People of the British Isles

Newsletter Issue 6 – March 2015

# Welcome

It is now nearly three years since our last newsletter. During that time we have continued to collect more samples from volunteers and enlarged our collection of photos for the face genetics project to nearly 2000. Most importantly, a subset of 2039 of our samples has been extensively analysed, and, after several revisions and adjustments, the results have appeared on 19th March as a major paper in the journal Nature. This is therefore a good opportunity to review the history of the PoBI project, take stock of what we have achieved so far and describe the key results of our Nature paper (http://nature.com/articles/doi:10.1038/nature14230).

In our first newsletter in 2006, we suggested that we were halfway through the anticipated 5 years of the project, and mentioned the channel 4 TV mini-series called 'The Face of Britain', which was broadcast the following year. By the time of our second newsletter in 2008, we had collected nearly 3000 samples, but it soon became clear that we were not even going to finish just the collection of samples within the 5 years of our first grant from the Wellcome Trust. Thus, by the time of the 3rd newsletter in 2009, we were able to report success in obtaining a further 5 years of funding from the Wellcome Trust and support for a new direction of our research, namely, to look into the genetics underlying facial features and certain other normal human traits, including taste, smell, hair colour and handedness. By the time of our 4th newsletter in 2011, we could report having submitted our first paper (published the following year – http://www.nature.com/ejhg/journal/v20/n2/full/ejhg2011127a.html ) describing the basic approach of our study and some very preliminary, but interesting genetic results. Our 5th newsletter in 2012 reported some of the early results of a comprehensive analysis of the data on more than 2000 samples. It included an initial map that already showed evidence of the remarkable concordance between genetic and geographical clustering of the samples. Now we can report the further development of this analysis, including the relationships of our UK clusters to the genetics of surrounding European groups, which tells us much about the history of the People of Britain from a genetic viewpoint. It has been a long journey to get to this point and we hope that you will think it has been worthwhile .Our success has depended on you, our volunteers, many of whom have come to us for a second visit. We are enormously grateful for all your efforts and interest.

## PoBI history: How did it all start, what were its aims and how was the study carried out?

I was asked to give the annual Sir John Rhy's memorial lecture to the British Academy (the UK's arts and

humanities equivalent of the Royal Society) in 1992, with the genetics of the Celts as my topic. This came as somewhat of a surprise because the lecture is usually given by a major Celtic scholar who would be a real expert on the Celtic people, their culture and their history. However, there was an interest in what genetics might be able to tell about the origins of the Celts in the UK. It seemed to me most unlikely that the Celts all came in a huge hoard from somewhere in central Europe, perhaps around 800 BC, and largely displaced the then indigenous population. Even the meagre genetic information available in 1992 did not support that view. The most likely hypothesis was that the Celtic culture was brought to the UK by a relatively small number of people and that what we call the Celts in the UK probably represent the populations closest to the early modern inhabitants of Britain, before the later invasions of Anglo-Saxons, and Vikings. This idea caught the attention of the BBC Horizon team, who asked me to present a programme on this topic. In the course of this project, we paid our first visit to Orkney to collect blood samples for a genetic study. This gave us the idea of doing a major study of the genetics of the British people by collecting samples systematically from throughout the UK.

The call for projects to celebrate the end of the millennium, which would be supported by a major national lottery, seemed an ideal opportunity for obtaining funding for such a study. How better to define the people of Britain at the end of the millennium than by a detailed map of their genetic distribution throughout the UK? Our stated aim in this application was “to create, as a capital asset, a collection of samples representative of the People of Britain that can be studied now and indefinitely into the future at the DNA level, that can be added to and so create a resource that defines the People of Britain as we know them today, that reaches back into the past, and that provides a resource for those who wish to ask in the future, "Where did they come from?” To do this, we would sample individuals from rural areas and limit “consideration to those whose parents and preferably grandparents come from the same area”, thus enabling us to “approach the characterisation of populations that existed before the major movements and admixtures of population groups that have taken place during most of this century.” Though this application was rejected, we eventually, some 7 years later in 2004, secured funding, in collaboration with Peter Donnelly, from the Wellcome Trust for the “People of the British Isles” (PoBI) project on the basis of the value of having a good UK control population for doing studies on inherited susceptibility to common chronic diseases. A better understanding of the origins of British diversity would be a by-product of the study.

