**Supplementary Appendix - A frameshift deletion in a sarcomere gene (MYL4) causes early onset familial atrial fibrillation**

**Table of contents**

*Page 2 Norwegian atrial fibrillation study population from The Tromsø Study*

*Page 2 Hong Kong atrial fibrillation study population*

*Page 3 Illumina genome wide genotyping*

*Page 3 Whole-genome sequencing and SNP imputations*

*Page 5 Long range phasing*

*Page 5 Genotype imputation*

*Page 7 In silico (genealogy-based) genotyping*

*Page 9 Case-control association testing*

*Page 11 Microsatellite genotyping*

*Page 13 Supplementary Table 1*

*Page 14 Supplementary Figure 1*

*Page 15 Supplementary Figure 2*

*Page 16 Supplementary references*

*Norwegian atrial fibrillation study population from The Tromsø Study:*

The Tromsø Study is a population-based prospective study with repeated health surveys in the municipality of Tromsø, Norway. More than 40,000 individuals have been examined. The population is followed on an individual level with registration and validation of diseases and death. An endpoint registry is kept for cardiovascular diseases (CVD). Discharge diagnosis lists of CVD are retrieved from the University Hospital of North Norway in Tromsø and medical records for all individuals with a CVD discharge diagnosis are reviewed. AF was registered between1986 and 2004 as part of the CVD endpoint registration in the Tromsø Study. Postoperative AF was excluded. Sex - and age matched controls without history of arrhythmias were drawn for each case of AF from the population based Tromsø 4 survey. All participants in the Tromsø Study gave informed written consent. The study was approved by the Regional Committee for Medical Research Ethics.

*Hong Kong atrial fibrillation study population:*

All subjects in the Hong Kong Study were of southern Han Chinese ancestry residing in Hong Kong. The cases consisted of 217 individuals from the Prince of Wales Hospital Diabetes Registry ([1](#_ENREF_1)) and 116 subjects from the Stroke Registry ([2](#_ENREF_2)). All AF diagnoses were confirmed with an ECG. The controls consisted of 2,836 subjects without evidence of AF. Informed consent was obtained for each participating subject. This study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong.

*Illumina genome-wide genotyping:*

The Icelandic chip-typed samples were assayed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. SNPs were excluded if they had (i) yield less than 95%, (ii) minor allele frequency less than 1% in the population or (iii) significant deviation from Hardy-Weinberg equilibrium in the controls (*P* < 0.001), (iv) if they produced an excessive inheritance error rate (over 0.001), (v) if there was substantial difference in allele frequency between chip types (from just a single chip if that resolved all differences, but from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. For the HumanHap series of chips, 304,937 SNPs were used for long range phasing, whereas for the Omni series of chips 564,196 SNPs were included. The final set of SNPs used for long-range phasing was composed of 707,525 SNPs.

*Whole-genome sequencing and SNP imputations:*

SNPs were imputed based on an extension of the previously described set of 2,636 Icelanders ([3](#_ENREF_3)) to 8,453 whole genome sequenced Icelanders. Twenty five point five million high quality and imputable SNPs and indels were identified in this set of sequenced individuals.

