

Complex spatio-temporal distribution and genogeographic affinity of mitochondrial DNA haplogroups in 24,216 Danes

Jonas Bybjerg-Grauholm^{1*}, Christian M Hagen^{1*}, Vanessa F Gonçalves², Marie Bækvad-Hansen¹, Christine S Hansen¹, Paula L Hedley¹, Jørgen K Kanters³, Jimmi Nielsen⁴, Michael Theisen¹, Ole Mors⁵, James Kennedy², Thomas D Als⁶, Alfonso B Demur⁷, Thomas M Werge⁷, Merete Nordentoft⁸, Anders Børghlum⁶, Preben Bo Mortensen⁹, David M Hougaard¹ & Michael Christiansen^{1,3#}

- 1) Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
- 2) Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada
- 3) Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark
- 4) Aalborg Psychiatric Hospital. Aalborg University Hospital, Aalborg, Denmark
- 5) Department of Clinical Medicine, Aarhus University, Århus, Denmark
- 6) Institute of Medical Genetics, Aarhus University, Århus, Denmark
- 7) Mental Health Centre, Sct Hans, Capital Region of Denmark, Denmark
- 8) Mental Health Centre, Capital Region of Denmark, Denmark
- 9) Center for Register Research, Institute of Economics, Aarhus University, Århus, Denmark

The study was conducted under the auspices of the iPSYCH study (www.iPSYCH.au.dk)

*JG and CMH contributed equally to the study.

Key words: mitochondrial DNA, haplogroup, population genetics, energy metabolism

Running title: *mtDNA haplogroups in 24,216 Danes*

#Correspondence:

Professor, chief physician,
Michael Christiansen, FRCPath, MD
Department for Congenital Disorders, Statens Serum Institut
And
Department of Biomedical Sciences, University of Copenhagen.
E-mail: mic@ssi.dk; Phone: +4520720463.

Abstract

Mitochondrial DNA (mtDNA) haplogroups (hgs) are evolutionarily conserved sets of mtDNA SNP-haplotypes with characteristic geographical distribution. Associations of hgs with disease and physiological characteristics have been reported, but have frequently not been reproducible. Using 418 mtDNA SNPs on the PsychChip (Illumina), we assessed the spatio-temporal distribution of mtDNA hgs in Denmark in DNA isolated from 24,642 geographically un-biased dried blood spots (DBS), collected from 1981 to 2005 through the Danish National Neonatal Screening program. Geno-geographic affinity (ancestry background) was established with ADMIXTURE using a reference of 100K+ autosomal SNPs in 2,248 individuals from nine populations. The hg distribution was typically Northern European, and hgs were highly variable based on median-joining analysis, suggesting multiple founder events. Considerable heterogeneity and variation in autosomal geo-geographic affinity was observed. Thus, individuals with hg H exhibited 95 %, and U hgs 38.2 % - 92.5 %, Danish ancestry. Significant clines between geographical regions and rural and metropolitan populations were found. Over 25 years, macro-hg L increased from 0.2 % to 1.2 % ($p = 1.1 \times 10^{-10}$), and M from 1 % to 2.4 % ($p = 3.7 \times 10^{-8}$). Hg U increased among the R macro-hg from 14.1 % to 16.5 % ($p = 1.9 \times 10^{-3}$). Geno-geographic affinity, geographical skewedness, and sub-hg distribution suggested that the L, M and U increases are due to immigration. The complex spatio-temporal dynamics and geno-geographic heterogeneity of mtDNA in the Danish population reflect repeated migratory events and, in later years, net immigration. Such complexity may explain the often contradictory and population-specific reports of mito-genomic association with disease.

Introduction

Mitochondria are subcellular organelles responsible for oxidative phosphorylation (OXPHOS), producing ~ 80% of the ATP in eukaryotic cells¹, apoptosis and cell-cycle regulation², redox- and calcium homeostasis³ as well as intracellular signaling⁴. Each mitochondrion contains 2 - 10 copies of a 16.6 kb double-stranded mtDNA containing 37 genes⁵. Thirteen genes code for proteins in the five enzyme complexes conducting OXPHOS, whereas twenty-two genes code for tRNAs and two for rRNAs, all involved in intra-mitochondrial translation⁶. The mitochondrial proteome comprises approximately 1200 proteins^{7; 8}, of which mtDNA genes encode ~ 1%. The mtDNA is maternally inherited⁹, exhibits a high mutation rate¹⁰, and does not undergo recombination⁵. Genetic variants in mtDNA – as well as variants in the nuclear genes encoding the mitochondrial proteome - have been associated with disease^{11; 12 13}. More than 150 mitochondrial syndromes¹⁴ have been associated with more than 300 variants^{7; 15}.

Geographically and population specific lineages of mtDNA, haplogroups (hgs), have become fixed¹⁷, through the processes of random genetic drift and selection as the human populations dispersed throughout the world¹⁶. The advent of high throughput DNA sequencing technology, as well as implementation of biobanking technologies¹⁸, has enabled the construction of a high-resolution phylogenetic matrilineal mtDNA tree, Figure 1¹⁹.

Mitochondrial hgs have been assigned a role as disease modifiers⁵. Particularly in neurological degenerative diseases such as Alzheimer's disease²⁰⁻²³ and Parkinson's disease²³⁻²⁵, but also in psychiatric disease²⁶ and cardiac diseases such as hypertrophic^{27; 28} and ischemic cardiomyopathy²⁹.

Supporting a role as disease modifiers, some mtDNA hgs have specific physiological characteristics, e.g. reduced or increased ATP synthesis rates^{30; 31}, and variation in methylation status of genes involved in inflammation and signaling^{32; 33}. The association of mtDNA SNPs and hgs with both diseases and functional characteristics of mitochondria, has led to a pathogenic paradigm³⁴ where variation in mitochondrial function is considered to be of paramount importance for development of disease. Specific hgs have also been associated with longevity³⁵ and likelihood of being engaged in endurance athletic activities³⁶. The clinical presentation of diseases caused by specific mtDNA variants depends, in some cases, on the hg background³⁷. However, some of these studies are contradictory, either because they have been too poorly powered³⁸, have not been carefully stratified with respect to sex, age, geographical background³⁹ or population admixture⁴⁰, or have used small areas of recruitment risking “occult” founder effects⁴¹. To circumvent some of these problems a recent large study on mtDNA SNPs identified a number of SNPs that were associated with several degenerative diseases⁴², however, the study pooled sequence information from a large geographical area, without correcting for potential population sub-structure.

Most countries have a complex history with repeated migrations⁴³ and several bottle-necks caused by disease, war and emigration^{44; 45}. These demographic events are reflected in the fine scale genetic structure within sub-populations⁴⁶. However, the significance of such events to countrywide mtDNA hg distribution has not yet been assessed. As mtDNA interact functionally with the nuclear genome, it is paramount to ensure that specific mtDNA hgs – which are a marker of matrilineal genetic origin – do not represent population sub-structure at the genomic level. In theory, mtDNA is inherited independently of the nuclear genome, but population admixture and

geographic isolation may result in linkage disequilibrium between mtDNA and the nuclear genome. Such a linkage disequilibrium might interfere with genetic association analysis.

Here we demonstrate the complexity of the spatio-temporal dynamics and genogeographic affinity of mtDNA hgs in 24,216 Danes, which were sampled at birth during a 25-year period. This number represents 1.6 % of the population. The sampling material was dried blood spots (DBSs) obtained as part of the Danish Neonatal Screening Program⁴⁷, the very nature of which makes sampling geographically un-biased. Array analysis was performed using the PsychChip (Illumina, CA, USA) typing 588,454 variants.

Materials and Methods

Ethics statement

This is a register-based cohort study solely using data from national health registries. The study was approved by the Scientific Ethics Committees of the Central Denmark Region (www.komite.rm.dk) (J.nr.: 1-10-72-287-12) and executed according to guidelines from the Danish Data Protection Agency (www.datatilsynet.dk) (J.nr.: 2012-41-0110). Passive consent was obtained, in accordance with Danish Law nr. 593 of June 14, 2011, para 10, on the scientific ethics administration of projects within health research. Permission to use the DBS samples stored in the Danish Neonatal Screening Biobank (DNSB) was granted by the steering committee of DNSB (SEP 2012/BNP).

Persons

As part of the iPSYCH (www.iPSYCH.au.dk) recruitment protocol, 24,651 singletons (47.1 % female), born between May 1 1981 and Dec 31 2005 were selected at random from the Danish Central Person Registry. The singletons had to have been alive one year after birth, and to have a mother registered in the Danish Central Person Registry. Furthermore, it should be possible to extract DNA from the DBS. DBS cards were obtained from the Danish Neonatal Screening Biobank at Statens Serum Institut ⁴⁸ and DNA was extracted and analysed as described below. At the time of analysis (2012), the mean age of females was 18.2 years (SD: 6.6 years) and for males 18.8 years (SD: 6.7 years). There was no bias in the geographical distribution of the birthplace of samples, Figure 2.

Genetic analysis

From each DBS card two 3.2-mm disks were excised from which DNA extracted using Extract-N-Amp Blood PCR Kit (Sigma-Aldrich, St Louis, MO, USA)(extraction volume: 200 μ L). The extracted DNA samples were whole genome amplified (WGA) in triplicate using the REPLig kit (Qiagen, Hilden, Germany), then pooled into a single aliquot. Finally, WGA DNA concentrations were estimated using the Quant-IT Picogreen dsDNA kit (Invitrogen, Carlsbad, CA, USA). The amplified samples were genotyped at the Broad Institute (MA, USA) using the Psychiatric Genetic Consortia developed PsychChip (Illumina, CA, USA) typing 588,454 variants. Following genotyping, samples with less than 97% call rate, as well as those where the estimated gender differed from the expected gender were removed from further analysis; altogether 435 (1.8 %) samples were removed due to problems with calling mtDNA variants. We then isolated the 418 mitochondrial loci and reviewed the genotype calls, before exporting into the PED/MAP format using GenomeStudio (Illumina, CA, USA). Samples were loaded into GenomeStudio (version 2011.a), a custom cluster was created using Gentrain (version 2), following automatic clustering all positions with heterozygotes were manually curated. The data was exported relative to the forward strand using PLINK Input Report Plug-in (version 2.1.3). Eigenvectors were calculated using PLINK (v1.90b3.31). PCA plots were created using the package ggplot2 (version 1.0.1) in R (version 3.1.3).

mtDNA SNPing

Haplotyping of mtDNA was performed manually using the defining SNPs reported in www.phylotree.org¹⁹. Hierarchical affiliation to macro-hg i.e. L0 – L6, M, N, R, and subsequent to hgs – units more distal in the cladogram, Figure 1, was performed. In some cases it was possible to establish affiliation to even sub-hgs. The call efficiencies of SNPs used in defining haplo- and sub-haplogroup affiliation are summarized in Suppl. Table 1.

Phylogenetic analyses

Phylogenetic analyses was performed by constructing median-joining networks with Network 4.6.1.3 (<http://www.fluxus-engineering.com>). Fasta converted sequences of mtDNA SNPs from each person were aligned, sequences were pre-processed with Star Contraction (Maximum star radius 5 for R, N, M and 1 for L), then Median Joining networks were constructed (using the network parameters: Epsilon 10, Frequency >1, active) followed by post-processing with a maximum parsimony algorithm (MP)^{49; 50}. Network Publisher were used to post-process the networks⁵¹.

Genogeographic affinity (nuclear genomic ancestry) analysis

Ancestry estimation was done using ADMIXTURE 1.3.0⁵². Briefly, a reference population consisting of Human Genome Diversity Project (HGDP) (<http://www.hagsc.org/hgdp/>) genotyping SNP data set, supplemented with representative samples of danes (716 individuals) and greenlanders (592 individuals) available at SSI from unrelated projects, was used. The final reference data set consisted of 103,268 SNPs and 2,248 individuals assigned to one of nine population groups: Africa, America, Central South Asia, Denmark, East Asia, non-Danish Europe, Greenland, Middle East and Oceania. K – number of clusters defined - was set to eight, based on principal component analysis clustering (data not shown).

Individuals belonging to individual mtDNA hgs or sub-hgs were merged with the reference population data set and analysed using ADMIXTURE. For prediction of the ancestry of individuals within the mtDNA hgs we created a random forest model⁵³ based on the reference data set, with the clusters Q1-8 as predictors and population groups as outcome. The prediction was thus supervised. Prediction was done in R version 3.2.2, using the Caret package. The distribution of the eight basic clusters in samples of different geographical origin is shown in Suppl Figure 1. As

expected, the African-characteristic cluster distribution plays a decreasing role when going from Africa over the Middle East to Central South Asia. Likewise, the Danish cluster distribution is very similar to that of Europe.

Statistics

The statistical significance of differences in mtDNA proportions was assessed using a permutation version of Fisher's exact test⁵⁴. Calculations were performed using R⁵⁵. To assess population stratification⁵⁶, principal component analysis (PCA) was performed.

Results

Distribution of macro-haplogroups

The mtDNA macro-hg distribution pattern is typically northern European⁵⁷ with > 90% belonging to the R macro-hg, 7.0 % belonging to N and 1.6 % to M of likely Near Eastern or Asian origin⁵⁷ and 0.7 % belonging to the combined L macro-hgs (L0-L6) of a likely African origin⁵⁸, Table 1. A PCA analysis (PC 1 versus PC 2) based on all the SNPs showed a clear separation between R and L, with N and M located intermediately, figure 3A. The PC1 seems to reflect time since branching, whereas PC2 reflects geographical distance.