There was some scepticism as to whether it would be possible to do such a study requiring the collection of thousands of samples from rural areas throughout the UK and all from people whose grandparents came from more or less the same area. We now have more than 4000 samples from volunteers who mostly satisfy these criteria. A wide variety of approaches has been used to attract our volunteers in each of the rural areas we have visited. These have included substantial help from the local news media, attending agricultural shows, contact with local history and genealogical societies, Women’s Institutes, the National Farmer’s Union, and with Rotary International, the Inner Wheel and Tangent. In many cases, we have also given talks to attract an interested audience and included a little free refreshment. In addition to forms for informed consent and basic demographic information, we collect blood samples as a source of DNA for the genetic studies. The blood is collected into

special tubes to keep the white cells, which contain the DNA, alive at room temperature. These tubes are rushed back to the laboratory where some of the white blood cells are separated out under sterile conditions, frozen down and stored in liquid nitrogen, while the residue is used for preparing DNA for immediate use. The frozen cells can be used in the future to make permanently growing cell lines that can provide an unlimited source of DNA for further research. We now have more than 1000 such cell lines and they have, for example, been very useful for obtaining whole genome DNA sequences.

Any two unrelated individuals will differ on average at about one in every 1000 of the DNA letters along their whole genome sequences of about 3000 million letters. Since the start of the PoBI project, the technology for looking for such DNA differences between individuals has advanced at an extraordinary pace, making it possible to look at far more differences than we originally imagined would be the case. Thus, it has become affordable to look at more than a million specific positions along the DNA where differences in sequence between individuals are commonly found. We can then use a recently developed and highly sophisticated statistical method, called fineSTRUCTURE, to assess the genetic similarities between individuals at, say, 500,000 of these variable positions. This analysis does not just count the number of these positions where two individuals have the same DNA letters, but also takes into account where these positions are along the chromosomes. Thus, whether or not two positions that are close together on the chromosome both differ or are both the same provides very valuable additional information on the extent of similarity between any two individuals at the DNA level. The output of the FineSTRUCTURE analysis leads to clustering people into groups such that the people within a group are genetically more similar to each other than they are to people in any other group. It also allows us to assess the relative extent of differences between people in different groups. Thus, it can show whether two particular groups of people are more similar to each other on average than they are to the people belonging to any other group.

## PoBI results: an outline of the Nature paper of 19th March 2015 – "The fine-scale genetic structure of the British population".

The published results are based on a FineSTRUCTURE analysis of 2039 PoBI samples from rural areas throughout the UK who had all four grandparents born within 80 km of each other. This analysis was overseen by Peter Donnelly, and mostly carried out by Stephen Leslie, when he was a post-doctoral fellow in my laboratory, and Garrett Hellenthal, working with Peter Donnelly and Simon Myers. Our approach to obtaining our volunteers effectively samples DNA from that of the grandparents. Since the average year of the grandparents' birth is 1885, we are effectively analysing the British population of the 19$^{th}$ century before the later major population movements associated with the industrial revolution, improved education and improved transport. Having obtained 17 clusters of individuals based solely on their genetic similarities, the map shown in Figure 1 is a plot of all the individuals at the mean position of their grandparents' place of birth, where membership of a group is indicated by the combination of the type of symbol and its colour. The black areas are where no samples were collected.

The first and most striking observation is the extraordinary correspondence between the genetic clusters and geographical location. Most obvious at a first glance is the distinctiveness of the genetic clusters in Orkney, which is perhaps not surprising given that Orkney was a Norse Earldom for several hundred years. Also striking is the clear separation of the Welsh clusters (green and pink squares, yellow inverted triangles) from the rest of the UK, and the extraordinary separation between Cornwall (pink crosses) and Devon (blue circles).