1. *Sample preparation.* Paired-end libraries for sequencing were prepared according to the manufacturer's instructions (Illumina). In short, approximately 5 μg of genomic DNA, isolated from frozen blood samples, were fragmented to a mean target size of 300 bp using a Covaris E210 instrument. The resulting fragmented DNA was end repaired using T4 and Klenow polymerases and T4 polynucleotide kinase with 10 mM dNTP followed by addition of an 'A' base at the ends using Klenow exo fragment (3′ to 5′-exo minus) and dATP (1 mM). Sequencing adaptors containing 'T' overhangs were ligated to the DNA products followed by agarose (2%) gel electrophoresis. Fragments of about 400 bp were isolated from the gels (QIAGEN Gel Extraction Kit), and the adaptor-modified DNA fragments were PCR enriched for ten cycles using Phusion DNA polymerase (Finnzymes Oy) and PCR primers PE 1.0 and PE 2.0 (Illumina). Enriched libraries were further purified using agarose (2%) gel electrophoresis as described above. The quality and concentration of the libraries were assessed with the Agilent 2100 Bioanalyzer using the DNA 1000 LabChip (Agilent). Barcoded libraries were stored at −20 °C. All steps in the workflow were monitored using an in-house laboratory information management system with barcode tracking of all samples and reagents.
2. *DNA sequencing.* Template DNA fragments were hybridized to the surface of flow cells (Illumina PE flowcell, v4) and amplified to form clusters using the Illumina cBot. In brief, DNA (3–10 pM) was denatured, followed by hybridization to grafted adaptors on the flowcell. Isothermal bridge amplification using Phusion polymerase was then followed by linearization of the bridged DNA, denaturation, blocking of 3´ ends and hybridization of the sequencing primer. Sequencing-by-synthesis was performed on Illumina GAIIx instruments equipped with paired-end modules. Paired-end libraries for whole-genome sequencing were sequenced using either 2×101 or 2×120 cycles of incorporation and imaging with Illumina sequencing kits, v4 or v5 (TruSeq). Each library or sample was initially run on a single lane for validation followed by further sequencing of ≥4 lanes with targeted raw cluster densities of 500–700 k/mm2, depending on the version of the data imaging and analysis packages. Imaging and analysis of the data was performed using SCS2.6 /RTA1.6, SCS2.8/RTA1.8 or SCS2.9&RTA1.9 software packages from Illumina, respectively. Real-time analysis involved conversion of image data to base-calling in real-time.
3. *Alignment.* Reads were aligned to NCBI Build 38 of the human reference sequence using Burrows-Wheeler Aligner (BWA) 0.7.10-r789 ([4](#_ENREF_4)). Alignments were merged into a single BAM file and marked for duplicates using Picard 1.55 (<http://picard.sourceforge.net/>). Only non-duplicate reads were used for the downstream analyses.
4. *Variant calling.* Variants were called using Genome Analysis Toolkit, (GenomeAnalysisTK) 1.2-29-g0acaf2d ([5](#_ENREF_5)), by applying base quality score recalibration, indel realignment and performing SNP and INDEL discovery and genotyping using standard hard filtering ([6](#_ENREF_6)). Variants were annotated using SNP effect predictor (snpEff) and Genome AnalysisToolkit 1.4-9-g1f1233b with only the highest-impact effect ([5](#_ENREF_5), [7](#_ENREF_7)).

*Long-range phasing:*

Long-range phasing of all chip-typed individuals was performed with methods described previously ([8](#_ENREF_8), [9](#_ENREF_9)). In brief, phasing was achieved using an iterative algorithm which phases a single proband at a time given the available phasing information about everyone else that shares a long haplotype identically by state with the proband. Given the large fraction of the Icelandic population that has been chip-typed, accurate long-range phasing is available genome-wide for all chip-typed Icelanders.

# *Genotype imputation:*

We imputed the SNPs identified and genotyped through sequencing into all Icelanders who had been phased with long-range phasing using the same model as used by IMPUTE ([10](#_ENREF_10)). The genotype data from sequencing can be ambiguous due to low sequencing coverage. In order to phase the sequencing genotypes, an iterative algorithm was applied for each SNP with alleles 0 and 1. We let *H* be the long-range phased haplotypes of the sequenced individuals and applied the following algorithm:

1. For each haplotype *h* in *H*, use the Hidden Markov Model of IMPUTE to calculate for every other *k* in *H*, the likelihood, denoted *γh,k*, of *h* having the same ancestral source as *k* at the SNP.
2. For every *h* in *H*, initialize the parameter *θh*, which specifies how likely the one allele of the SNP is to occur on the background of *h* from the genotype likelihoods obtained from sequencing. The genotype likelihood *Lg* is the probability of the observed sequencing data at the SNP for a given individual assuming *g* is the true genotype at the SNP. If *L*0, *L*1 and *L*2 are the likelihoods of the genotypes 0, 1 and 2 in the individual who carries *h*, then set .
3. For every pair of haplotypes *h* and *k* in *H* that are carried by the same individual, use the other haplotypes in *H* to predict the genotype of the SNP on the backgrounds of *h* and *k*: and . Combining these predictions with the genotype likelihoods from sequencing gives un-normalized updated phased genotype probabilities: , , and . Now use these values to update *θh* and *θk* to and .
4. Repeat step 3 when the maximum difference between iterations is greater than a convergence threshold *ε*. We used *ε*=10−7.

Given the long-range phased haplotypes and *θ*, the allele of the SNP on a new haplotype *h* not in *H*, is imputed as .