Distribution of haplogroups

The R macro-hg was dissolved into hgs as shown in Table 2. However, it was not possible, with the available SNPs, to differentiate the HV and P hgs from the R macro-hg (Suppl. Table 1, for details). Principle Component 1 versus 2, Figure 3B, demonstrated a clear clustering of mtDNA SNPs from persons belonging to each hg. The proximity of the clusters comprising U and K and H and V, respectively, is in accordance with the current phylogenetic mtDNA tree, Figure 1. Likewise, the median-joining graph of the R macro-hg, Figure 4A, based on all the called SNPs, disclosed a phylogenetic relationship compatible with that of Figure 1. In addition, all the hgs exhibit a considerable complexity, suggesting that each hg is the result of multiple migratory events and thus, multiple founder events. The sub-hg distribution of each of the hgs of the R macro-hg is shown in Table 3.

The L, N and M macro-hgs were infrequent, Table 1. M and N could be broken down into hgs as shown in Table 4. The complexity of these macro-hgs, and macro-hg L, as demonstrated by median-

joining phylogenetic analysis, Figure 4 B, C and D, suggest that these macro-hgs are composed of specific haplotypes from multiple immigrations over time. The number of DNA variations between individual haplotypes is so large that they can not represent development from a single founder within the time frame defined by the length of time in which Denmark has been populated.

Geno-geographical affinity of mtDNA haplogroups

An admixture analysis of the persons of different hgs was performed with results as shown in Table 5. The major hgs, H and its sub-hgs, have a 90-95% Danish ancestry and ~ 5% non-Danish European structure. However, there is a great variation between mtDNA hgs – most pronounced in hg U -, and most of the hgs have a notable – but varying – proportion of admixture from Europe, Middle East and Central South Asia, Table 5. This finding is compatible with the very complex M-J diagram of the Danish hgs, Figure 4A. The ancestry or geno-geographic affinity of the macro-hgs M and N differ even more, Table 5, with the M being of South East and East Asian affinity, and N of Danish and European affinity. Macro-hg L exhibited – surprisingly – a predominance of Middle Eastern and European genomic affinity, Table 5, where an African ancestry should be expected^{57; 59}.

Spatial distribution of mtDNA haplogroups

Denmark is divided into five geographical and administrative regions as shown in Figure 2. The samples for this study were obtained from all regions in Denmark and linked to the postal code of the birthplace. The frequency of the macro-hgs and the most frequent hgs is shown for each administrative region in Table 6. The most marked differences were the relatively low frequency of hg H, 43.4 %, in the Capital Region, compared to the North Denmark Region with 48.8 % hg H ($p < 0.0001$), and the higher frequencies of hgs L, M and R (in total 7.5 %) in the Capital Region compared to the North Denmark Region (in total 3.9 %) ($p < 0.0001$). Thus, in the Capital Region,

the L hgs have a frequency of 1.3 %, compared with 0.4 – 0.6 % in the other regions and the M hgs have a frequency of 2.6 % in the Capital region compared to 1.0 – 1.6 % in the other regions. As the L and M hgs are rare in the European population and very frequent in African and Asian populations, the noted difference probably reflects a higher proportion and preferential localization of non-ethnic Danes in the Capital region. A similar spatial difference was apparent when the hg distributions of persons from the Danish metropolitan areas, comprising the major cities, Copenhagen, Aarhus, Odense and Aalborg, were compared with the distributions in the remaining rural areas, Suppl Table 2. In the metropolitan areas, the combined frequency of L and M hgs was 4.4 % as compared to 1.9 % in the rural areas.

Temporal distribution of mtDNA haplogroups

The frequency of the major hgs H, J, T, K, V did not change significantly from year to year in the period from 1981 to 2005 (Results not shown).

The frequencies of M and L hgs, Figure 5, increased significantly over the period. The L hgs increased from a constant level of ~ 0.4 % from 1981 – 1995 to ~ 1.5 % in 2005. The M hgs rose from a constant level of ~ 1% from 1981 – 1991 to ~ 3 % in 2005. The proportion of L and M hgs increased significantly from the period 1981-1986 to the period 2000-2005, Table 7. The considerable diversity of the M macro-hg, Figure 4C and the extreme diversity of the L macro-hg, Figure 4D, as disclosed by the M-J-network, suggests that the hgs are the result of immigration from different source populations. The occurrence of new clusters on the PCA when the macro-hgs M and L from the period 1981-1986 were compared with the same macro-hgs in 2000-2005, Suppl. Figs. 2A & B, makes it likely that the increase in the proportion of both macro-hgs represent immigration.

The hg U also increased in proportion each year from 1981 to 2005, Figure 5, albeit not significantly. However, when analysing only the R-macro-hg, the proportion of hg-U increased significantly (data not shown). In the case of the U-hg, the PCA did not reveal the appearance of novel clusters when comparing 1981-1986 with 2000-2005, Suppl Fig 2C. Thus, there was no evidence for the introduction of novel U-hgs over the period. However, a more detailed analysis revealed that the increase in U-hgs was largely due to an increase in the infrequent sub-hgs U1, U6, U7, and U8, Suppl Table 3, as all these increased by more than 40%. An admixture analysis showed that whereas the Danish autosomal genomic affinity of the hgs U*, U2-U3, U4-9, U5a&b is in the range of 80.6% - 92.5 %, Table 5, comparable to that of the other major European hgs, the Danish affinity drops to 36.4% - 42.3 % for U1, U6 and U7. According to Table 5, the U1 and U7 have strong Middle Eastern (29.4% and 13.6%) and Central South Asian (14.7% and 39.8%) autosomal geno-geographic affinities, and the U6 exhibits a very strong Middle Eastern autosomal affinity (53.8%). This suggests that the increase in frequency of hg U is largely due to expansion of hgs brought into Denmark as a result of recent immigration.

Discussion

This study shows that the distribution of mtDNA hgs in Denmark is highly dynamic and complex. It comprises 1.6 % of the Danish population over a 25-year period, and is by far the largest performed of the distribution of mtDNA hgs in any country. The method of collecting stored DBS from the PKU biobank, where the coverage is $\sim 99\%$ ⁶⁰, enabled us to survey a true population based sample of mtDNA from persons, where the time and place of birth was known from national electronic registries. This is in contrast to most other population genetic studies of adults sampled in a specific bias-prone context, e.g. hospitalized patients or geographically biased samplings.

The distribution of macro-hgs, Table 1, is typical of Northern Europe^{61; 62}. The large majority of persons had hgs belonging to the R macro-hg, Table 2. The distribution of hgs within the R macro-hg, Table 2, is very similar to that previously described among 9000+ persons from the greater Copenhagen area⁶³ and a much smaller (~ 200 cases) Danish forensic control sample⁶⁴, as well as that found in 2000 Danish exomes from a mixed control and diabetes population⁶⁵. However, the latter study is seriously biased as it contains $\sim 50\%$ patients with type-2 diabetes, which have shown associations with mtDNA haplotypes⁶⁵. Despite this, the sub-hg distribution in macro-hg R is similar to that seen in the present study, Table 3, e.g. H1 and H3 constitute 37.6 % and 6.5 % of the H hgs in our study and 39.6 % and 5.2 % in the Li et al. study⁶⁵.

The mtDNA haplotyping was based on array data with only 418 mtDNA SNPs and as a consequence not all sub-hgs could be called. Ideally sub-haplotyping could be refined – without including sub-hg defining SNPs from Phylotree – by the application of clustering approaches. However, this was not attempted, as it would lead to results that could not be compared with other studies. The stringent

adherence to specific SNPs meant that a number of persons, normally in the order of 1-2% (see tables 2-4), were not assigned to a specific haplotype. An advantage of using a limited set of markers is that confounding due to private variants is avoided.

The complexity of each of the major hgs, as seen from the M-J-networks, Fig 4B-D, with multiple nodes and a span of many mutations between different leaves suggests that the hgs are not representative of single early founder event, as one mtDNA mutation is expected to occur per 3,000 yrs⁶⁶. A notable exception is hg A, with a total prevalence of ~ 0.5 %, where the largest node, figure 4B, is associated with “daughter” nodes only 1-2 mtDNA mutations from the major node. It was only possible to identify the sub-hgs A1 and A5, and they only constituted a small fraction of the total A hg. A likely source of the A hgs is the Inuit population from Greenland, that is now a self-ruling part of the Kingdom of Denmark. Several studies have established that the major hgs in Greenlandic Inuits is A2 followed by hg D3⁶⁷⁻⁷⁰, whereas other Inuit populations have other characteristic hgs^{70; 71}.

The lack of recombination, high mutation frequency, and fixation of mtDNA hgs has enabled the use of mtDNA in population genetics to study population ancestry, migrations, gene flow and genetic structure^{72; 73}. Thus, populations on different continents, e.g. Native Americans⁷⁴, Africans⁷⁵ and Europeans⁷⁶ were ascribed specific matrilineal mtDNA hg distributions^{62; 77}. However, recent studies using autosomal SNP markers have disclosed a considerable ancestral complexity underlying an mtDNA classification in specific admixed populations, and the prediction of a specific mtDNA hgs is not possible from a specific continental ancestry based on nuclear genetic markers⁷⁸. A problem with genetic analysis of admixed populations is the lack of temporal resolution. It is

often not possible to date a specific population split because differentiation between new and old migrations is impossible. Recent advances in sequencing ancient genomes⁷⁹ have made it possible to combine genetic information from ancient humans with archaeological information on the age of skeletal remains⁸⁰, thus constructing hg distribution maps with a temporal dimension for e.g. Ice age and Bronze age^{81; 82} Europe. Maps, that may help explain demic and cultural exchange^{70; 83-85}.

There is no solid evidence of the presence of humans in Denmark ^{86 87} prior to the Last Glacial Maximum (LGM)⁸⁸ (26.5 – 19 kYBP). At that time Denmark was covered in ice, except for the south-western part of Jutland⁸⁹, and following the retraction of the ice sheath, peopling became possible from the south⁸⁹⁻⁹². The first inhabitants documented were late Paleolithic hunters entering from southern Europe⁹³. These hunters are discernible from Bølling time⁸⁹ around 12,800 BC. The archaeological remains suggest their transient presence in seasonal hunting periods⁹⁴ until the Mesolithic around 9,700 BC. The Hamburgian, Federmesser, Bromme and Ahrensburgian material cultures^{89; 95}, well known from findings in Germany⁸⁹, are represented. The earliest anatomically normal humans (ANM), present from around 45 – 41.5 kYBP⁹⁶⁻⁹⁸, where ancient DNA studies have revealed the presence of mtDNA hgs M and pre-U2⁹⁹, have little similarity to present-day Europeans. However the Europeans from around 37 kYBP to 14 kYBP have left their mark on present-day Europeans⁸². From 14 kYBP the European population has a strong near-eastern component⁸². Around 7 kYBP the Neolithic transformation gradually started as the result of a demic dissemination of Neolithic Aegeans¹⁰⁰. Thus, a minimum of three ancestral populations, i.e. a western European hunter-gatherer, an ancient north Eurasian, and an early European farmer population are needed to explain present-day European autosomal genome compositions ¹⁰¹. In Denmark, where Paleolithic ancient genetic data have not been published, ancient mtDNA

haplotyping of three Neolithic corpses from 4.2 – 4 kYBP revealed two U4 and one U5a mtDNA hgs¹⁰² and later samples showed a mixture of hgs also found in northern Germany^{99; 102}. The presence of mtDNA clades deriving from Paleolithic and Neolithic Europeans in the extant Danish population is thus explained.

In the Bronze age, an influx of people from the Russian steppe and North Caucasus, bringing the Indo-European language and culture¹⁰³, resulted in the last major prehistoric demic change⁸¹ in Europe. In historic times the migrations have been many, particularly in the half-millennium following the fall of the Western Roman Empire¹⁰⁴. In Denmark, apart from continuous demic exchange with Southern Scandinavia and present-day Germany^{89; 92}, early historic time was mostly characterized by emigration, i.e. the Heruli, Cimbriae and Teutones, Burgundians and Vikings^{91; 92}. The first census held in 1769 AD reported the population of what is present day Denmark to be ca 800,000 persons¹⁰⁵. This number has increased to 5.6 mio in 2014, despite emigration of 287,000 persons between 1867 and 1914 (~10% of the population) largely to Northern America, and the same number from 1914 to 1968¹⁰⁶. After the Second World War, Denmark has seen a considerable immigration from other European countries, but also from Asian, Middle Eastern and African countries, Suppl Figure 3. This part of Danish history can explain the occurrence of a diversity of mtDNA clades deriving from a large number of geographic regions. The extant distribution of mtDNA hgs reported here, Tables 1-4, reflect the complicated and heterogeneous origins of the people currently inhabiting Denmark.

The distribution of H sub-hgs, Table 3, and the M-J graph of the H-hg, Figure 4A, with multiple major nodes and 5-10 variants between leaves, suggest that the H-hg lineages are the result of

repeated immigrations. Most likely from northern Europe, where a study of 39 prehistoric hg H samples has shown that the distribution of hg H differed between early and middle-to-late Neolithic groupings. Prior to Neolithicum H hgs have not been found in skeletal remains; an ensemble of Swedish Mesolithic hunter-gatherers all had U-hgs⁸³. Whereas H, H5 and H1 were found throughout Neolithicum, H5b, H10, H16, H23, H26, H46, H88, and H89 were seen in early Neolithic samples, and H2, H3, H4, H5a, H6, H7, H11, H13, H82, and H90 in middle-to-late Neolithic samples¹⁰⁷. The extant Danish H-hg distribution is compatible with contributions from throughout Neolithicum, but estimating when the hgs appeared in Denmark remains unfeasible, as the distribution is similar to that of northern Germany. However, not all carriers of an H-hg have a Danish autosomal geno-geographic affinity, table 5, compatible with a considerable recent immigration from countries with a European mtDNA distribution, Suppl Figure 3.

