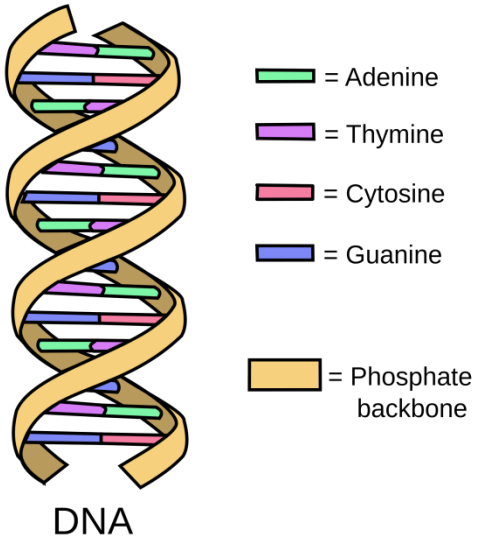
Use of DNA information in family research – information for IOWFHS members

What is DNA?

Since the discovery of deoxyribonucleic acid (DNA) in the 1950s, we have come to understand more about its role as a blueprint for all living things on this planet.

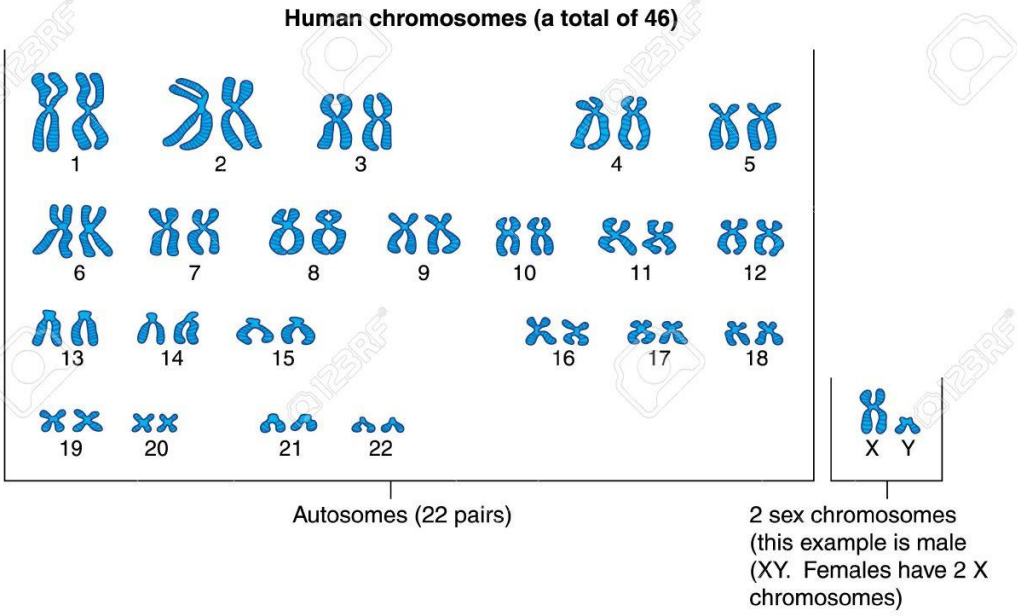
The discovery of DNA was credited to Watson and Crook at Cambridge University, but the reality was that the discovery was the result of collaboration between different groups of scientists in the UK and the US, piecing together their research findings.

DNA molecules are large and complex. We now know that the two chains of the DNA molecule are formed in a double helix pattern made of phosphate and sugar bound together by 4 different organic bases; cytosine (C), thymine (T), adenine (A) and guanine (G). It is the pattern of these four bases along the length of the molecule which forms a code, analogous to the binary digits in a computer program. This is the genetic code that determines the characteristics of every living thing.



A gene is a short section of DNA which has a specific function, e.g. the determination of eye colour. Genes are responsible for every characteristic you inherit, such as whether your hair is straight or curly, the colour of your skin or how tall you are.

In human beings the DNA is bundled up into 23 separate packages called chromosomes, each made up of an incredibly long strand of DNA. These chromosomes contain the instructions that tell your body how to grow, develop, function and give you all your individual traits. Every living cell in your body contains the same 23 chromosomes within its nucleus.



