# Online Resource 1:

Data Adjustments, Sample Characteristics, Distributional Properties of the Data, and Reliability of the measured phenotypes

**The following are provided in this appendix:**

*(A) Data adjustments*

*(B) Sample characteristics*

*(C) Distributional Properties of the Data*

*(D) Reliability of submaximal fitness traits in HERITAGE*

**(A) Data Adjustments**

Below is the method used for the covariate adjustments, followed by the table of results for the target ($\Delta VO_2 60$) and the correlated traits ($\Delta$WORK60, $\Delta$Q60)

Responses were computed as the differences in post-training vs baseline (pre-training) values, and then adjusted (in both mean and variance) for the effects of several covariates using a stepwise multiple regression procedure (stepwise forward with backward elimination). In summary, a response was regressed on the respective baseline value, the baseline weight and a cubic polynomial in age (age, age$^2$ and age$^3$), separately in four sex-by-generation groups (fathers, mothers, sons and daughters). Only significant terms (5%) were retained. The resulting squared residual from this mean regression was similarly adjusted for the same covariates (i.e. heteroskedasticity or variance regression step). The covariate-adjusted variable was the residual from the mean regression, divided by the square root of the predicted score from the second regression. A final standardization step ensured zero mean and unit variance. Values that were $\pm$ 4 standard deviations from the mean (on the raw scale) were temporarily set aside during model development. However, these outlying values were added back to the dataset and scores were computed for them in the final steps. Consequently, raw outlying values did not contribute to model development but were assigned scores in the end based on the models developed using the non-outlying data.

The computer program SAS was used to produce the covariate adjustments, the sample statistics and the frequency distributions.

**Table of Covariate Adjustments for Responses to Exercise Training: Significant Terms and % Variance Accounted For**

| | | Mean Regression | | Variance Regression | |
|---|---|---|---|---|---|
| | | Significant Terms | % Variance | Significant Terms | % Variance |
| $\Delta VO_2 60$ | Fathers | None | — | $Age^3$ | 5% |
| | Mothers | Baseline, Weight | 9% | None | — |
| | Sons | Baseline, Age | 12% | Weight | 8% |
| | Daughters | None | — | Weight | 3% |
| ΔWORK60 | Fathers | Weight, $Age^3$, Baseline | 17% | None | — |
| | Mothers | Baseline | 6% | None | — |
| | Sons | None | — | None | — |
| | Daughters | Baseline | 3% | None | — |
| ΔQ60 | Fathers | Baseline | 7% | None | — |
| | Mothers | Baseline | 17% | None | — |
| | Sons | Baseline | 8% | None | — |
| | Daughters | Baseline | 8% | Weight | 5% |

**(B) Sample Characteristics Statistics for Raw (Unadjusted) Data and Family Structures***

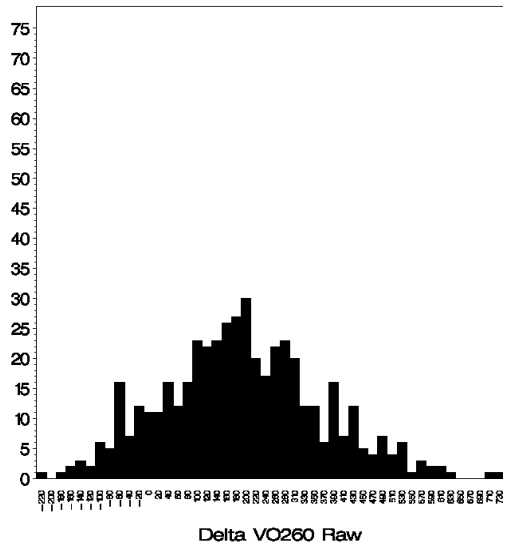
| Variable | Group | N | Mean | Standard Deviation | Gp Differences† |
|---|---|---|---|---|---|
| Age | (Skewness = 0.34 and Kurtosis = -1.36)‡ | | | | A, −, C, D |
| | Fathers | 92 | 53.34 | 5.34 | |
| | Mothers | 89 | 52.12 | 5.13 | |
| | Sons | 136 | 25.40 | 6.16 | |
| | Daughters | 158 | 25.51 | 6.44 | |
| Baseline Weight | (Skewness = 0.76 and Kurtosis = 0.56) | | | | A, B, C, D |
| | Fathers | 92 | 87.25 | 15.22 | |
| | Mothers | 89 | 72.52 | 13.58 | |
| | Sons | 136 | 82.37 | 16.39 | |
| | Daughters | 158 | 64.23 | 13.28 | |
| $\Delta VO_2 60$ | (Skewness = 0.13 and Kurtosis = 0.34 for adjusted) | | | | A, B, C, D |
| | Fathers | 92 | 198.60 | 169.02 | |
| | Mothers | 89 | 111.59 | 143.06 | |
| | Sons | 136 | 261.07 | 191.40 | |
| | Daughters | 158 | 191.77 | 142.62 | |
| ΔWORK60 | (Skewness = 0.34 and Kurtosis = 1.05 for adjusted) | | | | A, B, C, D |
| | Fathers | 92 | 24.43 | 14.10 | |
| | Mothers | 89 | 14.994 | 12.02 | |
| | Sons | 136 | 33.64 | 17.96 | |
| | Daughters | 158 | 23.43 | 12.36 | |
| ΔQ60 | (Skewness = 0.18 and Kurtosis = 0.19 for adjusted) | | | | A, −, C, D |
| | Fathers | 86 | 0.79 | 1.42 | |
| | Mothers | 81 | 0.42 | 1.20 | |
| | Sons | 130 | 1.18 | 1.63 | |
| | Daughters | 150 | 1.22 | 1.46 | |

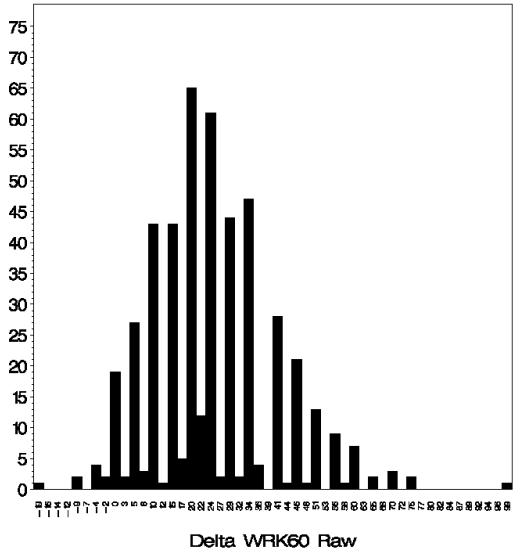
*Family structures: There are 475 individuals within 99 nuclear families (parents and offspring) having both marker data and $\Delta VO_2 60$ (response) phenotype data. There is an average of 4.8 members per family (on average 2 parents and 3 offspring). The number of offspring within families ranges from 1-5, for a total of 317 sibling pairs, 545 parent-offspring pairs, and 83 spouse pairs.

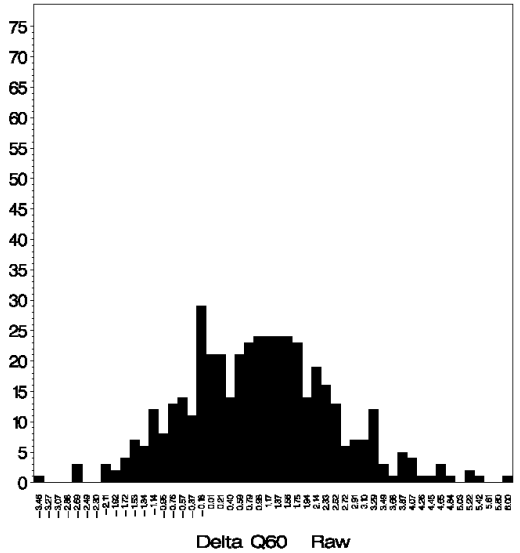
†Group Differences: Based on a comparison of raw means using the respective standard errors, there are group differences between fathers and mothers (label A), sons and daughters (label B), fathers and sons (label C), and mothers and daughters (label D).

‡Means and standard deviations are for the raw (unadjusted) data, and skewness and kurtosis is reported for the analysis variable (i.e. covariate-adjusted and standardized residuals) across the entire sample.