The most frequent Danish U-sub-hg is U5, Table 3, which is an old European mtDNA hg, with two major subclades, U5a and U5b, with coalescence time estimates of 16 – 20 kYBP and 20 – 24 kYBP, respectively¹⁰⁸. The U5-hg is the most frequent U-sub-hg after LGM⁹⁹ and the carriers have a Danish autosomal genomic ancestry around 90 %, and a European ancestry of 6.1 – 8.8%, Table 5. However, several of the less frequent U-sub-hgs have a considerable, ~ 40 – 60 % non-Danish and non-European geno-geographic affinity based on autosomal markers, with U7 related to South East Asia and U6 to the Middle East, respectively. In addition, J & T hgs, Figure 5, exhibit a relatively strong non-Danish autosomal genomic ancestry. The M-J graphs, Figure 4A, also disclose a considerable variation, much larger than could be attained merely within the timeframe where Denmark has been populated. This may be explained by an extensive immigration, and for U6 and

U7 in very recent time. This is also compatible with the rising frequency of these hgs during the 25-year study period, Figure 6 and Suppl Table 3.

The N-macro-hg exhibits a high (>90%) combined Danish and European genomic affinity, Table 5, suggesting that the major part has been in Denmark for long. The major N-hgs are the I-hg, which has been found in meso- and neolithic Scandinavians¹⁰² and hg-X¹⁰⁹ and hg-W, Table 4, and they exhibit, figure 4B, a very extensive heterogeneity. All three hgs are old and have a broad, low frequency, distribution in western Eurasia⁶¹, resulting from migratory events from the Near East and Central Asia. These events, viz. the significance for the presence of the hgs in Denmark, can not be temporally resolved.

Macro-hg-M, albeit infrequent (1.6 %), Table 1, has also increased in frequency recently, Figure 6, and exhibits extensive heterogeneity, Figure 4C, in the M-J graph, suggesting many migratory events of diverse origins. The largest group is M, that could not be decomposed further, but the second most frequent, D4 (14.2% of M-hgs), is most likely of central Asian origin¹¹⁰, and the M2, M3, M4 and M6 hgs, constituting 12.5% of M-hgs, are of Indian or Pakistani origin¹¹¹. These findings are compatible with the high genomic affinity (42.2%, Table 5) towards Central South Asia, and a recent entry to Denmark. There is a propensity for location in Metropolitan areas, Suppl Table 2, and the Capital region, Table 6, also compatible with recent immigration.

The low, but increasing, frequency of L-hgs (Table 1, Table 7 and Figure 6) predominantly L2 and L3 (Table 1), is also the result of multiple immigration events, as the M-J plot reveals an extensive

heterogeneity (Figure 4D). L2 and L3 hgs are frequent in south, west or east Africa, whereas the contribution from central Africa is small^{112; 113}. However, the admixture analysis, Table 5, suggests a much stronger autosomal genomic affinity to the Middle East (70.8 %) than to Africa (8.4%). However, this finding could be an artefact of the ADMIXTURE analysis, as the highly variable African genomic ancestry^{59 114} is represented by only 127 genomes (~ 6 % of the total reference set – see Materials and Methods), as compared to 178 genomes from the Middle East, making the definition of African ancestry imprecise. Alternatively, it could be caused by the extensive demic exchange that have occurred through time between the Near East and Northern Africa^{115; 116}. The number of children born in the period 1980 to 2005 with one or two African parents, See Suppl Figure 3, is compatible with the origin of the L hg persons being African. The L hgs are predominantly located in the Capital Region, Table 6, and in Metropolitan areas, Suppl Table 4, which is also compatible with recent immigration.

Immigration to Denmark increased from 1980, where 135,000 immigrants were registered, to 2005, where this number had risen to 345,000¹⁰⁵. In the same period the number of descendants of immigrants, i.e. persons that might turn up in this study, rose from 18,000 to 109,000. Roughly 50 % of the immigrants were from western countries¹⁰⁵, See Suppl Figure 3. This should give approximately 45,000 descendants of non-western immigrants over the time of study. This is within the order of magnitude to be expected from the frequency distribution of mtDNA hgs in Denmark. There is thus a reasonable concordance between the suspected number of immigrants, from the temporal and structural study of mtDNA hgs, and the registered births of persons with non-Danish parents.

Whereas the temporal change in mtDNA distribution can be explained by immigration, it is more difficult to explain the spatial clines, Table 6 and Suppl Table 2. The frequency of hg-H is higher in Northern Jutland than in other regions, particularly the Capital Region, and it is also higher in rural areas ($p = 0.0001$), Suppl Table 2. The difference cannot be explained by the comparatively much smaller differences in frequencies of hgs L and M. As the mobility of Danes was fairly restricted until around 1900⁹¹, it may represent differences caused by centuries of relative isolation of a population north of Limfjorden. In Slovenia historically confirmed geographically-based sub-stratification of the population¹¹⁷ has led to extreme differences in mtDNA distributions.

A recent fine-scale gDNA population structure study from the UK⁴⁶ revealed considerable geographical heterogeneity and enabled the identification of specific sources of admixture from continental Europe. A gDNA study of admixture in the Danish population¹¹⁸ showed considerably more homogeneity, but medieval admixture from Slavic tribes in North Germany, as well as a North-South gradient were discernible. Both of these studies limited the participants to persons with local grand-parents, whereas our present study was not directed towards previous generations, but rather present-time and prospective. In addition to the gDNA/mtDNA interaction, the demonstrated spatio-temporal dynamics of the mtDNA hg distribution should be taken into account when designing studies of mtDNA associations with disease, physiological characteristics and pharmacological effects. A prerequisite for genetic association studies is that the individuals are sampled from a homogenous population, or that cryptic population structure is corrected for, if not, false positive associations are likely^{119 120; 121}.

Thus, when studying bi-genomic, i.e. both nucleo- and mito-genomic, disease associations, our results indicate that it is necessary to compensate for gDNA and mtDNA genetic stratification, as well as the interaction between the two sources of variation and spatio-temporal clines. To our knowledge, this has never been done, suggesting that previous reports on disease associations and mtDNA haplogroups should be considered preliminary.

Acknowledgements:

The iPSYCH study was funded by The Lundbeck Foundation Initiative for Integrative Psychiatric Research. We further gratefully acknowledge the financial support of The Jascha Foundation, The Strategic Research Council (“Heart Safe”), The Augustinus Foundation, The Lundbeck Foundation (Grant no. R67-A6552), and Familien Hede Nielsens Fond. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation.

Legends to Tables and Figures

Table 1. Distribution of mtDNA macro hgs L0-L6, M, N and R. Incomplete is the number of individuals that could not be haplotyped using the haplotyping algorithm (see Materials and Methods).

Table 2. Distribution of mtDNA haplogroups constituting the macro-hg R. Incomplete is the number of individuals that could not be haplotyped using the haplotyping algorithm (see Materials and Methods).

Table 3. Distribution of mtDNA sub-haplogroups of H, V, J, T, K, and U, the most frequent European haplogroups. HV could not be defined and is included in the H hg. Not sub-haplotyped is the number of individuals that could not be subhaplotyped using the algorithm defined by the SNPs in Suppl Table 1.

Table 4. Distribution of mtDNA haplogroups contained within the N and M macro hgs. Not haplotyped is the number of individuals that could not be haplotyped using the algorithm defined by the SNPs in Suppl Table 1.

Table 5. Geno-geographic affinity (ancestry) of the autosomal genome of persons in the study broken down with respect to mtDNA macro-hg or hg. The number of persons in each geographical group in the reference population is given in the header.

Table 6. Distribution of the most frequent haplogroups in the five administrative regions of Denmark. Only persons born at locations with ≥ 20 births are included.

Table 7. Distributions of the most frequent haplogroups in the period 1981-1986 and 2000-2005.

Figure 1. Phylogenetic tree of mtDNA sequences, modified from (www.phylotree.org). MRCA: Most recent common ancestor.

Figure 2. Map of Denmark with major administrative regions and metropolitan areas. The population of each region or area, as well as the number of samples, is given on the map. Based on birth-locations with ≥ 20 included individuals

Figure 3. A. PCA of the macro-haplogroups L, M, N and R. **B.** PCA of the macro-haplogroup R, i.e. the major European haplogroups. **C.** PCA of the haplogroups belonging to macro-haplogroup N, and **D.** macro-haplogroup M, and **E.** PCA of the haplogroups belonging to macro-haplogroup L, i.e. the major African haplogroups. PC1: First principal component. PC2: Second principal component.

Figure 4. A. Median-joining (M-J) network of the macro-haplogroup R mtDNA sequences (2000 chosen at random). **B.** M-J network of the N and **C.** M haplogroups. **D.** M-J network of the L haplogroup sequences.

Figure 5. The proportion of haplogroup L, M and U as a function of birth year.

Tables

Table 1.

| Macro-hg | n | % |
|-------------------------|----------|----------|
| L0 | 26 | 0.1 |
| L1 | 13 | 0.1 |
| L2 | 54 | 0.2 |
| L3 | 71 | 0.3 |
| L4 | 16 | 0.1 |
| L5 | 2 | 0.0 |
| L6 | 0 | 0.0 |
| M | 394 | 1.6 |
| N | 1,700 | 7.0 |
| R | 21,940 | 90.6 |
| Total haplotyped | 24,216 | 100 |
| Incomplete | 426 | 1.7 |

Table 2.

| Hg | n | % |
|-------------------------|----------|----------|
| H | 10,912 | 50.0 |
| U | 3,237 | 14.8 |
| T | 2,323 | 10.6 |
| J | 2,195 | 10.1 |
| K | 1,754 | 8.0 |
| V | 785 | 3.6 |
| R | 570 | 2.6 |
| F | 50 | 0.2 |
| B | 10 | 0.0 |
| Total haplotyped | 21,836 | 100 |
| Incomplete | 104 | 0.5 |

Table 3

| Hg | Sub-hgs | n | % |
|-----------------------|-----------------------|----------|----------|
| H | H | 3,924 | 36.4 |
| | H1-H30b-H79a | 4,054 | 37.6 |
| | H2 | 1,058 | 9.8 |
| | H3 | 699 | 6.5 |
| | H4 | 220 | 2.0 |
| | H5-36 | 828 | 7.7 |
| | Number sub-haplotyped | 10,783 | 100 |
| | Not sub-haplotyped | 129 | 1.2 |
| U | U | 120 | 3.8 |
| | U1 | 66 | 2.1 |
| | U2-U3 | 596 | 18.8 |
| | U4-9 | 603 | 19.0 |
| | U5a | 1,113 | 35.1 |
| | U5b | 471 | 14.8 |
| | U6 | 26 | 0.8 |
| | U7 | 87 | 2.7 |
| | U8 | 90 | 2.8 |
| | Number sub-haplotyped | 3,172 | 100 |
| | Not sub-haplotyped | 65 | 2.0 |
| | K | K | 1 |
| K1 | | 1,385 | 81.8 |
| K2 | | 307 | 18.1 |
| K3 | | 1 | 0.1 |
| Number sub-haplotyped | | 1,694 | 100 |
| Not sub-haplotyped | | 60 | 3.4 |
| J | J | 2 | 0.1 |
| | J1 | 1,686 | 80.5 |
| | J2 | 407 | 19.4 |
| | Number sub-haplotyped | 2,095 | 100 |
| | Not sub-haplotyped | 100 | 4.6 |
| T | T* | 9 | 0.4 |
| | T1 | 396 | 17.1 |
| | T2 | 1,908 | 82.5 |
| | Number sub-haplotyped | 2,313 | 100 |
| | Not sub-haplotyped | 10 | 0.4 |

Table 4.

| Macro-hg | (Sub) hgs | n | % |
|-----------------|-------------------|----------|----------|
| N | A | 119 | 7.1 |
| | I | 655 | 39.2 |
| | N | 55 | 3.3 |
| | N1-5 | 111 | 6.6 |
| | N2 | 11 | 0.7 |
| | W | 319 | 19.1 |
| | X | 394 | 23.6 |
| | Y | 6 | 0.4 |
| | Number haplotyped | 1,670 | 100 |
| | Not haplotyped | 30 | 1.8 |
| M | C | 40 | 10.7 |
| | D4 | 53 | 14.2 |
| | D5 | 4 | 1.1 |
| | D6 | 2 | 0.5 |
| | E | 8 | 2.1 |
| | G | 10 | 2.7 |
| | M | 81 | 21.7 |
| | M1 | 30 | 8.0 |
| | M2 | 11 | 2.9 |
| | M3 | 24 | 6.4 |
| | M30 | 14 | 3.7 |
| | M4 | 3 | 0.8 |
| | M5 | 26 | 7.0 |
| | M6 | 9 | 2.4 |
| | M7 | 37 | 9.9 |
| | M8a | 7 | 1.9 |
| | M9 | 4 | 1.1 |
| | Z | 11 | 2.9 |
| | Number haplotyped | 374 | 100 |
| | Not haplotyped | 20 | 5,1 |

Table 5.

