Effects of mitochondrial haplogroups and variants on anthracycline associated cardiomyopathy

Jeremy Leipzig & Marijana Vukovic

3/7/2017

Contents

MitoChip and MiSeq analysis **1**

Population structure **2**

Initial findings . . . 2

Autosomal population background . . . 6

Left ventricular systolic dysfunction **7**

Distribution . . . 7

Macrohaplogroup analysis . . . 9

Haplogroup (letter-number-letter) analysis . . . 12

LVSD vs superpopulation and autosomal population . . . 16

Micro haplogroup . . . 19

Individual variants on LVSD **19**

Potentially deleterious mutations in cohort . . . 19

High LVSD variants . . . 20

mtGWAS . . . 23

GLM using both haplogroup and non-haplogroup defining variants . . . 23

GLM - non-haplogroup defining variants only . . . 23

GLMs . . . 24

Haplogroups of interest **25**

H . . . 25

T . . . 29

MitoChip and MiSeq analysis

97 subjects were run through the Illumina MitoChip v2. Calls were processed using ReseqChip to obtain consensus sequences. These were processed through Mitomaster to obtain variants and Haplogroup calls (via Haplogrep).

96 subjects were sequenced on an Illumina MiSeq and aligned by CAG using bwa-mem to the g1k_v37 reference. Samtools/bcftools was used to create a consensus sequence for reads mapping to MT. These sequences were also processed through Mitomaster.

7 subjects were run on both platforms.

Population structure

Initial findings

Concordance of sequence and microarray data

Sequence and microarray microhaplogroups were identical between the 7 samples run on both platforms

Duplicate microarray samples were removed for the remainder for the analysis.

Regno	cnt	samples	sources	mhaps
821630	2	821630, Mito_AML1031_PT1_F09_821630	seq, chip	U5a1a1d, U5a1a1d
822655	2	822655, Mito_AML1031_PT1_B02_822655	seq, chip	C, C
823157	2	823157, Mito_AML1031_PT1_H09_823157	seq, chip	K1a1b1a, K1a1b1a
825090	2	825090, Mito_AML1031_PT1_C10_825090	seq, chip	H1, H1
828528	2	828528, Mito_AML1031_PT1_E03_828528	seq, chip	H5, H5
829368	2	829368, Mito_AML1031_PT1_E07_829368	seq, chip	C, C
830330	2	830330, Mito_AML1031_PT1_C11_830330	seq, chip	H3a1, H3a1

Loci

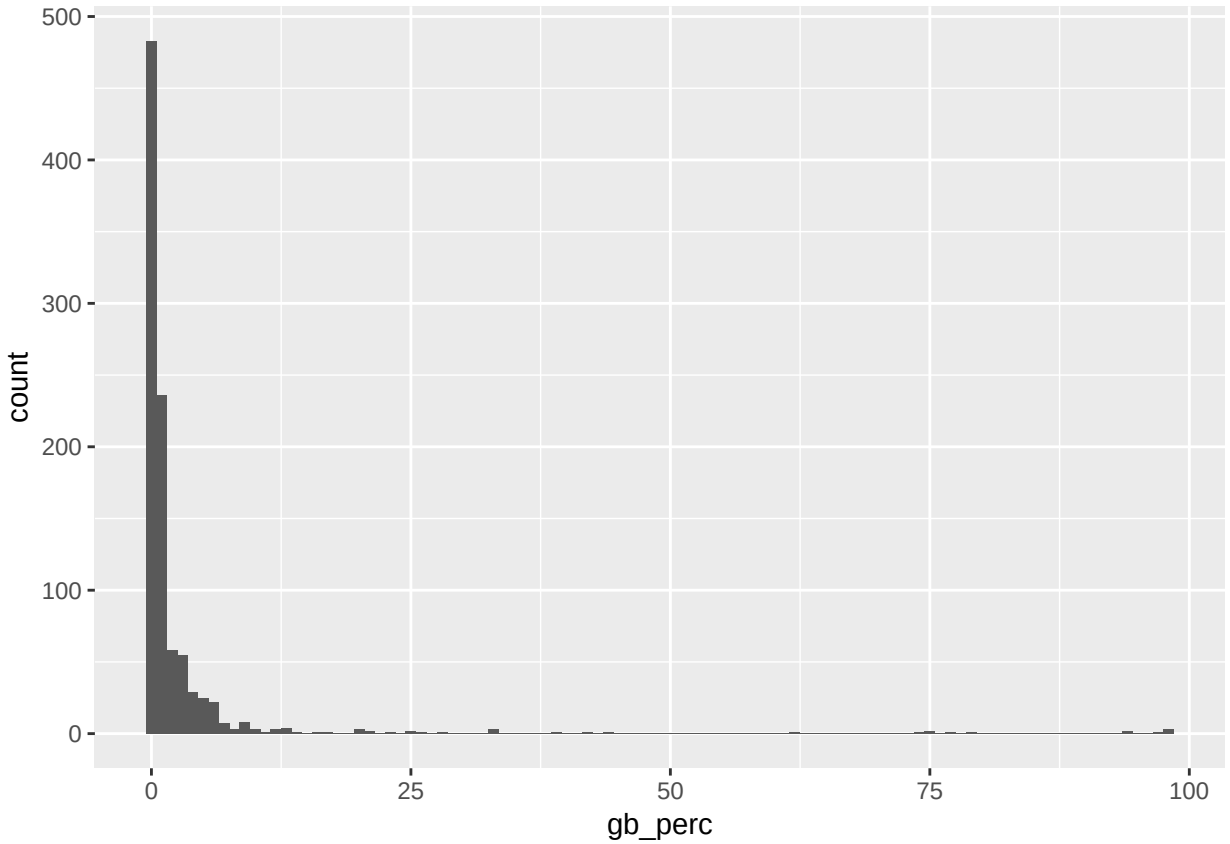
A total of 953 loci witnessed variants in at least one subject.

A word of caution

rCRS is a somewhat arbitrary reference with microhaplogroup H2a2a1. Variants from the rCRS do not imply deleteriousness, and in many cases reference alleles are risk factors for disease or other non-optimal outcomes. It is just as important to weight “protective variants” equally.

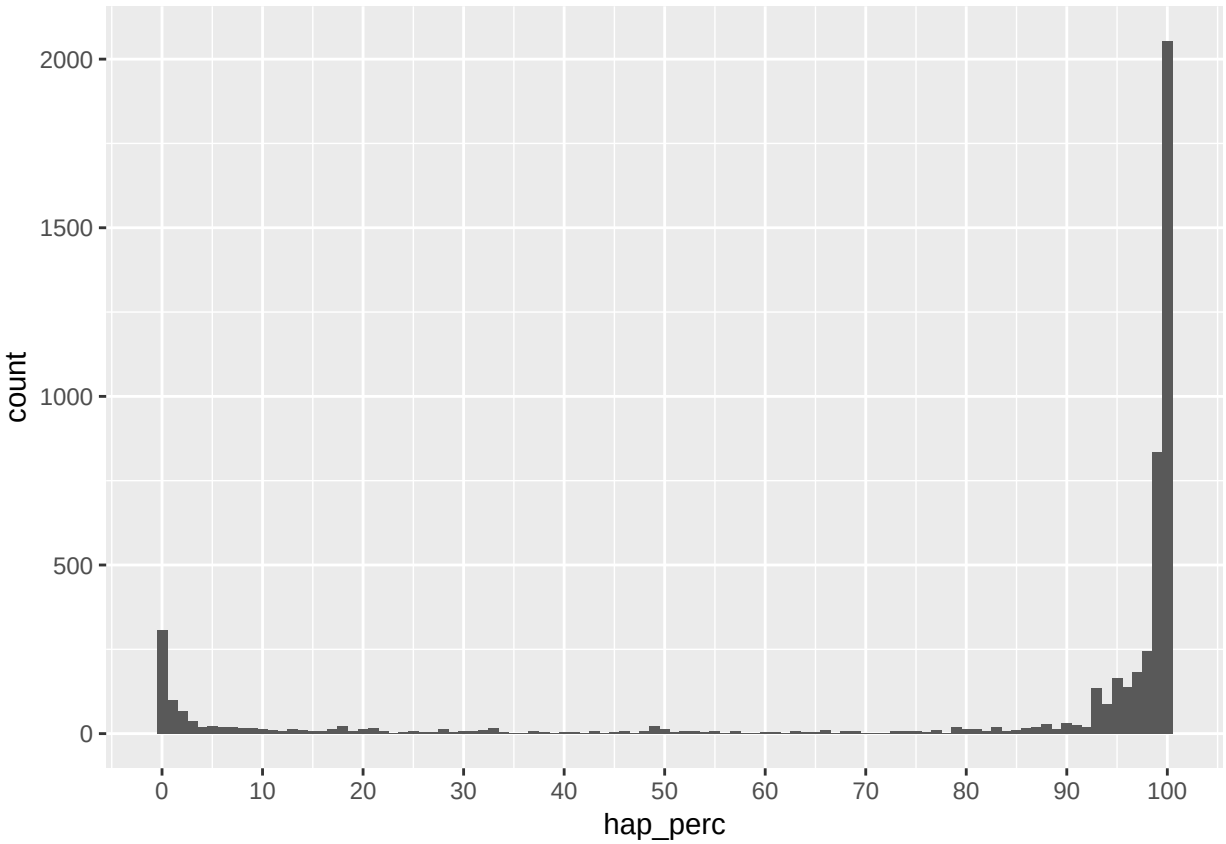
Variants by genbank frequency

Mitomaster computes frequency of each variant in its corpus of 30,589 GenBank sequences.



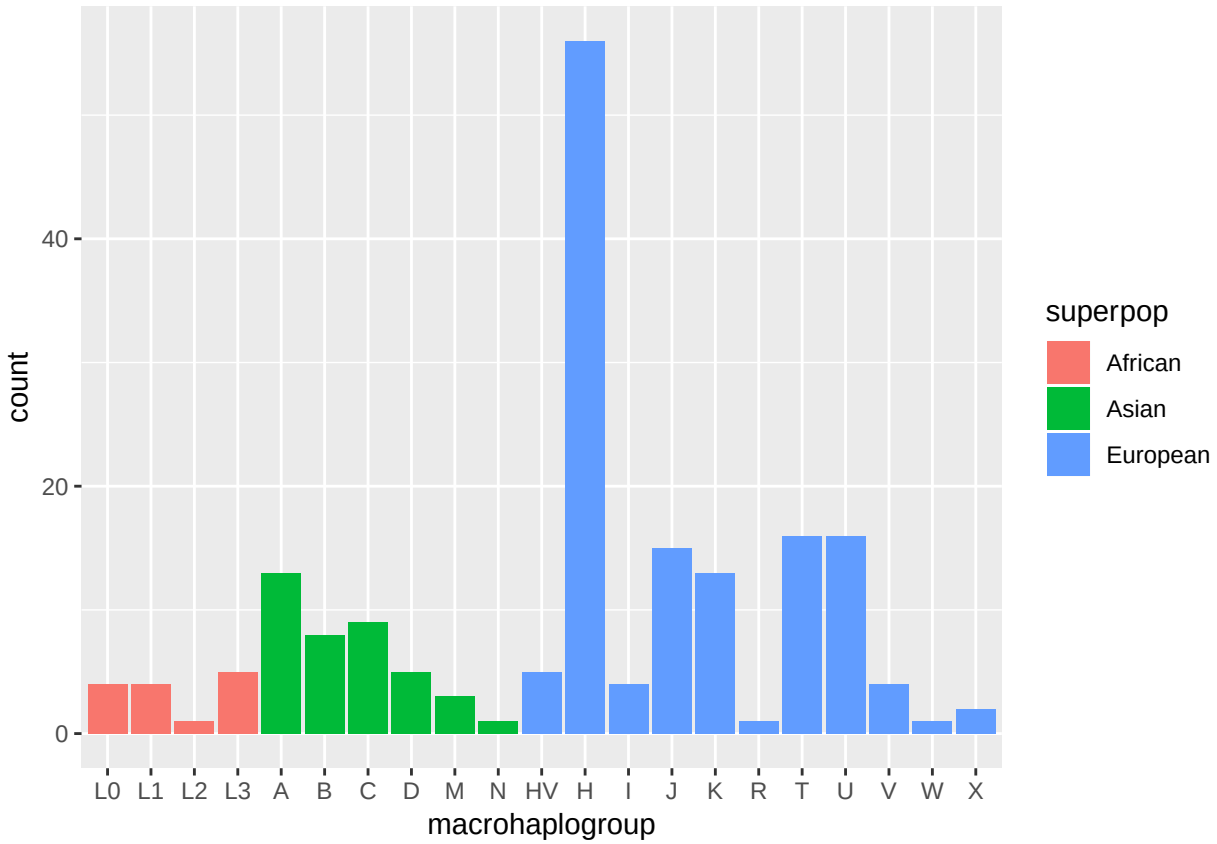
Variants by haplogroup frequency

Mitomaster also computes frequency within minor haplogroups. Frequencies near 100% are likely haplogroup-defining variants



625 of the variants are very rare (<1%) could be Mitochip artefacts, especially if they appear in more than one subject. Others are more common than not, and reflect the oddity of rCRS as a reference.