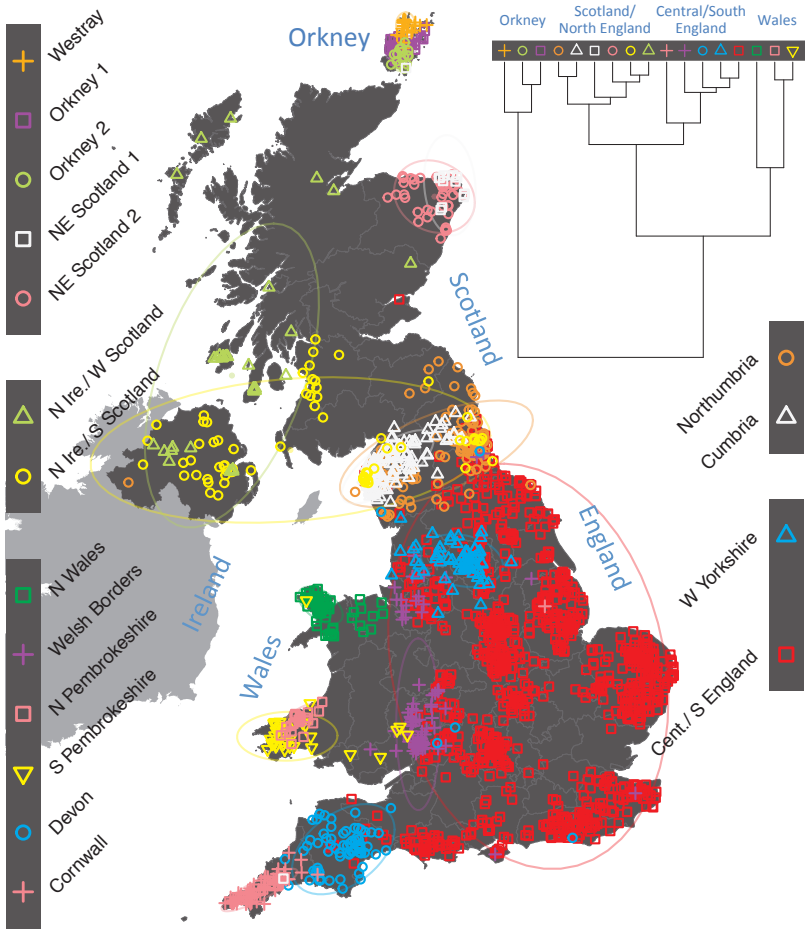


**Figure 1** A genetic map of the People of the British Isles (Figure 1 from the Nature paper)

*For each individual, the coloured symbol representing the genetic cluster to which an individual is assigned is plotted at the mean position of their grandparents' birthplaces. Cluster names are in the side-bar.*

By successively merging the most similar clusters one can obtain a hierarchical cluster tree as shown in the upper right hand part of Figure1. Here, the groups that are most similar have the shortest branches between them, as for example the three clusters in Orkney (purple squares, green circles and orange crosses), and the two in South Wales (pink squares and inverted yellow triangles). Interestingly, based on this hierarchical clustering, north and south Wales are about as distinct genetically from each other as are central and southern England from northern

England and Scotland, and the genetic differences between Cornwall and Devon are comparable to or greater than those between northern English and Scottish samples.

The most different of all the clusters from the rest of the UK are those found in Orkney, as already mentioned, which clearly corresponds to the existence of a Norse Viking Earldom in Orkney from 875 to 1472.

The next level of separation shows that Wales forms a distinct genetic group, followed by a further division between north and south Wales. This division corresponds well with the ancient kingdoms of Gwynedd (independent from the end of the Roman period to the13$^{th}$ century) in the north and Dyfed in the south. Subsequently, the north of England, Scotland, and Northern Ireland collectively separate from southern England. Then, at the next level, Cornwall forms a separate cluster quite distinct from Devon, followed by Scotland and Northern Ireland separating from northern England.