Genome is another term that is often used in this context. A genome is an organism's complete set of DNA, including all of its genes. Each genome contains all the information needed to build and maintain that organism. In humans, a copy of the entire genome—more than 3 billion DNA base pairs—is contained in each and every cell that has a nucleus.

The genome also includes other genetic material such as mitochondria. Mitochondria are a source of energy in cells and also have a role in gene expression and programmed cell death. Mitochondria are normally inherited exclusively from the mother; the mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization. The fact that mitochondrial DNA is maternally inherited makes it very useful in identifying the maternal line in DNA testing.

Why can DNA be a useful tool in family research?

Except for identical twins, each person's DNA is unique. This is why people can be identified using DNA fingerprinting. DNA can be cut up and separated, forming a sort of 'bar code' that is different from one person to the next.

It's the fact that DNA is inherited from generation to generation that makes it a useful tool in family research. Humans have 22 pairs of what are termed autosomal chromosomes in which one chromosome is inherited from the father and one from the mother. A child inherits one copy of each chromosome from each parent. However, although 50% of DNA is inherited randomly from each parent, the distribution of DNA from each grandparent is not uniform, meaning that a person may inherit more or less than 25% of their DNA from each grandparent. Go back several more generations and it's possible that a particular ancestor may not have contributed significantly to a descendant's DNA. At 10 generations back only a small fraction of

those ancestors will have contributed directly to an individual's DNA. This has an implication for autosomal DNA testing (see below).

It's also important for people to realise that siblings can inherit different combinations of the DNA segments form each parent and this can cause differences in the outcomes of some of the tests described below. (See for example the discussion at https://news.nationalgeographic.com/2018/03/dna-ancestry-test-siblings-different-results-genetics-science/)

The 23rd chromosome pair is the sex-linked X and Y chromosomes and these have a different inheritance pattern. One whole X chromosome is inherited from the mother, randomly from either of the two whole X chromosomes she inherited; one each from her mother and her father. Each individual inherits either a whole X or a whole Y chromosome from the father; if it is the X chromosome the person will be female, if the Y chromosome, the person will be male.

Some inherited genes are termed dominant – i.e. their function is expressed in the individual. Other genes may be termed recessive – i.e. they have the capability to perform a function, but are not expressed in the individual. However, when passed on to the next generation, a recessive gene may be expressed.

What do proprietary DNA test kits offer and which is best?

Most proprietary test kits offer one or a combination of the following type of DNA tests:

- Autosomal testing is the usual way into using DNA for most family research. Autosomal DNA tests normally focus on the 22 pairs of chromosomes not involved in determining a person's sex (i,e, all but the X and Y chromosomes). It is invaluable for matching strands of your DNA with those from cousins and other distant relatives. It is also used to estimate ethnicity. While autosomal testing can match your DNA to those of relatives, it will not show you which side of the family they are related to. This is why it can be useful to encourage each of your parents to be tested as well, if they are still alive, and as many known relatives as possible (the further back up the family line the better!) as each person will have inherited different amounts and different segments of DNA from each of the various shared ancestors.

- Y-DNA testing focuses on the Y chromosome which is only possessed by males, and is passed from father to son. It can therefore be used to trace the direct paternal line in your father's lineage. As only males can take a Y-DNA test, females wanting to make use of this test must use another male descended in her direct paternal line (e.g. a brother, father, cousin or uncle) as a proxy.

- mtDNA (mitochondrial DNA) tests look at the chromosomes of DNA in a person's mitochondria. These DNA strands are passed from mother to child and can be used to trace your direct maternal line. Both males and females can

use this type of test. There's very little chance that these strands could be altered, so your maternal line can be traced back quite far by this method.