The breakdown of macro haplogroups



Why are there so many Asians when that is not a self-reporting group? Here is the breakdown of Asian haplogroups:

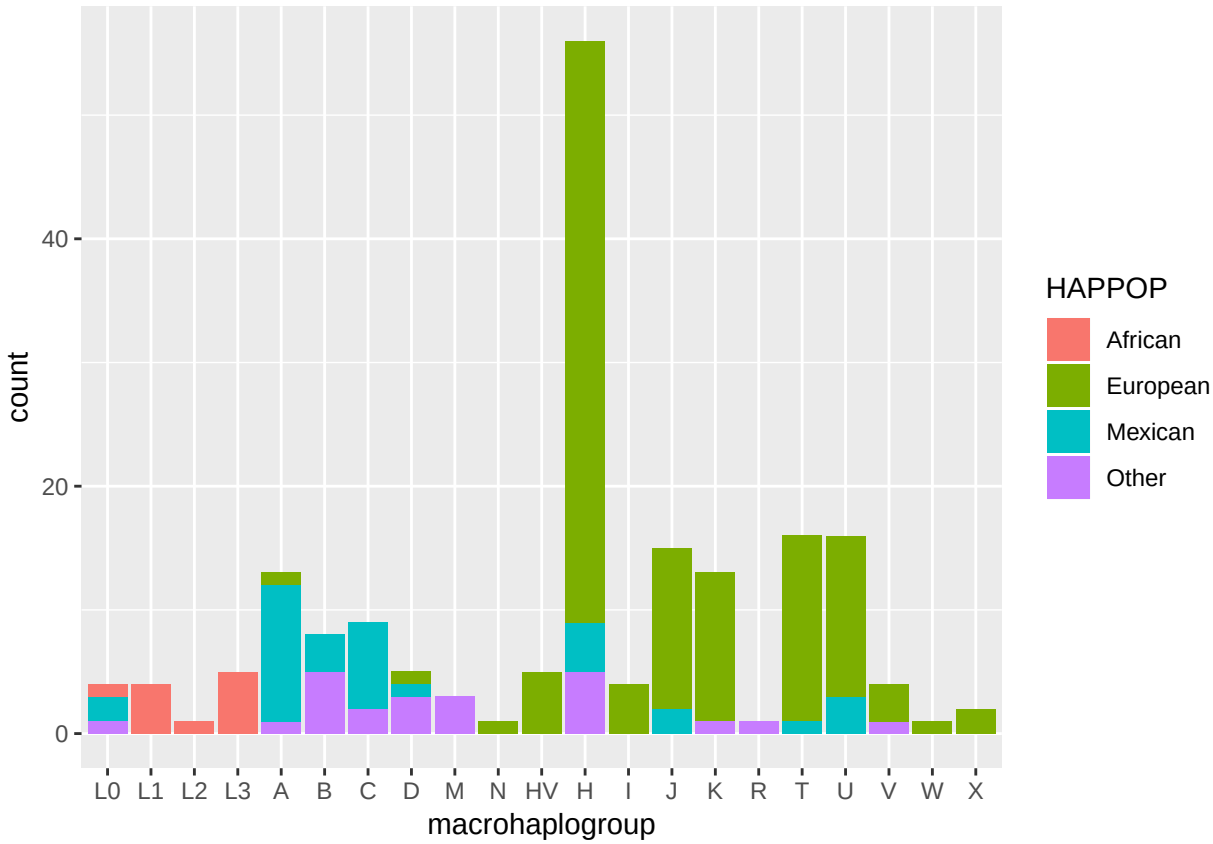
macrohaplogroup	count
A	13
B	8
C	9
D	5
M	3
N	1

What is the concordance with SVM PCA-component classified by Hapmap populations

These calls were made using the SVM's trained on HapMap and the following breakdown:

```
EUR = CEU + TSI
AFR = ASW + LWK + MKK + YRI
MEX = MXL
```

See https://github.research.chop.edu/vujkovicm/aplenc_gwas/blob/master/R/pop_strat.R for the code.



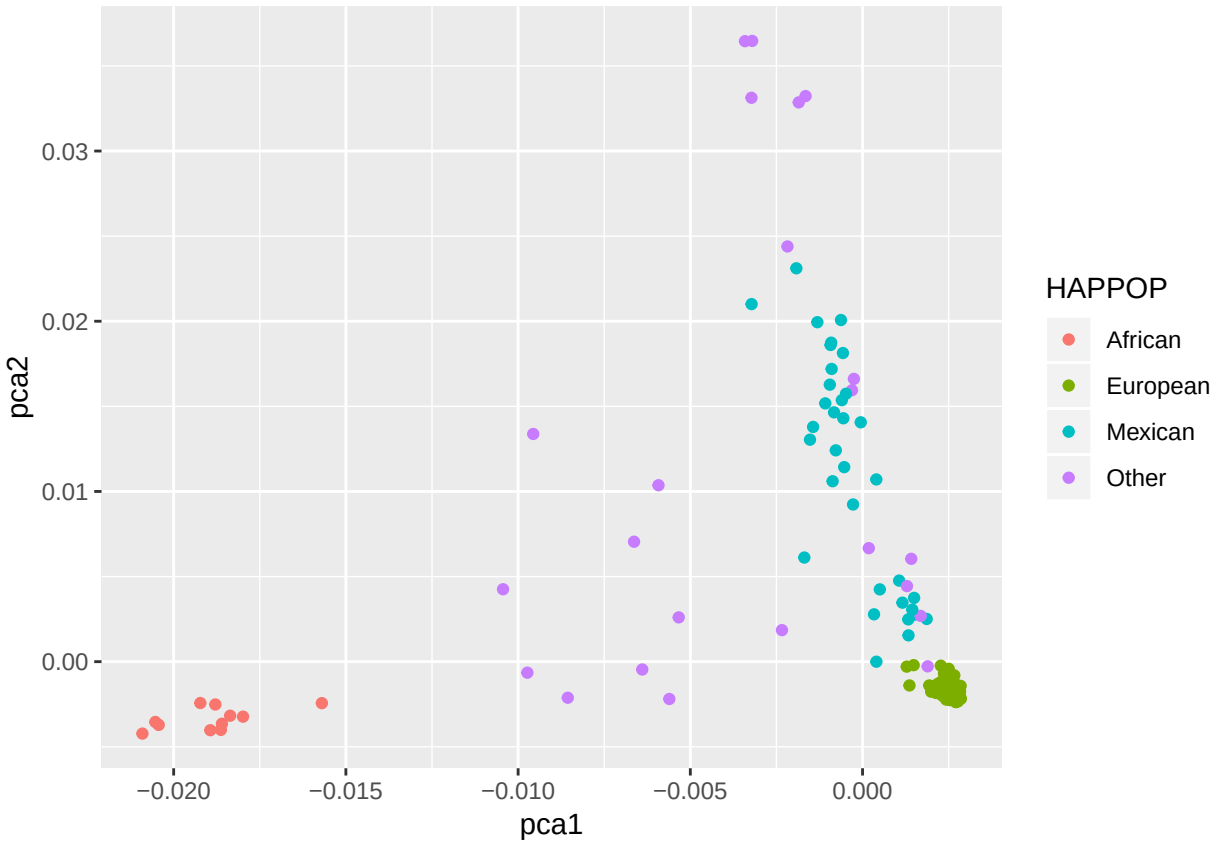
The prevalance of A haplogroups associated with Amerindian populations in Mexico is historically consistent. See http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3051415/

Autosomal population background

Classification

Three principal components from population-defining SNPs were provided as courtesy by CAG.

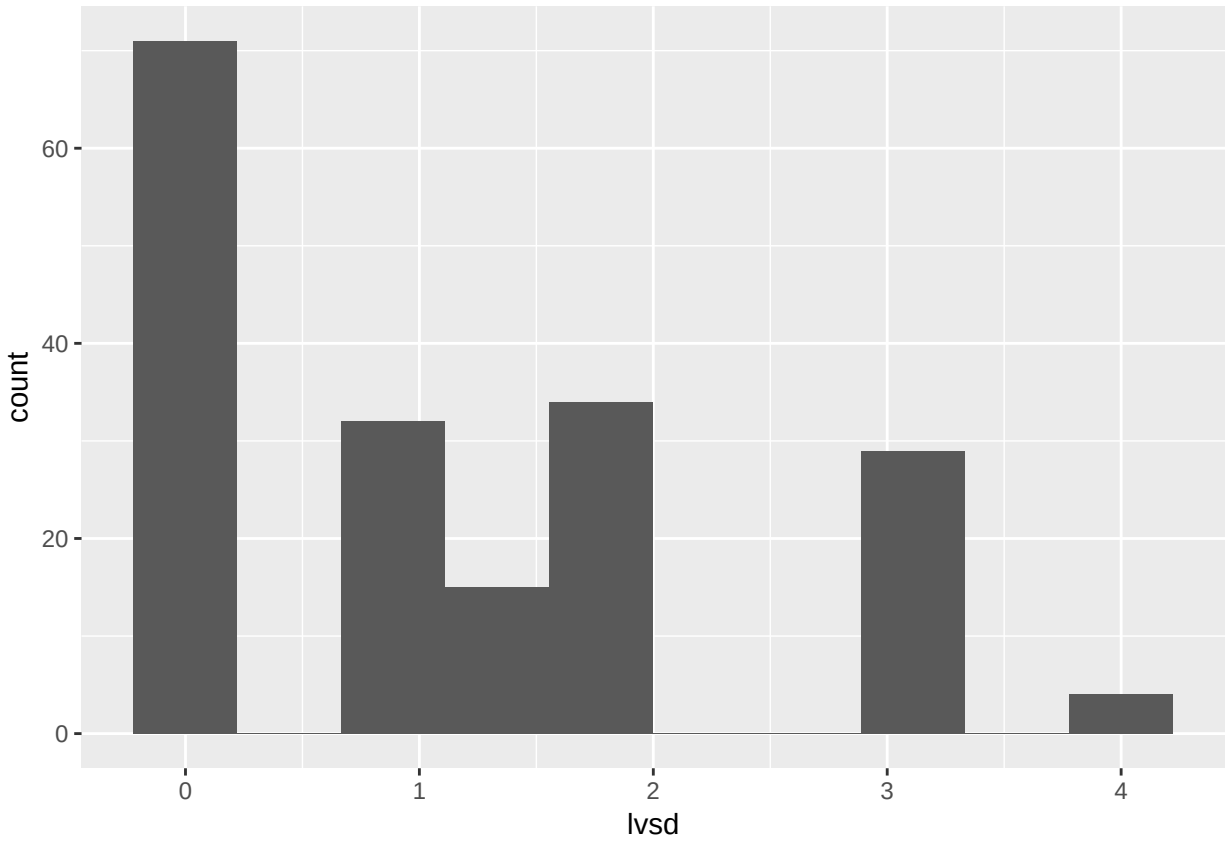
The first two PCs against self reported ethnicity:



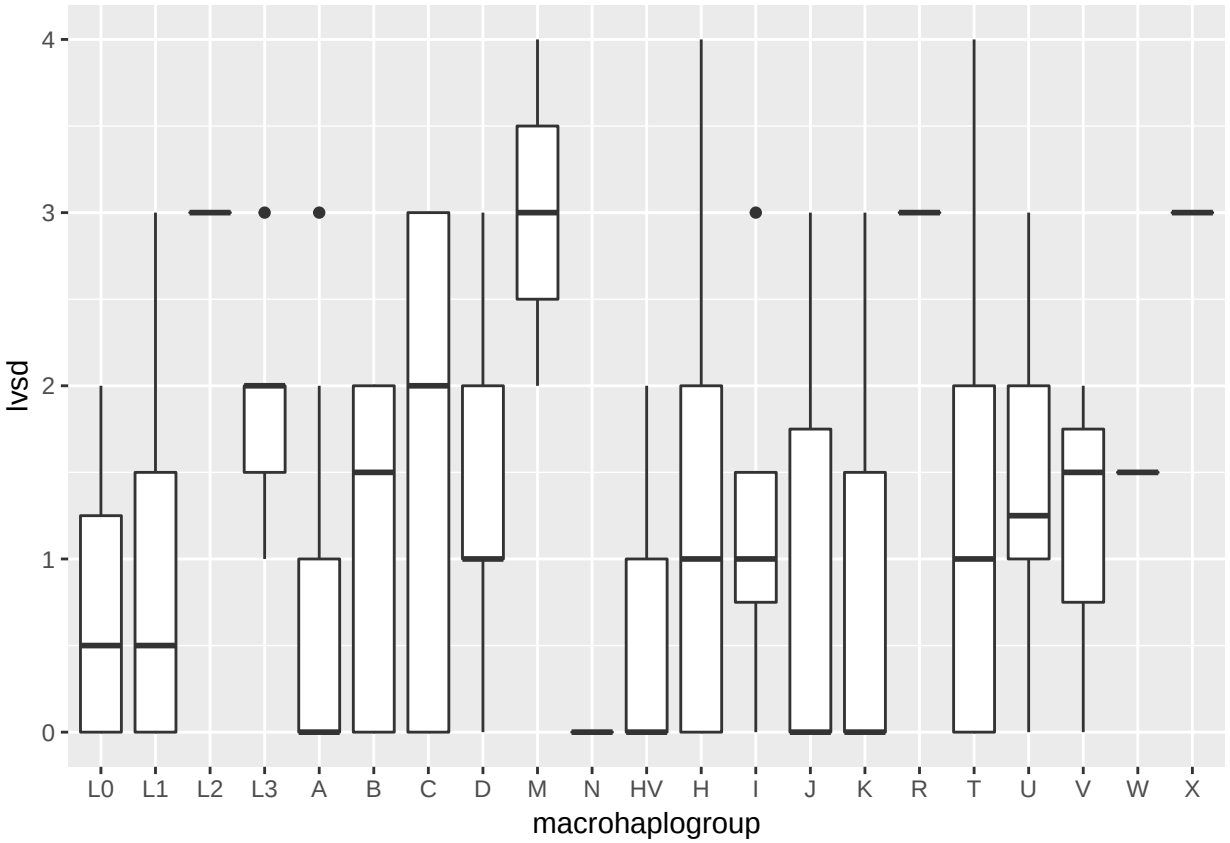
Left ventricular systolic dysfunction

Distribution

Cannot justify using parametric tests.