The above algorithm can easily be extended to handle simple family structures such as parent-offspring pairs and triads by letting the *P* distribution run over all founder haplotypes in the family structure. The algorithm also extends easily to the X-chromosome. If source genotype data are only ambiguous in phase, such as chip genotype data, then the algorithm is still applied, but all but one of the *L*s will be 0. In some instances, the reference set was intentionally enriched for carriers of the minor allele of a rare SNP in order to improve imputation accuracy. In this case, expected allele counts is biased toward the minor allele of the SNP. Call the enrichment of the minor allele *E* and let *θ*′ be the expected minor allele count calculated from the naïve imputation method, and let *θ* be the unbiased expected allele count, then and hence .

This adjustment was applied to all imputations based on enriched imputations sets. We note that if *θ*′ is 0 or 1, then *θ* will also be 0 or 1, respectively.

*In silico (genealogy-based) genotyping:*

In addition to imputing sequence variants from the whole-genome sequencing effort into chip-typed individuals, we also performed a second imputation step where genotypes were imputed into relatives of chip genotyped individuals, creating *in silico* genotypes. The inputs into the second imputation step are the fully phased (in particular every allele has been assigned a parent of origin) imputed and chip-type genotypes of the available chip-typed individuals. The algorithm used to perform the second imputation step consists of:

1. For each ungenotyped individual (the proband), find all chip-typed individuals within two meiosis of the individual. The six possible types of two meiosis relatives of the proband are (ignoring more complicated relationships due to pedigree loops): Parents, full and half siblings, grandparents, children and grandchildren. If all pedigree paths from the proband to a genotyped relative go through other genotyped relatives, then that relative is excluded. e.g. if a parent of the proband is genotyped, then the proband’s grandparents through that parent are excluded. If the number of meiosis in the pedigree around the proband exceeds a threshold (we used 12), then relatives are removed from the pedigree until the number of meiosis falls below 12, in order to reduce computational complexity.
2. At every point in the genome, calculate the probability for each genotyped relative sharing with the proband based on the autosomal SNPs used for phasing. A multipoint algorithm based on the hidden Markov model Lander-Green multipoint linkage algorithm using fast Fourier transforms is used to calculate these sharing probabilities ([11](#_ENREF_11), [12](#_ENREF_12)). First single point sharing probabilities are calculated by dividing the genome into 0.5cM bins and using the haplotypes over these bins as alleles. Haplotypes that are the same, except at most at a single SNP, are treated as identical. When the haplotypes in the pedigree are incompatible over a bin, then a uniform probability distribution was used for that bin. The most common causes for such incompatibilities are recombinations within the pedigree, phasing errors and genotyping errors. Note that since the input genotypes are fully phased, the single point information is substantially more informative than for unphased genotyped, in particular one haplotype of the parent of a genotyped child is always known. The single point distributions are then convolved using the multipoint algorithm to obtain multipoint sharing probabilities at the center of each bin. Genetic distances were obtained from the most recent version of the deCODE genetic map ([13](#_ENREF_13)).
3. Based on the sharing probabilities at the center of each bin, all the SNPs from the whole-genome sequencing are imputed into the proband. To impute the genotype of the paternal allele of a SNP located at , flanked by bins with centers at and . Starting with the left bin, going through all possible sharing patterns , let be the set of haplotypes of genotyped individuals that share identically by descent within the pedigree with the proband’s paternal haplotype given the sharing pattern and be the probability of at the left bin – this is the output from step 2 above – and let be the expected allele count of the SNP for haplotype . Then is the expected allele count of the paternal haplotype of the proband given and an overall estimate of the allele count given the sharing distribution at the left bin is obtained from . If is empty then no relative shares with the proband’s paternal haplotype given and thus there is no information about the allele count. We therefore store the probability that some genotyped relative shared the proband’s paternal haplotype, and an expected allele count, conditional on the proband’s paternal haplotype being shared by at least one genotyped relative: . In the same way calculate and . Linear interpolation is then used to get an estimates at the SNP from the two flanking bins:

If is an estimate of the population frequency of the SNP then is an estimate of the allele count for the proband’s paternal haplotype. Similarly, an expected allele count can be obtained for the proband’s maternal haplotype.