Some of the main kit providers are listed below:

Kit Provider	Y-DNA	Mt DNA	Auto-somal DNA	Ethnicity	Cousin Matching	Comments
23 and Me	Yes*	Yes	Yes	Yes	Yes	Optional health check provided.
Ancestry DNA	No	No	Yes	Yes – most extensive database available	Yes Largest database with family tree matching	Easily uploaded to other matching sites. No details of which chromosome/s matched.
Family Tree DNA	Yes*	Yes	Yes	Yes	Yes, and once you have linked a known close cousin, will identify other matches on that parent's side. Best test for matching through Y-DNA and mtDNA	Can upload from Ancestry and others (often free). Triangulation tool available. Tests are bought separately – the most expensive, especially for upgrading Y_DNA analysis.
Living DNA	Yes*	Yes	Yes	Yes, the most specific for those with British ancestry.	Available in 2018	All three tests are standard. Frequently updated as their database grows and their algorithm adjusted.
MyHeritage	No	No	Yes	Yes	Yes	Can upload from Ancestry and others. (currently free) Cheapest test, but smallest matching database

*Males only

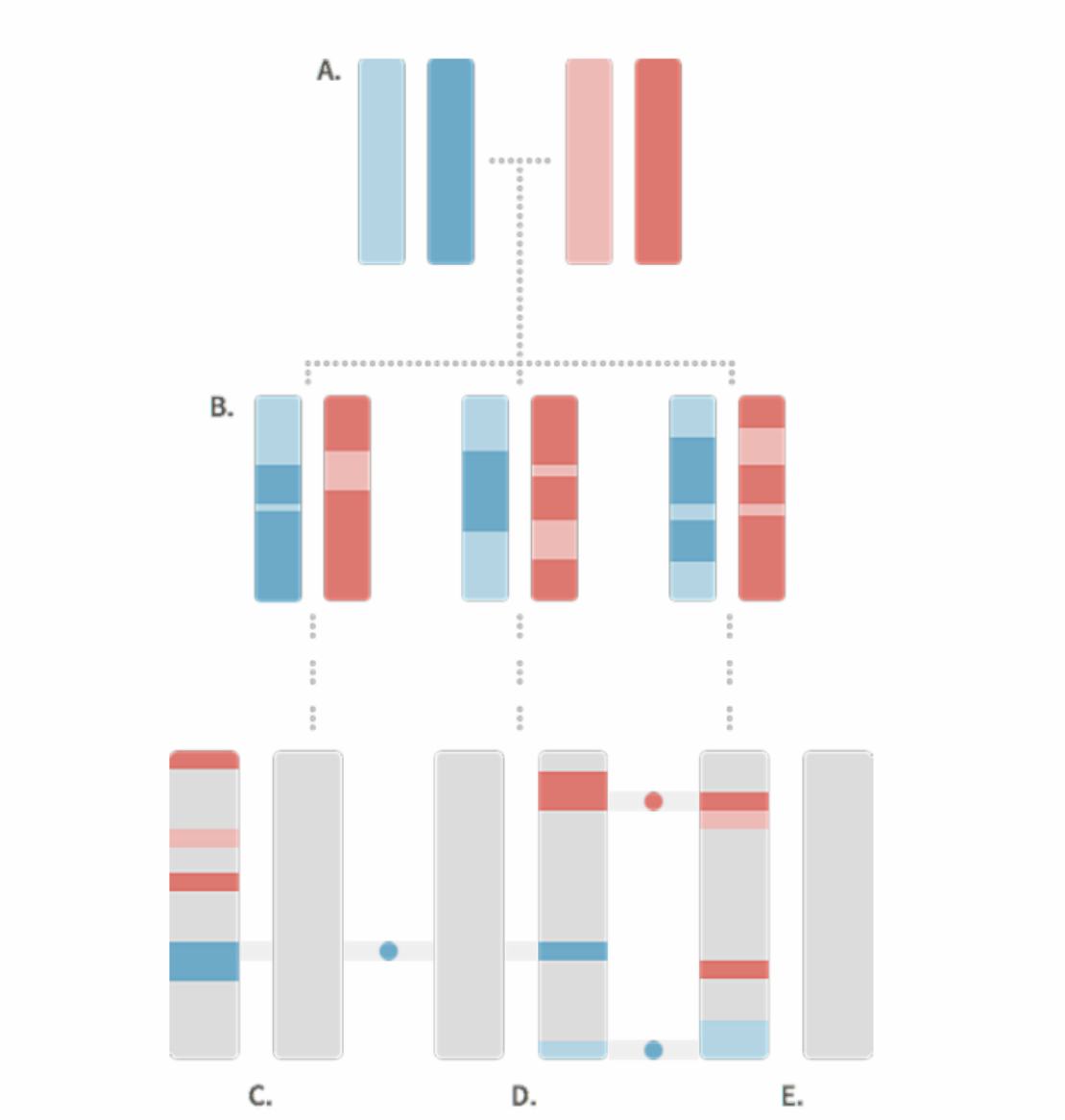
The best test for you depends on your objective. If you are looking to build or confirm your family tree, an autosomal DNA test would normally be the best option. This test can help identify other people who share one or more of your ancestors, back to about 10 generations. Beyond this point only a small fraction of your ancestors will have contributed directly to your DNA and it will be difficult to identify a particular shared ancestor by this method. If you want to look beyond about 10 generation or roughly