60
40
20
0
count
0
1
2
3
4
lvsd

Macrohaplogroup analysis



macrohaplogroup	cnt	mean_lvsd
N	1	0.0000000
HV	5	0.6000000
K	13	0.7307692
L0	4	0.7500000
A	13	0.7692308
J	15	0.9000000
L1	4	1.0000000
B	8	1.1250000
T	16	1.1875000
H	56	1.2410714
I	4	1.2500000
D	5	1.4000000
U	16	1.4062500
W	1	1.5000000
C	9	1.6666667
L3	5	1.9000000
L2	1	3.0000000
M	3	3.0000000
R	1	3.0000000
X	2	3.0000000
V	4	NA

Linear model

```
##
## Call:
## lm(formula = lvsd ~ macrohaplogroup, data = cpdf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6667 -0.9000 -0.1875  0.7589  2.8125
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.75000    0.56969   1.317   0.1898
## macrohaplogroupL1  0.25000    0.80566   0.310   0.7567
## macrohaplogroupL2  2.25000    1.27386   1.766   0.0792 .
## macrohaplogroupL3  1.15000    0.76431   1.505   0.1343
## macrohaplogroupA   0.01923    0.65146   0.030   0.9765
## macrohaplogroupB   0.37500    0.69772   0.537   0.5917
## macrohaplogroupC   0.91667    0.68468   1.339   0.1825
## macrohaplogroupD   0.65000    0.76431   0.850   0.3963
## macrohaplogroupM   2.25000    0.87021   2.586   0.0106 *
## macrohaplogroupN  -0.75000    1.27386  -0.589   0.5568
## macrohaplogroupHV -0.15000    0.76431  -0.196   0.8447
## macrohaplogroupH   0.49107    0.58968   0.833   0.4062
## macrohaplogroupI   0.50000    0.80566   0.621   0.5357
## macrohaplogroupJ   0.15000    0.64116   0.234   0.8153
## macrohaplogroupK  -0.01923    0.65146  -0.030   0.9765
## macrohaplogroupR   2.25000    1.27386   1.766   0.0792 .
## macrohaplogroupT   0.43750    0.63693   0.687   0.4931
## macrohaplogroupU   0.65625    0.63693   1.030   0.3044
## macrohaplogroupV   0.41667    0.87021   0.479   0.6327
## macrohaplogroupW   0.75000    1.27386   0.589   0.5568
## macrohaplogroupX   2.25000    0.98672   2.280   0.0239 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.139 on 164 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1548, Adjusted R-squared:  0.05169
## F-statistic: 1.502 on 20 and 164 DF,  p-value: 0.08673
```

Do the PCA's based on the microarrays help?

```
##
## Call:
## lm(formula = lvsd ~ macrohaplogroup + pca1 + pca2 + pca3, data = cpdf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6451 -0.8854 -0.1832  0.7770  2.7870
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)         0.58820    0.64384   0.914   0.3623
## macrohaplogroupL1   0.18828    0.93498   0.201   0.8407
## macrohaplogroupL2   2.16806    1.37375   1.578   0.1165
## macrohaplogroupL3   1.07574    0.92999   1.157   0.2491
## macrohaplogroupA    0.06747    0.70701   0.095   0.9241
## macrohaplogroupB    0.23737    0.74730   0.318   0.7512
## macrohaplogroupC    0.87464    0.75315   1.161   0.2472
## macrohaplogroupD    0.50636    0.87655   0.578   0.5643
## macrohaplogroupM    2.09012    0.94927   2.202   0.0291 *
## macrohaplogroupN   -0.51965    1.34086  -0.388   0.6989
## macrohaplogroupHV   0.06500    0.85661   0.076   0.9396
## macrohaplogroupH    0.69502    0.69466   1.001   0.3186
## macrohaplogroupI    0.72554    0.90056   0.806   0.4216
## macrohaplogroupJ    0.35143    0.74399   0.472   0.6373
## macrohaplogroupK    0.19255    0.75819   0.254   0.7998
## macrohaplogroupR    2.30424    1.32223   1.743   0.0833 .
## macrohaplogroupT    0.65549    0.74684   0.878   0.3814
## macrohaplogroupU    0.85483    0.73829   1.158   0.2486
## macrohaplogroupV    0.58175    0.91903   0.633   0.5276
## macrohaplogroupW    0.96474    1.33748   0.721   0.4718
## macrohaplogroupX    2.47013    1.06426   2.321   0.0215 *
## pca1              -15.04692   38.90261  -0.387   0.6994
## pca2               15.18466   17.28941   0.878   0.3811
## pca3               -2.44772   12.08871  -0.202   0.8398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.147 on 161 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1597, Adjusted R-squared:  0.03969
## F-statistic: 1.331 on 23 and 161 DF,  p-value: 0.1553
```

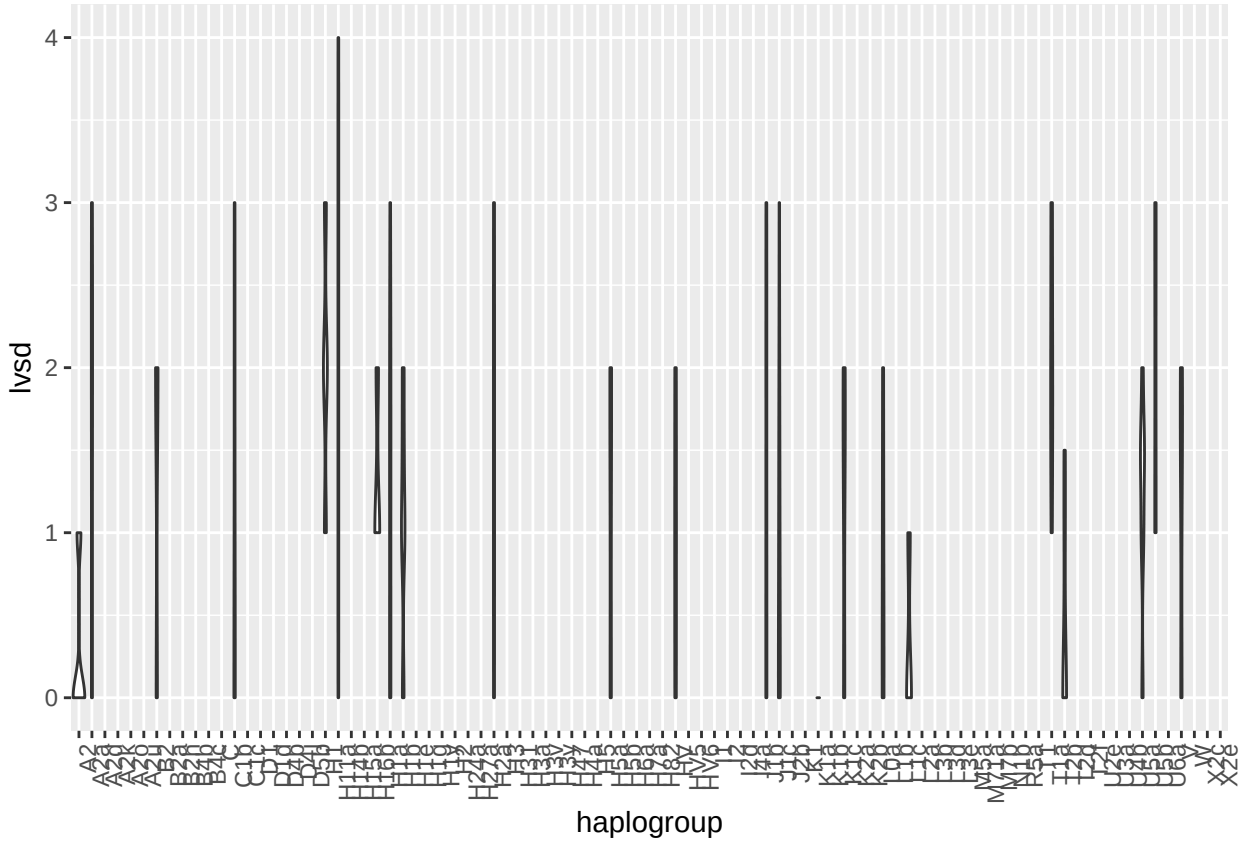
Nope.

All-by-all haplogroups kruskal-wallis

```
##
##  Kruskal-Wallis rank sum test
##
## data:  lvsd by macrohaplogroup
## Kruskal-Wallis chi-squared = 26.875, df = 20, p-value = 0.1388
```

Due to high variance, no macro haplogroup is significantly different

Haplogroup (letter-number-letter) analysis



haplogroup	mean_lvsd
A2d	0.0000000
A2k	0.0000000
B2h	0.0000000
D1	0.0000000
H14b	0.0000000
H1g	0.0000000
H2	0.0000000
H27a	0.0000000
H31	0.0000000
H3v	0.0000000
H3y	0.0000000
H5b	0.0000000
H7a	0.0000000
H82	0.0000000
HV5	0.0000000
HV6	0.0000000
I1	0.0000000
J2b	0.0000000
K1	0.0000000
K1a	0.0000000
K2a	0.0000000
N1b	0.0000000
T2d	0.0000000

haplogroup	mean_lvsd
T2f	0.0000000
U4b	0.0000000
A2	0.2500000
L1c	0.3333333
T2b	0.4166667
H1e	0.5000000
L0a	0.7500000
J1c	0.8500000
A2o	1.0000000
A2u	1.0000000
B2a	1.0000000
B4c	1.0000000
C1b	1.0000000
D1d	1.0000000
D5b	1.0000000
H47	1.0000000
H4a	1.0000000
HV	1.0000000
I2d	1.0000000
I4a	1.0000000
U2e	1.0000000
H1b	1.1000000
H5a	1.1000000
K1c	1.1666667
H1a	1.1875000
U5a	1.2142857
J1b	1.2500000
L3d	1.2500000
B2	1.3333333
H16b	1.3333333
A2a	1.4000000
H1y	1.5000000
H2a	1.5000000
U6a	1.5000000
W	1.5000000
B4b	2.0000000
D4b	2.0000000
H1	2.0000000
H11a	2.0000000
H6a	2.0000000
K1b	2.0000000
K2b	2.0000000
L3b	2.0000000
M71a	2.0000000
U3a	2.0000000
U5b	2.0000000
T1a	2.0833333
C	2.5000000
C1c	2.5000000
L3e	2.5000000
D4j	3.0000000
H15a	3.0000000

haplogroup	mean_lvsd
H24a	3.0000000
H3	3.0000000
H5	3.0000000
I2	3.0000000
L1b	3.0000000
L2a	3.0000000
M7b	3.0000000
R5a	3.0000000
X2c	3.0000000
X2e	3.0000000
H3a	4.0000000
M5a	4.0000000
T1	4.0000000
V	NA

Linear model

```
##
## Call:
## lm(formula = lvsd ~ haplogroup, data = cpdf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0000 -0.3333  0.0000  0.3333  2.1500
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.25000    0.51069   0.490  0.62558
## haplogroupA2a   1.15000    0.68516   1.678  0.09651 .
## haplogroupA2d  -0.25000    1.14193  -0.219  0.82717
## haplogroupA2k  -0.25000    1.14193  -0.219  0.82717
## haplogroupA2o   0.75000    1.14193   0.657  0.51289
## haplogroupA2u   0.75000    1.14193   0.657  0.51289
## haplogroupB2    1.08333    0.78009   1.389  0.16813
## haplogroupB2a   0.75000    0.88453   0.848  0.39860
## haplogroupB2h  -0.25000    1.14193  -0.219  0.82717
## haplogroupB4b   1.75000    1.14193   1.532  0.12869
## haplogroupB4c   0.75000    1.14193   0.657  0.51289
## haplogroupC     2.25000    0.88453   2.544  0.01256 *
## haplogroupC1b   0.75000    0.68516   1.095  0.27641
## haplogroupC1c   2.25000    0.88453   2.544  0.01256 *
## haplogroupD1   -0.25000    1.14193  -0.219  0.82717
## haplogroupD1d   0.75000    1.14193   0.657  0.51289
## haplogroupD4b   1.75000    1.14193   1.532  0.12869
## haplogroupD4j   2.75000    1.14193   2.408  0.01794 *
## haplogroupD5b   0.75000    1.14193   0.657  0.51289
## haplogroupH1    1.75000    0.72222   2.423  0.01726 *
## haplogroupH11a  1.75000    0.72222   2.423  0.01726 *
## haplogroupH14b -0.25000    1.14193  -0.219  0.82717
## haplogroupH15a  2.75000    1.14193   2.408  0.01794 *
## haplogroupH16b  1.08333    0.78009   1.389  0.16813
```

```
## haplogroupH1a   0.93750    0.62546   1.499  0.13718
## haplogroupH1b   0.85000    0.68516   1.241  0.21778
## haplogroupH1e   0.25000    0.88453   0.283  0.77807
## haplogroupH1g  -0.25000    0.88453  -0.283  0.77807
## haplogroupH1y   1.25000    1.14193   1.095  0.27641
## haplogroupH2   -0.25000    1.14193  -0.219  0.82717
## haplogroupH24a  2.75000    1.14193   2.408  0.01794 *
## haplogroupH27a -0.25000    1.14193  -0.219  0.82717
## haplogroupH2a   1.25000    0.78009   1.602  0.11236
## haplogroupH3    2.75000    1.14193   2.408  0.01794 *
## haplogroupH31  -0.25000    0.88453  -0.283  0.77807
## haplogroupH3a   3.75000    1.14193   3.284  0.00143 **
## haplogroupH3v  -0.25000    1.14193  -0.219  0.82717
## haplogroupH3y  -0.25000    1.14193  -0.219  0.82717
## haplogroupH47   0.75000    1.14193   0.657  0.51289
## haplogroupH4a   0.75000    1.14193   0.657  0.51289
## haplogroupH5    2.75000    1.14193   2.408  0.01794 *
## haplogroupH5a   0.85000    0.68516   1.241  0.21778
## haplogroupH5b  -0.25000    1.14193  -0.219  0.82717
## haplogroupH6a   1.75000    0.88453   1.978  0.05074 .
## haplogroupH7a  -0.25000    1.14193  -0.219  0.82717
## haplogroupH82  -0.25000    1.14193  -0.219  0.82717
## haplogroupHV    0.75000    0.78009   0.961  0.33875
## haplogroupHV5  -0.25000    1.14193  -0.219  0.82717
## haplogroupHV6  -0.25000    1.14193  -0.219  0.82717
## haplogroupI1   -0.25000    1.14193  -0.219  0.82717
## haplogroupI2    2.75000    1.14193   2.408  0.01794 *
## haplogroupI2d   0.75000    1.14193   0.657  0.51289
## haplogroupI4a   0.75000    1.14193   0.657  0.51289
## haplogroupJ1b   1.00000    0.72222   1.385  0.16938
## haplogroupJ1c   0.60000    0.60425   0.993  0.32322
## haplogroupJ2b  -0.25000    1.14193  -0.219  0.82717
## haplogroupK1   -0.25000    0.88453  -0.283  0.77807
## haplogroupK1a  -0.25000    0.78009  -0.320  0.74930
## haplogroupK1b   1.75000    1.14193   1.532  0.12869
## haplogroupK1c   0.91667    0.78009   1.175  0.24287
## haplogroupK2a  -0.25000    0.88453  -0.283  0.77807
## haplogroupK2b   1.75000    0.88453   1.978  0.05074 .
## haplogroupL0a   0.50000    0.72222   0.692  0.49041
## haplogroupL1b   2.75000    1.14193   2.408  0.01794 *
## haplogroupL1c   0.08333    0.78009   0.107  0.91515
## haplogroupL2a   2.75000    1.14193   2.408  0.01794 *
## haplogroupL3b   1.75000    1.14193   1.532  0.12869
## haplogroupL3d   1.00000    0.88453   1.131  0.26107
## haplogroupL3e   2.25000    0.88453   2.544  0.01256 *
## haplogroupM5a   3.75000    1.14193   3.284  0.00143 **
## haplogroupM71a  1.75000    1.14193   1.532  0.12869
## haplogroupM7b   2.75000    1.14193   2.408  0.01794 *
## haplogroupN1b  -0.25000    1.14193  -0.219  0.82717
## haplogroupR5a   2.75000    1.14193   2.408  0.01794 *
## haplogroupT1    3.75000    1.14193   3.284  0.00143 **
## haplogroupT1a   1.83333    0.65929   2.781  0.00653 **
## haplogroupT2b   0.16667    0.65929   0.253  0.80097
## haplogroupT2d  -0.25000    1.14193  -0.219  0.82717
```

