Ulster Medical Society

6 February 2020

Towards Personalised Medicine In Blood Cancers

Dr Jyoti Nangalia
Sanger Centre, Cambridge

Professor McMullin:

Good evening, everyone, to this lecture from the Ulster Medical Society. First of all, we're trying something tonight which we hope will work. We have the centre in Altnagelvin also on. Patrick Gelder very kindly volunteered that he would be there for each of these, so what we're trying tonight is, you saw them on screen, you can't see them now because that's a better idea, but they will hear the lecture and see the lecture live at the same time, so that's an interest, I think it will be a very interesting development. At the end, we'll be able to switch them on again, so that they can ask questions as well, so it's the first time for doing it, so hopefully it'll work well.

Tonight we have an absolutely wonderful speaker, who's going to talk to us about "Towards Personalised Medicine In Blood Cancers", carrying on the theme of how we're going to get, with new diagnostics, into particularly personalising diagnostics and then treatment, so I'm delighted that Dr Jyoti Nangalia has come today. Dr Jyoti Nangalia is CRUK Clinical Scientist at the Wellcome Sanger, and a consultant haematologist, and she is as young as she looks, because if we look back, she actually qualified in 2004 from Cambridge. She then was an academic trainee, and then did her PhD, and during her PhD, if you work in haematology, be very excited about the fact that she found the CALR mutation presented at first to the world at ASH in patients with myeloproliferative disorders, and she has gone on from there to be a clinical scientist. She has over 50 publications in major journals, the New England Journal of Medicine, the Lancet, and is a career scientist who is going to continue on this pathway, looking at sequencing technologies to understand the evolution of haematological cancers. So at that, I am delighted to welcome Jyoti Nangalia here tonight. Thank you.

Dr Nangalia:

Good evening. Thank you very much, Mary Frances, for that lovely and kind introduction. It's a real honour to be invited here today. Mary Frances is recognised and respected internationally in haematology, so to get the invitation from her to come here today is quite an honour, so thank you very much.

I'm going to give you a flavour of how we have been using genetic sequencing in the myeloproliferative neoplasms to work towards improving diagnostics for patients and improving our understanding of blood cancers, and then translating them into direct patient benefit, and there are going to be three stories that I talk about today, each looking, using genomics in a different way to help our understanding of blood cancers.

I love this picture, because I think bone marrow histology is beautiful, and I always think it would make really nice wallpaper. You have the red cells, the platelets, and the granulopoiesis occurring within the bone marrow, and then what's happening essentially is that you have stem cells; these are rare populations within the bone marrow. We don't really know how many there are within the bone marrow, maybe around 200,000 to 500,000 of these cells within our bone marrow, and they produce progenitors. Progenitors then have more proliferative capacity, but perhaps don't last as long as stem cells, and then these progenitors then lead to the mature finished products, so to speak, the white cells, the red cells, and the megakaryocytes and platelets. And obviously within the bone marrow, you see all the precursors leading to those differentiated lineages, and these are the fat spaces within the bone marrow. So I'm absolutely fascinated by haematopoiesis as a tissue, and what we study in the laboratory are the myeloproliferative neoplasms. I treat myeloproliferative neoplasms as a clinician, so I'm particularly attached to them and the patients that have these conditions, but from a scientific point of view, they're a really fascinating model system for studying cancer, because they are a really early form of cancer. Patients can have them for decades, and so you can study the natural evolution of these cancers over the decades, which you can't really do within the solid tumours. Often when you find a solid tumour, you really don't know how long it's been there for or what the precursor states of it were, and you can sample, because it's a liquid cancer, you can sample these cancers regularly and repeatedly, and often at the single cell level, so you can dissect out all the different types of cancer cells that might be there within a single patient's tumour, so as a model system, myeloproliferative neoplasms are unique in offering insights into cancer biology. What happens in the myeloproliferative neoplasms is the entire blood system seems to be retained. The stem cells are still producing the differentiated cells, but it's in overdrive, and so you get more of the cells than you ought to. The myeloproliferative neoplasms come in a number of different flavours. Over here, you can see what we call polycythaemia vera, where you have too many red cells, the blood becomes too thick. In a central thrombocythaemia, these little purple blobs here are

the platelets, you can produce too many platelets, because the overdrive is going down that pathway, and then in myelofibrosis, here is a picture of a bone marrow trephine with some scarring, some collagen scarring, which basically means it's an advanced form of MPNs, and often what happens in those conditions is, you start to lose that excessive overdrive, and you can then have lower blood counts, because everything's burnt out.

About 30,000 patients within the UK have MPNs. Again, we like it because it's the earliest stage of tumourigenesis, but many of us also treat patients with these conditions.