The split in the Northern Ireland group, one with the Scottish highlands and the other with the lowlands, suggests association with the people of Dalriada and with the Picts, respectively, a separation of clans that existed around 600 AD. The split in south Wales (pink squares and yellow inverted triangles) is suggestive of "Little England beyond Wales", as will be discussed later.

Particularly striking is the distribution of the large cluster of people (red squares) that covers most of eastern, central and southern England and extends up the east coast. This cluster contains almost half the individuals analysed (1006).

Several of the other genetic clusters show similar locations to the tribal groupings and kingdoms around at the time of the Saxon invasion (from the 5$^{th}$ century), suggesting that these tribes and kingdoms may have maintained a regional identity for many centuries. For example the Cumbrian cluster corresponds well to the kingdom of Rheged, West Yorkshire to the Elmet and Northumbria to the Bernicia (see Figure 2).

The existence of these largely quite well separated clusters suggests a remarkable stability of the British people over quite long periods of time. This is in marked contrast to what is often assumed.

It is important to emphasise that, although the genetic clustering found by FineSTRUCTURE analysis is quite clear and statistically very significant, it is based on very small genetic differences. Nevertheless it seems very likely that we would be able in many cases to assign an individual of unknown origin to their cluster location. This is illustrated by the fact that we found an individual sampled in the North east of Scotland who was thought to come from Blackburn near Manchester. It turned out, on further checking of the paper work, that this individual came from a Blackburn that was in Aberdeenshire, and not the better-known Blackburn in Lancashire.

To understand the relationships of the clusters to each other we need to consider the history of the British people and the possible contributions to their genetic makeup from the surrounding European countries.

The population of the UK has a relatively simple history compared to the rest of Europe. Great Britain was first

inhabited by modern humans after the last ice age, some 12,000 years ago, with the next major perturbation being the arrival of agriculture some 5,000 to 6,000 years ago. Even if agriculture came with some influx of people, the numbers were probably not large and by that time any genetic differences between the indigenous British populations and those at the origins of the farmers in the fertile-crescent would have been diluted out. The Roman occupation lasted from 43 – 410, followed soon after by the Saxons and related groups. The Vikings then came around 800-950, and finally there was the Norman invasion of 1066, after which there has been no further major invasion of Great Britain. Of all the migrations after the initial settlement, historical and archaeological evidence suggests that only the Saxon and Viking invasions are likely to have had a major impact on the genetics of the people of the British Isles.

In considering the contribution of human migration to changing culture, it is very important not to confuse the influence of a small elite group, such as the Normans, who may have a huge impact on culture and language but little impact on the genetics of the masses, from larger scale migrations, such as those of the Anglo-Saxons and Norse Vikings. History and archaeology focus on the elite groups, whereas genetics looks at the constitution of the whole of the population, including the peasants on the land who, after agriculture, were the main mass of the population until comparatively recently.

Data from a study of multiple sclerosis on 6,209 individuals from 10 different European countries were used to assess the extent to which these European countries might have contributed to the genetic composition of the British genetic clusters. These data were first analysed into genetic clusters using FineSTRUCTURE, just as for the British samples, though the sourcing of the European samples was much more coarse-grained that that done by PoBI. Most European countries were well separated into different genetic groups by this analysis, with some showing significant internal genetic heterogeneity, analogous to that found in the UK. A complex statistical procedure was then used to estimate how the British clusters could be expressed as mixtures of the European groups. Counting Norway as a single source, only 9 of the 51 European groups identified by the FineSTRUCTURE analysis contributed significantly to the British clusters. A summary of the results of this analysis is shown in Figure 3. Each pie chart represents one of the 17 British clusters and the relative contributions of the different European groups to that cluster are proportional to the sizes of the sectors in the pie chart, with the colour of the sector indicating its source.

The dark blue Norwegian contribution stands out clearly in the Orkney samples, as expected, but represents only about a 25% Norse Viking admixture. This shows that the Norse Vikings certainly did not wipe out the resident Pictish population and replace it, but rather intermarried significantly with it. There are also clear Norwegian contributions to all the Scottish and Northern Ireland samples, less to Northern England, even less to Wales and very small contributions elsewhere.