| Haplogroup | Reference population (n) | | | | | | | | |
|------------|--------------------------|-----------------|--------------|-------------------|--------------|---------------|--------------------------|-----------------|--------------|
| | Denmark (716) | Greenland (592) | Europe (161) | Middle East (178) | Africa (127) | America (108) | Central South Asia (210) | East Asia (241) | Oceania (39) |
| H* | 90.8 | 0 | 6.9 | 1.4 | 0 | 0 | 0.8 | 0 | 0 |
| H1 | 94.1 | 0 | 5.0 | 0.7 | 0 | 0 | 0.1 | 0 | 0 |
| H2 | 94.5 | 0.1 | 4.8 | 0.3 | 0 | 0 | 0.3 | 0 | 0 |
| H3 | 94.7 | 0 | 4.3 | 1.0 | 0 | 0 | 0 | 0 | 0 |
| H4 | 90.0 | 0 | 7.7 | 2.3 | 0 | 0 | 0 | 0 | 0 |
| H5'36 | 93.7 | 0 | 4.7 | 1.1 | 0 | 0 | 0.5 | 0 | 0 |
| J* | 0.0 | 0 | 50.0 | 50.0 | 0 | 0 | 0 | 0 | 0 |
| J1 | 90.0 | 0.1 | 6.9 | 1.9 | 0 | 0 | 1.1 | 0.1 | 0 |
| J2 | 89.4 | 0.2 | 5.9 | 4.2 | 0 | 0 | 0.2 | 0 | 0 |
| T* | 33.3 | 0 | 44.4 | 11.1 | 0 | 0 | 11.1 | 0 | 0 |
| T1 | 84.6 | 0 | 7.8 | 4.5 | 0 | 0 | 3.0 | 0 | 0 |
| T2 | 89.4 | 0.1 | 7.5 | 1.8 | 0 | 0.1 | 1.3 | 0 | 0 |
| K | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K1 | 91.8 | 0 | 6.4 | 1.0 | 0 | 0 | 0.8 | 0 | 0 |
| K2 | 92.5 | 0 | 6.5 | 0 | 0 | 0 | 1.0 | 0 | 0 |
| K3 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| U* | 92.5 | 0 | 6.7 | 0 | 0 | 0 | 0.8 | 0 | 0 |
| U1 | 38.2 | 0 | 17.6 | 29.4 | 0 | 0 | 14.7 | 0 | 0 |
| U2-U3 | 80.6 | 0 | 7.9 | 5.4 | 0 | 0 | 6.2 | 0 | 0 |
| U4-9 | 87.2 | 0 | 9.7 | 1.6 | 0 | 0 | 1.5 | 0 | 0 |
| U5a | 91.9 | 0 | 6.1 | 1.1 | 0 | 0 | 0.9 | 0 | 0 |
| U5b | 89.2 | 0 | 8.8 | 1.2 | 0 | 0 | 0.6 | 0.2 | 0 |
| U6 | 42.3 | 0 | 3.8 | 53.8 | 0 | 0 | 0 | 0 | 0 |
| U7 | 36.4 | 0 | 10.2 | 13.6 | 0 | 0 | 39.8 | 0 | 0 |
| U8 | 86.0 | 0 | 8.6 | 2.2 | 0 | 0 | 3.2 | 0 | 0 |
| Macro-hg | | | | | | | | | |
| L | 4.5 | 0.0 | 14.6 | 70.8 | 8.4 | 0.0 | 1.7 | 0.0 | 0.0 |
| M | 0.6 | 9.9 | 29.9 | 9.0 | 0.0 | 0.3 | 42.2 | 8.1 | 0.0 |
| N | 80.4 | 1.1 | 11.7 | 3.8 | 0.0 | 0.1 | 2.3 | 0.5 | 0.0 |

Table 6.

| Haplogroup | Capital Region | Region Zealand | Central Denmark Region | Region of South Denmark | North Denmark Region |
|-------------------|-----------------------|-----------------------|-------------------------------|--------------------------------|-----------------------------|
| | % (n) | % (n) | % (n) | % (n) | %(n) |
| H | 43.4 (2,962) | 46.9 (1,664) | 45.7 (2,584) | 45.6 (2,259) | 48.8 (1,284) |
| I | 2.6 (175) | 2.5 (89) | 3.2 (180) | 3.0 (148) | 2.3 (60) |
| J | 8.6 (586) | 8.8 (313) | 9.4 (534) | 9.4 (464) | 8.2 (215) |
| K | 6.9 (468) | 6.3 (224) | 8.1 (456) | 7.3 (364) | 8.7 (229) |
| L | 1.3 (88) | 0.5 (17) | 0.5 (31) | 0.4 (20) | 0.6 (17) |
| M | 2.6 (176) | 1.3 (45) | 1.0 (56) | 1.6 (79) | 1.0 (27) |
| N* | 1.3 (92) | 1.2 (43) | 1.2 (68) | 1.2 (61) | 1.3 (33) |
| R | 3.6 (244) | 2.1 (74) | 2.5 (144) | 2.4 (119) | 2.3 (61) |
| T | 9.2 (627) | 10.0 (354) | 9.3 (525) | 8.8 (434) | 9.0 (238) |
| U | 13.8 (939) | 13.8 (491) | 13.1 (744) | 13.4 (666) | 13.1 (346) |
| V | 3.4 (229) | 3.3 (116) | 3.4 (195) | 3.5 (175) | 2.8 (73) |
| W | 1.4 (96) | 1.6 (57) | 1.3 (74) | 1.4 (67) | 0.9 (24) |
| X | 2.0 (134) | 1.7 (60) | 1.2 (67) | 2.0 (101) | 0.9 (24) |
| NA | 0 (2) | 0.1 (3) | 0 (2) | 0 (1) | 0 (1) |
| Total: | 100 (6,818) | 100 (3,550) | 100 (5,660) | 100 (4,957) | 100 (2,632) |

Table 7.

| Haplogroup | 1981-1986 | 2000-2005 | P-Value |
|-------------------|------------------|------------------|----------------|
| | % (n) | % (n) | |
| H | 47.3 (2,162) | 44.3 (2,454) | 2,98E-03 |
| I | 2.8 (130) | 2.7 (150) | 0,71 |
| J | 8.9 (405) | 9.2 (511) | 0,53 |
| K | 7.9 (359) | 7.0 (387) | 0,10 |
| L | 0.2 (8) | 1.2 (67) | 1,14E-10 |
| M | 1.0 (44) | 2.4 (131) | 3,73E-08 |
| N* | 1.1 (50) | 1.3 (73) | 0,32 |
| R | 2.2 (102) | 2.9 (163) | 0,03 |
| T | 9.8 (447) | 8.5 (468) | 0,02 |
| U | 12.7 (579) | 14.3 (790) | 0,02 |
| V | 3.6 (163) | 3.1 (169) | 0,16 |
| W | 1.2 (56) | 1.4 (80) | 0,39 |
| X | 1.4 (64) | 1.6 (89) | 0,41 |
| NA | 0.0 (1) | 0.1 (3) | 0,63 |
| Total | 100 (4,570) | 100 (5,535) | - |

Figure 1.

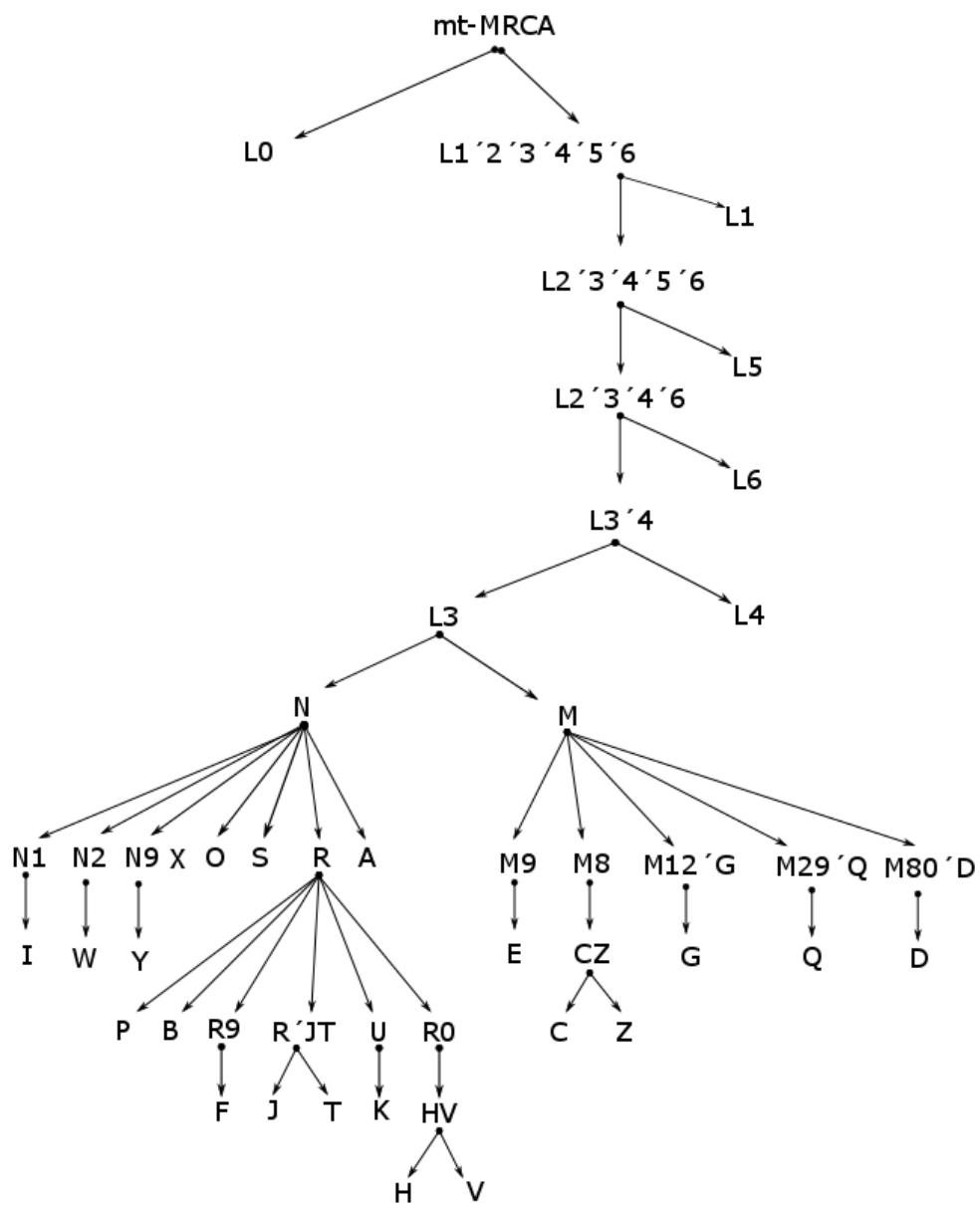


Figure 2.

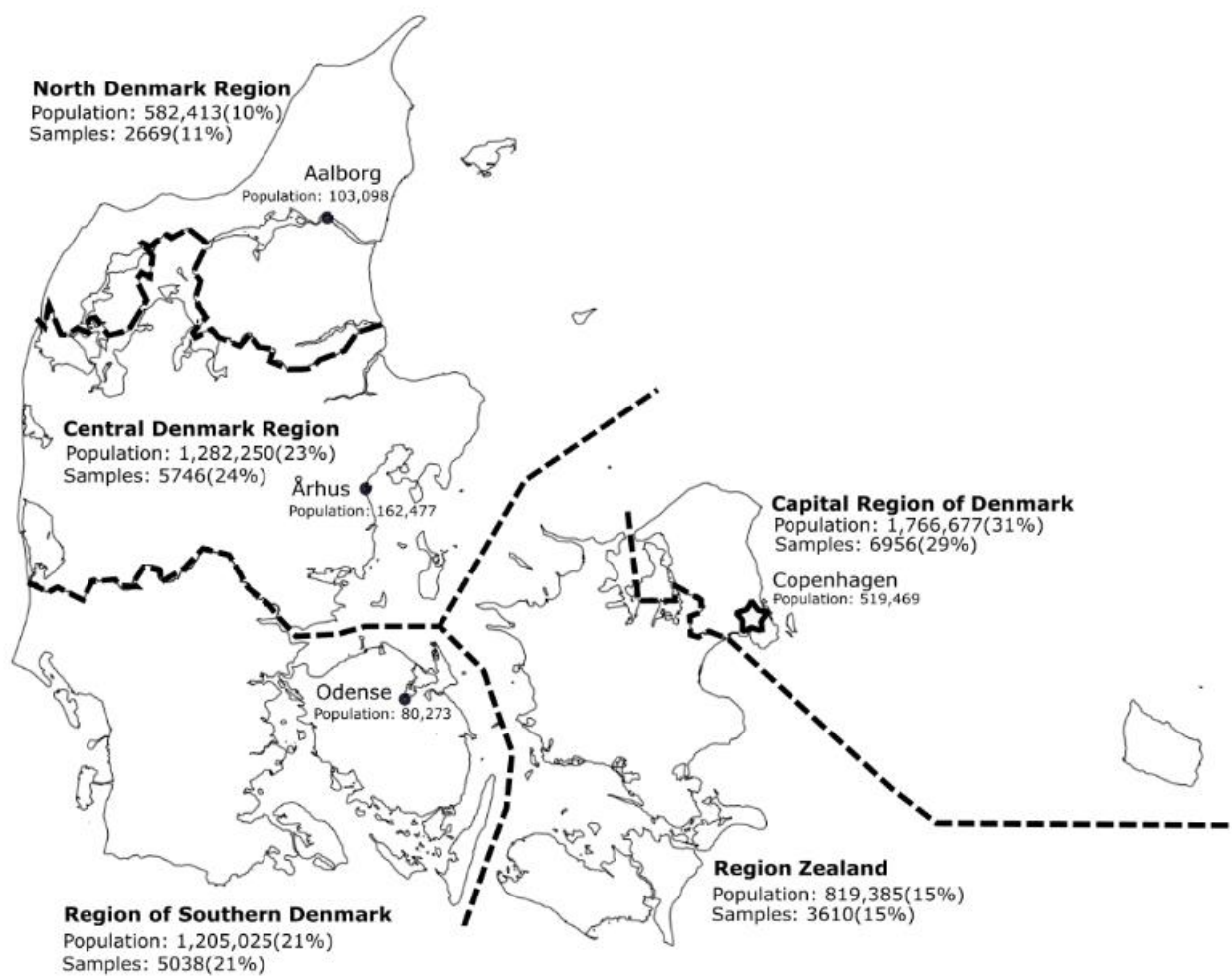


Figure 3.