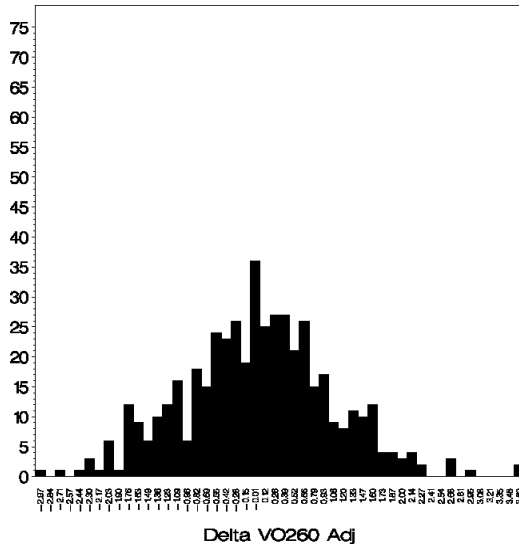
**(C) Distributional Properties for the Raw (Top Row) and Covariate Adjusted (Bottom Row) Response Phenotypes**

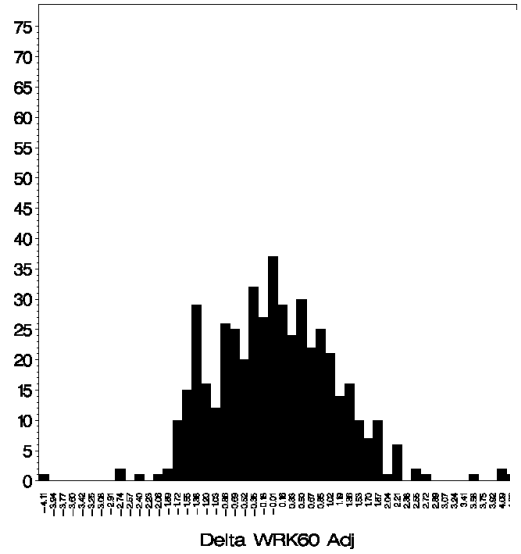


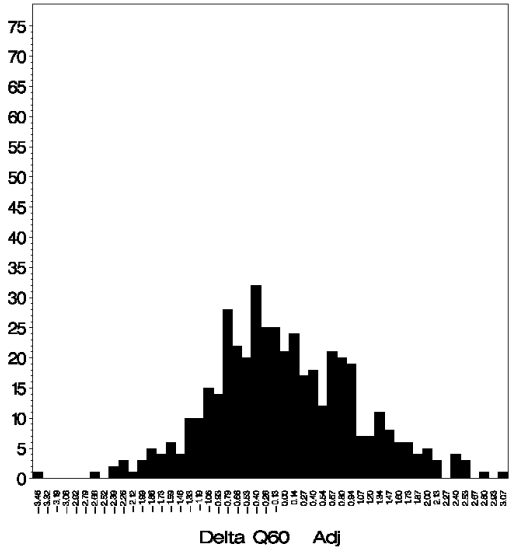




**(D) *Reliability of submaximal fitness traits in HERITAGE***



### *(D) Reliability of submaximal fitness traits in HERITAGE*

For reliability, coefficients of variation (CV) and intraclass correlation coefficients (ICC) were calculated using (i) 390 HERITAGE subjects to assess day-to-day variation across 2 days (reproducibility) and (ii) 55 subjects who were not enrolled in HERITAGE (but otherwise qualified) to assess intra-center quality control (ICQC) by testing each individual at each of the 5 clinical centers (Wilmore et al., 1998). For $\Delta VO_2 60$, the CVs and ICCs were 3.6% and 0.99 for reliability and for ICQC were 3.5% and 0.98. Similarly, for ΔQ60 the CVs and ICCs for reliability were 5.9% and 0.93 and for ICQC were 4.5% and 0.95. Thus, both within subject variation and measurement unreliabilities for day-to-day and across-center are generally small, particularly as compared with the between-subject variance in the responses to the submaximal exercise.

# Online Resource 2:

Hardy Weinberg Equilibrium (HWE) and Minor Allele Frequency (MAF) Plots

**The following are provided in this appendix:**

*(A) Scatter Plot of Association –log(p) by HWE –log(p)*

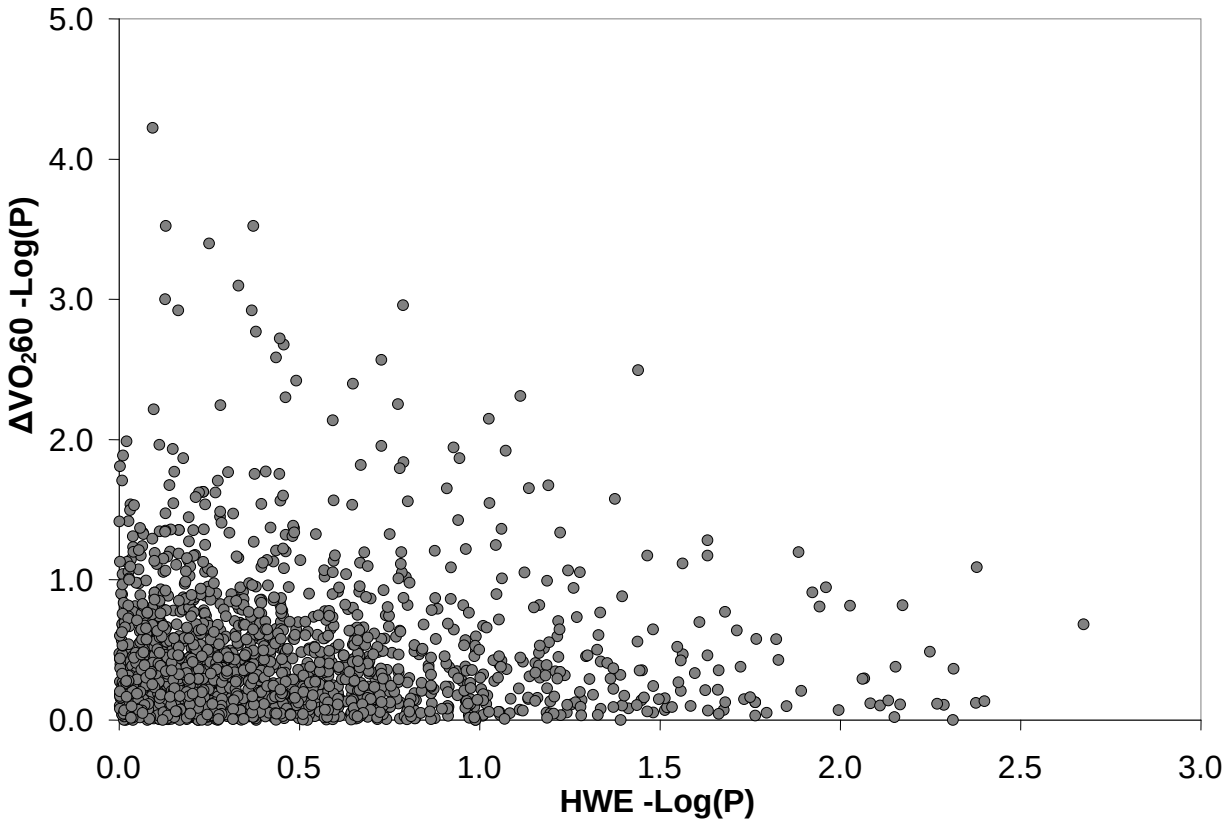
*(B) Scatter Plot of Association –log(p) by MAF*

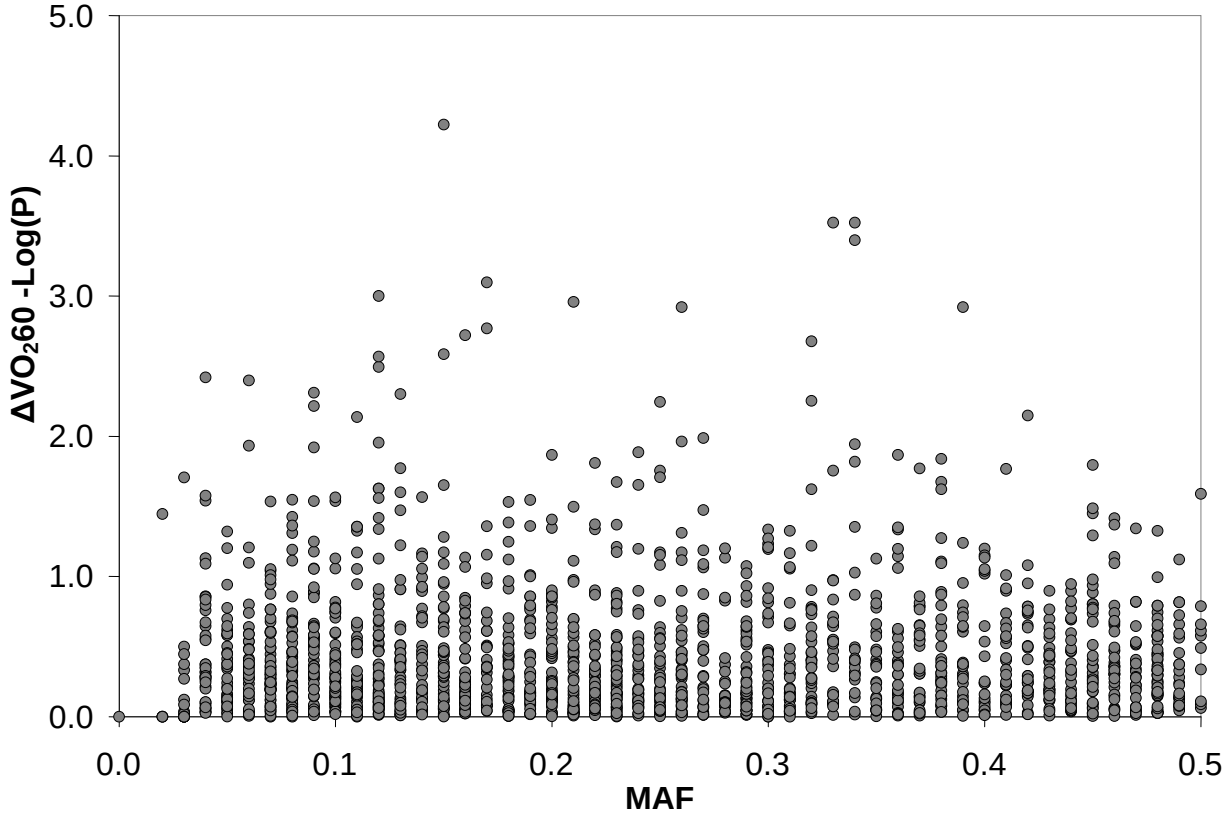
**FIGURE CAPTION:**

The top panel is a scatter plot of –Log(P) for Hardy-Weinberg Equilibrium (HWE) test on X-axis and –Log(P) for association (QTDT) test of $\Delta VO_2 60$. The minimum HWE P-value is 0.002. Note that HWE is not violated for any of the genetic associations that tested significant.