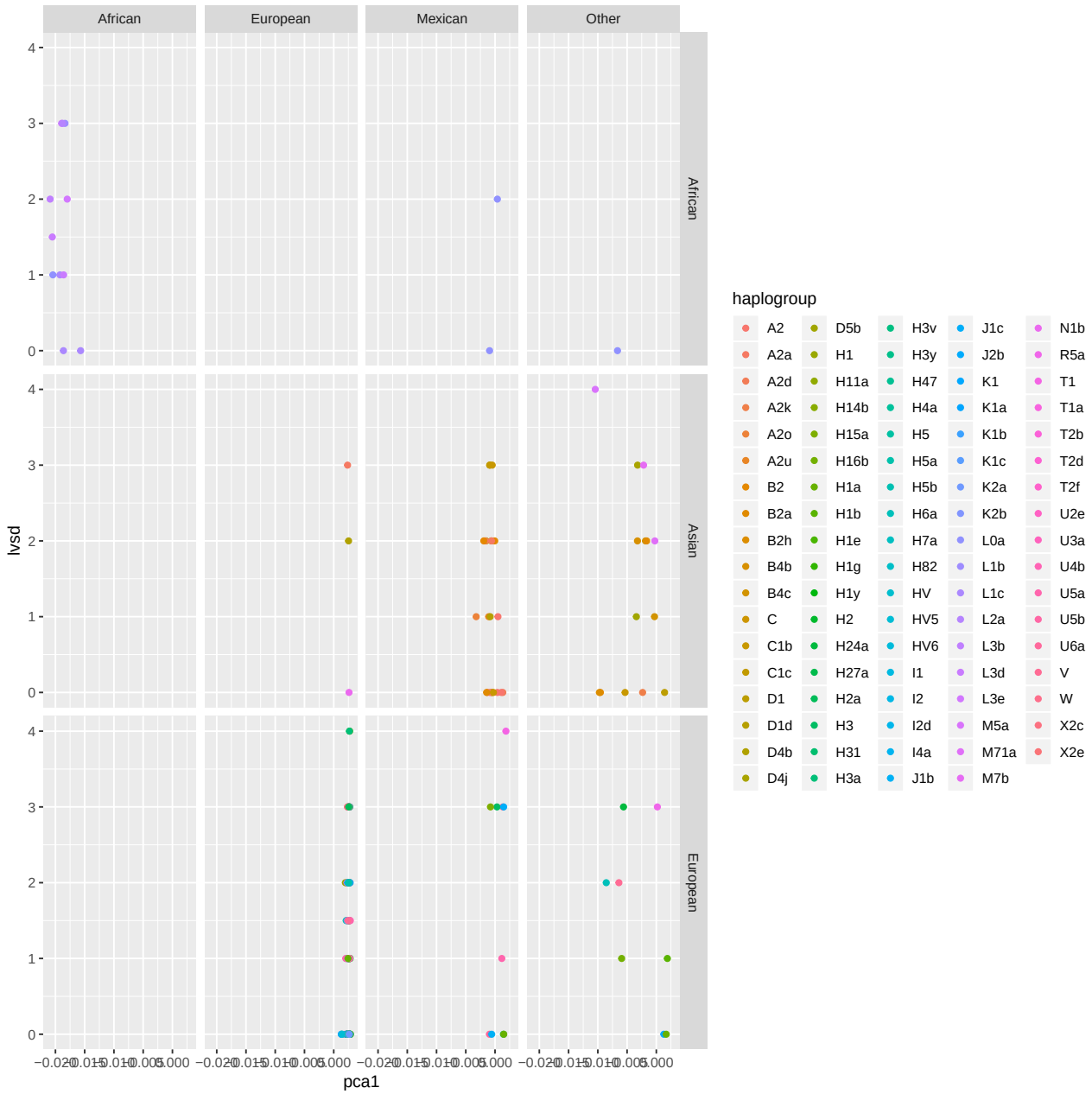
```
## haplogroupT2f  -0.25000    0.88453  -0.283  0.77807
## haplogroupU2e   0.75000    1.14193   0.657  0.51289
## haplogroupU3a   1.75000    1.14193   1.532  0.12869
## haplogroupU4b  -0.25000    1.14193  -0.219  0.82717
## haplogroupU5a   0.96429    0.64018   1.506  0.13528
## haplogroupU5b   1.75000    0.72222   2.423  0.01726 *
## haplogroupU6a   1.25000    0.88453   1.413  0.16084
## haplogroupV     0.91667    0.78009   1.175  0.24287
## haplogroupW     1.25000    1.14193   1.095  0.27641
## haplogroupX2c   2.75000    1.14193   2.408  0.01794 *
## haplogroupX2e   2.75000    1.14193   2.408  0.01794 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.021 on 96 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.2379
## F-statistic: 1.653 on 88 and 96 DF,  p-value: 0.008189
```

All-by-all haplogroups kruskal-wallis

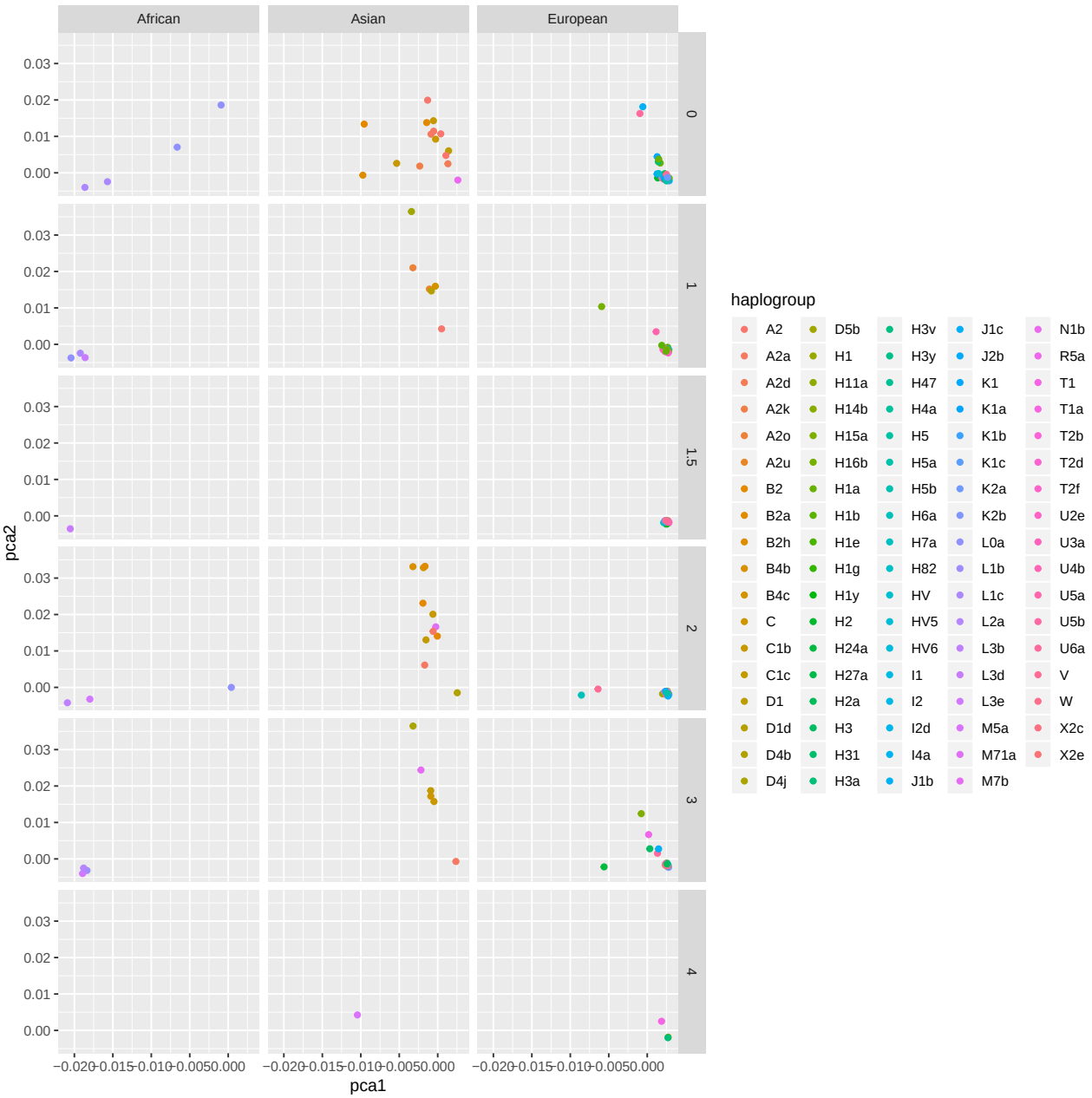
```
##
##  Kruskal-Wallis rank sum test
##
## data:  lvsd by haplogroup
## Kruskal-Wallis chi-squared = 106.96, df = 88, p-value = 0.0827
```

LVSD vs superpopulation and autosomal population

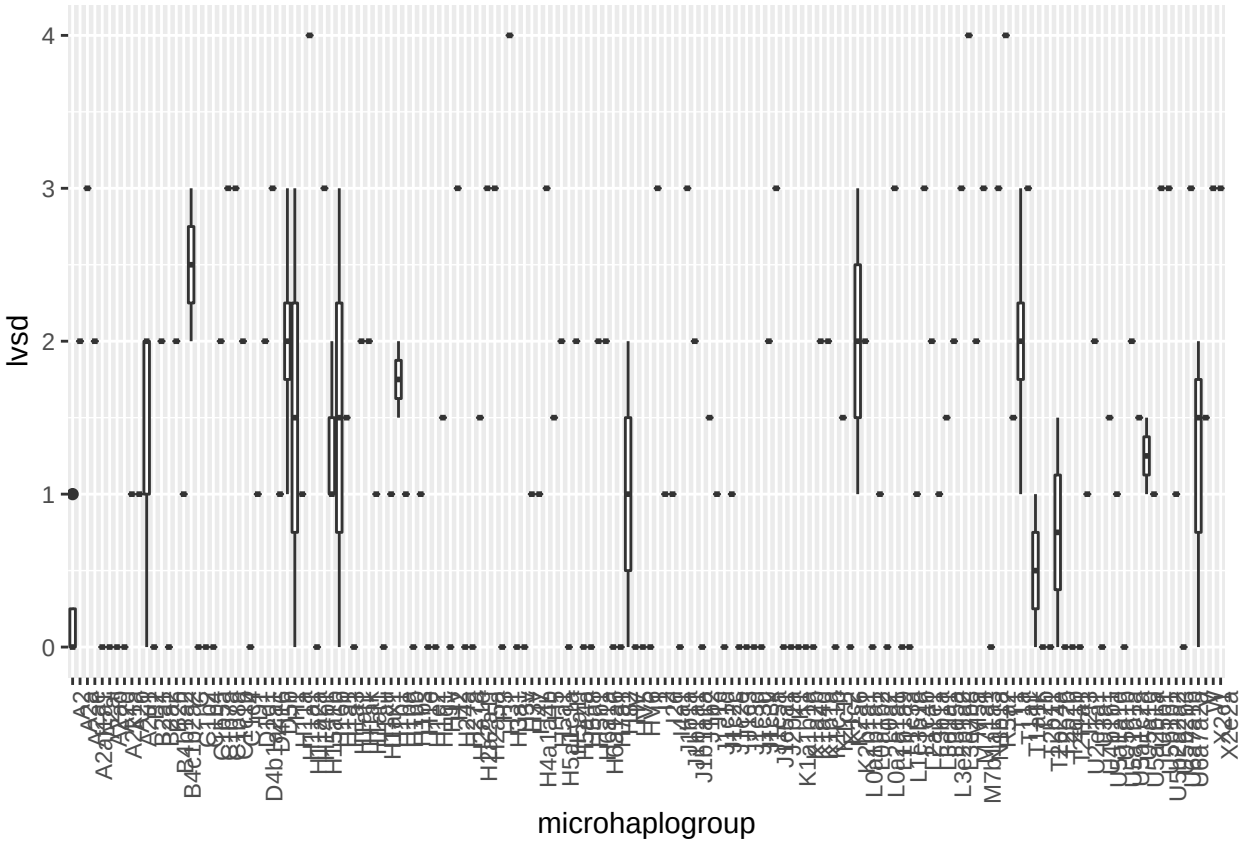
Autosomal background on x-facet, mitochondrial superpopulation on the y-facet. LVSD on the y-axis, SNP-chip PCA component 1 on the x-axis.



