**Supplementary materials**

**The proper implementation of SNP analysis results: building a perfect human**

1. **Downloading the data.**The data of Mike was downloaded from Google Disk.
2. **Converting to VCF file.**Downloading and preparing of the plink program:  
   wget <http://s3.amazonaws.com/plink1-assets/plink_linux_x86_64_20201019.zip>  
   unzip plink\_linux\_x86\_64\_20201019.zip  
   Creating a vcf file:  
   ~/Progs/plink --23file SNP\_raw\_v4\_Full\_20170514175358.txt --recode vcf --out snps\_clean --output-chr MT --snps-only just-acgt
3. **Annotation**Downloading and preparing SnpEff/SnpSift:  
   wget <https://snpeff.blob.core.windows.net/versions/snpEff_latest_core.zip>  
   unzip snpEff\_latest\_core.zip  
   Variant annotation:  
   java -jar ./snpEff/snpEff.jar GRCh37.75 snps\_clean.vcf > snps\_snpeff.vcfDownloading ClinVar variants:  
   wget <https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/clinvar.vcf.gz>  
   Comparing our snps with ClinVar variants:  
   java -jar ./snpEff/SnpSift.jar annotate clinvar.vcf snps\_clean.vcf > snps\_clean\_snpsift\_clinvar.vcf  
     
   Another option with VEP:  
   Annotation in VEP (snps\_clean.vcf as input file):  
   <http://grch37.ensembl.org/Homo_sapiens/Tools/VEP>  
   <http://grch37.ensembl.org/Homo_sapiens/Tools/VEP/Results?tl=Kdfwus6hKT0hgo8O-7103142>  
   Downloaded the results and saved as vep\_out.txt  
   Extraction of the SNPs with risk and pathogenic SNPs:  
   awk '($32!="-") ' vep\_out.txt | grep risk\_factor | cut -f 1-3 | sort | uniq > clin\_sig.txt  
   awk '($32!="-") ' vep\_out.txt | grep pathogenic | cut -f 1-3 | sort | uniq > pathogenic.txt  
   Summarized information from SnpEff/SnpSift and VEP is shown in table 1.  
     
   Table 1 - SNPs associated with risk or referred to as pathogenic (according to VEP).