The bottom panel is a similar plot showing the frequency distribution of the minor allele frequencies (MAF) on the X-axis and –Log(P) for association (QTDT) test of $\Delta VO_2 60$ on the Y-axis. Again, the MAF are typically in the middle ranges for the most significant genetic association tests.

**Hardy Weinberg Equilibrium (HWE) and Minor Allele Frequency (MAF) Plots**



## Online Resource 3:

*Genome-Wide Linkage Analysis for Response to Exercise Training in the HERITAGE Family Study for a Measure of Submaximal Exercise Capacity: Delta $VO_2$ at 60% of maximum ($\Delta VO_2 60$)*

**Genotypes:** The genome-wide linkage panel included 674 markers covering 22 autosomes. See Chagnon et al. (2000) for PCR conditions and genotyping methods. DNA sequencers from LI-COR were used to detect PCR products and genotypes were scored semi-automatically using the SAGA software. Mendelian transmission was checked for each marker and incompatible markers were re-genotyped (10%). Microsatellite markers mainly were selected from the Marshfield panel version 8a. Some restriction fragment length polymorphisms (RFLPs) were integrated, including candidate genes relevant for HERITAGE phenotypes. Map locations of the markers were taken from Build 35 of the National Center for Biotechnology Information (NCBI) physical map.

**Figure 1** shows the marker distribution across chromosomes for number of markers per chromosome (top left), maximum length per chromosome (bottom left), and density (top right). There was a mean spacing of 4.2 Mb on the physical map.

Analyses were performed using the computer program Merlin (Abecasis et al., 2002) under the regression-based procedure for linkage analysis that uses trait-squared sums and differences to predict IBD sharing between relative pairs (Sham et al., 2002).

**Figure 2** shows the linkage results for $\Delta VO_2 60$. The figure provides separate panels for each of the 22 autosomes (panels 1-22). Each panel depicts the marker locations along the X-axis and the LOD scores on the Y-axis.

As shown, the strongest linkage is seen on chromosome 13 (LOD score of 3.18, P-value = 0.0000649 at D13S787). The next-best signals are on chromosomes 4 (LOD score of 1.553, P = 0.0037, D4S403) and 15 (LOD score of 1.582, P = 0.0034, D15S120). These latter two signals are not significant at the genome-wide level. Details on Chromosome 13 are in **Figure 3**.

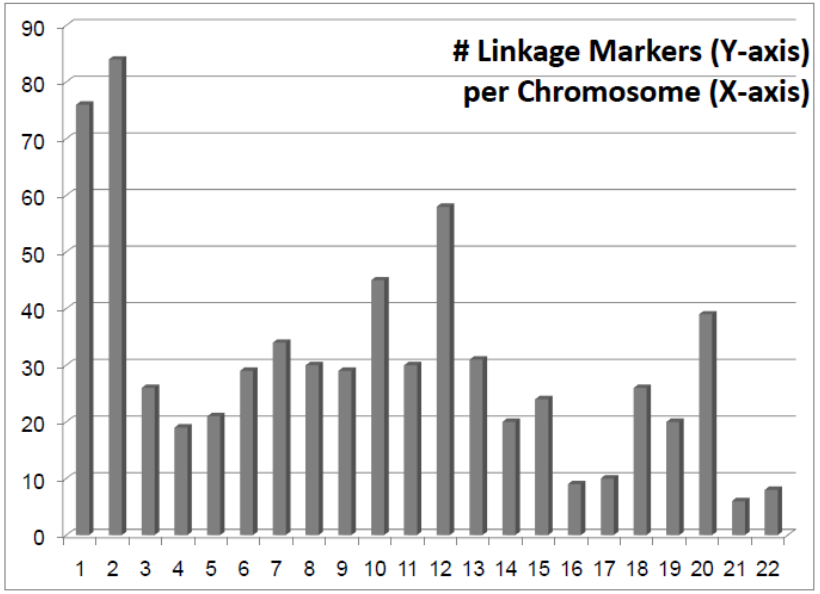
**Figure 3** is an enlarged version of the chromosome 13 LOD score plot. The vertical axis shows the marker location and the horizontal axis is the linkage LOD scores. The ideogram with banding patterns for chromosome 13 is shown on the far left, and the figure inset (bottom right) represents the magnified portion between ~20 and ~30 Mb. The vertical dashed lines on the inset figure (at 21.1 and 29.1 Mb) represent a 2-LOD drop interval.

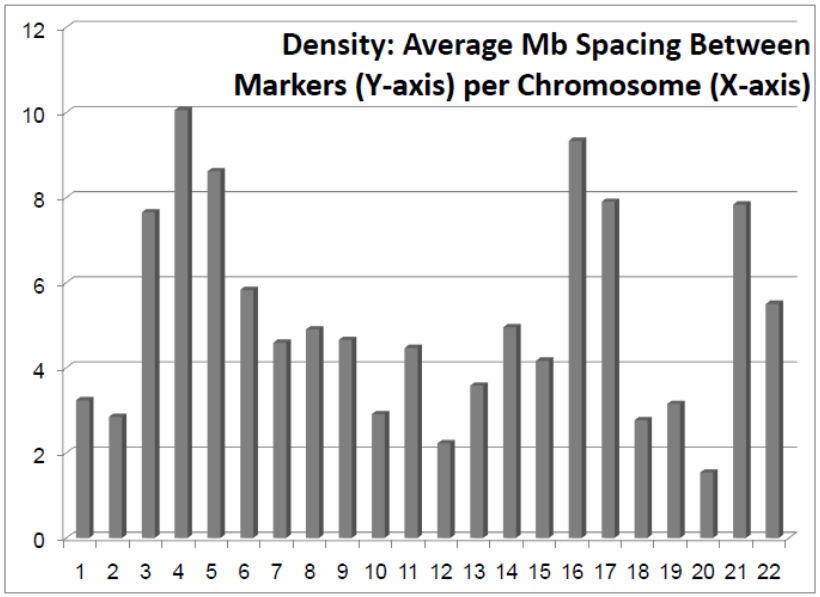
Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97-101.

Chagnon YC, Borecki IB, Perusse L, Roy S, Lacaille M, Chagnon M, Ho-Kim MA, Rice T, Province MA, Rao DC, Bouchard C. Genomewide search for genes related to the fat-free body mass in the Quebec family study. Metabolism 2000;49:203-207.

Sham PC, Purcell S, Cherny SS, Abecasis GR. Powerful regression-based quantitative-trait linkage analysis of general pedigrees. Am J Hum Genet 2002;71:238-253.