In 2005, there was a major breakthrough in this area, where the vast majority of patients, pretty much all patients with polycythaemia vera, and half of patients with ET[1], or central thrombocythaemia and myelofibrosis, that I'll call MF, were found to have a mutation in a gene called JAK2, the same mutation, despite the different clinical phenotypes, and whilst I won't go into that story today, although I remember it because I was an SHO in haematology when everyone was getting excited about it, and I was wondering what the excitement was about, what is quite remarkable about haematology research and blood cancer research is that within five years of the discovery of the mutation, we had agents that had already gone through clinical trials, that were published results, and we were using them in patients with myelofibrosis, to treat their symptoms, and that's really quite remarkable in terms of translating a clinical research finding, not only into a diagnostic test, which is another first-line test globally for these conditions, but then also into therapeutic benefit in such a small space of time, and that's real gratification for some of the academic work that we do.

Over the years, we've learnt about what it is about those JAK2 mutations that are driving these conditions, and essentially what's happening is that, when you're a stem cell or a progenitor cell, in order to know what type of end cell you want to become, be it a red cell or a leukocyte, or a lymphoid cell, you listen to signals from outside of the cell. So, if there's erythropoietin around, which signals to produce red cells, the erythropoietin docks on the erythropoietin receptor. Once it does that, JAK2 sits inside the cell, on the inside of the receptor, and it shuttles to the nucleus via the JAK-STAT signalling cascade, which is basically a cascade of enzymes that switch each other on, and then it signals to the nucleus to make a red cell. And essentially what happens with JAK2, the mutation switches the protein in a permanently on stage, so that even in the absence of erythropoietin, there is excessive, what we call JAK-STAT signalling, and there's excessive proliferation of cells when the body doesn't need it. And JAK2 sits downstream of many receptors on the cells, so not just the receptor for erythropoietin, which stimulates red cell production, but also thrombopoietin, which stimulates platelet production, which is why the same mutation can lead to different phenotypes in different patients.

Over the years, after the JAK2 mutation, we did discover, not our group but other groups, mutations in other parts of that cascade that led to the same outcome, so for example, in a rare number of patients, there are MPL mutations. Now, MPL is another work for the thrombopoietin receptor, and it switches that receptor in a permanently on stage, and these patients will always get essential thrombocythaemia, rather than polycythaemia, because what you've actually switched on there is not a generic protein that shuttles between all the receptors, but you've actually switched on the receptor for platelet production, and therefore you get ET.

So whilst we were beginning to, the landmark discovery of JAK2, we still didn't know what was causing the disease in half of patients with ET and MF, because only half of patients had a JAK2 mutation, and even though we did discover MPL mutations, they were only present in a handful of patients.

So, when I started my PhD in 2010, there were a number of unanswered questions in the field that I was particularly interested in. The first, which everyone was really interested in, is well, we have JAK2 mutations present in most patients, but what about the other half of patients with ET and MF?—and diagnosing these conditions in those patients was a challenge, because you have to rule out all the secondary causes of high platelet counts, be it infection or inflammation, bone marrow biopsies, it was quite a challenge. If you had a diagnostic marker, that would be excellent.

But the other thing as well is, we have these three morphological categories, and these phenotypic categories, but many of us as clinicians found patients that didn't really fit into one of those boxes. Maybe they had a high red cell count and high platelets, or maybe they had high platelets and a bit of fibrosis, so how do you call them? And also, within each category, no two patients were really the same. There was substantial heterogeneity, so if they all had the same mutation, well what's actually driving that heterogeneity between patients?—which is pretty important in this era of personalised medicine, we want to make sure that it's not a one-size-fits-all approach.

Not only were patients presenting with different features, but in terms of their outcome, it was different. Patients got thrombosis, other patients didn't. Some patients' disease progressed to acute leukaemia, and other patients' diseases didn't, so again, how do we know at diagnosis what's going to happen to a patient in the future?

[1] Essential Thrombocythaemia

Then the last question, which, I think all of our patients ask us really, is how long have I had it for? Did I just discover this six months before I presented to clinic, or was it six years ago, or have I had it for even longer than that? Is it something that I may be picked up from a family predisposition?—again, we didn't know, so these were the kinds of questions that I was interested in, and hopefully I'll give you an insight into each of those, so we can translate those questions into three, sort of research questions. One is detecting the novel cancer-driving mutations in MPNs, in those half of patients. The other is, how can we use our knowledge of the genomics of these conditions, and the clinical information from patients, to try and predict what's going to happen to patients in the future, and understand why they've presented with the disease that they have at diagnosis, and lastly, I want to talk about some recent work that is not yet published, but quite exciting, about the, how long have I had it for question.