| ID | Pos | N | Consequence | Significance by ClinVar | CLNDN from SnpEff intersected with ClinVar | Frequency (GnomAD) |
| --- | --- | --- | --- | --- | --- | --- |
| rs1024611 | 17:32579788-32579788 | G,  0/1 | None | Pathogenic, risk factor​ | Spina\_bifida,\_susceptibility\_to|Mycobacterium\_tuberculosis,\_susceptibility\_to|Coronary\_artery\_disease,\_modifier\_of|Coronary\_artery\_disease,\_development\_of,\_in\_hiv | 0.28232 |
| rs1049296 | 3:133494354-133494354 | T, 0/1 | TF : Missense Variant | Benign | Alzheimer\_disease,\_susceptibility\_to|Transferrin\_variant\_c1/c2|Atransferrinemia | 0.131941 |
| rs1169288 | 12:121416650-121416650 | C, 0/1 | HNF1A : Missense Variant | Benign​ | SERUM\_HDL\_CHOLESTEROL\_LEVEL,\_MODIFIER\_OF|Maturity-onset\_diabetes\_of\_the\_young,\_type\_3|Insulin\_resistance,\_susceptibility\_to|not\_specified|not\_provided | 0.354886 |
| rs12150220 | 17:5485367-5485367 | T,  0/1 | NLRP1 : Missense Variant | risk factor​ | Vitiligo-associated\_multiple\_autoimmune\_disease\_susceptibility\_1 | 0.327519 |
| rs13266634 | 8:118184783-118184783 | T, 0/1 | SLC30A8 : Missense Variant | risk factor​ | Diabetes\_mellitus\_type\_2,\_susceptibility\_to | 0.265589 |
| rs1801197 | 7:93055753-93055753 | G, 0/1 | CALCR : Missense Variant | risk factor​ | Bone\_mineral\_density\_quantitative\_trait\_locus\_15 | 0.357403 |
| rs1801274 | 1:161479745-161479745 | G, 0/1 | FCGR2A : Missense Variant | drug response​ | Lupus\_nephritis,\_susceptibility\_to|Pseudomonas\_aeruginosa,\_susceptibility\_to\_chronic\_infection\_by,\_in\_cystic\_fibrosis|Malaria,\_severe,\_susceptibility\_to|not\_specified|trastuzumab\_response\_-\_Efficacy | 0.493229 |
| rs1801275 | 16:27374400-27374400 | G, 0/1 | IL4R : Missense Variant | risk factor​ | Atopy,\_susceptibility\_to | 0.343606 |
| rs1801394 | 5:7870973-7870973 | G, 0/1 | MTRR : Missense Variant  FASTKD3 : 2KB Upstream Variant | drug response​ | Down\_syndrome,\_susceptibility\_to|Neural\_tube\_defects,\_folate-sensitive,\_susceptibility\_to|Gastrointestinal\_stromal\_tumor|Homocystinuria-Megaloblastic\_anemia\_due\_to\_defect\_in\_cobalamin\_metabolism,\_cblE\_complementation\_type|Disorders\_of\_Intracellular\_Cobalamin\_Metabolism|not\_specified|methotrexate\_response\_-\_Toxicity/ADR | 0.462196 |
| rs1801968 | 9:132580901-132580901 | G, 0/1 | TOR1A : Missense Variant | Benign​ | Dystonia\_1,\_torsion,\_modifier\_of|Dystonia\_1|not\_specified | 0.106569 |
| rs2004640 | 7:128578301-128578301 | T, 0/1 | IRF5 : Splice Donor Variant | Pathogenic, risk factor​ | Rheumatoid\_arthritis|Systemic\_lupus\_erythematosus\_10 | 0.430568 (ALFA) |
| rs2073658 | 1:161010762-161010762 | T, 0/1 | USF1 : Intron Variant | risk factor​ | Hyperlipidemia,\_familial\_combined,\_susceptibility\_to | 0.225855 |
| rs2184026 | 9:101304348-101304348 | T | GABBR2 : Intron Variant | no interpretation for the single variant​ |  | 0.195324 |
| rs2239704 | 6:31540141-31540141 | C | LTA : Intron Variant  LOC100287329 : Intron Variant | risk factor​ |  | 0.354616 |
| rs2241880 | 2:234183368-234183368 | G, 0/1 | ATG16L1 : Missense Variant  SCARNA5 : 2KB Upstream Variant | Benign​ | Inflammatory\_bowel\_disease\_10,\_susceptibility\_to|not\_specified | 0.430706 |
| rs2281845 | 1:201081943-201081943 | T, 0/1 | CACNA1S : 2KB Upstream Variant | Benign | Thyrotoxic\_periodic\_paralysis,\_susceptibility\_to,\_1|not\_provided | 0.494271 |
| rs231775 | 2:204732714-204732714 | G, 0/1 | CTLA4 : Missense Variant | Benign​ | Hashimoto\_thyroiditis,\_susceptibility\_to|TYPE\_1\_DIABETES\_MELLITUS\_12,\_SUSCEPTIBILITY\_TO|Thyroid-associated\_orbitopathy,\_susceptibility\_to|Celiac\_disease\_3|Systemic\_lupus\_erythematosus,\_susceptibility\_to|not\_specified | 0.424275 |
| rs4402960 | 3:185511687-185511687 | T, 0/1 | IGF2BP2 : Intron Variant | risk factor​ | Diabetes\_mellitus\_type\_2,\_susceptibility\_to | 0.376215 |
| rs4880 | 6:160113872-160113872 | G, 0/1 | SOD2 : Missense Variant | drug response​ | SUPEROXIDE\_DISMUTASE\_2\_POLYMORPHISM|Microvascular\_complications\_of\_diabetes\_6|cyclophosphamide\_response\_-\_Efficacy | 0.469164 |