*Case-control association testing:*

Logistic regression was used to test for association between SNPs and disease, treating disease status as the response and expected genotype counts from imputation or allele counts from direct genotyping as covariates. Testing was performed using the likelihood ratio statistic. When testing for association based on the *in silico* genotypes, controls were matched to cases based on the informativeness of the imputed genotypes, such that for each case controls of matching informativeness where chosen. Failing to match cases and controls will lead to a highly inflated genomic control factor, and in some cases may lead to spurious false positive findings. The informativeness of each of the imputation of each one of an individual’s haplotypes was estimated by taking the average of  
over all SNPs imputed for the individual, where is the expected allele count for the haplotype at the SNP and is the population frequency of the SNP. Note that and . The mean informativeness values cluster into groups corresponding to the most common pedigree configurations used in the imputation, such as imputing from parent into child or from child into parent. Based on this clustering of imputation informativeness, we divided the haplotypes of individuals into seven groups of varying informativeness which created 27 groups of individuals of similar imputation informativeness; seven groups of individuals with both haplotypes having similar informativeness, 21 groups of individuals with the two haplotypes having different informativeness, minus the one group of individuals with neither haplotype being imputed well. Within each group we calculate the ratio of the number of controls and the number of cases and choose the largest integer that was less than this ratio in all the groups. For example, if in one group there are 10.3 times as many controls as cases and if in all other groups this ratio was greater, then we would set and within each group randomly select ten times as many controls as there are cases.

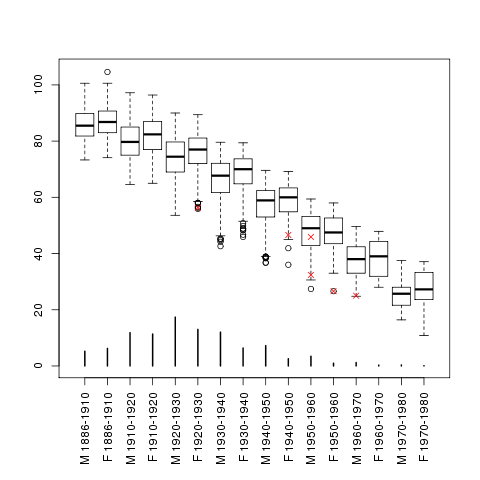
Two homozygous carriers were diagnosed with AF prior to the initiation of electronic medical recording in 1987, a male at 15 and female at 53 years of age. As the initial association test in our study was based on diagnoses extracted from electronic medical records (EMR), these two carriers were not handled as cases in the intial test. The AF diagnoses were subsequently discovered during detailed review of both paper and electronic medical records. In order to attach a significance level to this finding, we estimated the probability of these two individuals to be diagnosed with AF at the observed time in our Icelandic population. We then converted these probabilities to z-scores and combined them into a single z-score as follows. Only four male cases out of a total of 5,552 Icelandic AF cases were diagnosed before age 15 according to EMR. The number of Icelandic males born between 1982 and 1997 (so that their EMR diagnosis up to age 15 would be available to us) is 37,646 and of these, three were diagnosed with AF. Thus our data give a conservative estimate of one in ten thousand males being diagnosed with AF before age 15, corresponding to z\_1=-3.72. Data from the Framingham study suggest that less than one in a hundred women are diagnosed with AF before 53 years of age ([14](#_ENREF_14)) but having a sister with early-onset AF increases the probability of having early-onset AF substantially ([15](#_ENREF_15)). The probability of being diagnosed before 53, given that she had a sister with early-onset AF was estimated as five in a hundred corresponding to z\_2=-1.64. The combined z-score is then (z\_1+z\_2)/√2=-3.79, giving a two-sided P=1.5×10-4. Combined with the initial *P*-value of 1.7×10-12 (*Table S1*), the combined two-sided *P*-value for the association is 1.7×10-14. Adding these two cases to the set of early-onset AF cases and performing the test for association will create an inflated level of significance (*Table S1*, *P*=1.2×10-16), since homozygous carriers of the deletion would have a higher chance of having the early-onset AF diagnoses than controls.