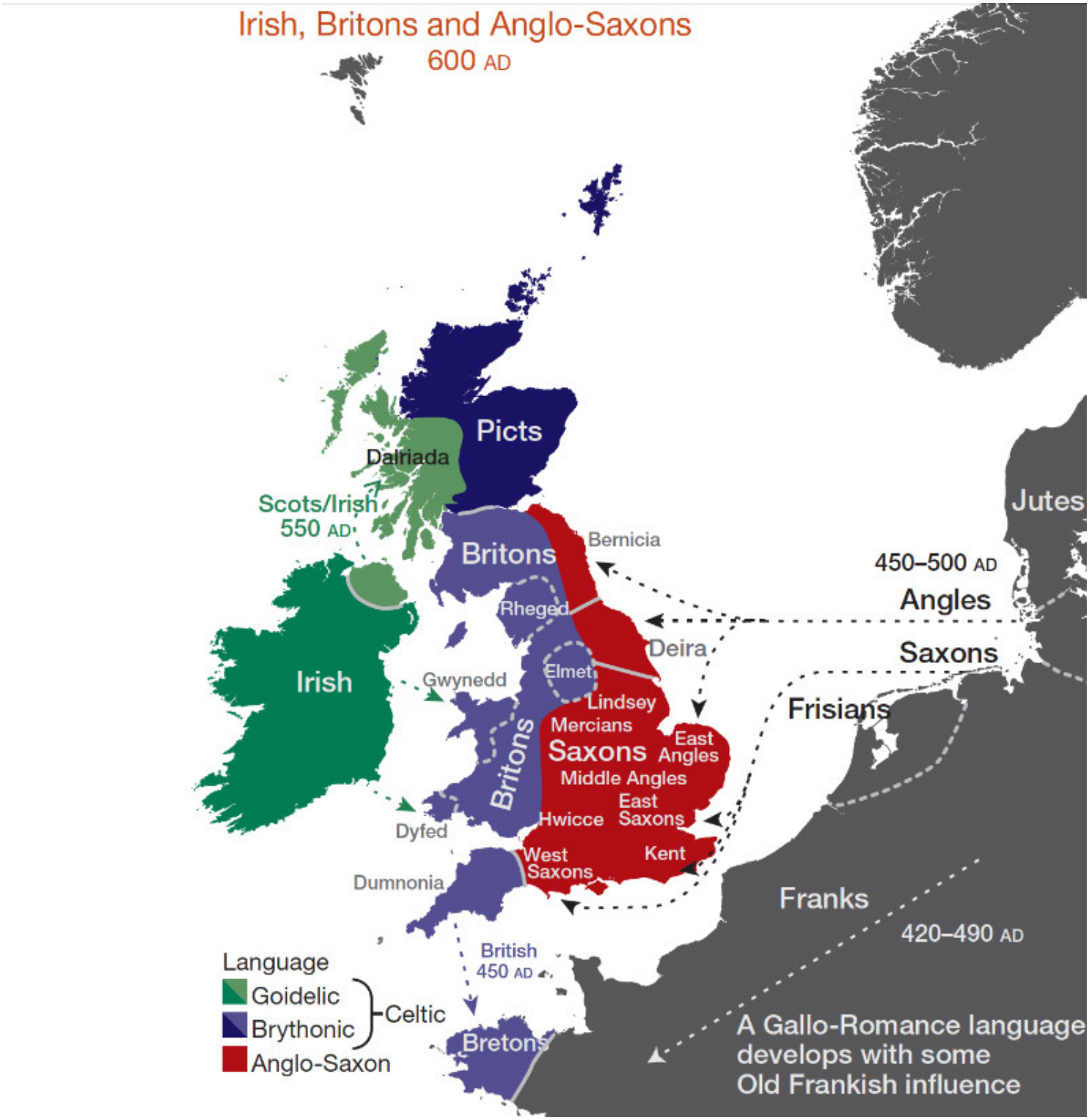


**Figure 2** The regions of ancient British, Irish and Saxon control (Figure 3c from the Nature paper)

The three Welsh clusters are the most distinctive and completely lack contributions from North and North West Germany (EU3 pink) and Northern France (EU17 red). They have the largest contributions from West Germany (EU6 medium green) and North West France (EU14 dark green). This configuration strongly suggests that the Welsh may be closest to the original settlers who came to Britain after the end of the ice age. While there is no clear 'Celtic Fringe', as is so often assumed, there is evidence of ancient British DNA in common with other British populations, especially in Scotland and Northern Ireland, but less in Cornwall, or Devon, in contrast to what might have been expected.