Mitochondrial superpopulation on the x-facet, LVSD on the y-facet. PCA's on the axes.


African
Asian
European
0
1
1.5
2
3
4
pca2
pca1
0.03
0.02
0.01
0.00
−0.020
−0.015
−0.010
−0.005
0.000
haplogroup
A2
A2a
A2d
A2k
A2o
A2u
B2
B2a
B2h
B4b
B4c
C
C1b
C1c
D1
D1d
D4b
D4j
D5b
H1
H11a
H14b
H15a
H16b
H1a
H1b
H1e
H1g
H1y
H2
H24a
H27a
H2a
H3
H31
H3a
H3v
H3y
H47
H4a
H5
H5a
H5b
H6a
H7a
H82
HV
HV5
HV6
I1
I2
I2d
I4a
J1b
J1c
J2b
K1
K1a
K1b
K1c
K2a
K2b
L0a
L1b
L1c
L2a
L3b
L3d
L3e
M5a
M71a
M7b
N1b
R5a
T1
T1a
T2b
T2d
T2f
U2e
U3a
U4b
U5a
U5b
U6a
V
W
X2c
X2e

Micro haplogroup



Individual variants on LVSD

Potentially deleterious mutations in cohort

Subjects with potentially deleterious mutations with Genbank frequency <= 1% are below.

patientphenotype	cnt	allele
ADPD / Possibly LVNC-cardiomyopathy associated	1	A3397G
Adult-Onset Dystonia	1	A3796G
BD-associated	1	C114T
Cardiomyopathy	1	G3337A
CPEO	1	T12311C
Cyclic Vomiting Syndrome with Migraine	1	C16176T
DEAF	1	T1005C
DEAF enhancer	1	T5655C
DEAF, possibly LVNC-associated	3	T961C
Hypertensive end-stage renal disease	1	A10086G
LHON	7	A8836G, C3497T, G15812A, T14325C, G3736A
LHON / LDYT / DEAF / hypertension helper mut.	1	G11696A
LHON / SNHL / DEAF	1	G7444A
LHON modulator	1	A15951G
LHON-like, LHON, MELAS	1	A13528G

patientphenotype	cnt	allele
LHON; PD	2	T11253C
MELAS / DEAF enhancer / hypertension / LVNC	1	T3308C
Poss. hypertension factor	1	G3277A
possible HCM susceptibility	4	T12477C, G13135A
Possibly DEAF-associated	4	T961G
Possibly LVNC-associated	3	T921C, A2755G
Prostate Cancer	4	A7158G, T7080C, C5911T
Prostate Cancer / enriched in POAG cohort	1	G6480A
Prostate Cancer / hypertension	2	G5913A
Varied familial presentation / spastic paraparesis	1	G4284A
DEAF helper mut.	1	T10454C
DMDF+HCM / GDM / possibly LVNC cardiomyopathy-associated	1	T3398C
Myopathy	1	T5567C
Possible LHON factor	2	T12811C, A13637G
Susceptibility to bullous pemphigoid	1	G8519A
Therapy-Resistant Epilepsy	1	C6489A
Thyroid Cancer Cell Line	1	A12634G

High LVSD variants

The following variants are found in at least 2 high (>=2.0) LVSD subjects but never in very low (<1.0) subjects

tpos	cnt	gb_avg	allele	locus	effect	haps
225	2	0.77	G225A	HVS2/OH/CSB1/ATT/D-Loop	non-coding	X2e, X2c
239	3	1.31	T239C	HVS2/OH/TFX/ATT/D-Loop	non-coding	T1, H6a, H6a
2387	2	0.24	T2387C	16S	rRNA	J1c, J1c
3915	2	1.30	G3915A	ND1	syn:G=>G	H6a, H6a
4164	2	1.04	A4164G	ND1	syn:M=>M	M7b, H5a
4727	2	0.80	A4727G	ND2	syn:M=>M	H6a, H6a
5198	2	0.28	A5198G	ND2	syn:L=>L	J1c, J1c
5773	2	1.10	G5773A	C/OL	tRNA	L3b, C1c
6216	3	0.56	T6216C	COI	syn:L=>L	U3a, C, B4b
6253	2	1.11	T6253C	COI	non-syn:M=>T	C1c, H15a
6371	2	1.06	C6371T	COI	syn:S=>S	X2e, X2c
6620	2	1.08	T6620C	COI	syn:G=>G	K1b, J1b
7768	2	1.85	A7768G	COII	syn:M=>M	U5b, U5b
8718	2	0.13	A8718G	ATPase6	syn:K=>K	M71a, X2e
9380	2	1.10	G9380A	COIII	syn:W=>W	H6a, H6a
9899	5	1.06	T9899C	COIII	syn:H=>H	T1a, T1, T1a, T1a, T1a
10192	2	0.19	C10192T	ND3	non-syn:S=>F	J1c, J1c
10819	2	2.10	A10819G	ND4	syn:K=>K	L3e, L3e
11002	2	0.95	A11002G	ND4	syn:Q=>Q	L3b, X2c
11253	2	0.62	T11253C	ND4	non-syn:I=>T	H6a, H6a
12633	5	1.21	C12633A	ND5	syn:S=>S	T1a, T1, T1a, T1a, T1a
13635	2	0.26	T13635C	ND5	syn:G=>G	C1b, R5a
13879	2	0.78	T13879C	ND5	non-syn:S=>P	J1b, J1b
13966	2	1.40	A13966G	ND5	non-syn:T=>A	X2e, X2c
14040	2	0.52	G14040A	ND5	syn:Q=>Q	R5a, L3b
14470	2	1.66	T14470C	ND6	syn:G=>G	X2e, X2c
15758	2	0.88	A15758G	Cytb	non-syn:I=>V	I2, H1a

tpos	cnt	gb_avg	allele	locus	effect	haps
15784	2	3.29	T15784C	Cytb	syn:P=>P	U5b, L2a
15930	3	2.55	G15930A	T/ATT	tRNA	C, C1c, C1c
16051	2	2.59	A16051G	ATT/D-Loop/HVS1	non-coding	K2b, C
16186	2	1.24	C16186T	ATT/D-Loop/HVS1/7S DNA	non-coding	T1a, T1a
16192	4	4.27	C16192T	ATT/D-Loop/HVS1/7S DNA	non-coding	U5a, H6a, A2a, J1b
16218	2	0.62	C16218T	ATT/D-Loop/HVS1/7S DNA	non-coding	HV, L0a
16239	2	0.56	C16239T, C16239G	ATT/D-Loop/HVS1/7S DNA	non-coding	H1, H3a
16256	2	3.60	C16256T	ATT/D-Loop/HVS1/7S DNA	non-coding	U5a, H3a
16270	6	5.28	C16270T	ATT/D-Loop/HVS1/7S DNA	non-coding	H1b, U5b, K2b, U5a, U
16482	2	0.77	A16482G	ATT/D-Loop/7S DNA	non-coding	H6a, H6a

The following variants are found in at least 2 very low (<1.0) LVSD subjects but never in high (>=2.0) subjects

tpos	cnt	gb_avg	allele	locus	effect	haps
64	5	2.9600000	C64T	HVS2/ATT/D-Loop/7S DNA	non-coding	A2, A2d, L0a, A2a
143	2	2.1900000	G143A	HVS2/OH/ATT/D-Loop/7S DNA	non-coding	A2, C1b
151	2	3.1800000	C151T	HVS2/OH/ATT/D-Loop/7S DNA	non-coding	L1c, L1c
228	2	2.7300000	G228A	HVS2/OH/CSB1/ATT/D-Loop	non-coding	J1c, J1c
236	2	1.3800000	T236C	HVS2/OH/TFX/ATT/D-Loop	non-coding	L0a, L0a
249	2	0.0850000	A249T, A249G	HVS2/OH/TFX/ATT/D-Loop	non-coding	C1b, H5b
316	2	1.6300000	G316A	HVS2/OH/ATT/D-Loop	non-coding	L1c, L1c
497	2	2.5800000	C497T	ATT/D-Loop	non-coding	K1a, K1a
930	4	2.0400000	G930A	12S	rRNA	T2b, T2b, T2b, T2
1048	2	3.6900000	C1048T	12S	rRNA	L0a, L0a
1692	2	0.2700000	A1692T, A1692G	16S	rRNA	U6a, U4b
2245	2	1.1100000	A2245G	16S	rRNA	L0a, L0a
3516	2	3.2700000	C3516A	ND1	syn:L=>L	L0a, L0a
4312	2	3.2800000	C4312T	I	tRNA	L0a, L0a
4561	2	0.7800000	T4561C	ND2	non-syn:V=>A	K2a, K2a
4586	2	1.5800000	T4586C	ND2	syn:A=>A	L0a, L0a
5147	5	4.1200000	G5147A	ND2	syn:T=>T	T2b, T2b, L0a, T2
5426	2	1.0500000	T5426C	ND2	syn:H=>H	T2b, T2f
5442	2	4.4300000	T5442C	ND2	non-syn:F=>L	L0a, L0a
5471	3	1.0600000	G5471A	ND2	syn:T=>T	U6a, N1b, H5b
5603	2	1.2900000	C5603T	A	tRNA	L0a, L0a
5951	2	1.4600000	A5951G	COI	syn:G=>G	L1c, L1c
6071	2	1.4700000	T6071C	COI	syn:V=>V	L1c, L1c
6185	2	3.3600000	T6185C	COI	syn:F=>F	L0a, L0a
6917	2	0.3600000	G6917A	COI	syn:V=>V	L1c, L1c
7389	3	2.4200000	T7389C	COI	non-syn:Y=>H	L1c, A2, L1c
7930	2	0.1300000	A7930T	COII	syn:G=>G	H31, H31
8251	3	6.4400000	G8251A	COII	syn:G=>G	N1b, L1c, I1
8428	2	1.1600000	C8428T	ATPase8	syn:F=>F	L0a, L0a
8460	2	0.6400000	A8460G	ATPase8	non-syn:N=>S	L0a, A2k
8566	2	1.2300000	A8566G	ATPase8/ATPase6	syn:Q=>Q	L0a, L0a
8602	2	0.1600000	T8602C	ATPase6	non-syn:F=>L	H1g, H1g
9042	2	3.2900000	C9042T	ATPase6	syn:H=>H	L0a, L0a
9055	3	4.9300000	G9055A	ATPase6	non-syn:A=>T	K1a, K1a, K2a
9072	2	1.4300000	A9072G	ATPase6	syn:S=>S	L1c, L1c
9347	2	3.2500000	A9347G	COIII	syn:L=>L	L0a, L0a