So, in my PhD, we took 150 patients, and we took blood samples and undertook a whole exome sequencing, which essentially means that you sequence all the genes within the DNA, within our genomes. We have three billion base pairs within our genome, but only 2% of that codes for proteins, and so back then, it would be much cheaper and more feasible to just sequence that 2%, so you'd capture all your genes, and when we did that, fortunately the most common mutation we found was JAK2, but, after much re-jigging of the data, we also found a large number of patients that had a mutation in calreticulin. Now, I'd never heard of this before, and what was particularly exciting is that, when we looked at which patients had the calreticular mutation, they were all the patients, predominantly almost 80% of the patients that didn't have the JAK2 and MPL mutation. Were they born with it, or was it something that they acquired during their life? and we looked at the mutations, and because the mutation led to a smaller-sized gene, because it was a deletion, we looked within the granulocytes of the patient, and the granulocytes confirmed that those patients had one normal copy and one shorter copy, which is here. The shorter copy runs a bit lower on the gel, whereas in the T cells, which aren't part of the tumour, they just have the wild [?] copy, suggesting that in this patient, the patient wasn't born with it, they acquired it in their stem cells, and those stem cells are now producing myeloid cells and the essential thrombocythaemia that the patient has, but it took many, many years for us to find the calreticular mutations, and that's really because of the way sequencing works. Back then, when you do next-generation sequencing, what you do is you take your DNA, you shatter it into fragments, and then you take your fragments and sequence from both ends, about 75 base pairs in into both ends. Now what we do is, we sequence 150 base pairs into both ends. Back then, when you did 75 into both ends, in order to identify the mutation, you need one DNA read to span the entire mutation, and often what was happening with the DNA reads, each of these lines, yellow and blue, is a run of DNA that we've sequenced, and then we've mapped it all to the region of CalA. Often what was happening is not many of the reads were getting to the other side, because it was a poorly-sequenced region, and as a result we missed it, but in this patient, there were one, two, three reads that spanned the deletion, and therefore we were able to identify it. When we found that they were real, we then went and looked in the patients that we thought ought to have them, and we found them in all of them as well, and in the vast majority of patients, the mutation was there, but the algorithms didn't pick it up, which is an example here, where you can't see this green band across the DNA, the algorithm can't find the mutation, even though it's there.

The other thing is, when you looked at JAK2, we had a lot of DNA sequencing from the region of JAK2, around 200 reads, whereas in the calreticulin region, we had about ten, and so if you had only 10% of the depth of sequencing, and you missed the mutation in 90% of your reads, you really hardly had any patients where you were able to pick it up. Fortuitously, there were two patients where we did pick it up, and then we managed to find that it was still there in the rest of them, just not picked up.

So what do calreticular mutations actually do, to drive myeloproliferative neoplasms? Well, we were all trying to figure it out at the time, and 85% of patients had either a deletion or an insertion, and the rest of the patients had variants of the same deletion and insertion. Now, in my mind, as a cancer geneticist, when you delete parts of a gene, you essentially ruin the function of that protein, and you sort of switch it off, and it doesn't work anymore, which sometimes happens with genes that are, tumour suppresser genes. If you switch them off, then you're perhaps more cancer-prone, but this was unusual in that every single deletion and every single insertion led to the same outcome to the protein. It led to, if you might remember with your DNA, every three bases codes for a codon, and that codon codes for an amino acid, what was happening is, no matter what your deletion was or your insertion was, you shifted the reading frame of your DNA by one base pair, so that no matter what it was, you ended up with this novel protein sequence within calreticulin, suggesting that this wasn't actually loss of function, but maybe gain of function, and some other groups have done some fascinating work over the last couple of years, to show that calreticulin actually has nothing to do with receptors and signalling. It sits on the endoplasmic reticulum, and it's a housekeeping pro-

tein. What it does is, when proteins are made, it helps those proteins fold into the right shape, and then shuttles them off to the surface. MPL, the receptor for thrombopoietin, that drives the platelet production, is a client of calreticulin. It goes into the endoplasmic reticulum to be folded, and in the process of doing that, it locks onto calreticulin, MPL gets switched on, and calreticulin then almost acts like a substrate for the receptor, switching it on, which is why these patients get high platelet counts, so that's quite exciting.