The small differences between South and North Pembrokeshire, especially the slightly larger contributions from Belgium (EU11 yellow) and Denmark (EU18 dark red) (matching Danish place names in South Pembrokeshire) are consistent with the suggestion that this group may represent the area that is sometimes called "Little England Beyond Wales". This is because the farmers settled there by Henry II probably mostly came from that part of Europe.

The most obvious contribution representing the Anglo-Saxons is EU3 (pink) from North and North West Germany. That is consistent with the lack of evidence for Anglo-Saxon incursions into Wales. Denmark (EU18 dark red) is another clear candidate for an Anglo-Saxon contribution. Based on these two contributions, the best estimates for the proportion of presumed Anglo-Saxon ancestry in the large eastern, central and southern England cluster (red squares) are a maximum of 40% and could be as little as 10%. This is strong evidence against an Anglo-Saxon wipe-out of the resident ancient British population, but clearly indicates extensive admixture between the incoming invaders and the indigenous people. The difference between Devon and Cornwall is most probably due to the greater Saxon influence in Devon, this being consistent with the slightly greater contributions of EU3 (pink) and EU18(dark red) to the makeup of the Devon cluster as compared to that in Cornwall.

The homogeneity of the east, central and southern British cluster (red squares) with no obvious differences in the Danish contribution (EU18 dark red) between them and the more northern English populations, strongly suggests that the Danish Vikings, in spite of their major influence through the "Danelaw' and many place names of Danish origin, contributed little of their DNA to the English population.

There is evidence for only a very small Spanish contribution to the PoBI samples, in contrast to what has been claimed by some authors.

The most intriguing European contribution is that from Northern France, (EU17 red). This clearly post- dates the original settlers since it is entirely absent from the Welsh samples. It is, however, widespread elsewhere, even right through the north of England and Scotland to Orkney. It is also especially prevalent in Cornwall and Devon. These results suggest a previously not described substantial migration across the channel after the original post-ice-age settlers but before Roman times. DNA from these migrants spread across England, Scotland, and Northern Ireland, but had little, if any, impact on Wales.

We have in this analysis of the PoBI samples achieved a much finer genetic structural analysis than has so far been achieved for any other population. The success of this very fine genetic structure analysis, creating an extraordinary genetic map of the British Isles, including Northern Ireland, can be attributed to a) the careful collection of the samples, b) the use of a large genome wide panel of genetic markers and c) the use of a novel sophisticated analysis of genetic differences and similarities.

Our thanks once again to all our volunteers without whom we could not have done this. Thanks to all the members of the PoBI group past and present, especially Bruce Winney, who did so much for organising the collections.

Thanks also to all our collaborators who helped us in our sampling and, of course, major thanks to the analysis team of Peter Donnelly, Stephen Leslie, Garrett Hellenthal, Simon Myers and their colleagues and to Mark Robinson for his incisive archaeological advice.