300 years, Y-DNA and mtDNA tests can provide accurate information about your paternal or maternal line respectively, potentially up to 100 generations and even further, depending on the specific test used.

Most providers will also enable you to download your DNA data for further analysis on sites such as GEDmatch and Geni.

Can you tell me more about matching relatives by DNA?

There's a good overview of how this is done this in Ancestry's DNA Matching white paper and the following illustration and explanation is adapted from that document. (https://www.ancestry.com/corporate/sites/default/files/AncestryDNA-Matching-White-Paper.pdf)



The illustration above depicts how the chromosomes of two common ancestors (A) are shared between three of their children (B). Chromosomes of the common ancestors (A) and their children (B) are shown. Those in blue are from the father; those in red are inherited from the mother. Although each child inherits 50% from the father and 50% from the mother, the actual distribution of the DNA within each chromosome is randomised.

Chromosome pairs for three 5th cousins (C, D and E) sharing the same two common ancestors (great-great-great-great-grandparents) are also depicted. In this case, these three 5th cousins have each inherited only a small proportion of their DNA from the two common ancestors. The DNA that is identical-by-descent (IBD) between distant cousins (C, D, E) is depicted as red or blue. Chromosomes of other intermediate

generations are not shown in the diagram. The blue and red circles indicate chromosome segments that are IBD between the indicated chromosomes.

Notice that because the transmission of DNA has repeated several times over several generations, DNA from different common ancestors (red and blue) can end up on the same chromosome of an individual. The grey portions of the chromosomes are inherited from other ancestors which may or may not contain segments that are IBD among the three 5th cousins.

While the three 5th cousins have all inherited some DNA from the common ancestors shown in the figure, only a few short segments of the chromosomes are actually identical in the same places on the chromosome of different cousins. In this example, we see that only 3 short chromosome segments, indicated by the blue and red circles, are IBD. One segment of DNA is shared by cousins C and D, and two segments are shared by cousins D and E. By contrast, cousins C and E, despite the fact that they are related through their great-great-great-great-grandparents (A), do not have any identical DNA that is IBD through these two common ancestors.

The first goal of DNA matching is accurately to identify the DNA segments on the 22 chromosome pairs that are identical-by-descent between pairs of individuals.

When linking DNA matches to family trees, you will often encounter a higher or lower generation difference for a cousin than you know is correct. This does not mean that the DNA analysis is wrong; instead it reflects the random element in DNA inheritance. It is also possible for known 4th cousins to have no autosomal match and more distant cousins to appear to be more closely related than they really are.

According to the test I took I am 50% Western European. How is this calculated and are the results reliable?

Ethnicity is estimated by comparing your autosomal DNA against DNA from a 'Reference population'. A reference population is a large number of people living in a particular region. Sometimes an additional filter may be applied to select those individuals whose origins are more likely to be representative of that region – e.g. by only selecting individuals whose grandparents were all born in the same region.

Because ethnicity tests are estimated from autosomal DNA, it is important to remember that the DNA from the majority of ancestors beyond 10 generations back in the past will neither be found in your DNA nor in that of the reference population. It is incorrect therefore to assume that our autosomal DNA can be used to identify the origins of our ancient forebears. To do this you would need to test the Y-Chromosome and/or the mitochondrial DNA.

Providers use different criteria to try to ensure that each person in the reference population is truly representative of that region. They use different methods (called algorithms), mathematical models and different reference populations to give a "best estimate". This is why results obtained from two different providers are very likely to be different.

It is important to remember that the percentage estimate for a specific ethnicity by a DNA testing company does not necessarily mean that this percentage of your DNA comes from that place. It simply means that your DNA shares DNA characteristics in common with that company's population sample. Because the methodology compares DNA in relatively modern populations this can sometimes give surprising results. For example, some British people have been known to obtain a small percentage Nigerian or "West African" in their ethnicity results – this is more likely to reflect European migration into West Africa showing up in the reference population (i.e. West Africans having European DNA) rather than vice versa.