A

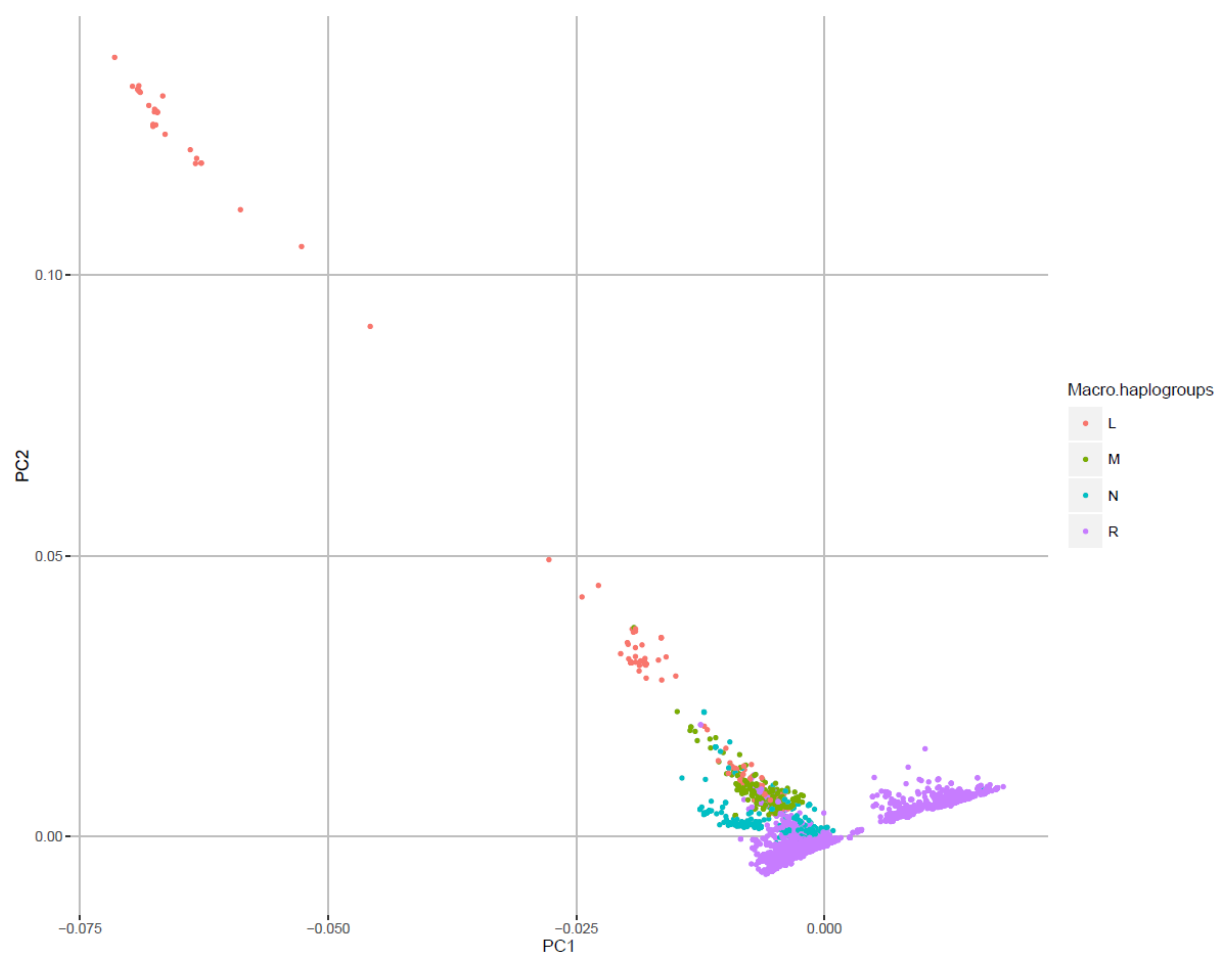


Figure 3.

B

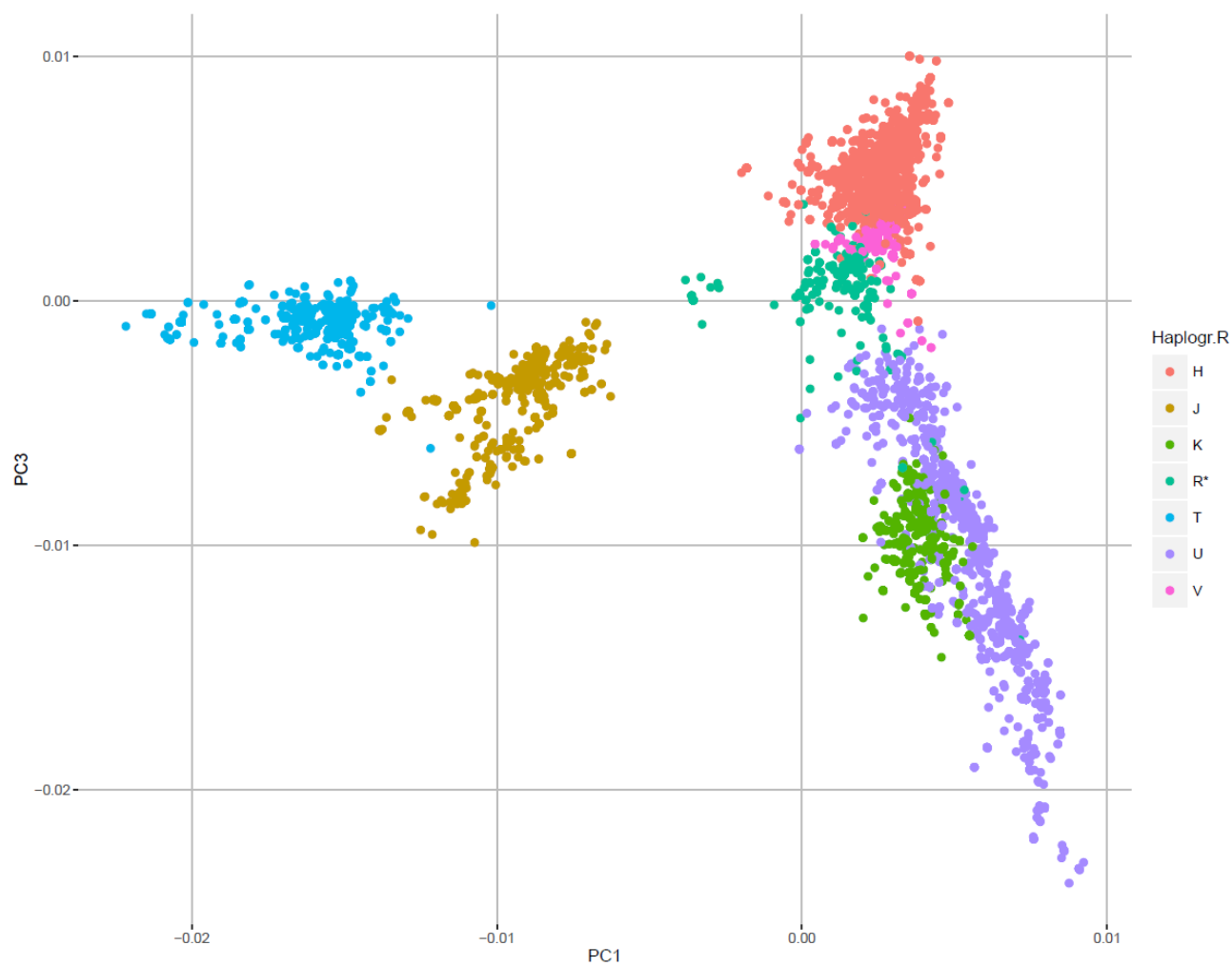


Figure 3.

C

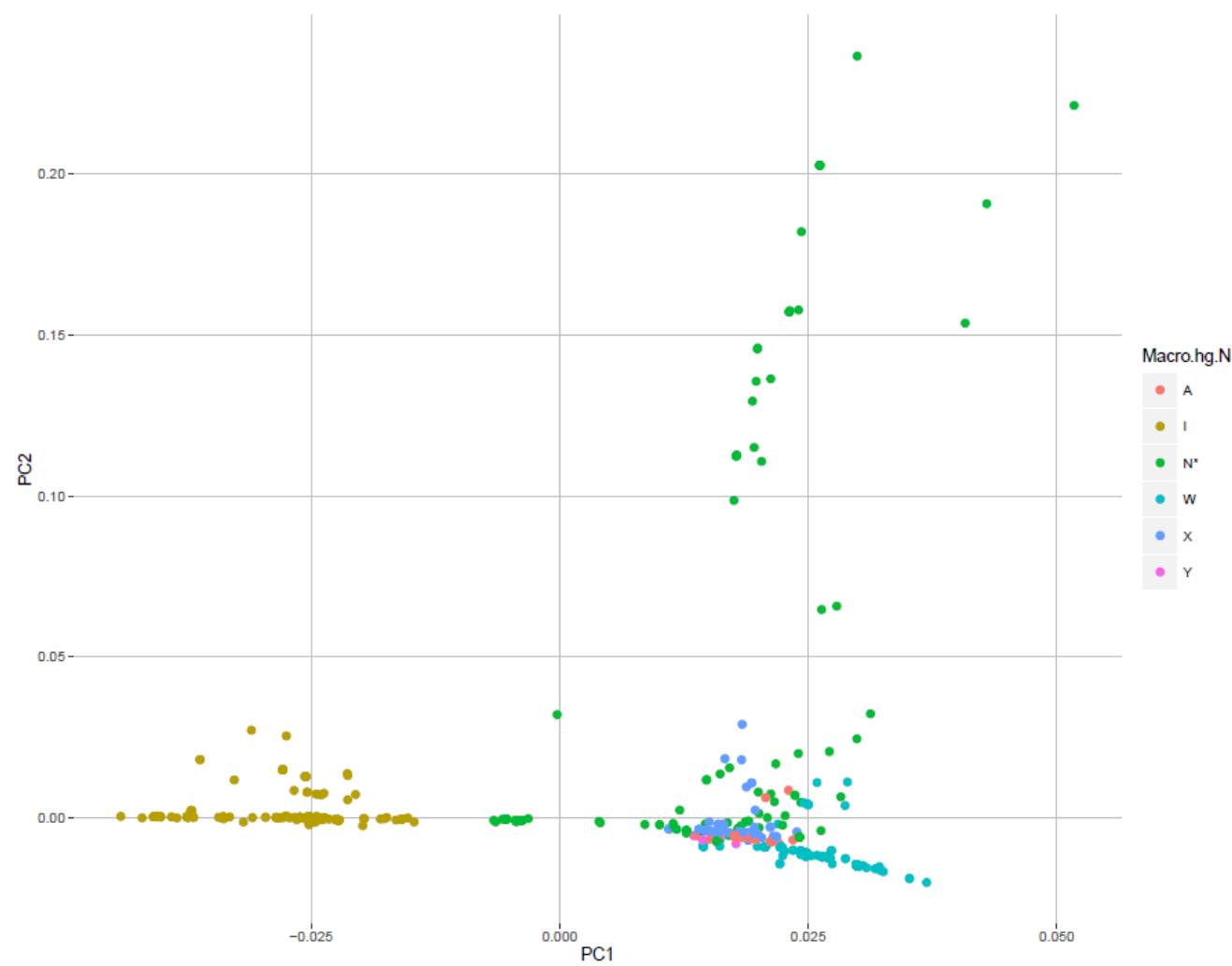


Figure 3.

D

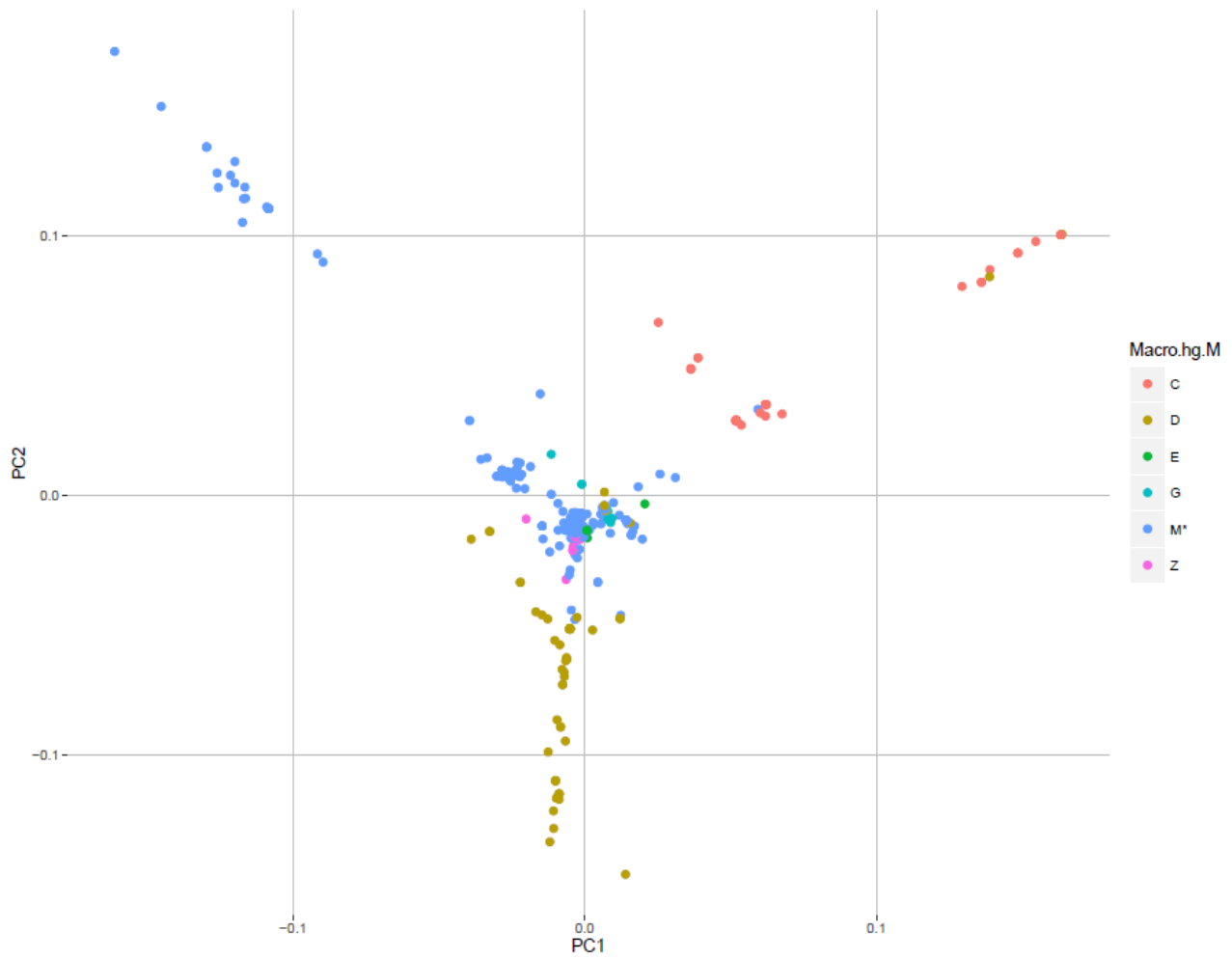


Figure 3.