The other exciting thing about calreticulin mutations is on the cell surface, so potentially it could be targeted with antibody-based treatments, and vaccines and things like that. Very soon, it became a new testing clinic, where if you wanted to screen for the patient having a calreticulin mutation, you sequenced the gene or part of it, and you checked its size, and if its size wasn't what you expected it to be, then you knew it either had a deletion or an insertion. Different hospitals have different tests, but very soon it got put into the national and international guidelines for diagnosis of MPNs, so that was quite gratifying. So now we have the situation where really there is this sense of completeness within the MPNs, unlike many cancers, where we pretty much have a diagnostic marker, a gene that can be tested for the vast majority of patients. There are maybe 10% of ETNMF patients where we don't yet have a diagnostic marker. They may still have a disease where there is a mutation to be found, or indeed there may be patients with secondary thrombocytosis due to infection or inflammation, or perhaps even familial thrombocytosis because it runs in the fami ly—we don't yet know.

So moving on to the next question, of how can we, now that we've done these large-scale sequencing studies, and we have all the mutations present in these cancers, how can we actually make more use of that information?—maybe to understand why patients present with the disease that they have, and what's going to happen to them in the future. So to do this, we thought, well, a couple of years ago we did 150 patients, now we're going to do 2,000, because sequencing is cheaper, and so for this, we took 2,000 patients from across the UK, many from Belfast, Ireland, Italy and Denmark, as part of some of the clinical trials that were ongoing at the time. We didn't sequence the whole exome, because when we first sequenced the whole exome, there were really only 30 or 40 genes that were mutated, so we thought, right, we'll just focus on that, because then we can sequence more patients for the same amount of money, so then we focused on about 35 genes, and we also looked for chromosomal changes, the kinds of things that you would normally do by fluorescent in-situ hybridization, or karyotyping, we were doing that genomically now, so you can look, if someone has, for example, a trisomy 21, you can pick it up genomically now, by looking at the germ-line polymorphisms present within those patients and seeing what the copy number is of those. So essentially we were doing copy-number calling, which is chromosome-level-wide calling, from the genomics, and what we did is, we then were building a model where we could correlate with baseline clinical findings, disease transformation and patient outcome.

I'll skirt over this, but this is to show the landscape of the mutations that we found, which we were becoming more and more familiar with. Essentially most patients have only a few mutations, maybe one to three, and then there's this long tail of rare mutations that are splattered across patients, and the difficulty is, how do you then understand the significance of a mutation, if it's so rarely found, and that's the challenge, and that's why you need thousands and thousands of patients to then understand what, if a gene has mutated say in four patients only, to then understand the risk associated with that. So these kinds of studies to be really holistic in incorporating all the mutations, we need very large numbers.

Then we were able to get at why patients present with different forms of a condition at diagnosis, so I hinted at the idea that the JAK2 mutation can cause both PV and ET, and indeed MF, because it sits downstream of many receptors, or what determines why one patient presents with PV, and another patient presents with ET? Well, we looked at all the factors that could do that, and the first thing is to say that these funny numbers and letters at the bottom, those are germ-line kind of natural variations within your genome, and if say my haemoglobin is slightly lower than the next person's, then if I was going to get an MPN, I was probably less likely to get polycythaemia vera, because my haemoglobin runs a bit lower anyway, and equally if someone has a slightly higher platelet count than the average, then they're probably more predisposed to a central thrombocythaemia, but we were able to find the exact polymorphisms that determine that, so what we're trying to do now with some people at the Sanger that really are into understanding why people's blood counts during health are shaped by their genetics, we're working with them to try and have a sort of baseline scoring system for what should their blood counts be to begin with, and is what they have normal or abnormal for them, which is quite interesting.

We then also found that, at the genomic level, the higher the burden of the JAK2 mutation, the more likely you are to have polycythaemia vera. Now, this has been known for a long time, but we were emphatically able to show that actually, it really was patients that had two mutant copies of JAK2 that were driven down the erythrocytosis route, and one mutant copy meant thrombocytosis, and that's probably because

the receptors have different thresholds for activation. The erythropoietin receptor probably needs a higher dose of mutant JAK2 for it to be activated. There's also some evidence that, with a single copy of JAK2, the thrombopoietin receptor gets activated, but then it gets degraded, and therefore you do end up having a bias in terms of the outcome of what that cell becomes, depending on the dose of the JAK2.

We found that when patients presented with myelofibrosis, then yes, there were some germ-line polymorphisms within their genome that determined that at the bottom here, if you were more prone to having low blood counts to begin with, but these patients generally had additional hits in their genome, additional mutations, that were then leading to clonal evolution of their tumour to a more aggressive form.