tpos	cnt	gb_avg	allele	locus	effect	haps
9716	2	1.0300000	T9716C	COIII	syn:G=>G	K2a, K2a
9755	3	3.1300000	G9755A	COIII	syn:E=>E	J1c, L0a, L0a
9818	2	1.5700000	C9818T	COIII	syn:H=>H	L0a, L0a
9966	2	0.6900000	G9966A	COIII	non-syn:V=>I	I1, L1c
10238	2	6.2600000	T10238C	ND3	syn:I=>I	N1b, I1
10586	2	1.8700000	G10586A	ND4L	syn:S=>S	L1c, L1c
10664	2	3.2800000	C10664T	ND4L	syn:V=>V	L0a, L0a
10771	2	0.1400000	A10771G	ND4	syn:L=>L	H31, H31
10915	2	3.9900000	T10915C	ND4	syn:C=>C	L0a, L0a
11176	2	1.4500000	G11176A	ND4	syn:Q=>Q	L0a, L0a
11302	2	0.2700000	C11302T	ND4	syn:L=>L	L1c, L1c
11314	2	0.2000000	A11314G	ND4	syn:E=>E	L1c, A2d
11641	2	1.2900000	A11641G	ND4	syn:M=>M	L0a, L0a
11812	7	3.3000000	A11812G	ND4	syn:L=>L	T2f, T2b, T2b, T2
12720	2	3.0500000	A12720G	ND5	syn:M=>M	L0a, L0a
12810	2	2.2900000	A12810G	ND5	syn:W=>W	L1c, L1c
13276	2	3.2400000	A13276G	ND5	non-syn:M=>V	L0a, L0a
13485	2	1.4200000	A13485G	ND5	syn:M=>M	L1c, L1c
13789	2	2.2700000	T13789C	ND5	non-syn:Y=>H	L1c, L1c
13934	5	1.1900000	C13934T	ND5	non-syn:T=>M	J1c, D1, J1c, J1c,
14000	2	1.4600000	T14000A	ND5	non-syn:L=>Q	L1c, L1c
14106	2	0.0800000	T14106C	ND5	syn:S=>S	A2, L0a
14178	3	2.7700000	T14178C	ND6	non-syn:I=>V	L1c, H27a, L1c
14233	7	3.6600000	A14233G	ND6	syn:D=>D	T2f, T2b, T2b, T2
14305	2	0.5600000	G14305A	ND6	syn:S=>S	K2a, K2a
14308	2	1.7400000	T14308C	ND6	syn:G=>G	L0a, L0a
14560	2	2.7900000	G14560A	ND6	syn:V=>V	L1c, L1c
14911	2	1.4300000	C14911T	Cytb	syn:Y=>Y	L1c, L1c
15136	2	1.2200000	C15136T	Cytb	syn:G=>G	L0a, L0a
15226	2	0.3300000	A15226G	Cytb	syn:L=>L	L1c, L1c
15431	2	1.5900000	G15431A	Cytb	non-syn:A=>T	L0a, L0a
15905	2	0.3900000	T15905C	T	tRNA	L1c, L1c
15924	4	4.0200000	A15924G	T	tRNA	N1b, K1a, I1, A2k
15978	2	0.3000000	C15978T	P/ATT	tRNA	L1c, L1c
16176	2	0.4600000	C16176G, C16176T	ATT/D-Loop/HVS1/7S DNA	non-coding	N1b, H82
16179	2	0.7200000	C16179T	ATT/D-Loop/HVS1/7S DNA	non-coding	K1c, B2
16188	2	0.9800000	C16188G	ATT/D-Loop/HVS1/7S DNA	non-coding	L0a, L0a
16209	2	2.7200000	T16209C	ATT/D-Loop/HVS1/7S DNA	non-coding	L1c, H1a
16215	2	0.2500000	A16215G	ATT/D-Loop/HVS1/7S DNA	non-coding	H1e, L1c
16220	2	0.1950000	A16220C, A16220G	ATT/D-Loop/HVS1/7S DNA	non-coding	H1b, H82
16222	2	0.8100000	C16222T	ATT/D-Loop/HVS1/7S DNA	non-coding	J1b, J1b
16224	4	5.0000000	T16224C	ATT/D-Loop/HVS1/7S DNA	non-coding	T2b, K1a, K1a, K2
16230	2	3.1600000	A16230G	ATT/D-Loop/HVS1/7S DNA	non-coding	L0a, L0a
16261	2	7.4700000	C16261T	ATT/D-Loop/HVS1/7S DNA	non-coding	J1b, H7a
16316	3	0.2933333	A16316T, A16316G	ATT/D-Loop/HVS1/7S DNA	non-coding	A2a, A2a, H27a
16317	3	0.0000000	A16317C	ATT/D-Loop/HVS1/7S DNA	non-coding	C1b, A2a, C1b

mtGWAS

GLM using both haplogroup and non-haplogroup defining variants

p.value < 0.05 before correction for multiple testing

tpos	hapsnp_p.val	tnt	qnt	calc_locus	gb_perc	allele	haps
225	0.0238794	G	A	HVS2/OH/CSB1/ATT/D-Loop	0.77	G225A	X2e
239	0.0243111	T	C	HVS2/OH/TFX/ATT/D-Loop	1.31	T239C	T1
437	0.0411979	C	T	OH/PL/TFL/ATT/D-Loop	0.07	C437T	A2a
487	0.0411979	A	G	ATT/D-Loop	0.02	A487G	A2a
2387	0.0324802	T	C	16S	0.24	T2387C	J1c, J1c
3565	0.0411979	A	G	ND1	0.10	A3565G	A2a
5046	0.0423120	G	A	ND2	2.15	G5046A	W
5102	0.0140978	A	G	ND2	0.01	A5102G	H11a
5147	0.0203247	G	A	ND2	4.12	G5147A	L0a, L0a
5198	0.0324802	A	G	ND2	0.28	A5198G	J1c, J1c
5252	0.0140978	G	A	ND2	0.34	G5252A	H11a
5461	0.0377832	C	T	ND2	0.03	C5461T	K2b
5985	0.0103637	G	A	COI	0.11	G5985A	T1
6446	0.0423120	G	A	COI	0.69	G6446A	L1b
7307	0.0411979	A	G	COI	0.02	A7307G	A2a
9458	0.0423120	C	T	COIII	0.17	C9458T	L1b
9899	0.0002890	T	C	COIII	1.06	T9899C	T1a, T1, T1a, T1a, T1a, T
10192	0.0324802	C	T	ND3	0.19	C10192T	J1c, J1c
11812	0.0058209	A	G	ND4	3.30	A11812G	C1b
14587	0.0140978	A	G	ND6	0.53	A14587G	H11a
15236	0.0411979	A	G	Cytb	1.10	A15236G	A2a
15412	0.0411979	T	C	Cytb	0.11	T15412C	A2a
16092	0.0140978	T	C	ATT/D-Loop/HVS1	1.47	T16092C	H11a
16163	0.0032930	A	G	ATT/D-Loop/HVS1/7S DNA/TAS	1.40	A16163G	L1c, T1
16239	0.0039954	C	G	ATT/D-Loop/HVS1/7S DNA	0.06	C16239G	H3a
16239	0.0039954	C	T	ATT/D-Loop/HVS1/7S DNA	1.06	C16239T	H1
16254	0.0140978	A	G	ATT/D-Loop/HVS1/7S DNA	0.13	A16254G	H11a
16270	0.0101111	C	T	ATT/D-Loop/HVS1/7S DNA	5.28	C16270T	H1b, K2b, H1b, K2b
16293	0.0247849	A	G	ATT/D-Loop/HVS1/7S DNA	2.28	A16293G	L0a, L0a
16317	0.0217274	A	C	ATT/D-Loop/HVS1/7S DNA	0.00	A16317C	C1b, A2a, C1b

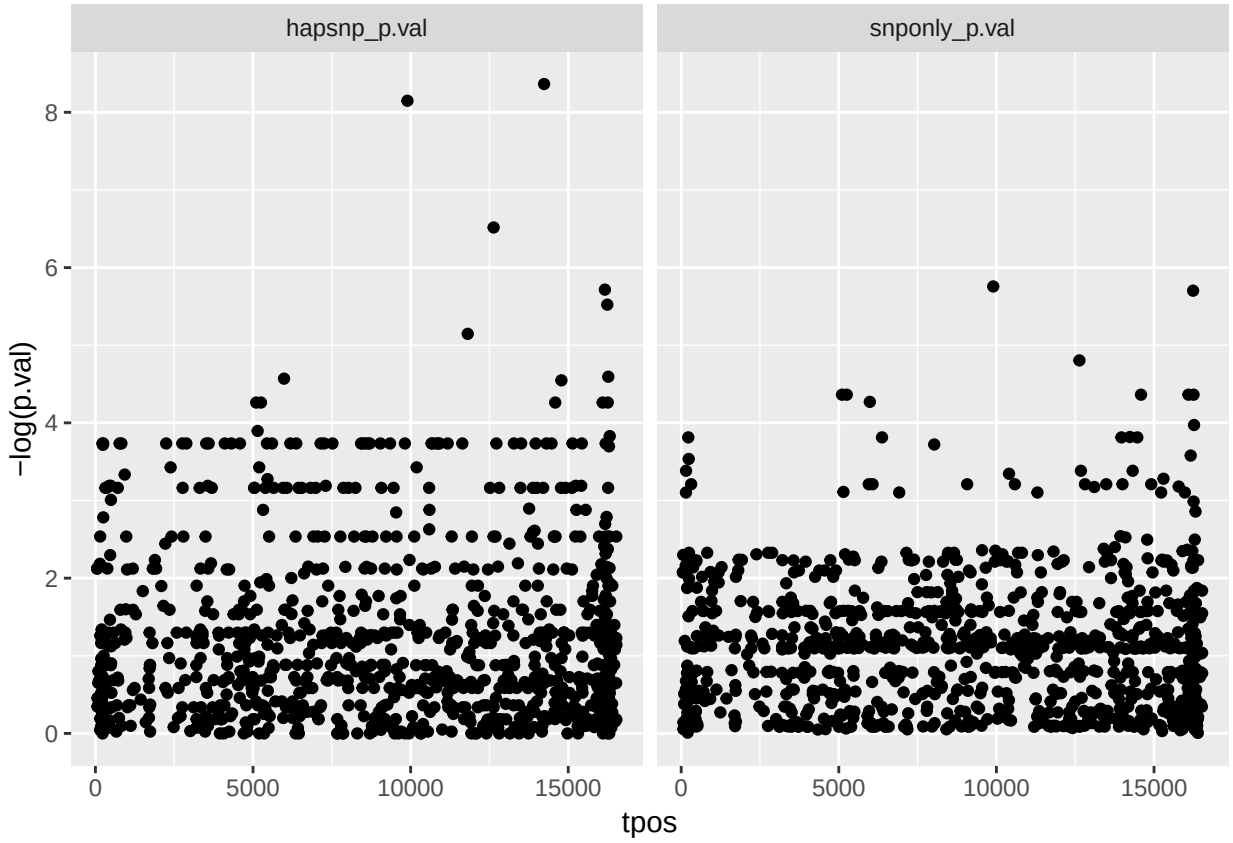
GLM - non-haplogroup defining variants only

p.value < 0.05 before correction for multiple testing

tpos	snponly_p.val	qnt	tnt	calc_locus	gb_perc	allele	haps
151	0.0449163	T	C	HVS2/OH/ATT/D-Loop/7S DNA	3.18	C151T	L1c, L1c
154	0.0339510	C	T	HVS2/OH/ATT/D-Loop/7S DNA	0.07	T154C	M5a
225	0.0220812	A	G	HVS2/OH/CSB1/ATT/D-Loop	0.77	G225A	X2e
239	0.0291555	C	T	HVS2/OH/TFX/ATT/D-Loop	1.31	T239C	T1
5102	0.0127407	G	A	ND2	0.01	A5102G	H11a
5147	0.0445695	A	G	ND2	4.12	G5147A	L0a, L0a
5252	0.0127407	A	G	ND2	0.34	G5252A	H11a
5985	0.0139742	A	G	COI	0.11	G5985A	T1

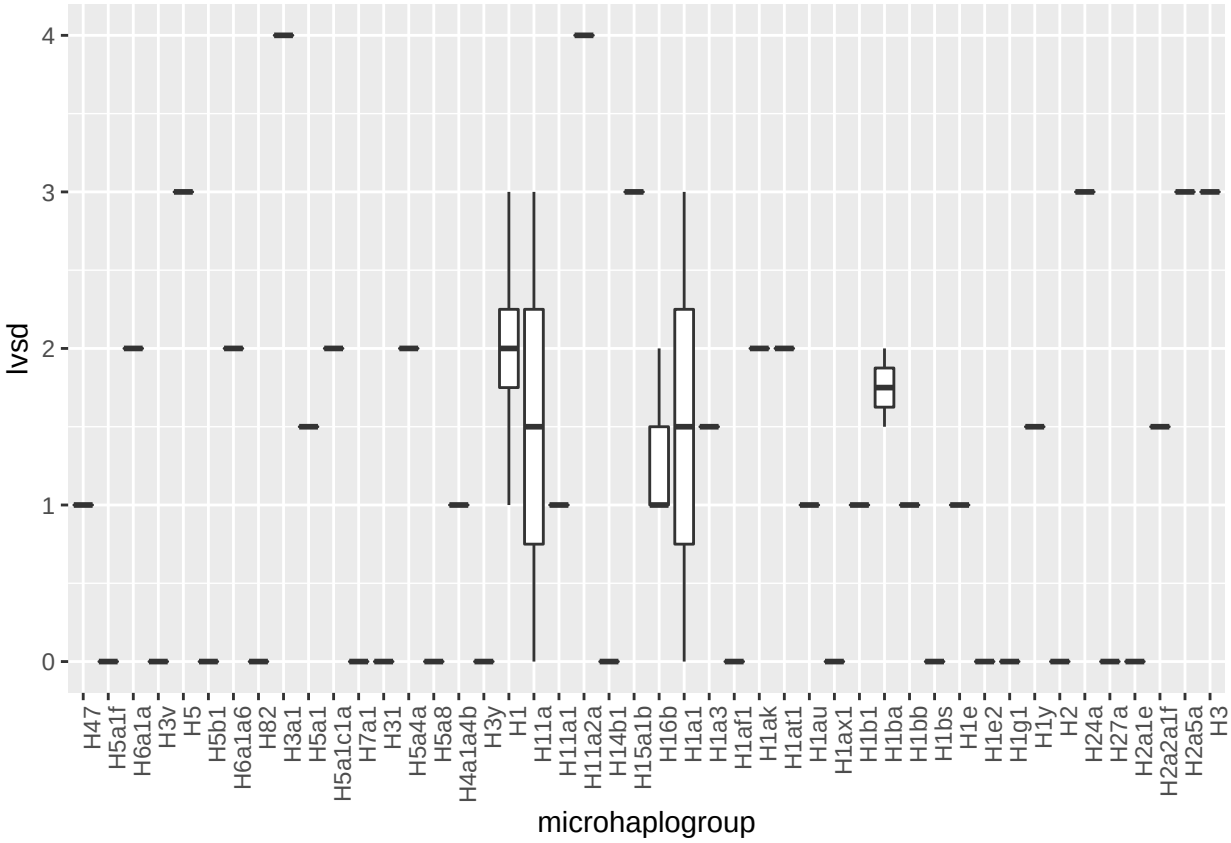
tpos	snponly_p.val	qnt	tnt	calc_locus	gb_perc	allele	haps
6917	0.0449163	A	G	COI	0.36	G6917A	L1c, L1c
9899	0.0031598	C	T	COIII	1.06	T9899C	T1a, T1, T1a, T1a, T1a,
11302	0.0449163	T	C	ND4	0.27	C11302T	L1c, L1c
12681	0.0339510	C	T	ND5	0.20	T12681C	M5a
13105	0.0418467	G	A	ND5	7.77	A13105G	D1d
14587	0.0127407	G	A	ND6	0.53	A14587G	H11a
15226	0.0449163	G	A	Cytb	0.33	A15226G	L1c, L1c
15301	0.0376168	A	G	Cytb	27.55	G15301A	B4c
15784	0.0416055	C	T	Cytb	3.29	T15784C	U5b
15978	0.0449163	T	C	P/ATT	0.30	C15978T	L1c, L1c
16092	0.0127407	C	T	ATT/D-Loop/HVS1	1.47	T16092C	H11a
16163	0.0279064	G	A	ATT/D-Loop/HVS1/7S DNA/TAS	1.40	A16163G	L1c, T1
16239	0.0033374	G	C	ATT/D-Loop/HVS1/7S DNA	0.06	C16239G	H3a
16239	0.0033374	T	C	ATT/D-Loop/HVS1/7S DNA	1.06	C16239T	H1
16254	0.0127407	G	A	ATT/D-Loop/HVS1/7S DNA	0.13	A16254G	H11a
16270	0.0188273	T	C	ATT/D-Loop/HVS1/7S DNA	5.28	C16270T	H1b, K2b, H1b, K2b

GLMs



Haplogroups of interest

H



Enrichment

Are there any differences in variants between the 10 >=3's and the 21 0's?