E

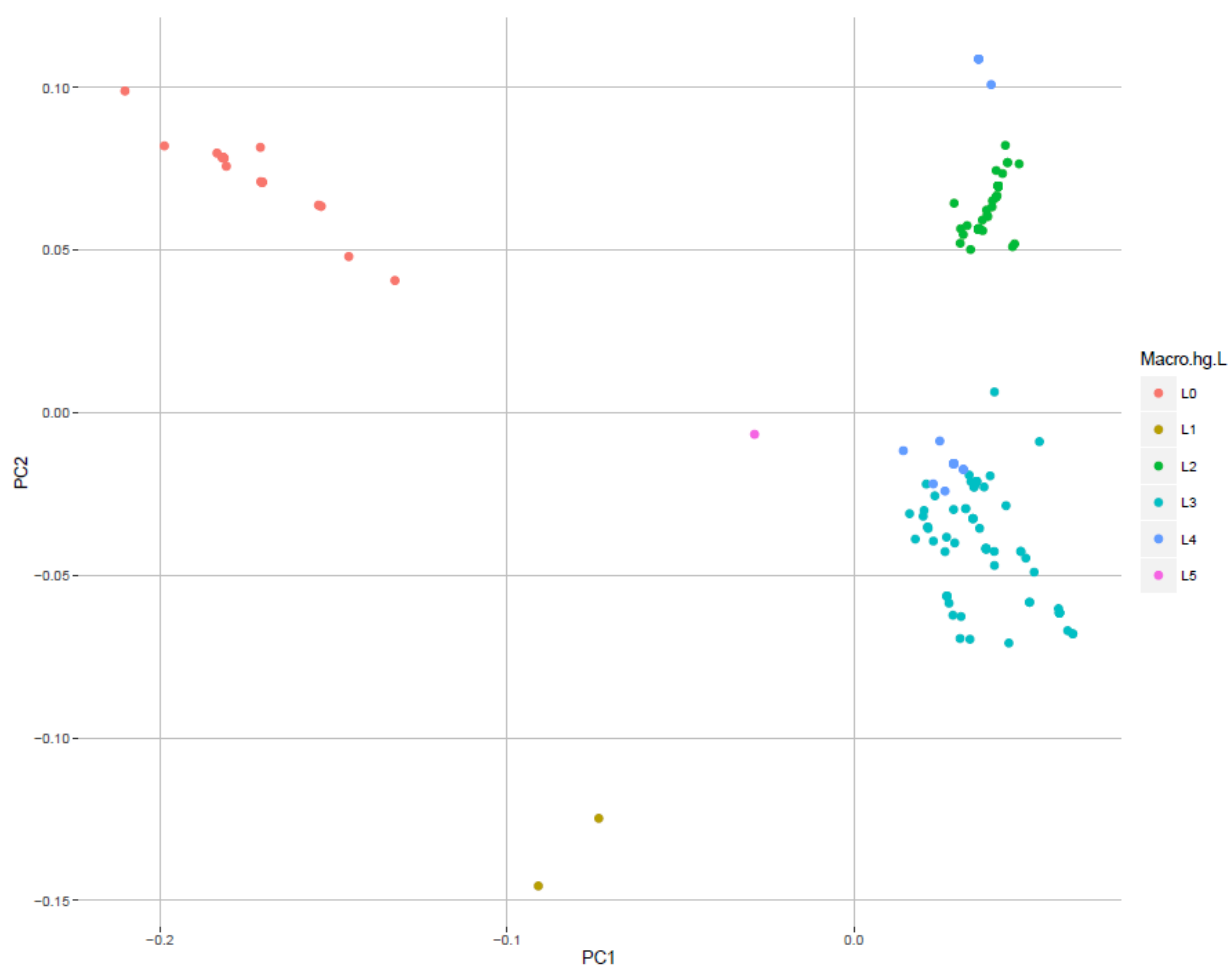


Figure 4.
A.

R haplogroups

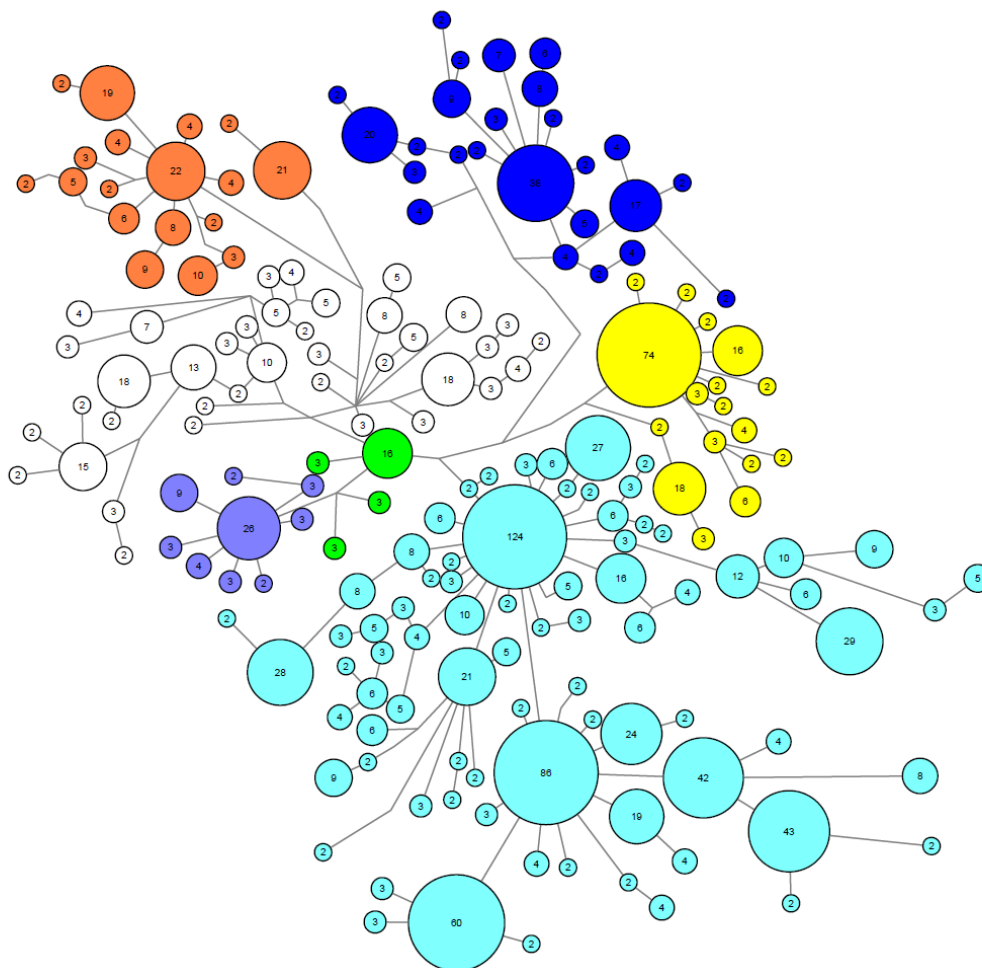


Figure 4.

B.

N haplogroups

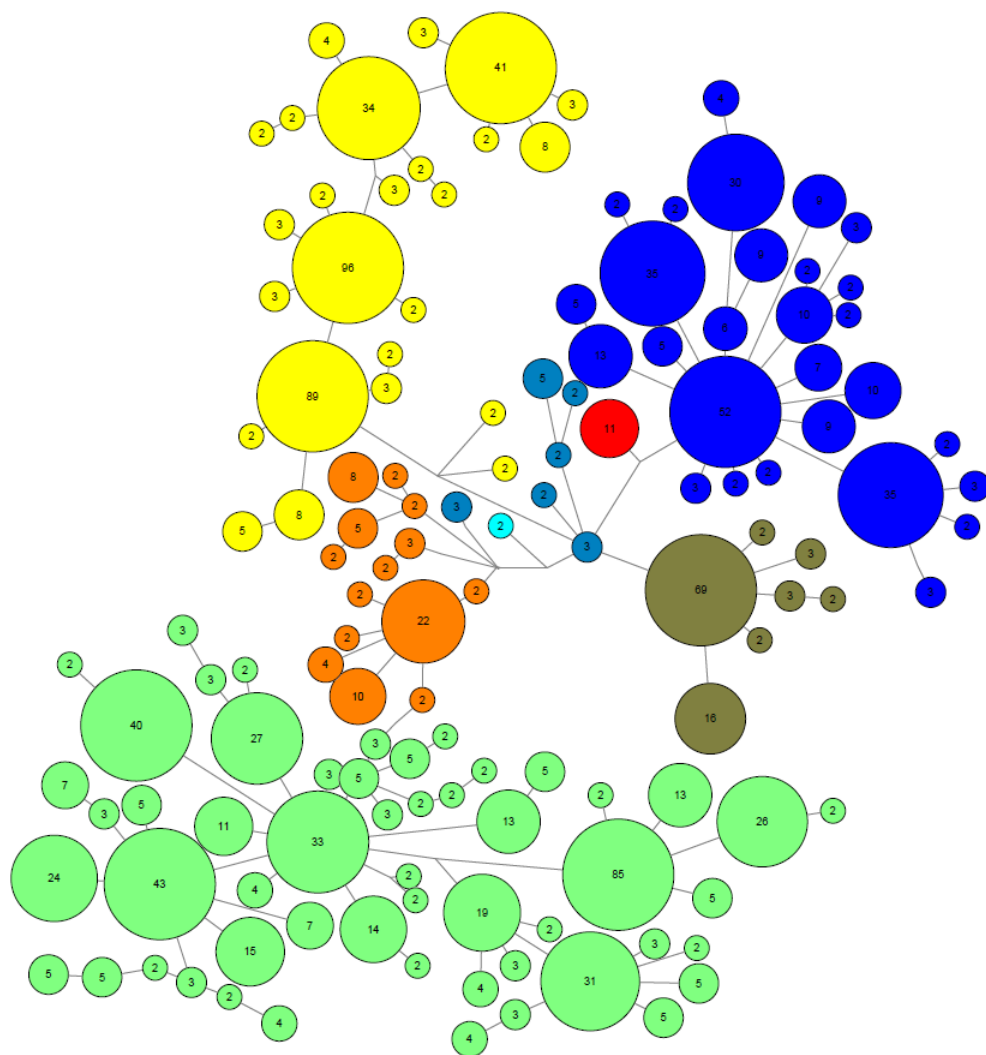


Figure 4.

C.

M haplogroups

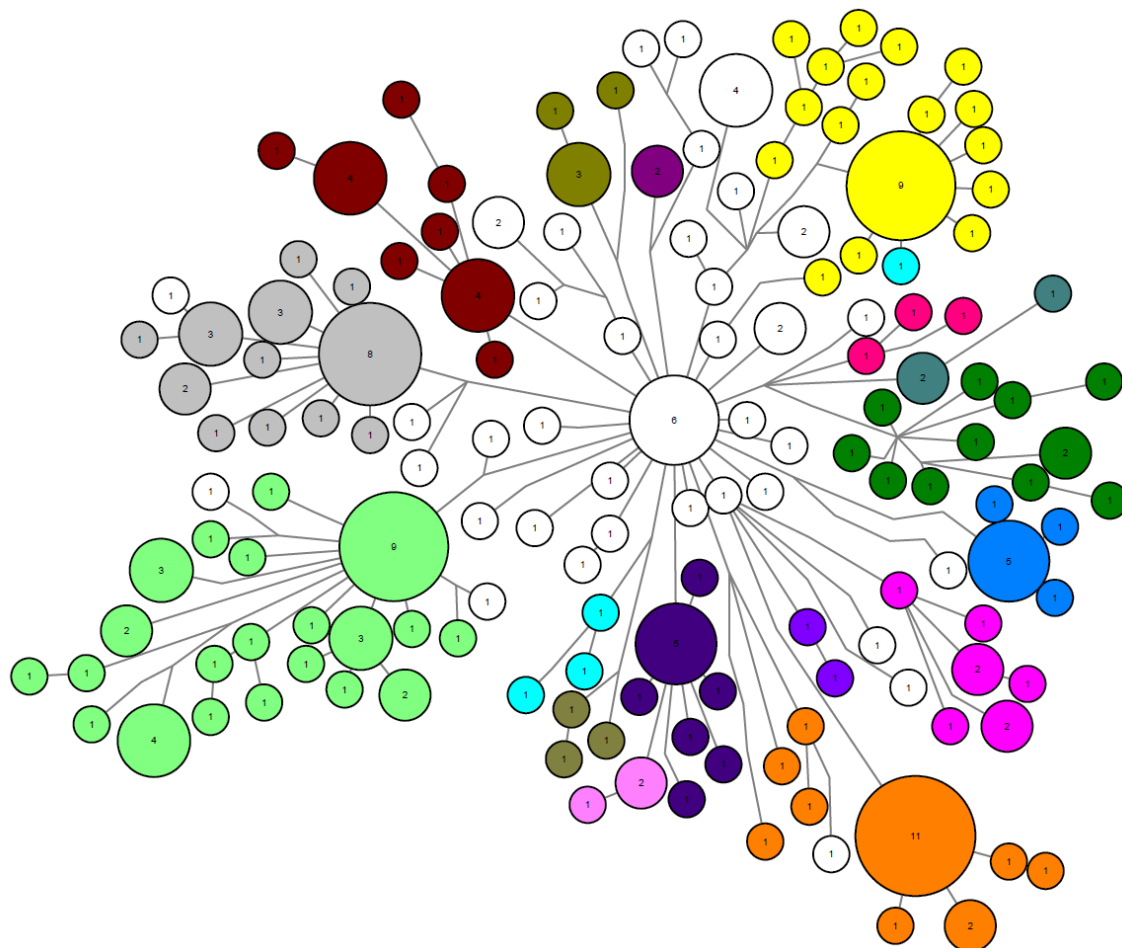


Figure 4.
D.