Bear in mind too that because of the random inheritance of DNA, siblings and other close relatives can often be assigned markedly different ethnicity estimates.

In general terms, the larger the reference database, the degree of rigour applied to screening the reference population, and the methodology employed to calculate ethnicity, the more reliable the estimate. However, ethnicity estimates are subject to a number of limitations and any quantitative information should be interpreted with particular caution.

University College London provides some useful information on its website to help non-scientists understand the limitations of this technique and warns against some of the more extreme claims which it refers to as 'genetic astrology' (see https://www.ucl.ac.uk/mace-lab/debunking).

Is it likely that the DNA of people who have a significant Isle of Wight ancestry is distinctive from other parts of Southern England?

This is a question which a number of members have expressed an interest in, and an article published in the May 2018 IOWFHS Journal 129 ('Island DNA' by Ian Roach) is relevant.

Members should avoid comparing ethnicity results obtained from proprietary DNA tests and drawing conclusions about the different ethnic groups involved in the history of the Isle of Wight. In the previous section we saw that the results are affected by the selection of reference database, the degree of rigour applied to screening the reference population, and the methodology employed to calculate ethnicity. Because ethnicity estimates are subject to so many limitations any quantitative information should be interpreted with caution.

In 2015 a landmark paper was published in association with the Wellcome Trust, through what is known as the People of the British Isles (POBI) project. In this study 2039 DNA samples were obtained from rural areas of the British Isles, from people whose four grandparents were all born within 80km (50 miles) of each other. After over a decade of sample collection and data analysis, the findings of the study were published in the Journal Nature in March 2015 (http://www.nature.com/articles/nature14230).

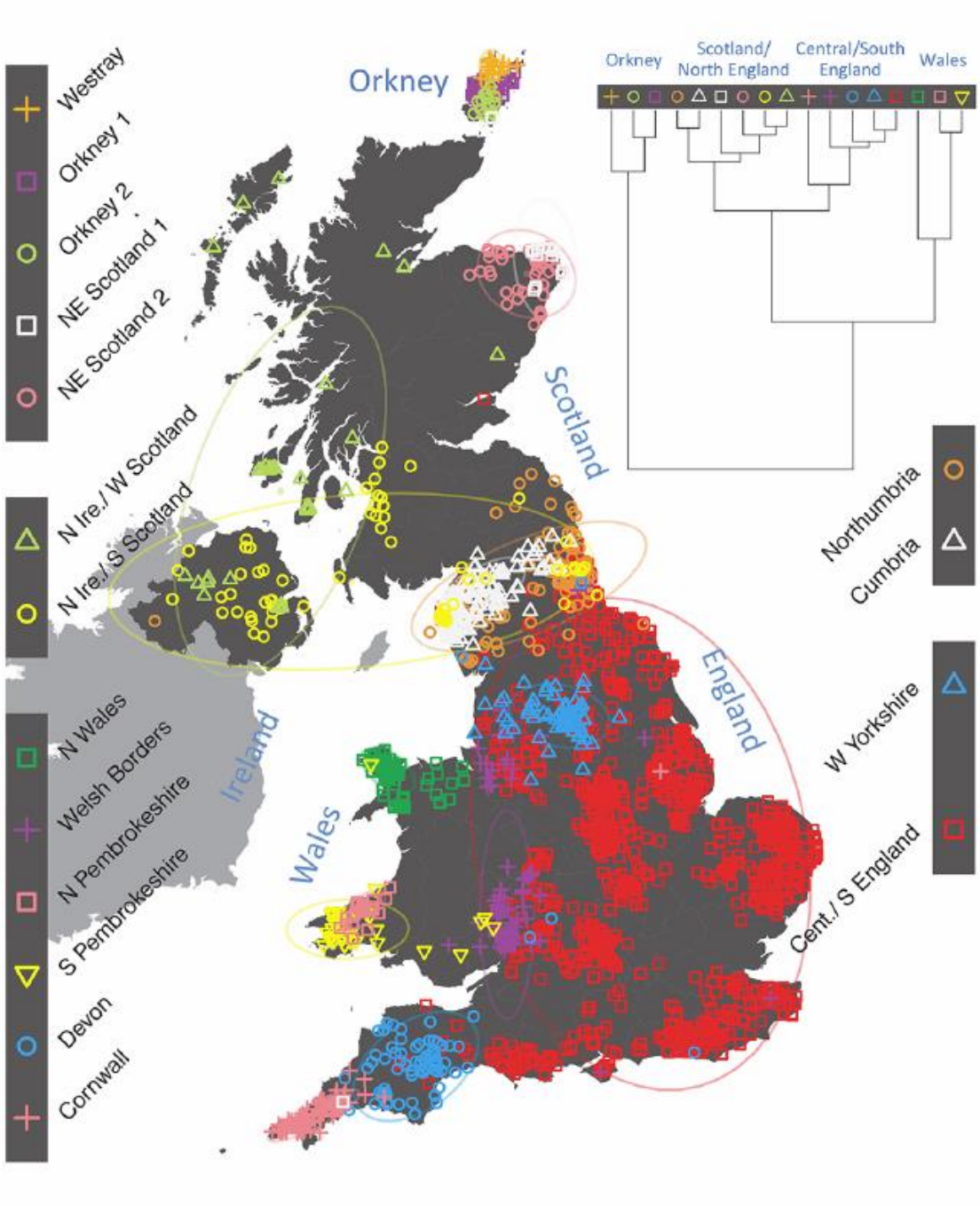
The researchers were able to obtain 17 clusters of individuals in the United Kingdom based solely on similarities in their DNA that matched their geographical locations. These clusters are shown in the map below.

