Appendix II - Analysis and Interpretation of Results

Matching and Identification of Genetic Connections - We know that the basic 37 marker Y-STR haplotype obtained for each man in the study can be used to identify genetic connections with other men similarly tested within a genealogical timescale, and to make a rough prediction of when their most recent common ancestor may have lived.

1. The first step is therefore to compare the Y-STR results of each participant with the remaining 500k men recorded within the FTDNA database and calculate the **genetic distance** for each comparison. **Genetic distance** (GD) is the term used to describe the number of differences or mutations between two sets of Y-chromosome DNA or mitochondrial DNA test results. A genetic distance of zero means that there are no differences in the two results and there is an exact match signifying a recent common ancestor for the participants. The interpretation of what the GD value means depends upon on the number of Y-STR markers tested and other factors. This is not an absolute science however.

2. The logic process used in this analysis is as follows:-
    a. Compare genetic distance between two men: The probability of them being related and sharing a common ancestor in the last 800-900 years is increased when:-
        i. There is a close distance between them (within normal meaningful tolerances)
        ii. They share the same or similar family name
        iii. They show the same, or some of the same, low-frequency or rare values for the same Y-STR marker (off-modal analysis)
        iv. If i and ii at least are true, then it is clear that they match each other and are part of the same male genetic family
    b. If two men are close in genetic distance i.e. they are a match, but do not share the same family name, then:-
        i. If they have ancestors who lived in the same town or region within the recordable past, there is a possibility that one of them is a non-paternal event (NPE).
        ii. If there is no geographical overlap in their ancestry in the preceding 900 years then it is likely that the match is an example of genetic convergence[16] and/or any connections would have occurred before family names were adopted.
        iii. A proportion of these matches will be totally random and meaningless. The lower the number of markers tested (i.e. 12 and 25) then the proportion of randomness will be very high. At 111 markers any match is much less likely to be random.

3. There are other factors which might influence the number of matches also:-
    a. The density of population of the haplogroup sector of the man being tested. I.e. In general the number of matches will be greater for a R1b man than for a I1 man.
    b. The length of time that the male ancestral line has stayed in one location. The longer a male line stays in one geographical location, then the greater the genetic distance there will be between their direct male descendants and the descendants of other early common ancestors. In other words, the further back in time that there was a genetic branch from early common ancestors in another place, then the greater the genetic distance will be between the two sets of present-day descendants (and the further apart their genetic matches will be).
    c. Sampling distortions in the FTDNA Y-STR database. There are more than 500.000 records in the FTDNA Y-STR database and inevitably there will still be some statistical distortions in the database coverage of the male genetic tree, compared with the actual world. Such a distortion may affect the number of matches up or down unpredictably and unknowingly.
    d. Rare or unusual marker values. The presence of a particular unusually high or low value for a marker might create an artificially high genetic distance to distort the calculation.

[16] Where sets of Y-STR values randomly converge towards each other over a long period of time, and they appear to be a more recent match than they actually are.

4. **Most Recent Common Ancestor**: Once a number of men with the same family name and Y-DNA signature have been identified, it then becomes possible to predict an approximate timescale in which their most recent common ancestor would have lived. This statistical analysis identifies the most probable date range for when the nearest common ancestor lived for the sample group. This obviously does not preclude the group having earlier common patriarchs, it only identifies the one most recent common male ancestor for that group of men tested.

5. The next stage in the analysis is to try and determine where on the overall male human genetic tree an individual is situated. The tree is divided into haplogroups, i.e. large numbers of men sharing the same relatively high-level genetic mutations, A position on the male human genetic tree is normally signified by a succession of unique genetic mutations (Y-SNP or Single nucleotide polymorphism), each one being at a lower level, and a more precise definition than the preceding one. The lower the level on the tree, then the more recently it is that the mutation occurred.

6. Further, deeper Y-DNA testing is capable of identifying the very lowest defining genetic mutation (Y-SNP) for an individual, but sometimes a reasonably accurate prediction can be made based on the Y-STR values and which will pinpoint the approximate region on the tree, without further testing.

7. Knowledge of an individual's position on the male genetic tree is useful in that it can be used to connect that individual with groups of other men in an earlier time and place. So, for example in this way it is possible to identify a historic ancestral origin in Scandinavia or with a group of Irish tribes or Scottish clans for example, at a time before family names had been adopted. Thus, for a Manx male line that we know must have arrived in the Isle of Man from elsewhere earlier, it becomes possible to identify part of their route by which they arrived and a possible timescale.

What Else can We Deduce?

By the selective testing of men with unique, indigenous, Manx names and our knowledge of history, we can already expect that:-

- Their family line will have lived on the IOM from at least 1200-1400 AD, and
- (The majority of, if not all,) their family names will have been formed and adopted on the island in a time period starting from 1050/1100AD onwards.

We also know from history that in general:-

- There were settlers from Scotland from 1000-1400 at least
- Norse-Gael settlers arrived in the period 900-1350 AD
- Scandinavian Viking settlers came from 850-1350AD
- Native Irish settlers came any time from 400AD onwards

So with that knowledge, we can be reasonably certain that any Hg I1 and Hg R1a men must generally have arrived 850-1350AD and Hg R1b and Hg I2 men any time from 400AD (or before) until 1400AD.