We then asked ourselves, if we ignore the fact that we know that these patients have either PV, ET or MF, and we look purely at the genomic level as to what groups of mutations are present within our 2,000 patients, so there we're asking our algorithms to say, right, what patterns of mutations can we find?—because no two patients are really the same, but broadly, what patterns, so which mutations are often found together more often than not?—perhaps which mutations are never found together?—and can you find me some genomic patterns such that the people within one group are very similar to each other, and there are distinct boundaries and patterns between the groups. So when we do that, we actually found far more than three categories. We found a group of patients that have the JAK2 mutation plus other changes, the MPL mutation and the CalA mutation, plus other changes. Interestingly, you have this group of patients that have mutations in a gene called TP53. Now, TP53 is a universally bad gene to be mutated in cancer, not only blood cancers but also solid tumours. It's a critical tumour-suppressor gene that prevents … normally it is the cell death signal, so if your DNA's damaged, it promotes cell death, so you get rid of that cell, whereas when TP53 isn't working, you seem to tolerate DNA damage, and the cell doesn't apoptose. So we found this group of patients where regardless of what mutations you had, if you had a mutation in TP53, it trumped everything else. Then we had a group of patients that had additional mutations, and these were of the patients that generally had myelofibrosis.

What was quite interesting is, of those categories of genomic groups, some of them were actually found in other blood cancers, not MPNs, such as myelodysplasia or acute myeloid leukaemia, particularly the group with the TP53 mutations, so it's almost suggesting that, at the genomic level, we probably have the same entity that potentially we're labelling differently, because histologically they're different, and these patients will then get different treatment, so it's kind of raised this concept and this idea that maybe genomic entities don't really match up with morphological and histological entities, and at some stage we need to try and understand why that is, and perhaps genomic entities transcend the morphological boundaries that we currently use.

So the next thing we wanted to do is, when we looked at genomic classification, we wanted to say, okay, right—how do we now work out what's going to happen to patients in the future, based on their clinical parameters at diagnosis, and all this genomic information that we have on the 40 genes, and the copy-number changes and their germ-line genome, and to do that we basically asked ourselves, well, what are we trying to model in terms of outcome of patients? So you sort of imagine a patient walking into clinic, they could either have ET or PV, which are chronic phase conditions, and that patient over time can either stay alive with their condition, they could die either with or from their condition, they could transform to myelofibrosis, or they can transform to leukaemia, so we have four journeys, four paths that we have to predict and model. They could actually arrive in myelofibrosis already, and then from that point they could stay alive, they could die, or they could transform to acute leukaemia, so we have another three things we have to model, so we basically decided that we were going to use 63 variables: genomics, clinical parameters, gender, blood count, splenomegaly, did they have thrombosis or not?—63 parameters, we're going to model each of those outcomes, and obviously each of those outcomes, the probability doesn't add up to 100, so then what you have to do, using maths that I can't claim to understand, but some of the statisticians understand as well, is you can compete each of risks against each other, so that overall, at any time in the future, you can predict, out of my 100% probability, what is the split amongst all of those journeys? So that's what we did, and then we asked ourselves the question, first of all, which, let's look at the population level, all 2,000 patients—which of those 63 variables are driving each of those outcomes?

So the first thing I want to show you is patients in chronic phase, if they die with their disease, the greatest risk factor is age, and I say this all the time when I give this talk, and Mary Frances has probably heard this so many times before, but when I show this, I have half the audience that kind of laugh, and go, you've just blown a huge amount of sequencing at the Sanger to show us that age is a risk factor for death, but I would argue that you can tell those patients that they're going to die with their condition and not from their condition, of old age, and their disease is not going to progress, which is actually critically important for patients that are going to have a blood cancer for potentially like 20 to 30 years of

their life, and some of these patients get diagnosed very young.

At the same time, if they have some of the other mutations, our model will be able to tell you whether those patients have a risk of transformation to myelofibrosis, here, and here you can see that the orange is the genomic component, if their disease is going to transform, so if they have any of these mutations, there's a risk associated with that, but also there's a risk associated, depending on what's your blood counts are, and whether you have an enlarged spleen and other things, suggesting that genomics isn't everything, and there are clinical elements of the disease that cannot be captured by genomics. They have, in their own right, prognostic value, which is why integration's important. Again, with AML transformation, a big chunk is the genomics.

One other thing I'd like to point out is, when you're looking at predicting risks associated with certain mutations, there will be the mutations in TP53, that are really poor prognostic markers, but are incredibly rare, and for a patient that's important, but for a population that might not be, and then there will be mutations that are incredibly common, but carry a tiny risk of progression, and for the patient, that's not so important, but for the population, that is, which is why, when you start to use risk-scoring systems, it's really important to not just focus on the bad genes, and see whether the patient has them, it's important to screen everything, because everything carries a risk, and it's about the cumulative risk, so that's what I'd say about that.