By successively merging the most similar clusters the team was able to obtain a hierarchical cluster tree, and this is shown above the upper right part of the map. Here, the groups that are most similar have the shortest branches between them, 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).

The map and the cluster tree gave some surprising results: For example, 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, with a Viking ancestry.

The next level of separation shows that Wales forms a distinct genetic group, followed by a further division between north and south Wales. At the next level, the north of England, Scotland, and Northern Ireland form another group which is separate from those in 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.



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

A further investigation was done to understand the possible contributions to the genetic makeup of each of the clusters from the surrounding European countries.

Data from a study of multiple sclerosis on 6,209 individuals from 10 different European countries[1] were used to assess the extent to which these European countries might have contributed to the genetic composition of the British clusters.

These data were first clustered into groups on the basis of genetic similarity. Out of 51 groups identified, only 9 contributed significantly to the British clusters. Their findings, which give a fascinating insight into the history of the British Isles, are summarised below:

- The population in Orkney emerged as the most genetically distinct, with 25% of DNA coming from Norwegian ancestors. This shows clearly that the Norse Viking invasion (9th century) did not simply replace the indigenous Orkney population..

- The three Welsh clusters are the most distinctive. The data strongly suggests that the Welsh may be closest to the original settlers who came to Britain after the end of the ice age.

- Contrary to expectation, there was not a single "Celtic" genetic group; instead Scotland, Northern Ireland, Wales and Cornwall are among the most different from each other genetically. The Cornish are much more similar genetically to other English groups than they are to the Welsh or the Scots.

- The most obvious contribution representing the Anglo-Saxons is from North and North West Germany and Denmark. 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 intermarrying with 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 North and North West Germany and Denmark 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 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.

[1] The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214–219 (2011).

- The analyses suggest there was a 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 impact in Wales..

(See more at https://peopleofthebritishisles.web.ox.ac.uk/population-genetics)

An important aspect for IW family history is that the Island was represented in this study, albeit by only a handful of subjects. Nevertheless, given the strict criteria applied to the selection of subjects in this study, we can have some confidence that those subjects would be truly representative of the Island. At the current time then, it seems that it is not possible to distinguish between the DNA of people with an Island family history to other people from southern, central and eastern England.

IOWFHS members might be interested in a piece of research submitted by James Rayner as his BA Thesis to the University of Iceland in 2015 which argues that the Isle of Wight dialect shows signs of Celtic influence. It is available online as a pdf document: https://skemman.is/bitstream/1946/20651/1/James_Rayner_BA_Thesis_Skemman.is_Feb_2015%20%283%29.pdf Some of the examples given in the thesis are debateable, and an impact on culture does not necessarily imply a shared genetic connection, but it is an interesting analysis nonetheless. The data from DNA studies such as PoBI do not currently support a Celtic ancestry in the Island population.