Enriched in high LVSD (>1fc)

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
55	T	C	ATT/D-Loop/7S DNA	0.20	1	0	Inf	T55C
57	T	C	HVS2/ATT/D-Loop/7S DNA	0.20	1	0	Inf	T57C
73	A	G	HVS2/ATT/D-Loop/7S DNA	73.71	1	1	1.070389	A73G
189	A	C	HVS2/OH/ATT/D-Loop/7S DNA	1.41	1	1	1.070389	A189C
189	A	G	HVS2/OH/ATT/D-Loop/7S DNA	5.51	1	1	1.070389	A189G
379	A	G	OH/mt4H/ATT/D-Loop	0.01	1	0	Inf	A379G
961	T	C	12S	0.99	2	2	1.070389	T961C
961	T	G	12S	0.38	2	2	1.070389	T961G
1187	T	C	12S	0.08	1	0	Inf	T1187C
1822	T	C	16S	0.57	1	0	Inf	T1822C
1842	A	G	16S	0.33	1	0	Inf	A1842G
1926	A	G	16S	0.03	1	0	Inf	A1926G
3333	C	T	ND1	0.14	1	0	Inf	C3333T

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
3579	A	G	ND1	0.05	1	0	Inf	A3579G
3666	G	A	ND1	2.77	1	0	Inf	G3666A
5102	A	G	ND2	0.01	1	0	Inf	A5102G
5252	G	A	ND2	0.34	1	0	Inf	G5252A
6253	T	C	COI	1.11	1	0	Inf	T6253C
6365	T	C	COI	0.32	1	1	1.070389	T6365C
6629	A	G	COI	0.27	1	0	Inf	A6629G
6776	T	C	COI	1.94	2	2	1.070389	T6776C
7220	T	C	COI	0.07	1	0	Inf	T7220C
8448	T	C	ATPase8	0.43	2	1	2.070389	T8448C
8540	T	C	ATPase8/ATPase6	0.02	1	0	Inf	T8540C
8745	A	G	ATPase6	0.02	1	0	Inf	A8745G
9181	A	G	ATPase6	0.22	1	0	Inf	A9181G
10364	G	A	ND3	0.07	1	0	Inf	G10364A
10653	G	A	ND4L	0.05	1	0	Inf	G10653A
11410	T	C	ND4	0.16	1	0	Inf	T11410C
11914	G	A	ND4	11.02	1	0	Inf	G11914A
13404	T	C	ND5	0.18	1	0	Inf	T13404C
13708	G	A	ND5	6.82	1	0	Inf	G13708A
13759	G	A	ND5	2.93	2	1	2.070389	G13759A
14070	A	G	ND5	0.34	1	0	Inf	A14070G
14587	A	G	ND6	0.53	1	0	Inf	A14587G
14953	C	T	Cytb	0.12	1	0	Inf	C14953T
16092	T	C	ATT/D-Loop/HVS1	1.47	1	0	Inf	T16092C
16140	T	C	ATT/D-Loop/HVS1/7S DNA	1.30	1	0	Inf	T16140C
16162	A	G	ATT/D-Loop/HVS1/7S DNA/TAS	1.43	1	1	1.070389	A16162G
16184	C	T	ATT/D-Loop/HVS1/7S DNA	0.72	1	0	Inf	C16184T
16209	T	C	ATT/D-Loop/HVS1/7S DNA	2.72	1	1	1.070389	T16209C
16239	C	G	ATT/D-Loop/HVS1/7S DNA	0.06	2	0	Inf	C16239G
16239	C	T	ATT/D-Loop/HVS1/7S DNA	1.06	2	0	Inf	C16239T
16254	A	G	ATT/D-Loop/HVS1/7S DNA	0.13	1	0	Inf	A16254G
16256	C	T	ATT/D-Loop/HVS1/7S DNA	3.60	1	0	Inf	C16256T
16293	A	G	ATT/D-Loop/HVS1/7S DNA	2.28	3	1	2.655352	A16293G
16311	T	C	ATT/D-Loop/HVS1/7S DNA	20.66	3	1	2.655352	T16311C

Enriched in low LVSD (<-1fc)

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
150	C	T	HVS2/OH/ATT/D-Loop/7S DNA	11.81	0	1	-Inf	C150T
152	T	C	HVS2/OH/ATT/D-Loop/7S DNA	25.84	1	6	-1.514573	T152C
204	T	C	HVS2/OH/ATT/D-Loop	6.07	0	1	-Inf	T204C
242	C	T	HVS2/OH/TFX/ATT/D-Loop	0.39	0	1	-Inf	C242T
249	A	G	HVS2/OH/TFX/ATT/D-Loop	0.17	0	1	-Inf	A249G
249	A	T	HVS2/OH/TFX/ATT/D-Loop	0.00	0	1	-Inf	A249T
408	T	A	OH/PL/ATT/D-Loop	0.21	0	1	-Inf	T408A
575	C	T	D-Loop	0.11	0	1	-Inf	C575T
951	G	A	12S	0.79	0	1	-Inf	G951A
1503	G	A	12S	0.28	0	1	-Inf	G1503A
1719	G	A	16S	5.07	0	1	-Inf	G1719A
2352	T	C	16S	2.62	0	1	-Inf	T2352C
2581	A	G	16S	0.29	0	1	-Inf	A2581G

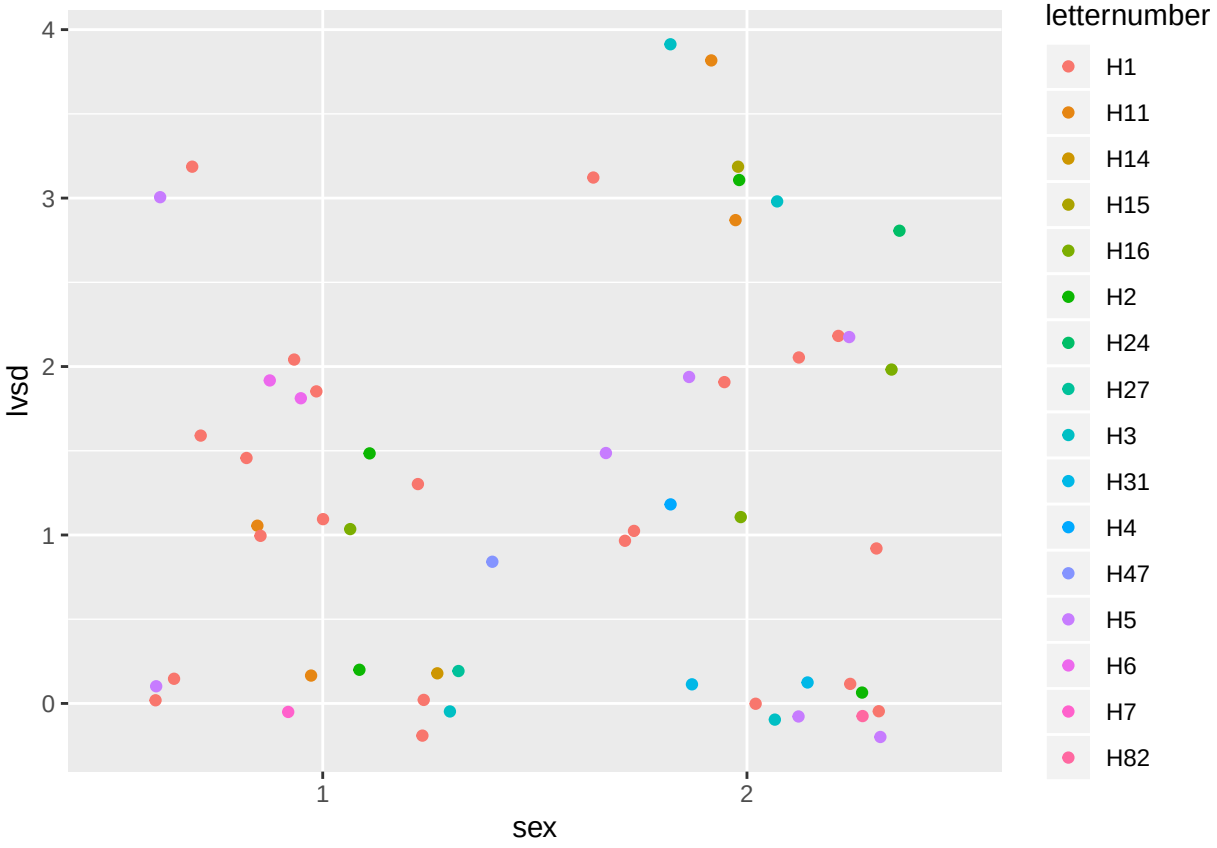
tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
2882	T	C	16S	0.01	0	1	-Inf	T2882C
3197	T	C	16S	4.21	0	1	-Inf	T3197C
3397	A	G	ND1	0.30	0	1	-Inf	A3397G
3906	T	C	ND1	0.03	0	1	-Inf	T3906C
4227	A	G	ND1	0.10	0	1	-Inf	A4227G
4336	T	C	Q	1.04	0	2	-Inf	T4336C
4512	G	A	ND2	0.06	0	1	-Inf	G4512A
4736	T	C	ND2	0.17	0	1	-Inf	T4736C
4793	A	G	ND2	0.85	0	1	-Inf	A4793G
5054	G	C	ND2	0.07	0	1	-Inf	G5054C
5460	G	A	ND2	6.25	0	1	-Inf	G5460A
5471	G	A	ND2	1.06	0	1	-Inf	G5471A
5480	A	G	ND2	0.18	0	1	-Inf	A5480G
6126	A	G	COI	0.01	0	1	-Inf	A6126G
7471	C	T	S(UCN)	0.05	0	1	-Inf	C7471T
7645	T	C	COII	0.28	0	1	-Inf	T7645C
7930	A	T	COII	0.13	0	2	-Inf	A7930T
8429	C	T	ATPase8	0.04	0	1	-Inf	C8429T
8602	T	C	ATPase6	0.16	0	2	-Inf	T8602C
9163	G	A	ATPase6	0.06	0	1	-Inf	G9163A
9391	C	T	COIII	0.07	0	1	-Inf	C9391T
9489	G	A	COIII	0.02	0	1	-Inf	G9489A
9581	T	C	COIII	0.14	0	1	-Inf	T9581C
10217	A	G	ND3	0.12	0	1	-Inf	A10217G
10493	T	C	ND4L	0.27	0	1	-Inf	T10493C
10771	A	G	ND4	0.14	0	2	-Inf	A10771G
11101	A	G	ND4	0.14	0	1	-Inf	A11101G
11719	G	A	ND4	75.36	0	1	-Inf	G11719A
11864	T	C	ND4	0.17	0	1	-Inf	T11864C
11969	G	A	ND4	1.51	0	1	-Inf	G11969A
14097	C	T	ND5	0.07	0	1	-Inf	C14097T
14178	T	C	ND6	2.77	0	1	-Inf	T14178C
14212	T	C	ND6	1.86	0	2	-Inf	T14212C
14325	T	C	ND6	0.10	0	1	-Inf	T14325C
14359	C	T	ND6	0.00	0	1	-Inf	C14359T
14401	A	G	ND6	0.02	0	1	-Inf	A14401G
14497	A	G	ND6	0.05	0	1	-Inf	A14497G
14857	T	C	Cytb	0.06	0	1	-Inf	T14857C
14861	G	A	Cytb	0.27	0	1	-Inf	G14861A
14871	T	C	Cytb	0.02	0	1	-Inf	T14871C
15430	C	T	Cytb	0.01	0	1	-Inf	C15430T
15511	T	C	Cytb	0.30	0	1	-Inf	T15511C
15646	C	T	Cytb	0.06	0	1	-Inf	C15646T
15817	A	G	Cytb	0.11	0	1	-Inf	A15817G
15833	C	T	Cytb	0.66	0	1	-Inf	C15833T
15992	A	G	P/ATT	0.01	0	1	-Inf	A15992G
16093	T	C	ATT/D-Loop/HVS1	5.66	0	2	-Inf	T16093C
16126	T	C	ATT/D-Loop/HVS1/7S DNA	11.58	0	1	-Inf	T16126C
16129	G	A	ATT/D-Loop/HVS1/7S DNA	13.07	0	2	-Inf	G16129A
16129	G	C	ATT/D-Loop/HVS1/7S DNA	0.69	0	2	-Inf	G16129C
16176	C	G	ATT/D-Loop/HVS1/7S DNA	0.21	0	1	-Inf	C16176G
16176	C	T	ATT/D-Loop/HVS1/7S DNA	0.71	0	1	-Inf	C16176T