| rs4961 | 4:2906707-2906707 | T, 0/1 | ADD1 : Missense Variant | drug response​ | Hypertension,\_salt-sensitive\_essential,\_susceptibility\_to|furosemide\_and\_spironolactone\_response\_-\_Efficacy | 0.1699 |
| rs5174 | 1:53712727-53712727 | T, 0/1 | LRP8 : Missense Variant | risk factor​ | Myocardial\_infarction\_1 | 0.288519 |
| rs5186 | 3:148459988-148459988 | C, 0/1 | AGTR1 : 3 Prime UTR Variant | Benign​ | Hypertension,\_essential,\_susceptibility\_to|Renal\_dysplasia | 0.196517 |
| rs61747071 | 16:53720436-53720436 | T, 0/1 | RPGRIP1L : Missense Variant | Benign​ | Retinitis\_pigmentosa\_in\_ciliopathies,\_modifier\_of|Nephronophthisis|Joubert\_syndrome|Joubert\_syndrome\_7|Meckel\_syndrome,\_type\_5|Meckel-Gruber\_syndrome|Nephronophthisis\_8|not\_specified | 0.043491 |
| rs6265 | 11:27679916-27679916 | T, 0/1 | BDNF : Missense Variant  BDNF-AS : Non Coding Transcript Variant | Benign​ | Memory\_impairment,\_susceptibility\_to|not\_specified | 0.151869 |
| rs6280 | 3:113890815-113890815 | T, 0/1 | DRD3 : Missense Variant | Benign​ | Essential\_tremor,\_susceptibility\_to|Schizophrenia,\_susceptibility\_to|Hereditary\_essential\_tremor\_1|not\_specified | 0.439651 |
| rs6504649 | 17:48437456-48437456 | G, 0/1 | XYLT2 : Missense Variant | Benign​ | Spondyloocular\_syndrome,\_autosomal\_recessive|Pseudoxanthoma\_elasticum,\_modifier\_of\_severity\_of | 0.330802 |
| rs699 | 1:230845794-230845794 | G, 0/1 | AGT : Missense Variant | Benign​ | not\_specified|not\_provided | 0.424542 |
| rs763110 | 1:172627498-172627498 | T | FASLG : 2KB Upstream Variant | risk factor​ |  | 0.458616 |
| rs7794745 | 7:146489606-146489606 | T, 0/1 | CNTNAP2 : Intron Variant | risk factor​ | Autism\_15 | 0.49374 |
| rs909253 | 6:31540313-31540313 | G, 0/1 | LTA : Intron Variant  LOC100287329 : Intron Variant | risk factor​ | Myocardial\_infarction|Psoriatic\_arthritis,\_susceptibility\_to | 0.395429 |
| rs10151259 | 14:21790040-21790040 | T, 0/1 | RPGRIP1 : Missense Variant | Benign​ | Leber\_congenital\_amaurosis\_1|Cone-rod\_dystrophy\_13|Leber\_congenital\_amaurosis\_6|not\_specified|none\_provided|not\_provided | 0.227667 |
| rs1042503 | 12:103246700-103246700 | T, 0/1 | PAH : Synonymous Variant | Benign​ | Phenylketonuria|not\_specified|not\_provided;CLNDNINCL=Phenylketonuria | 0.223999 |
| rs11085825 | 19:13007458-13007458 | T, 0/1 | GCDH : Intron Variant | Benign​ | Glutaric\_aciduria,\_type\_1|not\_provided | 0.300492 |
| rs11558492 | 1:231408091-231408091 | G, 0/1 | GNPAT : Missense Variant | Benign​ | Rhizomelic\_chondrodysplasia\_punctata\_type\_2|not\_specified|not\_provided | 0.14603 |
| rs16879498 | 6:49580247-49580247 | T, 0/1 | RHAG : Missense Variant | Benign | Rh-null,\_regulator\_type | 0.038606 |
| rs1800435 | 9:116153891-116153891 | G, 0/1 | ALAD : Missense Variant | Benign​ | AMINOLEVULINATE\_DEHYDRATASE,\_ALAD\*1/ALAD\*2\_POLYMORPHISM|Porphobilinogen\_synthase\_deficiency | 0.056802 |
| rs2301612 | 9:136301982-136301982 | G, 0/1 | ADAMTS13 : Missense Variant | Benign | Upshaw-Schulman\_syndrome|not\_specified | 0.343915 |
| rs6339 | 1:156848946-156848946 | T, 0/1 | NTRK1 : Missense Variant | Benign | Familial\_medullary\_thyroid\_carcinoma|Hereditary\_insensitivity\_to\_pain\_with\_anhidrosis|not\_specified|not\_provided | 0.037198 |
| i5006568 | 2:71829924-71829924 | G,  0/1 | DYSF: Missense Variant | Benign | Autosomal\_recessive\_limb-girdle\_muscular\_dystrophy\_type\_2B|Qualitative\_or\_quantitative\_defects\_of\_dysferlin|Miyoshi\_muscular\_dystrophy\_1|not\_specified|DYSF-Related\_Disorders|not\_provided | No info in dbSNP |
| i6015729 | 1:156848918-156848918 | T,  0/1 | NTRK1: Missense Variant | Benign | Familial\_medullary\_thyroid\_carcinoma|Hereditary\_insensitivity\_to\_pain\_with\_anhidrosis|not\_specified|not\_provided | No info in dbSNP |
| i6058764 | 16:27356203-27356203 | G, 0/1 | IL4R: Missense Variant | Pathogenic,\_protective | Acquired\_immunodeficiency\_syndrome,\_slow\_progression\_to|Atopy,\_resistance\_to | No info in dbSNP |
| i6060296 | 19:13010520-13010520 | G,  0/1 | GCDH: Missense Variant | Benign | Glutaric\_acidemia|Glutaric\_aciduria,\_type\_1 | No info in dbSNP |