**Figure 1:** Marker Distributions



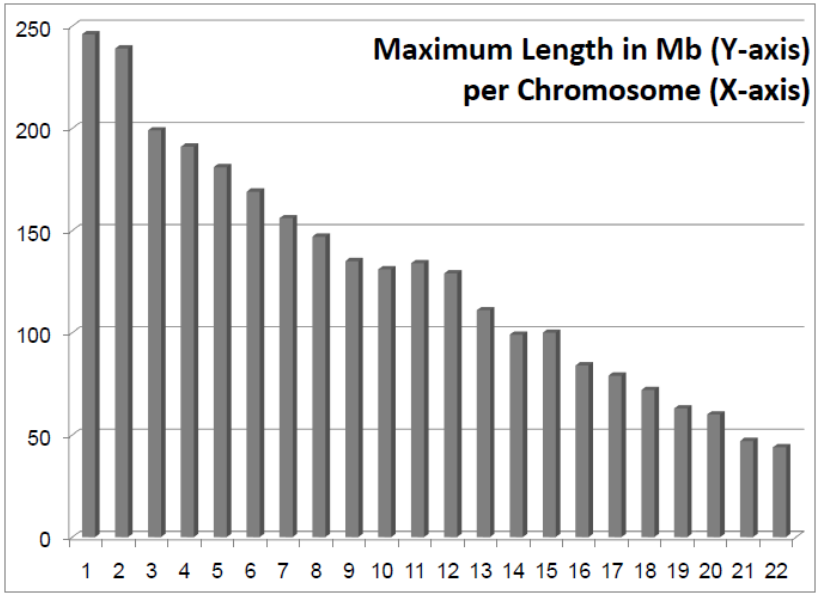



**Figure2:** Linkage Results

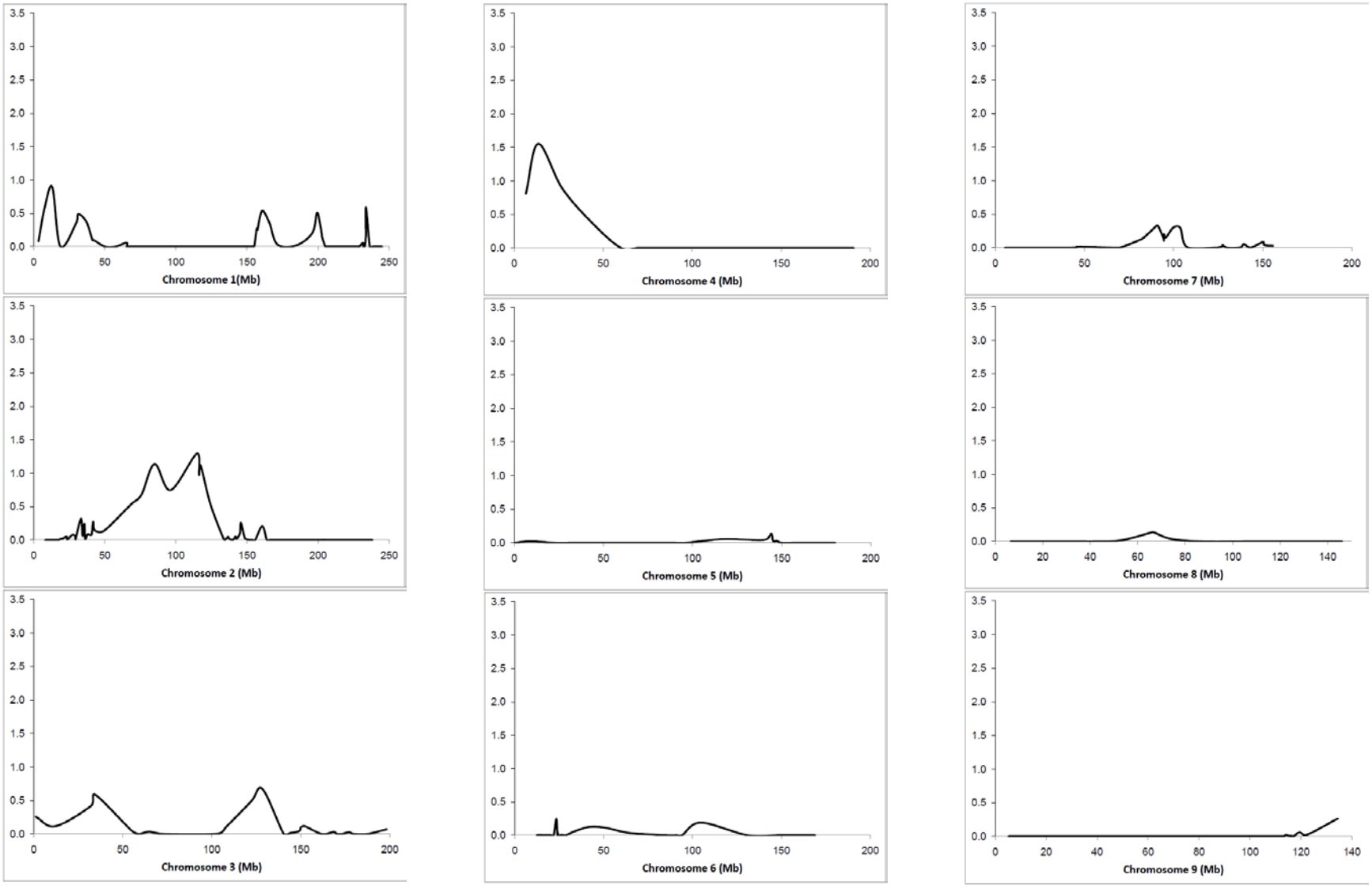


**Figure2:** Linkage Results

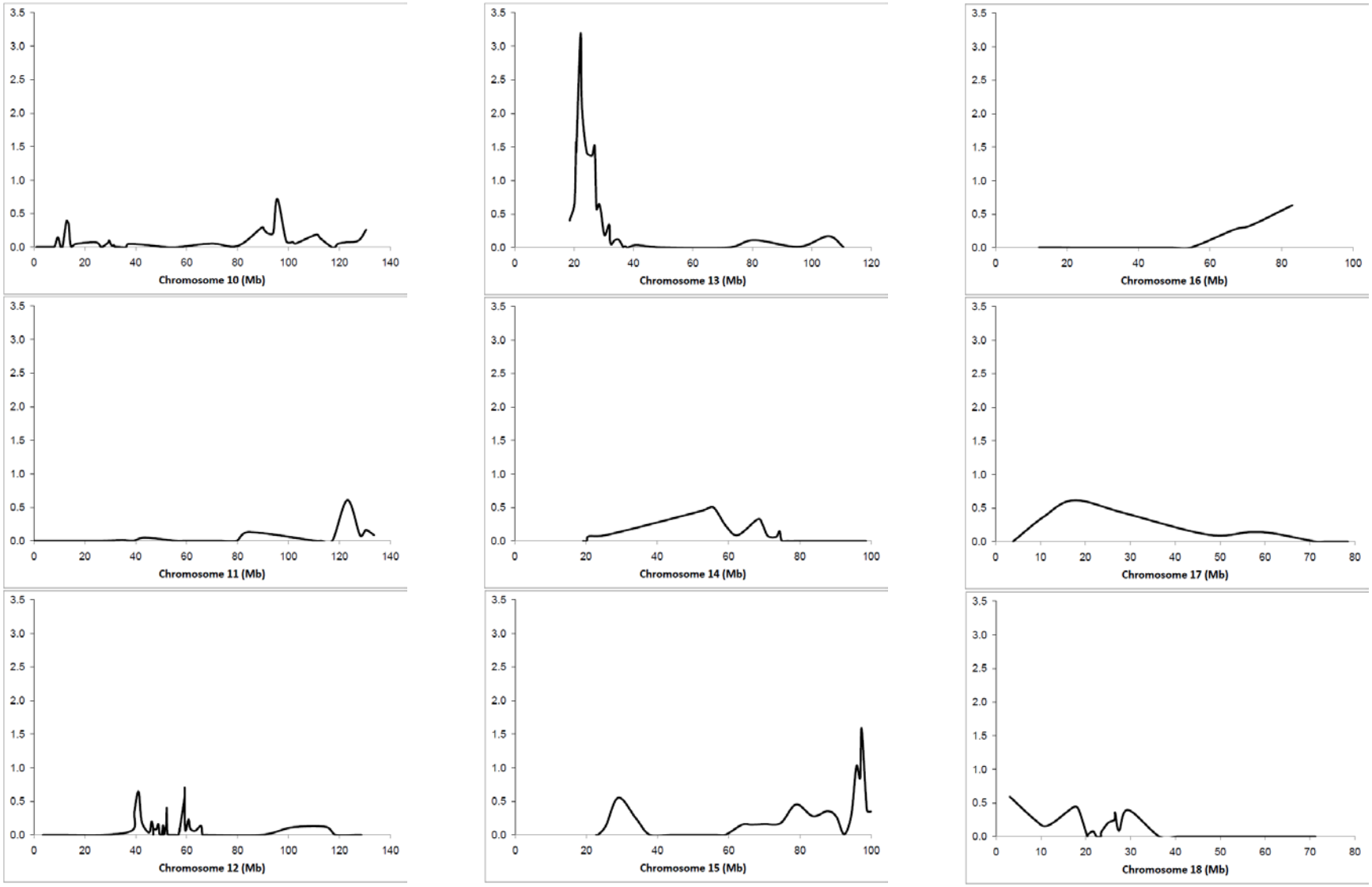


**Figure2:** Linkage Results

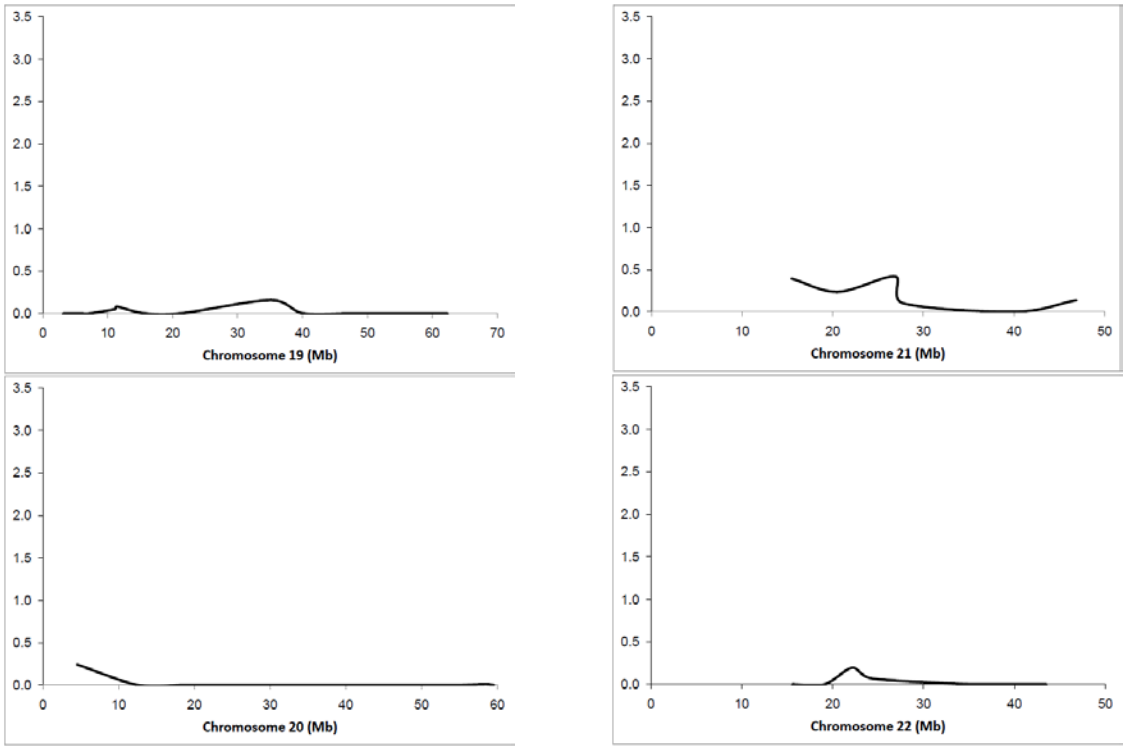
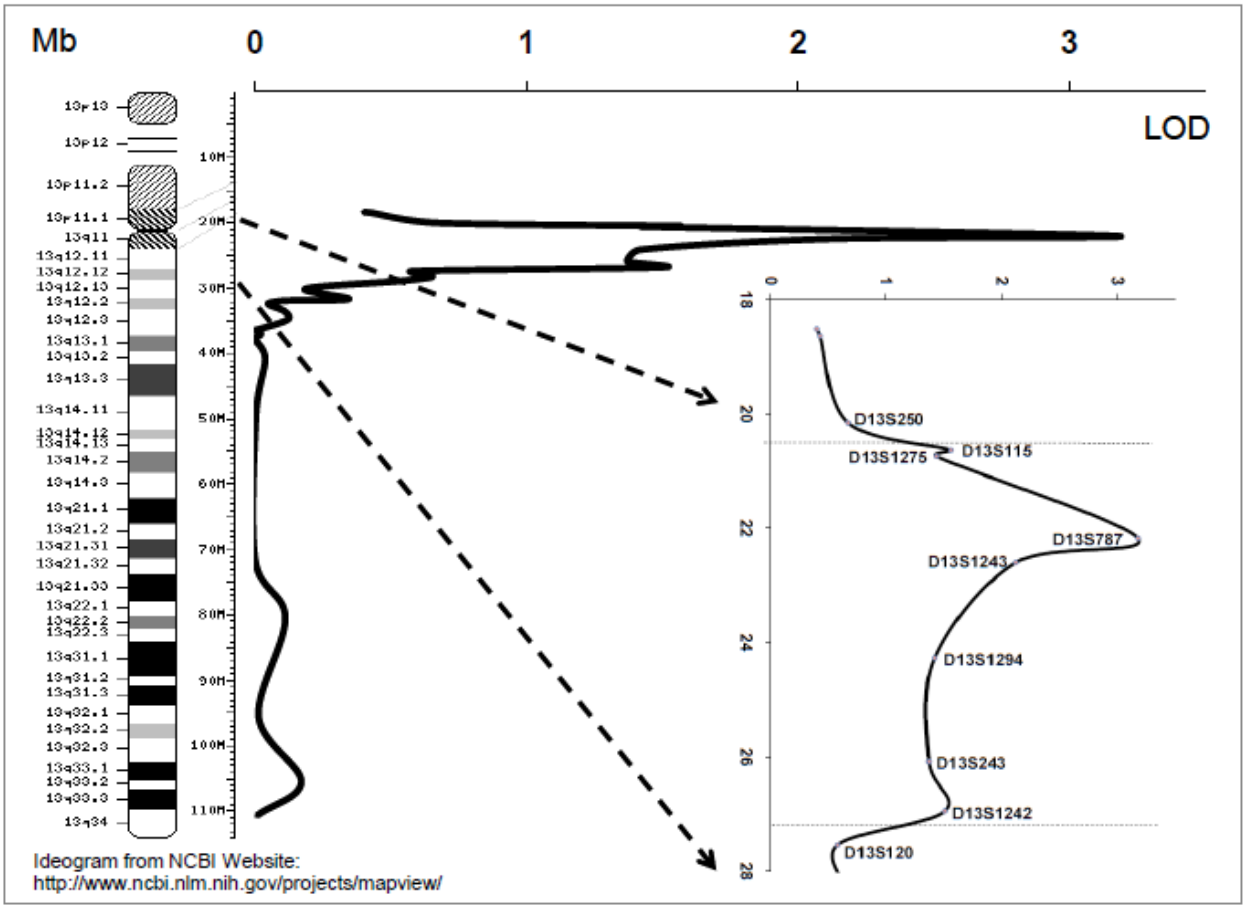


**Figure3:** Chromosome 13 Details



**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions

Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | | | | | 1 | | RS7332465 | 21.567707 | | | Green = SNPs in Haplotypes |
| 0.5 | | | | | 1 | | RS1408051 | 21.725752 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS2025302 | 21.637748 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS2208999 | 21.638245 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS4523799 | 21.639300 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS9578443 | 21.641154 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS11148350 | 21.641718 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS17378638 | 21.644539 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS9316815 | 21.644953 | | | Outside of designated region |
| 0.5 | | | | | 1 | | RS7325032 | 21.645054 | | | Outside of designated region |
| 1 | 21,982k | 22,228k | 13q12.11 | | | 1 | RS9578482 | 22.068536 | | | Blue for single-SNP analysis |
| 1 | | | | | | 1 | RS7342452 | 22.109944 | | | |
| 1 | | | | | | 1 | RS4436648 | 22.118204 | | | |
| ***1*** | | | | | | ***1*** | ***RS9578484*** | ***22.121090*** | | ***Near FTHL7 46,911 bp*** | ***P = 0.002 ($VO_2$) 0.01 (WORK)*** |
| 1 | | | | | | 1 | RS2334946 | 22.126060 | | | |
| 1 | | | | | | 1 | RS1034200 | 22.126691 | | | |
| 1 | | | | | | 1 | RS2151499 | 22.127153 | | | |
| 1 | | | | | | 1 | RS9550855 | 22.134792 | | | |
| 1 | | | | | | 1 | RS9510292 | 22.149737 | | | |
| ***1*** | | | | | | ***1*** | ***RS9506903*** | ***22.150146*** | | ***Near FTHL 17,855 bp*** | ***P = 2.0x10$^{-4}$ ($VO_2$ and WORK)*** |
| 1 | | | | | | 1 | RS17266994 | 22.155044 | | | |
| 1 | | | | 1 | | | FTHL7 | 22.168001 | 22.168910 | ferritin, heavy polypeptide-like 7 | |
| 1 | | | | 2 | | | LOC646164 RP11-274P12.2 | 22.174879 | 22.178345 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 39 pseudogene | |
| 2 | 22.279k | 22.712k | 13q12.12 | 3 | | | LOC100131224 | 22.275516 | 22.275767 | DISCONTINUED: similar to PRO1477 | |
| 2 | | | | 4 | | | LOC401730 RP11-363G2.2 | 22.308934 | 22.310381 | inositol polyphosphate multikinase pseudogene | |
| 2 | | | | 5 | | | LOC100129167 | 22.322179 | 22.322730 | similar to Rieske (Fe-S) domain containing | |
| 2 | | | | 6 | | | LOC646201 | 22.369480 | 22.370661 | brain abundant, membrane attached signal protein 1 pseudogene | |