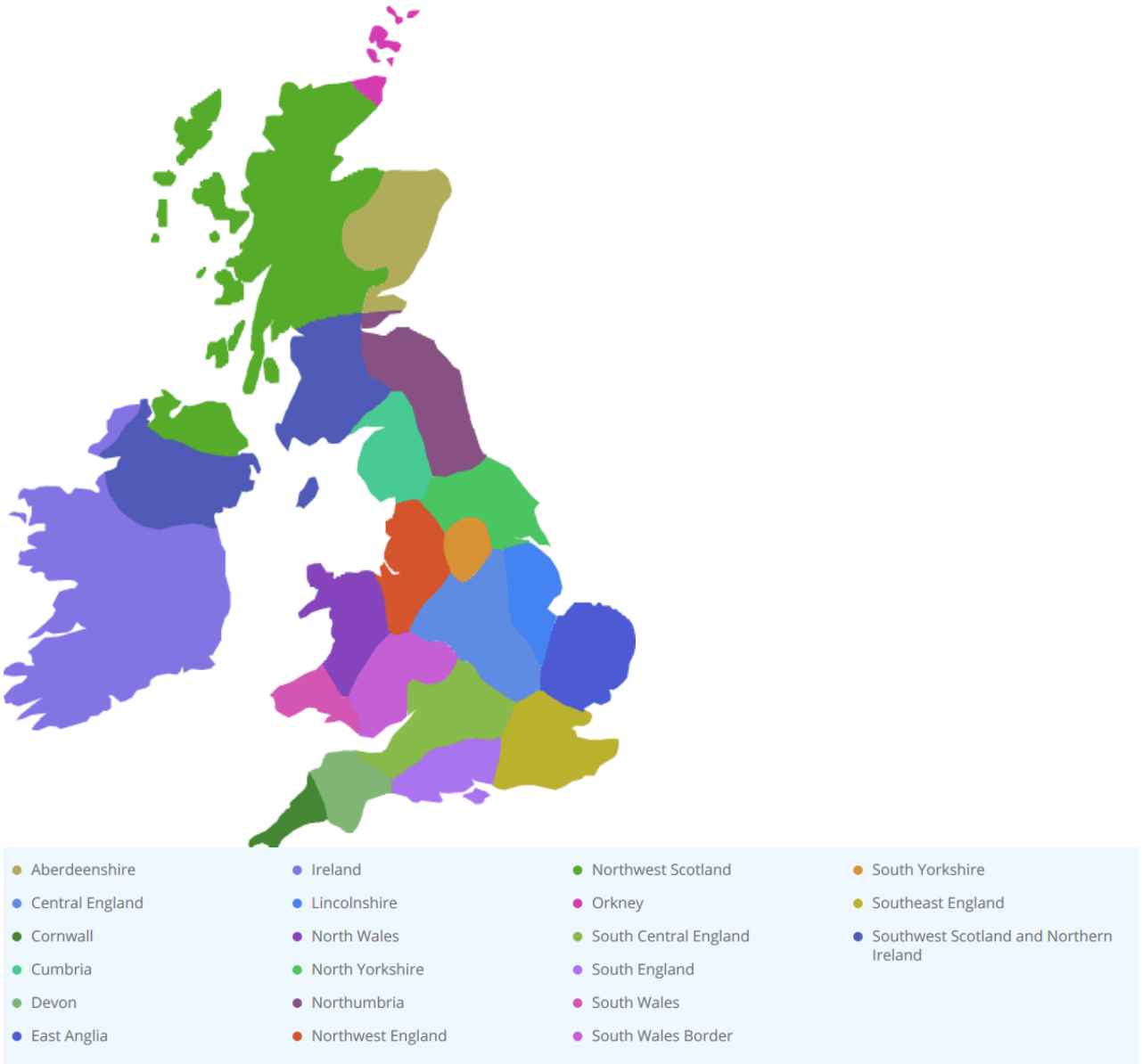
NB Since the publication of the POBI study, a study using a similar approach to create an Irish DNA atlas of the British Isles was published in 2017. (https://www.nature.com/articles/s41598-017-17124-4)

Although DNA science cannot currently distinguish any difference between the Isle of Wight and the rest of Southern England, is it likely that we may be able to do so in the future?