1. **MtDNA haplogroup**Determination of the haplogroup on <https://dna.jameslick.com/mthap/>:  
   mthap version 0.19b (2015-05-11); haplogroup data version PhyloTree Build 17 (2016-02-18) +mods  
   raw data source SNP\_raw\_v4\_Full\_20170514175358.txt (14MB)  
   Found 3270 markers at 3268 positions covering 19.7% of mtDNA.  
   Markers found:  
   **HVR2:** 152C 263G  
   **CR:** 750G 1438G 4769G 8860G  
   **HVR1:  
   Best match: H(T152C)**  
   **HVR2: 152C 263G  
   CR: 750G 1438G 4769G 8860G 15326G  
   HVR1: -  
   Marker path from rCRS to haplogroup H(T152C):  
   H2a2a1(rCRS) ⇨ 263G ⇨ H2a2a ⇨ 8860G 15326G ⇨ H2a2 ⇨ 750G ⇨ H2a ⇨ 4769G ⇨ H2 ⇨ 1438G ⇨ H ⇨ 152C ⇨ H(T152C)  
   Untested marker: 15326G** (which means testing service did not test this position).
2. **Y-chromosome haplogroup**Determination of the haplogroup on <https://ytree.morleydna.com/extractFromAutosomal>  
   MorleyDNA.com Y-SNP Terminal Subclade Predictor, using [**Experimental tree ( 26 July 2013)**](https://ytree.morleydna.com//experimental-phylogeny) and [**ISOGG's tree**](http://www.isogg.org/tree/index.html) **( 19 July 2013)** : **R1a1a**.