| ***2*** | | | | | | ***1*** | ***RS4770336*** | ***22.378267*** | | ***Near LOC646201 7,606 bp Near LOC646208 9,523 bp*** | ***P = 0.003 ($VO_2$) 0.02 (WORK)*** |
| 2 | | | | | | 1 | RS971206 | 22.381330 | | | |
| 2 | | | | 7 | | | LOC646208 | 22.387790 | 22.389285 | nuclear undecaprenyl pyrophosphate synthase 1 pseudogene | |
| ***2*** | | | | | | ***1*** | ***RS9506969*** | ***22.394997*** | | ***Near LOC646208 5,712 bp*** | ***P =0.3 ($VO_2$) 0.02 (WORK)*** |
| 2 | | | | | | 1 | RS9510428 | 22.397333 | | | |
| ***2*** | | | | | | ***1*** | ***RS2019091*** | ***22.437060*** | | ***Near LOC646208 47,775*** | ***P = 0.002 ($VO_2$ and WORK)*** |
| 2 | | | | | 2 | | RS9580493 | 22.459107 | | | |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | | | | | 2 | | RS9510493 | 22.459759 | | | |
| 2 | | | | | 2 | | RS9552790 | 22.464147 | | | |
| 2 | | | | | 2 | | RS9578521 | 22.471959 | | | |
| 2 | | | | | 2 | | RS9506995 | 22.473155 | | | |
| 2 | | | | | | 1 | RS9510516 | 22.476548 | | | Both single-SNP and haplotype |
| 2 | | | | | 2 | | RS9510516 | 22.476548 | | | Both single-SNP and haplotype |
| 2 | | | | | 2 | | RS17078046 | 22.484251 | | | |
| 2 | | | | | 2 | | RS1926845 | 22.484809 | | | |
| 2 | | | | | 2 | | RS9550895 | 22.485273 | | | |
| 2 | | | | | | 1 | RS9580513 | 22.488702 | | | |
| 2 | | | | | 3 | | RS9552800 | 22.497673 | | | |
| 2 | | | | | 3 | | RS7998662 | 22.497986 | | | |
| 2 | | | | | 3 | | RS9550898 | 22.498118 | | | |
| 2 | | | | | 3 | | RS7993846 | 22.498969 | | | |
| 2 | | | | | 3 | | RS9510528 | 22.499288 | | | |
| 2 | | | | | 3 | | RS7999743 | 22.499733 | | | |
| 2 | | | | | 3 | | RS17357143 | 22.504391 | | | |
| 2 | | | | | 3 | | RS12430695 | 22.505940 | | | |
| 2 | | | | | | 1 | RS17078161 | 22.548298 | | | |
| ***2*** | | | | | | ***1*** | ***RS7335200*** | ***22.555518*** | | ***Near LOC100130029 50,899*** | ***P = 0.002 ($VO_2$) 0.007 (WORK)*** |
| 2 | | | | | | 1 | RS7319068 | 22.579661 | | | |
| 2 | | | | 8 | | | LOC100130029 | 22.606417 | 22.606706 | similar to high mobility group AT-hook 1 | |
| 2 | | | | | 4 | | RS1415130 | 22.651879 | | | |
| 2 | | | | | 4 | | RS4770402 | 22.652744 | | | |
| 2 | | | | 9 | | | SGCG | 22.653091 | 22.797304 | sarcoglycan, gamma (35kDa dystrophin-associated glycoprotein) | This gene encodes gamma-sarcoglycan, one of several sarcolemmal transmembrane glycoproteins that interact with dystrophin. The dystrophin-glycoprotein complex (DGC) spans the sarcolemma and is comprised of dystrophin, syntrophin, alpha- and beta-dystroglycans and sarcoglycans. The DGC provides a structural link between the subsarcolemmal cytoskeleton and the extracellular matrix of muscle cells. Defects in the encoded protein can lead to early onset autosomal recessive muscular dystrophy, in particular limb-girdle muscular dystrophy, type 2C (LGMD2C). |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions

Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | | | | | 4 | | RS4770403 | 22.653127 | | | 5' UTR |
| 2 | | | | | 4 | | RS3794370 | 22.654627 | | | Intron |
| 2 | | | | | 4 | | RS512444 | 22.656237 | | | Intron |
| 2 | | | | | 4 | | RS7322327 | 22.657216 | | | Intron |
| 2 | | | | | 4 | | RS3794368 | 22.659882 | | | Intron |
| 2 | | | | | 4 | | RS578196 | 22.660272 | | | Intron |
| 2 | | | | | 4 | | RS12871155 | 22.661240 | | | Intron |
| 3 | 22.735k | 23.767k | 13q12.12 | 10 | | | SACS | 22.800965 | 22.905841 | spastic ataxia of Charlevoix-Saguenay (sacsin) | This gene consists of nine exons including a gigantic exon spanning more than 12.8k bp. It encodes the sacsin protein, which includes a UBQ region at the N-terminus, a HEPN domain at the C-terminus and a DnaJ region upstream of the HEPN domain. The gene is highly expressed in the central nervous system, also found in skin, skeletal muscles and at low levels in the pancreas. Mutations in this gene result in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), a neurodegenerative disorder characterized by early-onset cerebellar ataxia with spasticity and peripheral neuropathy. Alternatively spliced transcript variants encoding different isoforms have been found, but the full-length nature of these variants has not been determined. May function in chaperone-mediated protein folding |
| ***3*** | | | | | | ***1*** | ***RS9510730*** | ***22.892263*** | | | ***Intron P =0.003 ($VO_2$) 0.08 (WORK) 0.04 (Q)*** |
| 3 | | | | | | 1 | RS17330522 | 22.962269 | | | |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | | | | 11 | | | TNFRSF19 | 23.042723 | 23.148232 | tumor necrosis factor receptor superfamily, member 19 | The protein encoded by this gene is a member of TNF-receptor superfamily. This receptor is highly expressed during embryonic development. It has been shown to interact with TRAF family members, and to activate JNK signaling pathway when overexpressed in cells. This receptor is capable of inducing apoptosis by a caspase-independent mechanism, and it is thought to play an essential role in embryonic development. Alternatively spliced transcript variants encoding distinct isoforms have been described. Can mediate activation of JNK and NF-kappa-B. May promote caspase-independent cell death |
| 3 | | | | 12 | | | MIPEP | 23.202328 | 23.361559 | mitochondrial intermediate peptidase | The product of this gene performs the final step in processing a specific class of nuclear-encoded proteins targeted to the mitochondrial matrix or inner membrane. This protein is primarily involved in the maturation of oxidative phosphorylation (OXPHOS)-related proteins. This gene may contribute to the functional effects of frataxin deficiency and the clinical manifestations of Friedreich ataxia. Cleaves proteins, imported into the mitochondrion, to their mature size |
| ***3*** | | | | | | ***1*** | ***RS9551012*** | ***23.308057*** | | | ***Intron P =0.004 ($VO_2$) 0.03 (WORK)*** |
| 3 | | | | 13 | | | PCOTH | 23.361028 | 23.364242 | prostate collagen triple helix | |
| 3 | | | | 14 | | | LOC387911 C1QTNF9B | 23.363428 | 23.369125 | similar to hypothetical protein MGC48915; C1q and tumor necrosis factor related protein 9B | |
| 3 | | | | 15 | | | FLJ46358 | 23.413070 | 23.420113 | FLJ46358 protein | |
| 3 | | | | | 5 | | RS12876596 | 23.419485 | | | Intron |
| 3 | | | | | 5 | | RS2147995 | 23.423604 | | | |
| 3 | | | | | 5 | | RS2765114 | 23.426164 | | | |
| 3 | | | | | 5 | | RS2765119 | 23.428895 | | | |
| 3 | | | | | 5 | | RS2765122 | 23.429398 | | | |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | | | | | 5 | | RS942873 | 23.429781 | | | |
| 3 | | | | | 5 | | RS1886796 | 23.430522 | | | |
| 3 | | | | | 5 | | RS4769296 | 23.431038 | | | |
| 3 | | | | | 5 | | RS7330124 | 23.436001 | | | |
| 3 | | | | | 6 | | RS2810699 | 23.464156 | | | |
| 3 | | | | | 6 | | RS2810700 | 23.465372 | | | |
| 3 | | | | | 6 | | RS2765158 | 23.467307 | | | |
| 3 | | | | | 6 | | RS7338032 | 23.467748 | | | |
| 3 | | | | | 6 | | RS724932 | 23.468111 | | | |
| 3 | | | | | 6 | | RS4770541 | 23.468384 | | | |
| 3 | | | | | 6 | | RS2182228 | 23.472686 | | | |
| 3 | | | | | 6 | | RS9510987 | 23.473243 | | | |
| 3 | | | | | 6 | | RS17079726 | 23.473370 | | | |
| 3 | | | | | 6 | | RS17079728 | 23.473845 | | | |
| 3 | | | | | 6 | | RS12429956 | 23.476832 | | | |
| 3 | | | | | | 1 | RS9553166 | 23.565701 | | | |
| 3 | | | | | | 1 | RS17080058 | 23.582320 | | | |
| 3 | | | | | 7 | | RS10735697 | 23.584077 | | | |
| 3 | | | | | 7 | | RS1409032 | 23.584133 | | | |
| 3 | | | | | 7 | | RS1409030 | 23.584523 | | | |
| 3 | | | | | 7 | | RS9551063 | 23.588862 | | | |
| ***3*** | | | | | | ***1*** | ***RS2104257*** | ***23.591742*** | | ***Near LOC100128337 1,421 bp*** | ***P =0.001 ($VO_2$) 0.13 (WORK) 0.035 (Q) Located in "linkage" QTL region Both single-SNP and haplotype*** |
| 3 | | | | | 7 | | RS2104257 | 23.591742 | | | Both single-SNP and haplotype Other SNPs in this haplotype are in pseudogene (below) |
| 3 | | | | | 7 | | RS17354429 | 23.592347 | | | |
| 3 | | | | 16 | | | LOC100128337 RP11-309I15.2 | 23.593163 | 23.594927 | similar to RanBP7/importin 7 pseudogene | The actual Importin 7 (on chromosome 7) functions as nuclear import cofactor, and has been implicated in the control of multiple signal transduction pathways, including the direct nuclear import of the activated (phosphorylated) form of MAP Kinase |
| 3 | | | | | 7 | | RS17458394 | 23.594846 | | | ???? |
| 3 | | | | | 7 | | RS1923909 | 23.597119 | | | |
| 3 | | | | | 7 | | RS1923912 | 23.598152 | | | |
| 3 | | | | | 7 | | RS7317009 | 23.603613 | | | |
| 3 | | | | 17 | | | SPATA13 | 23.632861 | 23.779212 | spermatogenesis associated 13 | |
| 3 | | | | | | 1 | RS9511156 | 23.695786 | | | Intron |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions

Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.5 | | | | 18 | | | PABPC3 | 24.568276 | 24.570704 | Homo sapiens poly(A) binding protein, cytoplasmic 3 (PABPC3), mRNA | |
| 3.5 | | | | | | 1 | RS9511655 | 24.609337 | | | |
| 3.5 | | | | | | 1 | RS9551180 | 24.615548 | | | |
| 3.5 | | | | 19 | | | FAM123A | 24.640672 | 24.643857 | Homo sapiens family with sequence similarity 123A (FAM123A), transcript variant 2, mRNA. | |
| 5 | 25,375k | 25,670k | 13q12.13 | 20 | | | ATP8A2 | 24.844209 | 25.493420 | ATPase, aminophospholipid transporter-like, class I, type 8A, member 2 | |
| 5 | | | | | 8 | | RS9581435 | 25.182795 | | | Intron |
| 5 | | | | | 8 | | RS3132360 | 25.195093 | | | Intron |
| 5 | | | | | 8 | | RS11618222 | 25.197649 | | | Intron |
| 5 | | | | | 8 | | RS9553651 | 25.199550 | | | Intron |
| 5 | | | | | 8 | | RS10781952 | 25.201970 | | | Intron |
| 5 | | | | | 8 | | RS977655 | 25.202569 | | | Intron |
| 5 | | | | | 8 | | RS4770868 | 25.203574 | | | Intron |
| 5 | | | | | 8 | | RS9511901 | 25.203602 | | | Intron |
| 5 | | | | | 8 | | RS9511904 | 25.205457 | | | Intron |
| 5 | | | | | 9 | | RS3783139 | 25.438642 | | | Intron |
| 5 | | | | | 9 | | RS912514 | 25.442369 | | | Intron |
| 5 | | | | | 9 | | RS11616429 | 25.450606 | | | Intron |
| 5 | | | | | 9 | | RS9553692 | 25.451107 | | | Intron |
| 5 | | | | | 9 | | RS975508 | 25.463586 | | | Intron |
| 5 | | | | | 9 | | RS975507 | 25.463645 | | | Intron |
| 5 | | | | 21 | | | TMEM46 SHISA2 | 25.516735 | 25.523198 | transmembrane protein 46; shisa homolog 2 (Xenopus laevis) | |
| 5 | | | | | | 1 | RS1960061 | 25.568374 | | | |
| ***5*** | | | | | | ***1*** | ***RS2038798*** | ***25.570294*** | | ***Near LOC100129595 19,388 bp*** | ***P =0.003 ($VO_2$) 0.004 (WORK)*** |
| 5 | | | | 22 | | | LOC100129595 RP11-380N8.1 | 25.589682 | 25.673433 | hypothetical LOC100129595; prune pseudogene | |
| ***4*** | ***26.180k*** | ***26.780k*** | ***13q12.13*** | | | ***1*** | ***RS2441075*** | ***26.199854*** | | ***Near GPR12 27,487 bp*** | ***P = 0.002 ($VO_2$) 0.01 (WORK)*** |
| ***4*** | | | | | | ***1*** | ***RS9512361*** | ***26.206857*** | | ***Near GPR12 20,484 bp*** | ***P = 0.004 ($VO_2$) 0.02 (WORK)*** |
| 4 | | | | 23 | | | GPR12 | 26.227341 | 26.232922 | G protein-coupled receptor 12 | |
| 4 | | | | | | 1 | RS2182880 | 26.268669 | | | |
| 4 | | | | | | 1 | RS7332384 | 26.271154 | | | |
| 4 | | | | | | 1 | RS17566649 | 26.327202 | | | |
| ***4*** | | | | | | ***1*** | ***RS1576168*** | ***26.400279*** | | ***Near LOC100129306 89,409 bp*** | ***P = 0.001 ($VO_2$) 0.015 (WORK)*** |
| ***4*** | | | | | | ***1*** | ***RS7326591*** | ***26.419810*** | | ***Near LOC100129306 69,878 bp*** | ***P = 0.002 ($VO_2$) 0.0004 (WORK)*** |
| 4 | | | | 24 | | | LOC100129306 RPS20_15_1306 | 26.489688 | 26.490046 | similar to hCG1644665 | |
| 4 | | | | | | 1 | RS2479561 | 26.497574 | | | |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | | | | 25 | | | USP12 | 26.540438 | 26.644029 | ubiquitin specific peptidase 12 | |
| 4 | | | | 26 | | | RPS21 | 26.723692 | 26.728702 | ribosomal protein L21 | Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L21E family of ribosomal proteins. It is located in the cytoplasm. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. |
| 4 | | | | 27 | | | SNORD102 | 26.727201 | 26.727272 | small nucleolar RNA, C/D box 102 | |
| 4 | | | | 28 | | | SNORA27 | 26.727538 | 26.727663 | small nucleolar RNA, H/ACA box 27 | |
| 4 | | | | 29 | | | RASL11A | 26.742464 | 26.745827 | RAS-like, family 11, member A | RASL11A is a member of the small GTPase protein family with a high degree of similarity to RAS (see HRAS, MIM 190020) proteins |
| 4 | | | | | | 1 | RS17523405 | 26.777607 | | | |
| 4 | | | | | | 1 | RS9512630 | 26.816639 | | | |
| 4 | | | | | | 1 | RS1040988 | 26.875300 | | | |
| 4 | | | | | | 1 | RS17085633 | 26.920617 | | | |
| 6 | 27,134k | 27,540k | 13q12.2 | 30 | | | POLR1D | 27.094003 | 27.139548 | polymerase (RNA) I polypeptide D, 16kDa | |
| 6 | | | | 31 | | | LOC100128903 (NPM1P4) | 27.168527 | 27.169341 | hypothetical LOC100128903 (nucleophosmin 1 (nucleolar phosphoprotein B23, numatrin) pseudogene 4) | |
| 6 | | | | | 10 | | RS1326383 | 27.250558 | | | |
| 6 | | | | | 10 | | RS1326384 | 27.254322 | | | |
| 6 | | | | | 10 | | RS1231051 | 27.256840 | | | |
| 6 | | | | | 10 | | RS1231054 | 27.261442 | | | |
| 6 | | | | 32 | | | GSX1 | 27.264780 | 27.266089 | GS homeobox 1 | |
| 6 | | | | | 10 | | RS3742112 | 27.265275 | | | Intron |
| 6 | | | | | 10 | | RS1231064 | 27.277709 | | | |
| 6 | | | | | 10 | | RS10162043 | 27.283757 | | | |
| 6 | | | | | 10 | | RS1616483 | 27.291672 | | | |
| ***6*** | | | | | | ***1*** | ***RS9554166*** | ***27.295996*** | | ***Near GSX1 29,907 bp*** | ***P = 0.002 ($VO_2$) 0.03 (WORK) Single-SNP and haplotype*** |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 10 | | RS9554166 | 27.295996 | | | Single-SNP and haplotype |
| 6 | | | | | 11 | | RS944496 | 27.327194 | | | |
| 6 | | | | | 11 | | RS9512900 | 27.327738 | | | |
| 6 | | | | | 11 | | RS9554178 | 27.329736 | | | |
| 6 | | | | | 11 | | RS4771183 | 27.331193 | | | |
| 6 | | | | | 11 | | RS2154072 | 27.335571 | | | |
| 6 | | | | | 11 | | RS7983897 | 27.339031 | | | |
| 6 | | | | | 11 | | RS1756957 | 27.344006 | | | |
| 6 | | | | | 11 | | RS2996450 | 27.348418 | | | |
| 6 | | | | | 11 | | RS9581925 | 27.356869 | | | |
| 6 | | | | 33 | | | PDX1 | 27.392168 | 27.398451 | pancreatic and duodenal homeobox 1 (GSF; IPF1; IUF1; IDX-1; MODY4; PDX-1; STF-1; PDX1) | The protein encoded by this gene is a transcriptional activator of several genes, including insulin, somatostatin, glucokinase, islet amyloid polypeptide, and glucose transporter type 2. The encoded nuclear protein is involved in the early development of the pancreas and plays a major role in glucose-dependent regulation of insulin gene expression. Defects in this gene are a cause of pancreatic agenesis, which can lead to early-onset insulin-dependent diabetes mellitus (NIDDM), as well as maturity onset diabetes of the young type 4 (MODY4). |
| 6 | | | | 34 | | | LOC432369 ATP5EP2 | 27.417343 | 27.417710 | ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit pseudogene 2 | |
| 6 | | | | 35 | | | LOC100132234 | 27.425669 | 27.427602 | similar to hCG2020055 | |
| 6 | | | | 36 | | | CDX2 | 27.434278 | 27.441317 | caudal type homeobox 2 | The level and beta-cell specificity of insulin gene expression are regulated by a set of nuclear proteins that bind to specific sequences within the promoter of the insulin gene (INS; MIM 176730) and interact with RNA polymerase to activate or repress transcription. The proteins LMX1 (MIM 600298) and CDX3 are homeodomain proteins that bind an A/T-rich sequence in the insulin promoter and stimulate its transcription (German et al., 1994 [PubMed 7698771]) |
| 6 | | | | 37 | | | PRHOXNB | 27.450243 | 27.460774 | parahox cluster neighbor | |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | | | | 38 | | | FLT3 | 27.475411 | 27.572729 | fms-related tyrosine kinase 3 | This gene encodes a class III receptor tyrosine kinase that regulates hematopoiesis. The receptor consists of an extracellular domain composed of five immunoglobulin-like domains, one transmembrane region, and a cytoplasmic kinase domain split into two parts by a kinase-insert domain. The receptor is activated by binding of the fms-related tyrosine kinase 3 ligand to the extracellular domain, which induces homodimer formation in the plasma membrane leading to autophosphorylation of the receptor. The activated receptor kinase subsequently phosphorylates and activates multiple cytoplasmic effector molecules in pathways involved in apoptosis, proliferation, and differentiation of hematopoietic cells in bone marrow. Mutations that result in the constitutive activation of this receptor result in acute myeloid leukemia and acute lymphoblastic leukemia. |