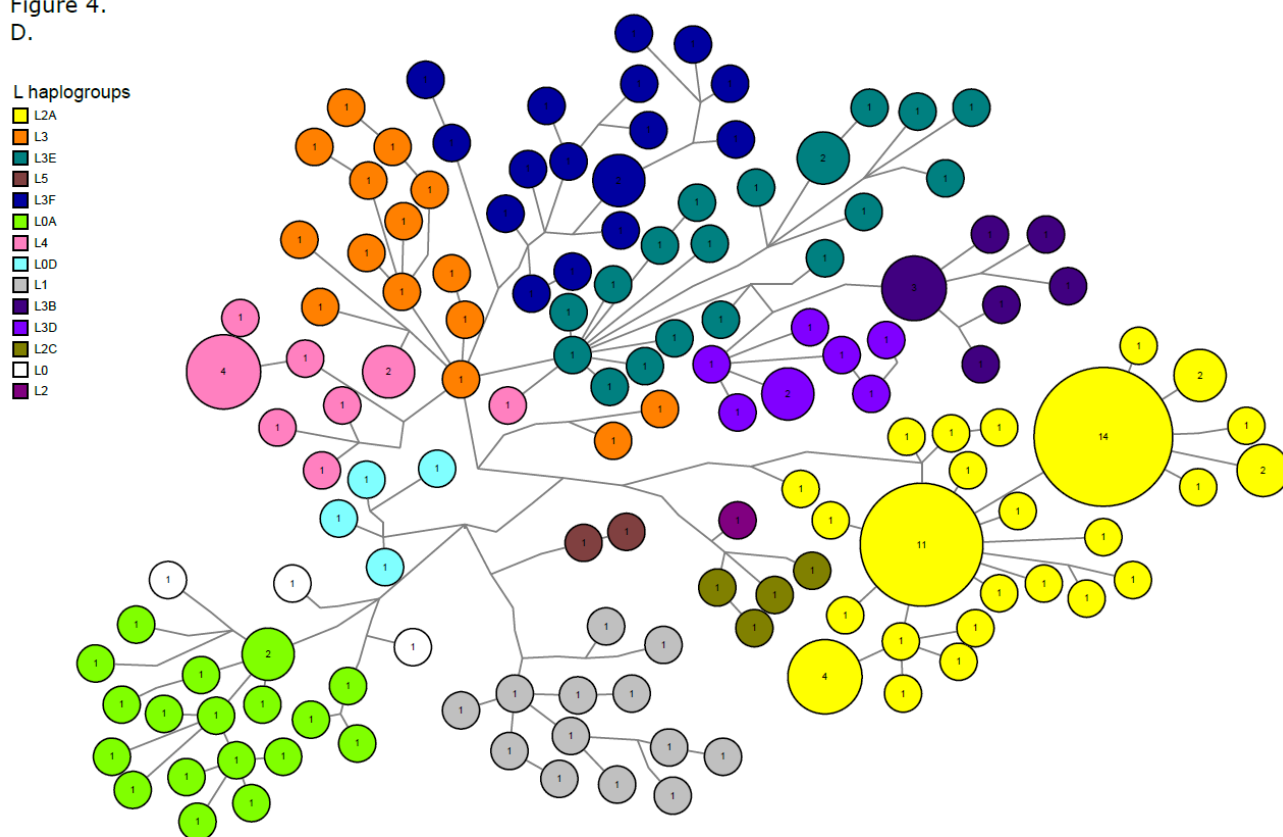
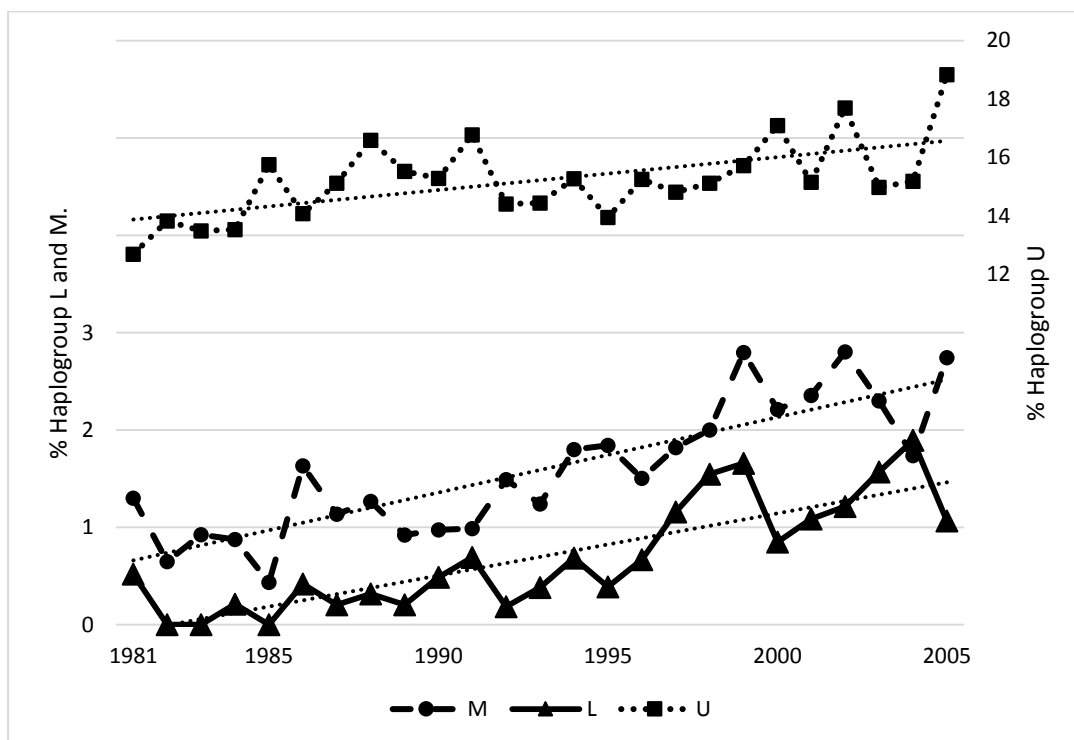


Figure 5.



References

1. Papa, S., Martino, P.L., Capitanio, G., Gaballo, A., De Rasmio, D., Signorile, A., and Petruzzella, V. (2012). The oxidative phosphorylation system in mammalian mitochondria. *Adv Exp Med Biol* 942, 3-37.
2. Antico Arciuch, V.G., Elguero, M.E., Poderoso, J.J., and Carreras, M.C. (2012). Mitochondrial regulation of cell cycle and proliferation. *Antioxid Redox Signal* 16, 1150-1180.
3. Carafoli, E. (2010). The fateful encounter of mitochondria with calcium: how did it happen? *Biochimica et biophysica acta* 1797, 595-606.
4. Zhang, F., Zhang, L., Qi, Y., and Xu, H. (2016). Mitochondrial cAMP signaling. *Cell Mol Life Sci*.
5. Chinnery, P.F., and Hudson, G. (2013). Mitochondrial genetics. *Br Med Bull* 106, 135-159.
6. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics* 23, 147.
7. Calvo, S.E., and Mootha, V.K. (2010). The mitochondrial proteome and human disease. *Annu Rev Genomics Hum Genet* 11, 25-44.
8. Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., et al. (2008). A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112-123.
9. Pyle, A., Hudson, G., Wilson, I.J., Coxhead, J., Smertenko, T., Herbert, M., Santibanez-Koref, M., and Chinnery, P.F. (2015). Extreme-Depth Re-sequencing of Mitochondrial DNA Finds No Evidence of Paternal Transmission in Humans. *PLoS Genet* 11, e1005040.
10. Brown, W.M., George, M., Jr., and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76, 1967-1971.
11. Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J., 2nd, and Nikoskelainen, E.K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242, 1427-1430.
12. Holt, I.J., Harding, A.E., and Morgan-Hughes, J.A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331, 717-719.
13. Bourgeron, T., Rustin, P., Chretien, D., Birch-Machin, M., Bourgeois, M., Viegas-Pequignot, E., Munnich, A., and Rotig, A. (1995). Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nat Genet* 11, 144-149.
14. Vafai, S.B., and Mootha, V.K. (2012). Mitochondrial disorders as windows into an ancient organelle. *Nature* 491, 374-383.
15. Koopman, W.J., Willems, P.H., and Smeitink, J.A. (2012). Monogenic mitochondrial disorders. *N Engl J Med* 366, 1132-1141.
16. Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., and Wallace, D.C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303, 223-226.
17. Wallace, D.C., Brown, M.D., and Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238, 211-230.
18. Bowton, E., Field, J.R., Wang, S., Schildcrout, J.S., Van Driest, S.L., Delaney, J.T., Cowan, J., Weeke, P., Mosley, J.D., Wells, Q.S., et al. (2014). Biobanks and electronic medical records: enabling cost-effective research. *Science translational medicine* 6, 234cm233.

19. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human mutation* 30, E386-394.
20. Chinnery, P.F., Taylor, G.A., Howell, N., Andrews, R.M., Morris, C.M., Taylor, R.W., McKeith, I.G., Perry, R.H., Edwardson, J.A., and Turnbull, D.M. (2000). Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. *Neurology* 55, 302-304.
21. Carrieri, G., Bonafe, M., De Luca, M., Rose, G., Varcasia, O., Bruni, A., Maletta, R., Nacmias, B., Sorbi, S., Corsonello, F., et al. (2001). Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. *Hum Genet* 108, 194-198.
22. van der Walt, J.M., Dementieva, Y.A., Martin, E.R., Scott, W.K., Nicodemus, K.K., Kroner, C.C., Welsh-Bohmer, K.A., Saunders, A.M., Roses, A.D., Small, G.W., et al. (2004). Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett* 365, 28-32.
23. Coskun, P., Wyrembak, J., Schriener, S.E., Chen, H.W., Marciniack, C., Laferla, F., and Wallace, D.C. (2012). A mitochondrial etiology of Alzheimer and Parkinson disease. *Biochim Biophys Acta* 1820, 553-564.
24. Hudson, G., Nalls, M., Evans, J.R., Breen, D.P., Winder-Rhodes, S., Morrison, K.E., Morris, H.R., Williams-Gray, C.H., Barker, R.A., Singleton, A.B., et al. (2013). Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. *Neurology* 80, 2042-2048.
25. Pyle, A., Foltynie, T., Tiangyou, W., Lambert, C., Keers, S.M., Allcock, L.M., Davison, J., Lewis, S.J., Perry, R.H., Barker, R., et al. (2005). Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. *Ann Neurol* 57, 564-567.
26. Verge, B., Alonso, Y., Valero, J., Miralles, C., Vilella, E., and Martorell, L. (2011). Mitochondrial DNA (mtDNA) and schizophrenia. *Eur Psychiatry* 26, 45-56.
27. Hagen, C.M., Aidt, F.H., Hedley, P.L., Jensen, M.K., Havndrup, O., Kanters, J.K., Moolman-Smook, J.C., Larsen, S.O., Bundgaard, H., and Christiansen, M. (2013). Mitochondrial haplogroups modify the risk of developing hypertrophic cardiomyopathy in a Danish population. *PLoS One* 8, e71904.
28. Castro, M.G., Huerta, C., Reguero, J.R., Soto, M.I., Domenech, E., Alvarez, V., Gomez-Zaera, M., Nunes, V., Gonzalez, P., Corao, A., et al. (2006). Mitochondrial DNA haplogroups in Spanish patients with hypertrophic cardiomyopathy. *Int J Cardiol* 112, 202-206.
29. Fernandez-Caggiano, M., Barallobre-Barreiro, J., Rego-Perez, I., Crespo-Leiro, M.G., Paniagua, M.J., Grille, Z., Blanco, F.J., and Domenech, N. (2012). Mitochondrial haplogroups H and J: risk and protective factors for ischemic cardiomyopathy. *PLoS One* 7, e44128.
30. Kenney, M.C., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., Del Carpio, J.C., Nesburn, A.B., Boyer, D.S., et al. (2014). Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. *Biochim Biophys Acta* 1842, 208-219.
31. Larsen, S., Diez-Sanchez, C., Rabol, R., Ara, I., Dela, F., and Helge, J.W. (2014). Increased intrinsic mitochondrial function in humans with mitochondrial haplogroup H. *Biochim Biophys Acta* 1837, 226-231.
32. Atilano, S.R., Malik, D., Chwa, M., Caceres-Del-Carpio, J., Nesburn, A.B., Boyer, D.S., Kuppermann, B.D., Jazwinski, S.M., Miceli, M.V., Wallace, D.C., et al. (2015). Mitochondrial DNA variants can mediate methylation status of inflammation, angiogenesis and signaling genes. *Hum Mol Genet* 24, 4491-4503.
33. Kenney, M.C., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., Caceres-del-Carpio, J., Nesburn, A.B., Boyer, D.S., et al. (2014). Inherited mitochondrial DNA