Then, on the per patient level, what we can do now, and we do this online with a tool, is we create these personalised graphs for patients at diagnosis. If patients come in and have a gene sequencing panel, we're moving towards that at the moment, it's not present and accessible everywhere, and we're trying to figure out, as a community, when it is and isn't warranted, you can then type in the disease, the gender, the age, the blood counts, the mutations that the patient has, and then you get a graph, and the black line is the survival curve over 25 years, and then if you're alive, it tells you in what state, and if you've died, it tells you of what cause.

So now our challenge is, now that we can diagnose, and we can personalise prognosis, now our challenge is, we can now identify those patients that aren't going to do well at diagnosis. How then can we use these models to think of new strategies for treatment to potentially try and alter their outcome?—and that's the next step, and in acute myeloid leukaemia, they've actually done that. They've managed to use clinical trial information, do the genomic sequencing, look at the treatment that the patient's got and their outcome, and actually then predict, you get two graphs, you get a graph with a transplant, after first remission, and a graph of what the survival might look like without the transplant, so you can not only then do diagnosis classification, prognosis, but also therapy, in terms of clinical decision-making.

So that's where we're at in terms of personalising, using genomics to personalise medicine within MPNs, and I also wanted to briefly talk about the thing that I'm doing right now, which is perhaps not as translational in the immediate instance, in terms of how long have I had it for, but I think it's critically important to us maybe detecting these cancers at an earlier stage, and understanding the trajectories of these cancers over the lifetime of the patient, and in terms of the kind of primary prevention and monitoring and early detection of cancers.

So, in order to answer the question, how long have I had it for, we took advantage of a couple of strategies that we have within the research environment, so what we do is, we take blood from MPN patients. We separate the blood out so that you can grow progenitors from the blood on a dish, to the point where no progenitor cell is touching another progenitor cell, and where that single cells lands on the dish, it sticks to the dish and it grows into a little clone, such as here, this is a red cell clone, and the nice thing about this clone is it is derived from one cell, so what you can then do is pick these colonies from a tumour, and essentially what you have is a single-cell readout of the tumour, but you're not dealing with single cells, you're dealing with a nice big lump of cells, which is always easier to deal with than a single cell, so what you're doing is taking a tumour and separating it into its individual members, but in a way in which you can still handle the samples, a bit like taking the royal family tree and finding all the individual members of that tree, and what we're trying to do here now is use whole genome sequencing and the mutations to then reconstruct that tree, and work out how all those cells within that tumour relate to one another, which two clones are really distantly related, maybe which two clones are only more recently diverged, a bit like siblings versus third cousins, so what we do is take a cancer, split it into its individual components, grow it up so we can handle it, and then try and reconstruct almost the family tree of that tumour. Why do we want to do that?—well, that can help us with tiny mutations, so imagine one stem cell has acquired a thousand mutations over its life, every time that cell divides it gets a mutation because we have three billion nucleotides to replicate with each cell division, and it's not a perfect process. We think our stem cells are probably acquiring 20 mutations a year, probably 50 in the first year of life, because growth is so rapid, and thereafter maybe 20. Now, these aren't mutations to worry about, these are just like random, a rain shower on the genome. It doesn't actually cause any ill effects, but they can be used as lineage tracing, so

if you have one stem cell that's acquired a thousand mutations, and another stem cell from the same person that has a thousand mutations, if those two stem cells share 200 of those mutations, and then have their own lot of 800 mutations, then you can tell that those 200 mutations occurred first, and then that cell divided into two, and those two cells then go up 800 mutations each, so by looking at what mutations are shared and not, a bit like the family tree with the royal family, you can work out when things happened.

So what we do is, this is just to show you that when we do this on a per stem cell basis, the number of mutations in haematopoietic stem cells goes up with age. That's the strongest risk factor—it's not smoking, it's not UV light or anything like that. In haematopoietic systems, it's well protected from all the DNA-damaging processes. The thing that basically damages our haematopoietic stem cells is cell division that occurs over age.

So now that we know that these mutations are occurring constantly with time, and we can time the mutations relative to one another, all of a sudden you can then put an absolute number, and in terms of years, in the lifetime of the individual, on mutations. So say you have cell one that has these mutations, and cell two that has these, and cell three that has these, you can see that all three of them have A, and these two have D, so when you draw a family tree, A occurred first, then that cell split, then you got D, then that cell split, so you can use mutations to draw a family tree, and as I showed you before, these mutations are acquired constantly during life, so you can imagine here, there's one mutation per year, therefore A occurred in the first year of life, D occurred in the second year, and these occurred in the last four years, you can't tell which, so with both approaches, you can time mutations.