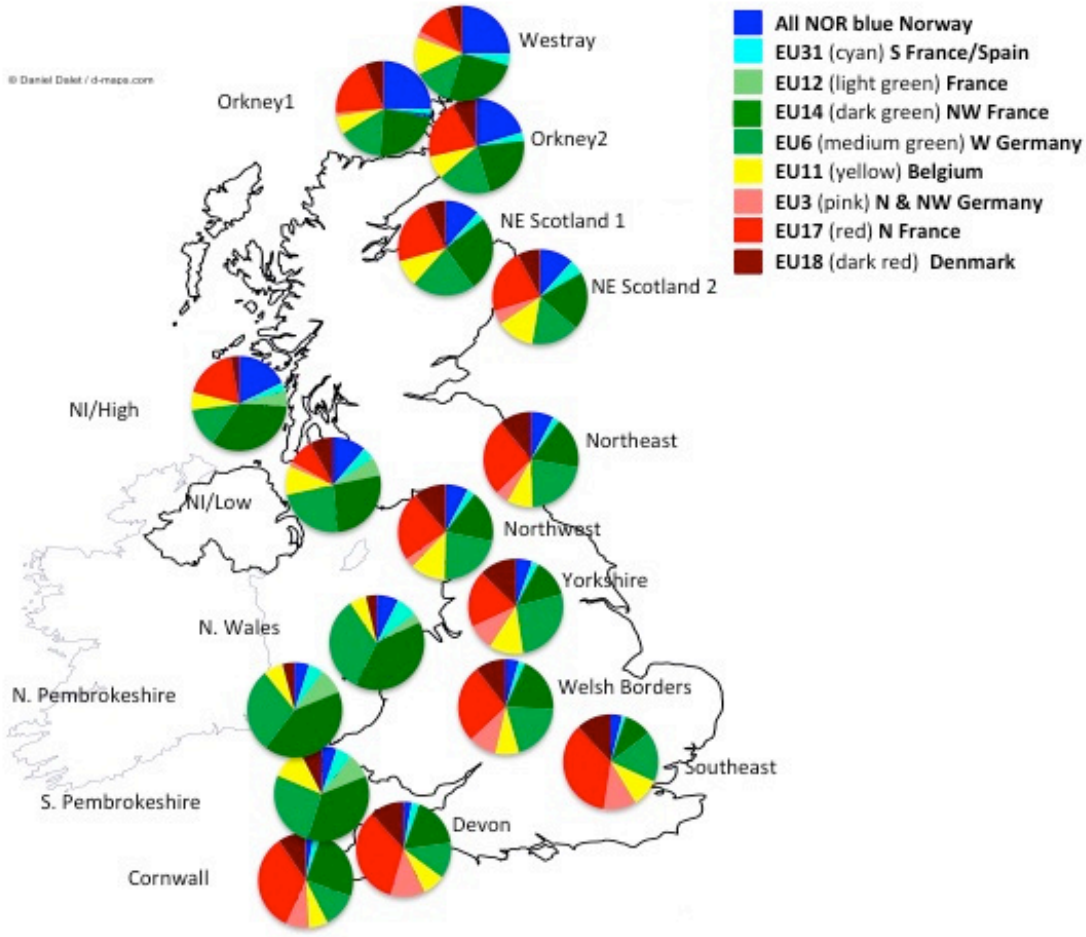


**Figure 3** European ancestry profiles of the UK clusters (modified from extended data Figure 6 of the Nature paper)

*Each pie chart represents one of the 17 British clusters and the relative contributions of the different European groups to that cluster are proportional to the sizes of the sectors in the pie chart, with the colour of the sector indicating its source.*

# The Genetics of Variation in Facial Features and other Normal Variation

In 2009, as already mentioned, we obtained 5 years of funding from the Wellcome Trust for the expansion of the project into the analysis of the genetics underlying normal genetic variation in humans, especially of facial features but also including skin tone, hair and eye colour, height, sleep patterns and smell and taste perception. This is being done by returning to our UK volunteers (including some new recruits) and combining a questionnaire with taking measurements. The major aim of this new part project is to understand how genetics controls differences between people's faces. The technology is now available to capture extremely effective 3D representations of faces, rendered as 90,000 data points: a far better platform for assessing complicated facial morphology than 2D photographs. Though there are substantial statistical challenges in analysing such big datasets, it should be possible to find specific major genetic differences responsible for particular facial differences, by comparing the DNA of individuals with carefully chosen particular facial signatures. We have made good progress on this front, having already accumulated more than 2000 images from volunteers on whom we have the genetic information, and hope to publish our initial set of findings in the summer.


**Figure 4** "Sir Walter Bodmer and his PoBI research team".