- variants can affect complement, inflammation and apoptosis pathways: insights into mitochondrial-nuclear interactions. *Hum Mol Genet* 23, 3537-3551.
34. Wallace, D.C. (2005). A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* 39, 359-407.
 35. Durieux, J., Wolff, S., and Dillin, A. (2011). The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* 144, 79-91.
 36. Castro, M.G., Terrados, N., Reguero, J.R., Alvarez, V., and Coto, E. (2007). Mitochondrial haplogroup T is negatively associated with the status of elite endurance athlete. *Mitochondrion* 7, 354-357.
 37. Ji, Y., Zhang, A.M., Jia, X., Zhang, Y.P., Xiao, X., Li, S., Guo, X., Bandelt, H.J., Zhang, Q., and Yao, Y.G. (2008). Mitochondrial DNA haplogroups M7b1'2 and M8a affect clinical expression of leber hereditary optic neuropathy in Chinese families with the m.11778G-->a mutation. *Am J Hum Genet* 83, 760-768.
 38. Samuels, D.C., Carothers, A.D., Horton, R., and Chinnery, P.F. (2006). The power to detect disease associations with mitochondrial DNA haplogroups. *Am J Hum Genet* 78, 713-720.
 39. Raule, N., Sevini, F., Santoro, A., Altilia, S., and Franceschi, C. (2007). Association studies on human mitochondrial DNA: methodological aspects and results in the most common age-related diseases. *Mitochondrion* 7, 29-38.
 40. Biffi, A., Anderson, C.D., Nalls, M.A., Rahman, R., Sonni, A., Cortellini, L., Rost, N.S., Matarin, M., Hernandez, D.G., Plourde, A., et al. (2010). Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. *Am J Hum Genet* 86, 904-917.
 41. Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., et al. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75, 910-918.
 42. Hudson, G., Gomez-Duran, A., Wilson, I.J., and Chinnery, P.F. (2014). Recent Mitochondrial DNA Mutations Increase the Risk of Developing Common Late-Onset Human Diseases. *Plos Genetics* 10.
 43. Manco, J. (2015). *Ancestral Journeys*.(London: Thames & Hudson).
 44. Cunliffe, B. (2008). *Europe between the oceans*.(New Haven and London: Yale University Press).
 45. Bellwood, P. (2013). *The Global prehistory of Human Migration*.(West Sussex, UK: Wiley Blackwell).
 46. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Wellcome Trust Case Control, C., et al. (2015). The fine-scale genetic structure of the British population. *Nature* 519, 309-314.
 47. Norgaard-Pedersen, B., and Hougaard, D.M. (2007). Storage policies and use of the Danish Newborn Screening Biobank. *J Inherit Metab Dis* 30, 530-536.
 48. Agerbo, E., Mortensen, P.B., Wiuf, C., Pedersen, M.S., McGrath, J., Hollegaard, M.V., Norgaard-Pedersen, B., Hougaard, D.M., Mors, O., and Pedersen, C.B. (2012). Modelling the contribution of family history and variation in single nucleotide polymorphisms to risk of schizophrenia: a Danish national birth cohort-based study. *Schizophr Res* 134, 246-252.
 49. Bandelt, H.J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16, 37-48.
 50. Bandelt, H.J., Forster, P., Sykes, B.C., and Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141, 743-753.
 51. WWW.fluxus-engineering.com

52. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655-1664.
53. Breiman, L. (2001). Random forests. *Mach Learn* 45, 5-32.
54. Raymond, M., and Rousset, F. (1995). An exact test for population differentiation. *Evolution* 49, 1280-1283.
55. Team, R.C. (2015). R: A language and environment for statistical computing. In. (Vienna, Austria, R Foundation for Statistical Computing.
56. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *Plos Genet* 2, e190.
57. Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet* 6, 3.
58. Campbell, M.C., and Tishkoff, S.A. (2010). The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* 20, R166-173.
59. Campbell, M.C., Hirbo, J.B., Townsend, J.P., and Tishkoff, S.A. (2014). The peopling of the African continent and the diaspora into the new world. *Curr Opin Genet Dev* 29, 120-132.
60. Pihl, K., Larsen, T., Jonsson, L., Hougaard, D., Krebs, L., Norgaard-Pedersen, B., and Christiansen, M. (2008). [Quality control of prenatal screening]. *Ugeskrift for laeger* 170, 2691-2695.
61. Pinhasi, R., Thomas, M.G., Hofreiter, M., Currat, M., and Burger, J. (2012). The genetic history of Europeans. *Trends Genet* 28, 496-505.
62. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., and Richards, M.B. (2010). The archaeogenetics of Europe. *Curr Biol* 20, R174-183.
63. Benn, M., Schwartz, M., Nordestgaard, B.G., and Tybjaerg-Hansen, A. (2008). Mitochondrial haplogroups: ischemic cardiovascular disease, other diseases, mortality, and longevity in the general population. *Circulation* 117, 2492-2501.
64. Mikkelsen, M., Sorensen, E., Rasmussen, E.M., and Morling, N. (2010). Mitochondrial DNA HV1 and HV2 variation in Danes. *Forensic Sci Int Genet* 4, e87-88.
65. Li, S., Besenbacher, S., Li, Y., Kristiansen, K., Grarup, N., Albrechtsen, A., Sparso, T., Korneliusen, T., Hansen, T., Wang, J., et al. (2014). Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. *European journal of human genetics : EJHG* 22, 1040-1045.
66. Fu, Q., Mittnik, A., Johnson, P.L., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23, 553-559.
67. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *American journal of human genetics* 67, 718-726.
68. Helgason, A., Palsson, G., Pedersen, H.S., Angulalik, E., Gunnarsdottir, E.D., Yngvadottir, B., and Stefansson, K. (2006). mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. *Am J Phys Anthropol* 130, 123-134.
69. Starikovskaya, Y.B., Sukernik, R.I., Schurr, T.G., Kogelnik, A.M., and Wallace, D.C. (1998). mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. *American journal of human genetics* 63, 1473-1491.
70. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliusen, T.S., Gronnow, B., Appelt, M., Gullov, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. *Science* 345, 1255832.

71. Volodko, N.V., Starikovskaya, E.B., Mazunin, I.O., Eltsov, N.P., Naidenko, P.V., Wallace, D.C., and Sukernik, R.I. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *American journal of human genetics* 82, 1084-1100.
72. Underhill, P.A., and Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41, 539-564.
73. Pakendorf, B., and Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6, 165-183.
74. Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *American journal of human genetics* 53, 563-590.
75. Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S., and Wallace, D.C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *American journal of human genetics* 57, 133-149.
76. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L., and Wallace, D.C. (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144, 1835-1850.
77. Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet* 6, 3.
78. Emery, L.S., Magnaye, K.M., Bigam, A.W., Akey, J.M., and Bamshad, M.J. (2015). Estimates of continental ancestry vary widely among individuals with the same mtDNA haplogroup. *American journal of human genetics* 96, 183-193.
79. Orlando, L., Gilbert, M.T., and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 16, 395-408.
80. Morozova, I., Flegontov, P., Mikheyev, A.S., Bruskin, S., Asgharian, H., Ponomarenko, P., Klyuchnikov, V., ArunKumar, G., Prokhortchouk, E., Gankin, Y., et al. (2016). Toward high-resolution population genomics using archaeological samples. *DNA Res* 23, 295-310.
81. Allentoft, M.E., Sikora, M., Sjogren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167-172.
82. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwangler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200-205.
83. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207-211.
84. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536, 419-424.
85. Skoglund, P., Malmstrom, H., Raghavan, M., Stora, J., Hall, P., Willerslev, E., Gilbert, M.T., Gotherstrom, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336, 466-469.
86. Larsen, N.K., Knudsen, K.L., Krohn, C.F., Kronborg, C., Murray, A.S., and Nielsen, O.B. (2009). Late Quaternary ice sheet, lake and sea history of southwest Scandinavia - a synthesis. *Boreas* 38, 732-761.

87. Egeland, C.P., Nielsen, T.K., Byo, M., Kjaergaard, P.C., Larsen, N.K., and Riede, F. (2014). The taphonomy of fallow deer (*Dama dama*) skeletons from Denmark and its bearing on the pre-Weichselian occupation of northern Europe by humans. *Archaeol Anthropol Sci* 6, 31-61.
88. Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The Last Glacial Maximum. *Science (New York, N Y)* 325, 710-714.
89. Price T., D. (2015). *Ancient Scandinavia: An Archaeological History From The First Humans To The Vikings.*(New York: Oxford University Press).
90. Andersson, M., Karsten, P., Knarrestrom, B., Svensson, M. (2004). *Stone Age Scania. Significant Places Dug and Read by Contract Archaeology.*(Lund: National Heritage Board. Archaeological Excavations, UV Syd.).
91. Derry, T.K. (2000). *History of Scandinavia: Norway, Sweden, Denmark, Finland, and Iceland.*(Minneapolis, US: University of Minneapolis Press).
92. Randsborg, K. (2009). *The Anatomy of Denmark. Archaeology and History from the Ice Age to the Present.*(London: Bristol Classical Press).
93. Torroni, A., Bandelt, H.J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B., et al. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American journal of human genetics* 62, 1137-1152.
94. Petersen, V.P., Johansen, L. (1996). Tracking Late Glacial reindeer hunters in Eastern Denmark. *Nationalmuseets Arbejdsmark*, 80-97.
95. Holm, J. (1993). Settlements of the Hamburgian and Feddermesser culture at Slotseng, South Jylland. *J Dan Arch* 10, 7-19.
96. Mellars, P. (2006). A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439, 931-935.
97. Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Trinkaus, E., Chandler, B., Groning, F., Collins, C., Hillson, S., et al. (2011). The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479, 521-524.
98. Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., et al. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479, 525-528.
99. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwangler, A., Wissing, C., et al. (2016). Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr Biol* 26, 827-833.
100. Hofmanova, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Diez-Del-Molino, D., van Dorp, L., Lopez, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6886-6891.
101. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409-413.
102. Melchior, L., Lynnerup, N., Siegismund, H.R., Kivisild, T., and Dissing, J. (2010). Genetic diversity among ancient Nordic populations. *PloS one* 5, e11898.
103. Fortson, B., W. (2009). *Indo-European Language and Culture: An Introduction.*(London: Wiley-Blackwell).

104. Wickham, C. (2010). *The inheritance of Rome: A history of Europe from 400 to 1000.* (London: Penguin).
105. (2015). *Danmarks Statistik.* In. (
106. Hvidt, K. (1976). *Danes go west. A book about the emigrations to America.* (Copenhagen: Rebild National Park Society,).
107. Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Sarkissian, C.D., Ganslmeier, R., Friederich, S., et al. (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* 4, 1764.
108. Malyarchuk, B., Derenko, M., Grzybowski, T., Perkova, M., Rogalla, U., Vanecek, T., and Tsybovsky, I. (2010). The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PloS one* 5, e10285.
109. Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H.-V., Parik, J., Loogvali, E.-L., Derenko, M., Malyarchuk, B., et al. (2003). Origin and diffusion of mtDNA haplogroup X. *American journal of human genetics* 73, 1178-1190.
110. Comas, D., Plaza, S., Wells, R.S., Yuldaseva, N., Lao, O., Calafell, F., and Bertranpetit, J. (2004). Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *European journal of human genetics : EJHG* 12, 495-504.
111. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26.
112. Gonder, M.K., Mortensen, H.M., Reed, F.A., de Sousa, A., and Tishkoff, S.A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24, 757-768.
113. Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol* 24, 2180-2195.
114. Campbell, M.C., Hirbo, J.B., Townsend, J.P., and Tishkoff, S.A. (2014). The peopling of the African continent and the diaspora into the new world. *Curr Opin Genet Dev* 29, 120-132.
115. Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., Lopez-Soto, M., Carracedo, A., et al. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res* 22, 821-826.
116. Busby, G.B.J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., Zemunik, T., Hayward, C., Toncheva, D., Karachanak-Yankova, S., et al. (2015). The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Curr Biol* 25, 2518-2526.
117. Zupan, A., Hauptman, N., and Glavac, D. (2016). The maternal perspective for five Slovenian regions: The importance of regional sampling. *Ann Hum Biol* 43, 57-66.
118. Athanasiadis, G., Cheng, J.Y., Vilhjalmsen, B.J., Jorgensen, F.G., Als, T.D., Le Hellard, S., Espeseth, T., Sullivan, P.F., Hultman, C.M., Kjargaard, P.C., et al. (2016). Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity. *Genetics* 204, 711-722.
119. Catelli, M.L., Alvarez-Iglesias, V., Gomez-Carballa, A., Mosquera-Miguel, A., Romanini, C., Borosky, A., Amigo, J., Carracedo, A., Vullo, C., and Salas, A. (2011). The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome. *BMC Genet* 12, 77.
120. Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E., et al. (2005). Population structure, differential bias and

genomic control in a large-scale, case-control association study. *Nature genetics* 37, 1243-1246.

121. Marchini, J., Cardon, L.R., Phillips, M.S., and Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature genetics* 36, 512-517.