We did that in MPNs, we took colonies, we grew up colonies. Each red dot is a red colony, as it happens, from a patient who was diagnosed with ET at age 20, and we sampled them aged 23, and these colonies here all have JAK2, and these colonies here don't, so that's just normal blood, and the JAK2 mutation is shared amongst those colonies in addition to another 100 mutations, which means that the JAK2 mutation occurred within the first, probably three to four years of life, because the first 50 mutations occur by the time the baby's born, we know from other work that's happening in our labs, so this patient probably had their mutation during infancy, which we weren't expecting to find. So we thought to ourselves, this is unpublished work, it's not gone through peer review, so take it with a pinch of salt if you want, but this is what we're trying to publish at the moment. We weren't sure what to make of that. We thought, well this patient was a young patient that had an MPN. Maybe there was some sort of germ-line predisposition. Let's look at an older patient, because most patients with MPN are older, and this patient with a central thrombocythaemia was diagnosed at age 65. We sampled them almost 20 years later, that's when the sample was available, and this patient acquired JAK2 by 200 or 300 mutations of life, which again is decades and decades before the patient presented to clinic, and again probably by their teen years.

So what we're now trying to do is look at the pattern of these branches, whether they're gradually expanding versus expanding quite quickly. You can actually do a lot of mathematical modelling on the pattern of the tree, to try and understand, well, what was the rate of growth of that stem cell over time? Did it expand over five years or ten years? Was it increasing by 10% a year? What are the kernel dynamics, and that's much of the sort of mathematical work that we're doing at the moment, to try and understand basically how cancer originates, and using MPNs as a model, so the genomic seems to keep giving for us, in terms of its applications to understanding the disease.

So in summary, I've shown you how we've used genetic sequencing, initially whole exome sequencing, to aid diagnosis, and hopefully future treatment. We then used targeted sequencing of specific genes of interest, to characterise the genomic sub-types of MPN, and build a personalised prognostic model for patients, that's present online, and then we've moved to whole genome sequencing, to utilise this idea of the mutation clock that we have within our bone marrow, to look at the absolute timing of driver mutations, and also to understand the dynamics of kernel expansion, and what does that mean in terms of what a driver mutation is doing to a stem cell, in terms of determining how quickly that tumour grows.

So I'd like to thank my small and growing group at the Sanger, and Peter Campbell and Tony Green's group, as well as all of our clinical collaborators who send us so generously their samples, and most of all Mary Frances, for her invitation to allow me to share this with you guys today. Thank you.

Professor McMullin:

Thank you very much, that was absolutely wonderful and beautifully explained. Questions?—yes, Peter?

Peter:

Thanks very much. That was really clear, and a marvellous lecture, so thank you so much for that. I was really struck by the clinical heterogeneity, why patients might develop splenomegaly or have a very weird red cell count. What other factors do you think are in the mix, making that happen?—you alluded a little bit more to it than the genetic mutation, so is it something else? Is there something else that you can

study, like the genetic phenomena that explains why those occurrences are so different?

Dr Nangalia:

Yeah, that's a very good point, so at the moment we know that driver mutations play a role, the dose of JAK2 to maybe other mutations. We know that age and gender play a role. Maybe the males will more likely have polycythaemia vera, higher red cell counts, and that's probably to a certain extent hormonally-driven. Age, we don't understand quite why maybe we're less able to produce platelets with age perhaps, but still able to produce red cells, I don't know, and then our germ-line background in terms of where we sit within the normal population in terms of our blood counts; but we also know that, if you take into account all of those factors together, you can't explain all the variability as you alluded to.

I think a couple of things that we haven't looked at is micro-environment, inflammatory milieu. We don't really understand what drives extra-medullary haematopoesis and splenomegaly. Sometimes patients don't necessarily have fibrosis in the marrow when they then develop their large spleens, so we don't really know whether it's purely the degree of proliferative drive, or something else, and then what we're doing on the epigenetic level at the moment is, we are undertaking, we're looking at the methylation patterns within DNA at the moment. There is this idea, there are two more ideas: one is, maybe stem cells are inherently biased towards what kind of cell they're going to produce, like a platelet-biased stem cell or a red cell-biased stem cell. The idea is more about a platelet-biased stem cell. Maybe if you had a mutation in a cell that was already biased towards a certain lineage, you could go down that route, but we're looking at the epigenetic level, we think that perhaps there are certain epigenetic landscapes within cells before the mutations hits, that are fertile ground for certain kinds of downstream paths, so what we're doing within these phylogenetic trees is looking to see whether we see the same methylation patterns in the clones that have expanded versus the ones that haven't, and thereby it might not even be the driving mutation, it might be the permissive state and the fertile ground of the cell that then allows a driving mutation to have a certain consequence. So yes, I'm sure you're right—I don't know how yet, but we're looking into at least the epigenome at the moment.

Peter:

[?] you gave the risk factor being increasing age, be an increasing change in methylation with the age as well?—having the same problem happening?