The answer to this question is almost certainly yes, but as the differences are likely to be relatively small, it will be crucial to build a large enough DNA sample and reference sets whilst maintaining tight criteria for the selection of subjects.

There is a team at Living DNA based in Frome Somerset (which has links with some of the researchers in the POBI study) which is interested in exploring these distinctions further. They are encouraging people to take a DNA test with them, but are also interested in receiving data from individuals who have obtained proprietary DNA results through other providers. For inclusion in the database, this team follows the strict criterion applied in the POBI study whereby all 4 grandparents need to be born within a 50 mile radius.

Living DNA – UK and Ireland regional breakdown



They have subdivided the UK and Ireland into 21 distinct regions based on DNA. This is already a refinement on the POBI study, but the Isle of Wight is still not distinct from Hampshire, Wiltshire and Dorset. (See https://www.livingdna.com/en-gb/family-ancestry)

Living DNA calls this the One Family research project and it is actively seeking more data to strengthen its findings and database. The aim of this research is to achieve the greatest level of differentiation possible, not just for the UK but for all other countries around the world.

Like the researchers working on the POBI project, Living DNA asks that all participants provide information about their parents' and grandparents' places of birth. Those whose grandparents were born within a 50 mile radius will be eligible to be used in the research, and even those who don't qualify are still encouraged to upload their DNA data to the project.

IOWFHS members, especially those with four grandparents born on the Island, can find out more about how to participate in this project via this link: https://www.livingdna.com/en-gb/one-family/research

Is there more that I can do with my DNA data?

So, you've done a DNA test with one of the providers; is there more that you can do with the data? Well actually there are a number of options which can help you to maximise the value and usefulness of the data you have.

First you will need to download your raw DNA data to a separate file. Your provider should have a simple way of doing this. With Ancestry, for example, you simply go from the DNA Home Page to 'Settings' and on the Settings page there's a panel on the right hand side with the heading 'Actions'. One of the Actions is 'Download RAW DNA data'. This creates a file which can easily be uploaded into a third party's site.

- Use of GEDmatch for processing and comparisons: It is worth registering and submitting data to GEDmatch (which is a free to use platform available at https://www.gedmatch.com/select.php) which allows you to undertake cousin matching through their database and targeted ethnicity testing with several different databases, each with their own "reference populations" which give a variety of answers. GEDmatch has a number of useful free tools, and more advanced tools for a nominal fee, designed to help your research.

- DNA Painter is a chromosome mapping tool that helps to make sense of DNA testing. By mapping segments of DNA to chromosomes, it's possible to see which ancestors gave us which pieces of DNA, and thus how new matches are related. Find out more at https://dnapainter.com/

- As mentioned already, Living DNA continues to build its database of the UK and other countries, and during 2018 it will be possible to register with them and upload raw DNA data files to their One Family research project which will offer cousin matching. See https://www.livingdna.com/en-gb/one-family/research

- DNA can be uploaded to Family Tree DNA to access its Family Finder cousin matching service and it has a useful suite of genealogical tools to explore.

- For those with an Irish heritage, the Irish yDNA Project welcomes males who are doing or who have done a y-chromosome DNA test (yDNA) and who also have Irish ancestry on their own paternal line to submit data. (https://www.familytreedna.com/groups/ireland-heritage/about/background)

- My Heritage is building its database and welcomes DNA uploads from other tests. This is a relatively new site which is becoming more user-friendly and useful.

- If you have a tree on Geni - you can also upload your DNA and link it to your tree for matches.

- Other specialised databases which welcome new relevant participants include a number of region-specific closed Facebook groups (eg, for Scotland, Ireland, Essex) which help you research your connection to DNA matches who share

that ancestry. They mostly use a comparison tool called Matchbox which compares the group's database of members against the GEDMatch one-to-many match data you upload. There are also more specialised closed groups for researching specific chromosomes and Y-DNA.

Prepared by Dr Steve Stotesbury

with valuable input from

Claudia Boorman
Sue Boothroyd
Susan Johnstone
Stella Ridley
Ian Roach
Pete Spencer