying to make sense of the raw data that we generate for a genome and just to go through those in a step-wise way, the first thing that we do is to try and align the individual pieces of data that we use back to a reference. So, this is a little bit like putting a jigsaw together. We have a general picture of what we think the reference genome looks like, and that's changed drastically over time. Using that reference genome, we try to use the individual data points, in this case sequence and reads that we get off the sequences, to piece together what that genome looks like for that individual.

The second part of that process, having put that picture together, is that not everyone's genome is the same. We try to detect differences in a really robust way and so those differences may look very, very different. Some of them are a single Nucleotides level, so a single letter is changed, and we swap an A to a T. Some of them may be much more large scale, so large regions, maybe a million individual letters are missing from one genome versus another. We use different computer programmes to try and detect these different types of variants. So that's two phases, alignment and variant detection.

After that is then trying to make sense of the variation that we've detected. So, the first thing we do is try and annotate each of those differences with useful information. That can be the frequency with which we've seen them before or the frequency with which they're seen in other populations. It could be their predicted consequence on genes and what that actually means at a molecular level. Also, we provide information for our users and as part of a team that I'm part of, the rare disease team, that helps our users who in many cases are clinical scientists, to provide clinical interpretation of that variant, and in some cases provide a genetic diagnosis for individuals that have been referred.

**Naimah:** Ok, then a little bit about you. So how did you get into this job?

**Jamie:** My story will be very different to somebody else that could be sitting in this chair given this podcast, and I think we all come from really varied backgrounds and really varied experiences. The thing that I was really driven by and inspired by was quite early in my career seeing the real-life impact of what this area of work could do. So, in a very specific example, was able to look at the genome of a family that was suffering from a particular condition. The young lad was suffering from a particular condition and saw the impact that a genetic diagnosis had on his management in the clinic. In this case, this meant that we could detect signs of a disease that was going to develop in later years and, as a result, do something about that. So specifically provide a transplant on the basis of the genetic information that had been found and the clinical indications that were then investigated afterward. Seeing how that could have such an impact on an individual family, and how we could develop ways to look at the data that enabled that impact, was inspiring and really motivating for me.

**Naimah:** What advice would you give other people who would like to become a bioinformatician or thinking about moving into the field?

**Jamie:** I think there's no fixed pathway to a role here, and as we've already touched upon, there are lots of varied roles so people would naturally come from very diverse paths. I think if you're in the early stages of your career and thinking about going to university, thinking about scientific principles, something that really interests you, whether that's biology, maths, physics, computer science, developing that kind of core set of skills to think logically, to look at really complex problems and reduce them to simple things that can be tested, can be falsified, can be investigated in reproducible ways. I think that core set of thinking is really apparent across everyone here who's a Bioinformatician at Genomics England. So learning those general processes is important, but specifically what you specialise in, I think that undergraduate level doesn't matter. Later on, there are specific Master's courses, there are various companies where real-life experience may be really beneficial, so there are medicine masters and bioinformatics masters, and having the opportunity to work with these big data sets I think is really quite key in those later stages.

**Naimah:** That was Jamie Ellingford giving us a brief overview of the role and skills of a bioinformatician. If you've got any other questions, please feel free to get in touch with us at info@genomicsengland.co.uk. Thank you for listening.