*Microsatellite genotyping:*

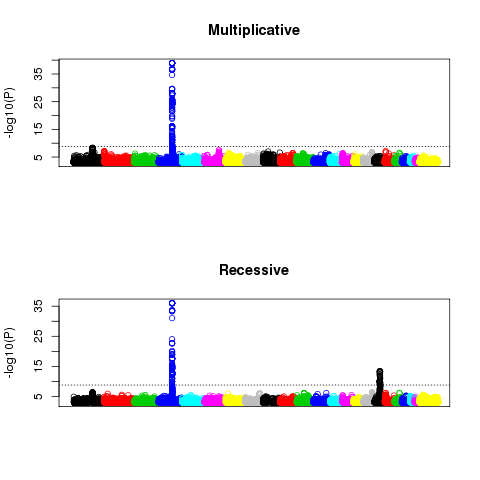
The PCR amplifications were set up and pooled using Zymark SciClone ALH 500 robots. The reaction volume was 5 μl, and, for each PCR, 20 ng of genomic DNA was amplified in the presence of 2 pmol of each primer, 0.14 U AmpliTaq Gold, 0.33 mmol/liter dNTPs, and 3.3 mmol/liter MgCl2. The PCR conditions were 95°C for 10 minutes, then stepdown 4 cycles of 15 s at 94°C, 30 s at 63°C -2,5°C per cycle, and 30 s at 72°C, 11 cycles of 15 s at 94°C, 30 s at 55°C, and 1 min at 72°C, at last 22 cycles of 15 s at 89°C, 30 s at 55°C, and 1 min at 72°C. The PCR products were supplemented with an internal size standard, and the fragments were separated and detected on an Applied Biosystems model 3730 sequencer, using Genescan (v. 3.0) peak-calling software. Alleles were called using an internal allele-calling program ([16](#_ENREF_16)).

**Table S1.** Association results for the recessive model for sequence variants showing genome-wide significant association with early-onset AF. Association results are given for discharge diagnoses of early-onset AF and discharge diagnoses combined with the early-onset AF diagnoses from medical records for the two homozygous carriers of c.234delC in *MYL4*. Odds ratios (OR) correspond to the minor allele. Positions are based on build 38, rs# are based on dbSNP build 144, minor allele frequency (MAF) is given in percentage points based on the Icelandic sequencing data and info measures the quality of the variant imputations on a scale of 0 to 1. The position of c.234delC is 47,219,973.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Discharge diagnoses** | | | | **All diagnoses** | | | |  | |  | |  | |
| **Position** | | **P** | | **OR** | | **P** | | **OR** | | **rs#** | | **MAF (%)** | | **Info** | |
| 44,004,779 | | 2.0E-12 | | 122.8 | | 1.4E-16 | | 158.2 | | rs754480096 | | 0.45 | | 1.00 | |
| 44,046,638 | | 1.3E-11 | | 119.1 | | 9.1E-16 | | 155.2 | | rs548925021 | | 0.45 | | 0.99 | |
| 44,309,184 | | 3.7E-11 | | 122.4 | | 1.1E-14 | | 155.0 | | rs572147699 | | 0.45 | | 0.99 | |
| 44,426,848 | | 3.8E-14 | | 132.4 | | 3.8E-14 | | 132.4 | | rs7213877 | | 0.49 | | 0.95 | |
| 44,877,815 | | 1.5E-12 | | 123.3 | | 1.0E-16 | | 159.0 | | - | | 0.48 | | 1.00 | |
| 45,135,106 | | 1.1E-13 | | 117.9 | | 1.1E-13 | | 117.9 | | rs28588212 | | 0.51 | | 1.00 | |
| 45,227,818 | | 3.9E-12 | | 112.3 | | 3.1E-16 | | 146.0 | | rs747955585 | | 0.55 | | 1.00 | |
| 45,320,758 | | 5.0E-10 | | 16.1 | | 1.8E-12 | | 19.4 | | - | | 2.13 | | 1.00 | |
| 45,655,853 | | 1.5E-11 | | 98.1 | | 1.5E-15 | | 129.4 | | - | | 0.61 | | 1.00 | |
| 45,961,779 | | 1.4E-12 | | 123.9 | | 9.4E-17 | | 159.5 | | - | | 0.53 | | 0.99 | |
| 46,265,623 | | 4.8E-11 | | 104.6 | | 7.9E-15 | | 132.8 | | - | | 0.49 | | 0.91 | |
| 46,766,338 | | 6.2E-11 | | 83.2 | | 8.4E-15 | | 111.7 | | - | | 0.61 | | 0.99 | |
| 47,219,973 | | 1.7E-12 | | 120.6 | | 1.2E-16 | | 155.8 | | rs776650813 | | 0.58 | | 0.95 | |
| 47,324,483 | | 1.7E-12 | | 121.6 | | 1.2E-16 | | 157.1 | | rs751157601 | | 0.55 | | 0.97 | |
| 47,483,198 | | 1.7E-10 | | 72.7 | | 2.9E-14 | | 99.2 | | - | | 0.88 | | 0.84 | |
| 47,639,508 | | 1.9E-12 | | 121.0 | | 1.3E-16 | | 156.4 | | rs765960243 | | 0.56 | | 0.97 | |
| 47,796,934 | | 2.0E-12 | | 119.6 | | 1.4E-16 | | 154.7 | | rs762024721 | | 0.55 | | 0.98 | |
| 47,946,476 | | 3.7E-10 | | 67.7 | | 7.1E-14 | | 94.6 | | rs777703284 | | 0.78 | | 0.99 | |