**Online Resource 4:** Listing of Known Genes, Significant Single-SNPs and SNPs Within Haplotypes: The following is provided in this appendix: (A) Table of SNPs and genes in regions 1-6, with Mb locations and descriptions
Blue is for single SNPs, Green is for Haplotypes, and Red is for Genes

| Region | Start | End | Cytogenic | gene | Hap | SNP | Genes / SNPs | Start | End | Description | Function / Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6.5 | | | | 39 | | | FLT1 | 27.857688 | 27.967265 | fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor) | This gene encodes a member of the vascular endothelial growth factor receptor (VEGFR) family. VEGFR family members are receptor tyrosine kinases (RTKs) which contain an extracellular ligand-binding region with seven immunoglobulin (Ig)-like domains, a transmembrane segment, and a tyrosine kinase (TK) domain within the cytoplasmic domain. This protein binds to VEGFR-A, VEGFR-B and placental growth factor and plays an important role in angiogenesis and vasculogenesis. Expression of this receptor is found in vascular endothelial cells, placental trophoblast cells and peripheral blood monocytes. Multiple transcript variants encoding different isoforms have been found for this gene. Isoforms include a full-length transmembrane receptor isoform and shortened, soluble isoforms. The soluble isoforms are associated with the onset of pre-eclampsia. Receptor for VEGF, VEGFB and PGF. Has a tyrosine-protein kinase activity. The VEGF-kinase ligand/receptor signaling system plays a key role in vascular development and regulation of vascular permeability. Isoform SFIT1 may have an inhibitory role in angiogenesis. |
| 6.5 | | | | | | 1 | RS2281827 | 27.899721 | | | Intron |
| 6.5 | | | | | | 1 | RS4771249 | 27.911414 | | | Intron |
| 6.5 | | | | | | | LOC646630 hCG_1790806 | 28.681732 | 28.682013 | coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene | |
| 6.5 | | | | 40 | | | PAN3 | 28.748417 | 28.869472 | PAN3 poly(A) specific ribonuclease subunit homolog (S. cerevisiae) | Functions in cytoplasmic mRNA decay. As part of the Pan nuclease complex, recruits polyadenylate-binding protein which in turn stimulates PAN2 nuclease activity |
| 6.5 | | | | | | 1 | RS11840712 | 28.787442 | | | Intron |
| 6.5 | | | | 41 | | | LOC341784 | 29.172954 | 29.174341 | eukaryotic initiation factor 4AI pseudogene | |

**Online Resource 5**: Summary of association results.

| Region | Total | | | Genotyped SNPs | | | | | | Haplotype | | Gene | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Start bp | End bp | N bp | N | Density | $VO_2$60 | | Correlated Traits | | N | SNPs | N | SNPs in Genes | Genes w / ≥1 SNP |
| | | | | | | N | % | N | % | | | | | |
| 1 | 21982611 | 22227181 | 244571 | 55 | 4446.75 | 15 | 27.27 | 13 | 11.82 | 0 | 0 | 2 | 0 | 0 |
| 2 | 22269773 | 22711813 | 442041 | 187 | 2363.86 | 22 | 11.76 | 22 | 5.88 | 3 | 26 | 7 | 7 | 1 |
| 3 | 22735627 | 23766138 | 1030512 | 333 | 3094.63 | 25 | 7.51 | 37 | 5.56 | 3 | 30 | 8 | 5 | 5 |
| 4 | 26180602 | 26777607 | 597006 | 166 | 3596.42 | 22 | 13.25 | 39 | 11.75 | 0 | 0 | 7 | 0 | 0 |
| 5 | 25377364 | 25660364 | 283001 | 86 | 3290.71 | 8 | 9.30 | 13 | 7.56 | 2 | 15 | 3 | 15 | 1 |
| 6 | 27134975 | 27537249 | 402275 | 91 | 4420.60 | 7 | 7.69 | 6 | 3.30 | 2 | 18 | 9 | 1 | 1 |

**Abbreviations / Definitions:** bp is base pair; N bp is number of base pairs, SNPs, Haplotypes or Genes; Density is average number of base pairs per typed SNP in the region; N and % are number and percent of typed SNPs that show significant (adjusted $P \leq 0.05$) association with submaximal capacity traits ($\Delta VO_2$60 or correlated traits ΔWORK60 and ΔQ60); Haplotype SNPs is number of SNPs contributing to any haplotype in the region; SNPs in Genes is number of significant (adjusted $P \leq 0.05$) SNPs (single SNP or haplotype analysis) located within genes; and Genes w / > 1 SNP represents the number of genes that have at least 1 significant (adjusted $P \leq 0.05$) SNP. Regions 1 through 4 have a preponderance of single-SNP signals. However, regions 2 and 3 have the best density of genotyped SNPs, more haplotypes, and more known genes.

**Online Resource 6**. Results of stepwise regression analysis for predictors of $\Delta VO_2 60$ using the SNPs that were significant in single-SNP (N=16) and in haplotype analyses (N=27)

A. Stepwise regression model for predictors of $\Delta VO_2 60$ using significant SNPs from single SNP association analyses

| SNP | Partial $R^2$ | Model $R^2$ | C(p) | F value | P value |
|---|---|---|---|---|---|
| rs9506903 | 0.0373 | 0.0373 | 74.3519 | 17.49 | <.0001 |
| rs7326591 | 0.0278 | 0.0652 | 61.2153 | 13.4 | 0.0003 |
| rs9554166 | 0.0206 | 0.0858 | 52.0143 | 10.12 | 0.0016 |
| rs9510730 | 0.0183 | 0.1041 | 44.0576 | 9.16 | 0.0026 |
| rs9551012 | 0.0185 | 0.1226 | 36.0074 | 9.42 | 0.0023 |
| rs2441075 | 0.0154 | 0.138 | 29.6399 | 7.96 | 0.005 |
| rs9578484 | 0.0127 | 0.1507 | 24.7299 | 6.66 | 0.0102 |
| rs9506969 | 0.0165 | 0.1671 | 17.7793 | 8.78 | 0.0032 |
| rs7335200 | 0.0128 | 0.18 | 12.8087 | 6.93 | 0.0088 |
| rs1576168 | 0.0073 | 0.1873 | 10.836 | 3.97 | 0.0468 |

B. Stepwise regression model for predictors of $\Delta VO_2 60$ using significant SNPs from haplotype association analyses

| SNP | Partial $R^2$ | Model $R^2$ | C(p) | F value | P value |
|---|---|---|---|---|---|
| rs9510516 | 0.0214 | 0.0214 | 9.4666 | 9.54 | 0.0021 |
| rs9554166 | 0.0148 | 0.0362 | 4.7311 | 6.71 | 0.0099 |
| rs9506995 | 0.011 | 0.0472 | 1.7429 | 5.01 | 0.0256 |
| rs9552790 | 0.009 | 0.0562 | -0.3495 | 4.14 | 0.0424 |
| rs4770403 | 0.0084 | 0.0646 | -2.1507 | 3.87 | 0.0497 |

C. Stepwise regression model for predictors of $\Delta VO_2 60$ using significant SNPs from combined single SNP and haplotype analysis association analyses

| SNP | Partial $R^2$ | Model $R^2$ | C(p) | F value | P value |
|---|---|---|---|---|---|
| rs9506903 | 0.0358 | 0.0358 | 73.134 | 15.83 | <.0001 |
| rs9510516 | 0.0281 | 0.064 | 60.6267 | 12.78 | 0.0004 |
| rs7326591 | 0.0232 | 0.0872 | 50.6453 | 10.79 | 0.0011 |
| rs9578484 | 0.0198 | 0.107 | 42.4579 | 9.36 | 0.0024 |
| rs2104257 | 0.0177 | 0.1247 | 35.3303 | 8.53 | 0.0037 |
| rs9551012 | 0.0142 | 0.1389 | 29.9893 | 6.96 | 0.0086 |
| rs2441075 | 0.0178 | 0.1566 | 22.8352 | 8.84 | 0.0031 |
| rs9510730 | 0.0104 | 0.1671 | 19.4598 | 5.24 | 0.0225 |
| rs9506995 | 0.0102 | 0.1773 | 16.1955 | 5.19 | 0.0233 |
| rs9506969 | 0.0093 | 0.1866 | 13.4021 | 4.77 | 0.0296 |
| rs9554166 | 0.0092 | 0.1958 | 10.641 | 4.78 | 0.0294 |
| rs7332384 | 0.009 | 0.1931 | 10.0457 | 4.64 | 0.0318 |
| rs9552790 | 0.0079 | 0.201 | 7.9687 | 4.12 | 0.0431 |