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
16215	A	G	ATT/D-Loop/HVS1/7S DNA	0.25	0	1	-Inf	A16215G
16220	A	C	ATT/D-Loop/HVS1/7S DNA	0.33	0	2	-Inf	A16220C
16220	A	G	ATT/D-Loop/HVS1/7S DNA	0.06	0	2	-Inf	A16220G
16261	C	T	ATT/D-Loop/HVS1/7S DNA	7.47	0	1	-Inf	C16261T
16316	A	G	ATT/D-Loop/HVS1/7S DNA	0.88	0	1	-Inf	A16316G
16316	A	T	ATT/D-Loop/HVS1/7S DNA	0.00	0	1	-Inf	A16316T
16357	T	C	ATT/D-Loop/HVS1/7S DNA	1.09	0	1	-Inf	T16357C
16390	G	A	ATT/D-Loop/7S DNA	5.82	0	1	-Inf	G16390A

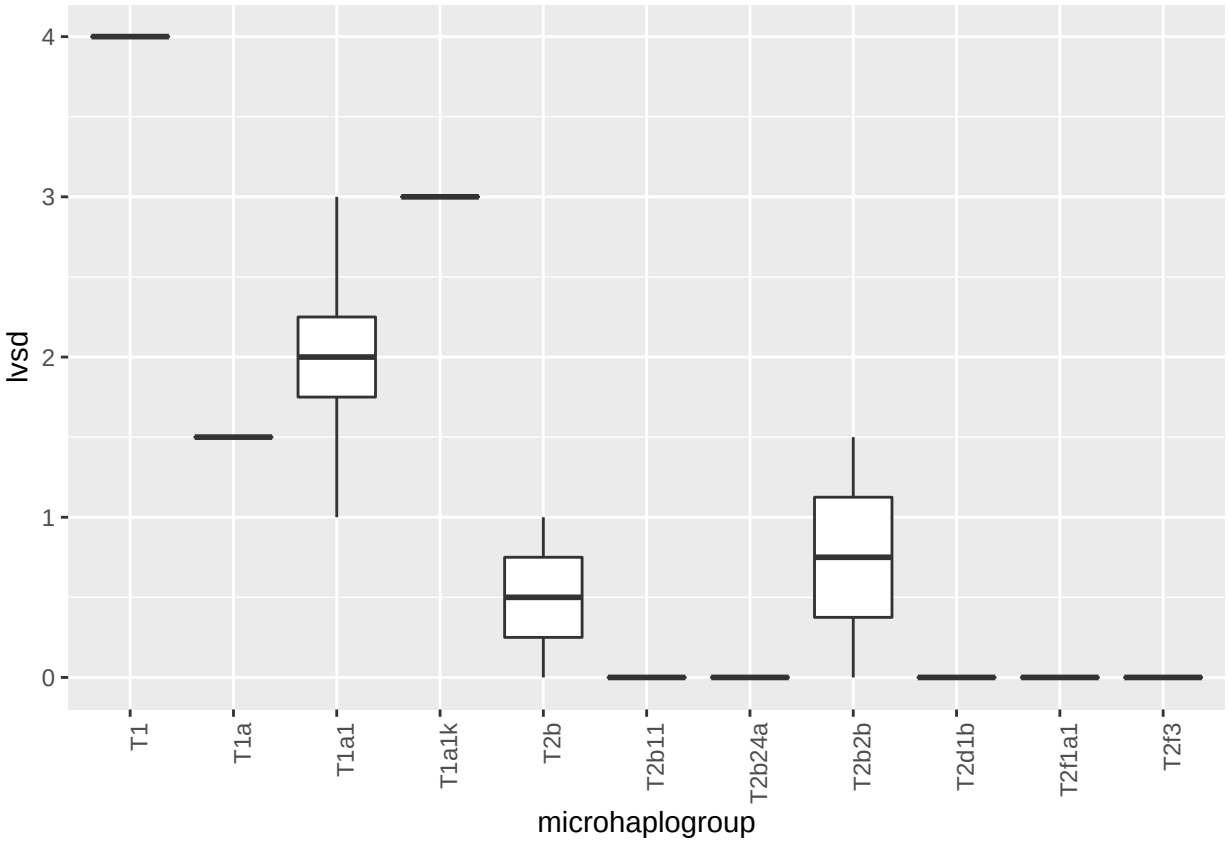
Sex differences in H

H has differences in the number of high LVSD subjects by gender. There is no discernable interaction with any H subclades.

```
## # A tibble: 4 x 3
##   sex   `lvsd >= 3`   cnt
##   <fct> <lgl>       <int>
## 1 1     FALSE          24
## 2 1     TRUE            2
## 3 2     FALSE          22
## 4 2     TRUE            8
```



T



Enrichment

Are there any differences in variants between the 3 >=3's and the 7 0's?

Enriched in high LVSD (>1fc)

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
152	T	C	HVS2/OH/ATT/D-Loop/7S DNA	25.84	3	2	1.807355	T152C
195	T	C	HVS2/OH/ATT/D-Loop	20.51	2	2	1.222392	T195C
239	T	C	HVS2/OH/TFX/ATT/D-Loop	1.31	1	0	Inf	T239C
469	C	T	ATT/D-Loop	0.04	1	0	Inf	C469T
5985	G	A	COI	0.11	1	0	Inf	G5985A
9899	T	C	COIII	1.06	3	0	Inf	T9899C
12633	C	A	ND5	1.21	3	0	Inf	C12633A
12633	C	T	ND5	0.28	3	0	Inf	C12633T
16163	A	G	ATT/D-Loop/HVS1/7S DNA/TAS	1.40	3	0	Inf	A16163G
16186	C	T	ATT/D-Loop/HVS1/7S DNA	1.24	1	0	Inf	C16186T
16189	T	C	ATT/D-Loop/HVS1/7S DNA	25.41	1	0	Inf	T16189C
16311	T	C	ATT/D-Loop/HVS1/7S DNA	20.66	1	1	1.222392	T16311C

Enriched in low LVSD (<-1fc)

tpos	tnt	qnt	calc_locus	gb_perc	hot	cold	enrichment	allele
194	C	T	HVS2/OH/ATT/D-Loop	1.73	0	1	-Inf	C194T
207	G	A	HVS2/OH/ATT/D-Loop	4.69	0	1	-Inf	G207A
321	T	C	HVS2/OH/HPR/ATT/D-Loop	0.04	0	1	-Inf	T321C
930	G	A	12S	2.04	0	4	-Inf	G930A
3204	C	T	16S	0.37	0	1	-Inf	C3204T
3398	T	C	ND1	0.36	0	1	-Inf	T3398C
5147	G	A	ND2	4.12	0	4	-Inf	G5147A
5277	T	C	ND2	0.26	0	1	-Inf	T5277C
5426	T	C	ND2	1.05	0	2	-Inf	T5426C
5460	G	A	ND2	6.25	0	1	-Inf	G5460A
5747	A	G	OL	0.09	0	1	-Inf	A5747G
6489	C	A	COI	0.21	0	1	-Inf	C6489A
7295	A	G	COI	0.02	0	1	-Inf	A7295G
8537	A	G	ATPase8/ATPase6	0.07	0	1	-Inf	A8537G
8572	G	A	ATPase8/ATPase6	0.28	0	1	-Inf	G8572A
8854	G	A	ATPase6	0.08	0	1	-Inf	G8854A
9233	T	C	COIII	0.13	0	1	-Inf	T9233C
11242	C	G	ND4	0.06	0	1	-Inf	C11242G
11812	A	G	ND4	3.30	0	7	-Inf	A11812G
12373	A	G	ND5	0.15	0	1	-Inf	A12373G
13260	T	C	ND5	0.21	0	1	-Inf	T13260C
13708	G	A	ND5	6.82	0	1	-Inf	G13708A
14233	A	G	ND6	3.66	0	7	-Inf	A14233G
15028	C	A	Cytb	0.20	0	1	-Inf	C15028A
15043	G	A	Cytb	22.77	0	1	-Inf	G15043A
16224	T	C	ATT/D-Loop/HVS1/7S DNA	5.00	0	1	-Inf	T16224C
16278	C	T	ATT/D-Loop/HVS1/7S DNA	10.27	0	1	-Inf	C16278T
16296	C	T	ATT/D-Loop/HVS1/7S DNA	2.25	0	1	-Inf	C16296T
16304	T	C	ATT/D-Loop/HVS1/7S DNA	6.57	0	1	-Inf	T16304C

T1 vs T2

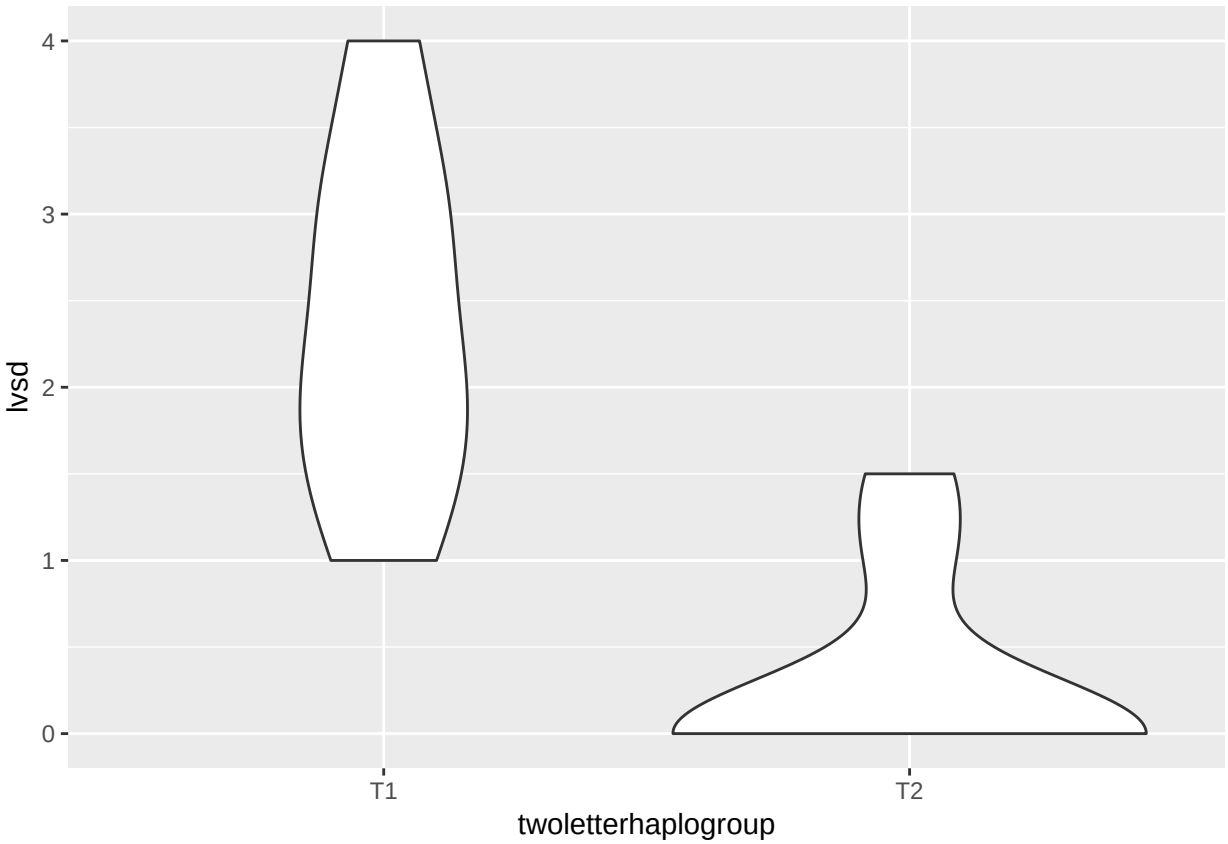
T1 and T2, appear to segregate strongly with regard to LVSD.

The major subclades of T are distinguished by just a few loci.

T1 is defined by C12633A A16163G T16189C

T2 is defined by A11812G A14233G (C16296T)

See http://www.ncbi.nlm.nih.gov/pubmed/19434233



T subclade summary

twoletterhaplogroup	cnt	mean	range	sd
T1	7	2.3571429	1-4	1.0293317
T2	9	0.2777778	0-1.5	0.5651942

Logistic regression assuming lvsd>1 is affected

```
##
## Call:
## glm(formula = affected ~ twoletterhaplogroup, family = binomial(link = "logit"),
##     data = Tdf)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9728  -0.4854  -0.4854   0.5553   2.0963
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)              1.792      1.080   1.659   0.0971 .
## twoletterhaplogroupT2   -3.871      1.514  -2.557   0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 21.930  on 15  degrees of freedom
## Residual deviance: 12.021  on 14  degrees of freedom
## AIC: 16.021
##
## Number of Fisher Scoring iterations: 4
```

Logistic regression by the `outcome` variable (ignoring -9's)

```
##
## Call:
## glm(formula = status ~ twoletterhaplogroup, family = binomial(link = "logit"),
##     data = Tdfo)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.51678  -0.51678  -0.25837   0.00005   2.03933
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)              20.57    7238.39   0.003    0.998
## twoletterhaplogroupT2   -22.51    7238.39  -0.003    0.998
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 19.4081  on 13  degrees of freedom
## Residual deviance:  6.0283  on 12  degrees of freedom
## AIC: 10.028
##
## Number of Fisher Scoring iterations: 19
```

Next steps

- Parsable Phylotree
- Build mixed model, test at different levels (macro, two-letter, three-letter)
- Look at heteroplasmies
- Investigate burden of variants - especially in ND4 and ND6
- Develop RF or other ML approaches for classification