****

**Figure S1.** The age at AF discharge diagnosis stratified by sex and birth cohort. Discharge diagnoses were collected from 1982 to 2012. ‘M’ and ‘F’ indicate male and female cases. For example, the cohort ‘M 1910-1920’ includes male cases born between 1910 and 1920. The vertical lines starting from zero indicate the fraction of cases in percentages that belong to the cohort. The lower edges of the boxes, the thick horizontal lines inside the boxes, and the upper edges of the boxes indicate the lower quartile, median and upper quartile ages within each cohort. The whiskers extend to the most outlying ages, except if they are further away from the median than 1.5 times the inter quantile range (IQR). Ages further away from the median than 1.5 times the IQR are indicated with circles. The six homozygous carriers of c.234delC in *MYL4* with AF discharge diagnoses are indicated with red ‘x’s.



**Figure S2.** The results of the genome-wide association scans for early-onset AF under the multiplicative and recessive models (*P*<0.001). Autosomal chromosomes 1 through 22 and the X chromosome are shown in differing colors and the markers within a chromosome are plotted as a function of their build 38 order. The y-axis shows –log10 of the *P*-value for each marker. The horizontal line indicates our threshold for genome-wide significance significance (P < 0.05/25.5 million = 2.0 ×10-9).

References

1. Yang X, So WY, Kong AP, Ho CS, Lam CW, Stevens RJ, Lyu RR, Yin DD, Cockram CS, Tong PC, Wong V, Chan JC. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes care* 2007; 30(1):65-70.

2. Baum L, Wong KS, Ng HK, Tomlinson B, Rainer TH, Chan DK, Thomas GN, Chen X, Poon P, Cheung WS, Woo KS. Methylenetetrahydrofolate reductase gene A222V polymorphism and risk of ischemic stroke. *Clin Chem Lab Med* 2004; 42(12):1370-1376.

3. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, Besenbacher S, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Stacey SN, Frigge ML, Holm H, Saemundsdottir J, Helgadottir HT, Johannsdottir H, Sigfusson G, Thorgeirsson G, Sverrisson JT, Gretarsdottir S, Walters GB, Rafnar T, Thjodleifsson B, Bjornsson ES, Olafsson S, Thorarinsdottir H, Steingrimsdottir T, Gudmundsdottir TS, Theodors A, Jonasson JG, Sigurdsson A, Bjornsdottir G, Jonsson JJ, Thorarensen O, Ludvigsson P, Gudbjartsson H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Arnar DO, Magnusson OT, Kong A, Masson G, Thorsteinsdottir U, Helgason A, Sulem P, Stefansson K. Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics* 2015; 47(5):435-444.

4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25(14):1754-1760.

5. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20(9):1297-1303.

6. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2010; 43(5):491.

7. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ( 1118) ; iso-2; iso-3. *Fly (Austin)* 2012; 6(2):80-92.

8. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature genetics* 2008; 40(9):1068-1075.

9. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Parental origin of sequence variants associated with complex diseases. *Nature* 2009; 462(7275):868-874.

10. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; 39(7):906-913.

11. Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* 1987; 84(8):2363-2367.

12. Kruglyak L, Lander ES. Faster multipoint linkage analysis using Fourier transforms. *J Comput Biol* 1998; 5(1):1-7.

13. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, Gudjonsson SA, Frigge ML, Helgason A, Thorsteinsdottir U, Stefansson K. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 2010; 467(7319):1099-1103.

14. Lloyd-Jones DM, Wang TJ, Leip EP, Larson MG, Levy D, Vasan RS, D'Agostino RB, Massaro JM, Beiser A, Wolf PA, Benjamin EJ. Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. *Circulation* 2004; 110(9):1042-1046.

15. Arnar DO, Thorvaldsson S, Manolio TA, Thorgeirsson G, Kristjansson K, Hakonarson H, Stefansson K. Familial aggregation of atrial fibrillation in Iceland. *European heart journal* 2006; 27(6):708-712.

16. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. *Nat Genet* 2002; 31(3):241-247.