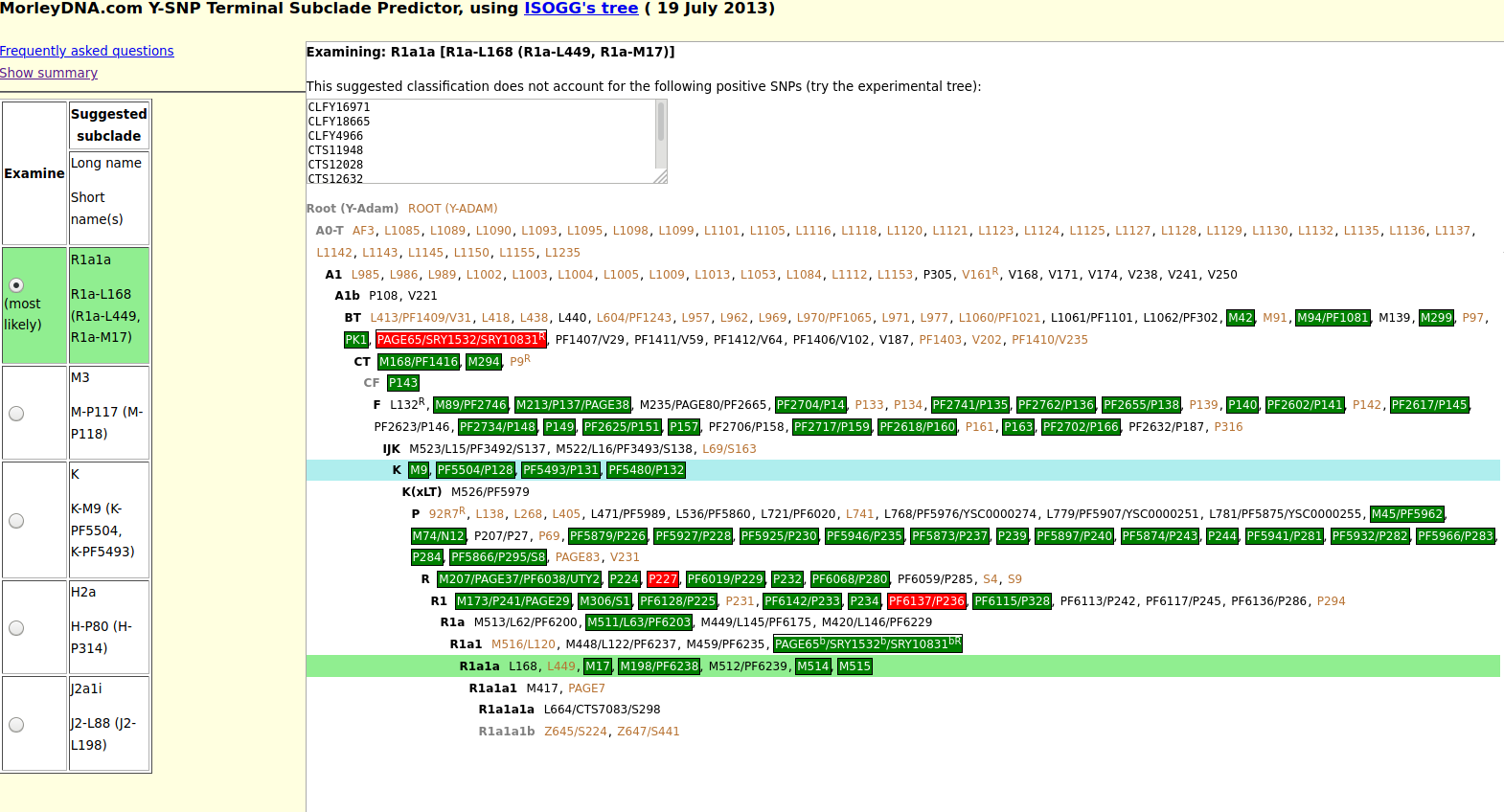


Figure 1. ISOGG tree for Y-DNA haplotype

This suggested classification does not account for the following positive SNPs:

CLFY16971

CLFY18665

CLFY4966

CTS11948

CTS12028

CTS3229

CTS4557

CTS5808

CTS5815

CTS6327

CTS8626

F1857

FB1947

FB988

FGC12056

FGC199

FGC24468

FGC38

FGC54

FGC77

FN1317

FN495

FN591

FN833

L1346

L486

L88

L882

M1185

M1188

M1189

M1219

M1226

M1239

M1272

M141

M2695

M288

M3639

M3683

M3698

M3707

M3735

M3749

M3771

M5705

M672

M696

M74

M9118

M9336

M96

N12

P102

P103

P117

P222

P239

P32

P38

P80

PAGES00026

PAGES00029

PAGES00037

PAGES00038

PAGES00081

PF1580

PF2587

PF2660

PF2669

PF2673

PF2679

PF2686

PF2692

PF2703

PF2720

PF2730

PF2771

PF2775

PF3558

PF3708

PF518

PF5505

PF5506

PF5861

PF5863

PF5870

PF5896

PF5901

PF5915

PF5930

PF5934

PF5951

PF5963

PF5977

PF6036

PF6050

PF6126

PF6147

PF6234

PF6240

PF6276

PF:14231292(A|G)

PF:14432928(A|G)

PF:14639427(C|A)

PF:14692227(C|T)

PF:14958218(C|T)

PF:14993358(G|A)

PF:15027529(T|G)

PF:15086183(C|G)

PF:16492547(T|C)

PF:16971648(G|T)

PF:18026855(A|G)

PF:18109555(A|C)

PF:18656508(C|G)

PF:18865298(T|C)

PF:18914441(T|C)

PF:19179540(A|G)

PF:21409706(C|G)

PF:21410840(T|G)

PF:22190371(G|A)

PF:22739367(C|T)

PF:23476936(T|G)

PF:28590278(A|G)

PF:28733101(G|C)

PF:2887824(A|C)

S118

S24307

S26596

S26597

S26598

S27130

S27150

S27154

S27177

S27244

S27310

S27390

S27470

S27500

S27554

S27700

U250

YSC0000222

YSC0001265

Z11996

Z8133

Z8143

Z8146