Dr Nangalia:

Yes, and I think you've recently published work on how their methylome does change with age, and it does impact on phenotypes within MPNs, so we're looking at that, but sort of at the clonal level, to see whether there are certain methylation backdrops that permit a certain outcome.

Professor Patrick Morrison:

A very interesting talk, thank you. Your slide, about two-thirds of the way through, with pie charts, where you kind of show the age, maybe it was this one, you've really looked at the affected patients, but it would be really interesting to look at say a couple of thousand controls, and then follow them up for 20 years, and see how some of them then develop the condition, and see if there's a key factor, and whether it's diet or environment or some other change? I don't know if it's possible to do that, because at the moment, your family trees are essentially looking at affected patients. The key thing is why … you could have a picture like that if you're healthy, and then in ten years' time, suddenly get the condition?

Dr Nangalia:

Well, we can tell from the trees when the clone expanded, because the branches within the trees show us historical events, so in those patients that we know, as you say, have almost been pre-selected because they have an MPN, we can look at the trajectory of the growth of that clone, so we can translate those trees into a line graph, which I haven't shown you yet, because we're not quite sure we're doing it right yet, in terms of the growth of that clone. I think you're entirely right—what we need to know is, do we see that pattern because those patients were predestined in some way to get an MPN, or is this how MPNs occur?

We've looked in clonal haematopoiesis, and in clonal haematopoiesis, as far as we can tell, the driver mutations are occurring pretty early as well. The BBC picked up on a story that came out yesterday, with the pan-cancer genome, the 2,000 genomes, and again there we're seeing that many of the driver mutations that are driving cancer are occurring decades before the presentation of the disease. Does that mean that those mutations sit around doing nothing for a long time, and then there's a secondary effect, like a bottleneck infection, whatever, ageing? Or is it that it is growing, and it's just doing it very slowly? I don't think we know that yet.

Professor McMullin:

Altnagelvin, did that work?

Patrick Gelder:

Yes, that worked very well.

Professor McMullin:

And have you any questions?

Patrick Gelder:

I'm still thinking! A fascinating talk, yeah.

Professor McMullin:

It's useful that we've been able to do that. So I think we're probably all most fascinated by the fact that you can go back, and these things are absolutely in childhood. Do we all have them? I used to spend a lot of time growing colonies for my own blood, way back.

Dr Nangalia:

Well I have … we have with tumours as well. It's like the fittest kind of clonal landscape, no driver mutations, yeah. I think, we're trying to build simulators where we're building bone marrow, stem cells kind of dividing, and then we're introducing driver mutations. Most of the time when you do that, the driver mutations don't stick around. The cell dies, or it doesn't divide, and we find we're having to introduce them multiple times to get it to fix within the population, so my sense, and this is complete handwaving with no data to support it whatsoever, and in two weeks' time I might be showing data that will change what I think; my sense is that you've got an expanding bone marrow, and so there are risk periods within your life. When your bone marrow's expanding, that is the stage when you get a driver mutation which is chance, then it's perhaps more likely to fix within the population. What it does then, it's really hard to know, because at that stage you have multiple other normal stem cells, and I have this genuine belief that normal stem cells are best adapted to making blood, and abnormal stem cells only kind of come to the fore when the normals are suffering, so I have a feeling that maybe it's the expansion of the bone marrow itself, that then allows some of these mutations, if they occur at certain windows, to then stick within the population. They probably don't do very much until later on in life, when perhaps the normal stem cells are suffering a bit, and perhaps aren't quite as fit, and then there is a drive to produce more blood, and then those cells suddenly realise that they are fitter and able to do that, so that's what I feel, but again there's no data to back it up.

Phil:

In spite of the 79 year old lady with her risk factors for transformation or death, do you offer this to all your patients?

Dr Nangalia:

No, so this is what we are debating at the moment. I would love to, I'd love to do whole genome sequencing on other patients. I think we are heading towards an era where we will be doing these kinds of things in more and more patients. At the moment, we do run prognostic models in all our patients with myelofibrosis, because it helps us ascertain whether they might be a candidate for a transplant or not. These models are more accurate, and I think it will take time for them to come into the clinic, because they require much more sequencing, so I would love to see them being used in clinic. What we do with the information that we're getting, we don't yet know. They would need to be tested in some sort of a trial.

Phil:

So that [?] those results?

Dr Nangalia:

No, this is all research. Patients, this patient died at five years, from myelofibrosis, which was the highest risk factor that she had, so we kind of built the model, and then assessed it in comparison to actual outcomes, and that's how we trained the model.

Professor McMullin:

Okay, well if there's nobody else, then I think we should repair for a cup of tea downstairs, so thank you very much for an excellent talk.