AJHG, Volume *86*

Supplemental Data

Distinct Variants at *LIN28B* Influence Growth in Height from Birth to Adulthood

Elisabeth Widén, Samuli Ripatti, Diana L. Cousminer, Ida Surakka, Tuuli Lappalainen, Marjo-Riitta Järvelin, Johan G. Eriksson, Olli Raitakari, Veikko Salomaa, Ulla Sovio, Anna-Liisa Hartikainen, Anneli Pouta, Mark I. McCarthy, Clive Osmond, Eero Kajantie, Terho Lehtimäki, Jorma Viikari, Mika Kähönen, Chris Tyler-Smith, Nelson Freimer, Joel N. Hirschhorn, Leena Peltonen, and Aarno Palotie

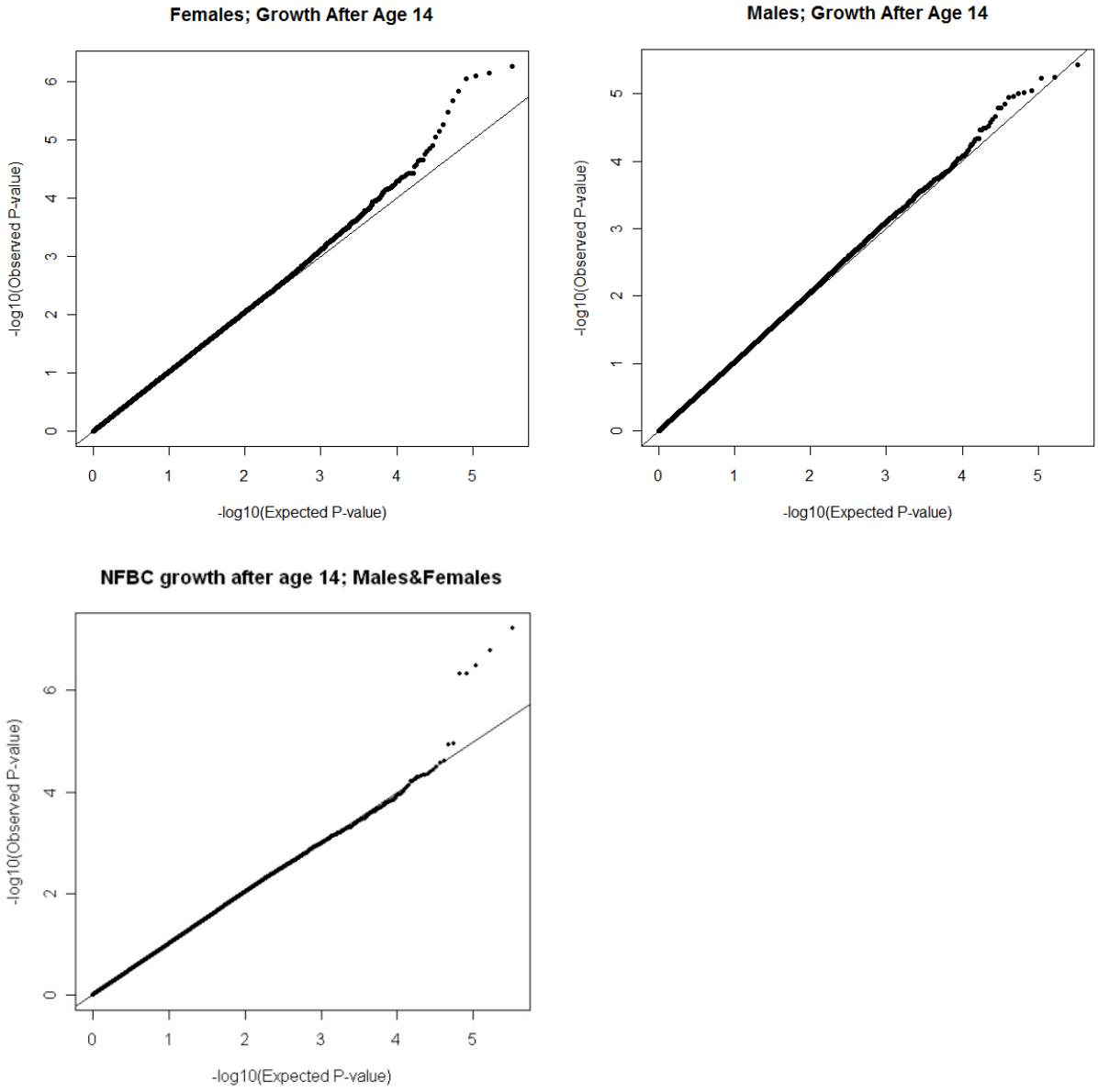

Figure S1. Quantile-Quantile Plots of the Ordered p Value Distributions for the Association Analyses of Pubertal Height Growth in the Northern Finland Birth Cohort 1966 (NFBC1966)

The grey lines represent the expected p-value distribution under the null hypothesis. Ordered observed p-values are shown with solid circles. NFBC1966 study subjects were genotyped on Illumina Infinium 370CNV Duo arrays. All individuals in the study had genotyping call rates >95%. Markers were excluded from the analysis if the call rate was <95%, if there was evidence of distortion of Hardy-Weinberg equilibrium by p-value <0.0001, or if the minor allele frequency was <1%. Applying these exclusion criteria, data was obtained for 329 097 markers. Within-cohort multidimensional scaling (MDS) was carried out using Plink[1].The data were analyzed by linear regression as implemented in PLINK[1] covering the autosomes, assuming additive inheritance and including individual-specific scores of the first two dimensions of the multidimensional scaling IBS analysis as covariates. The phenotype distribution for the increase

in height between ages 14 and 31 was normalized by logarithm-transformation and sex-specific Z-scores computed from the normalized phenotype were used as input in the association analysis. The genomic inflation factor λ was 1.04 in both sexes. The association results from males and females were combined into fixed effects meta-analysis with reciprocal weighting on the square of standard errors of the effect size estimates, using the MetABEL package for the R software (http://mga.bionet.nsc.ru/~yurii/ABEL/, http://www.r-project.org/). Details of the genotyping procedure are described by Sabatti et al (2009).[2] The estimated power for detecting an allele with a effect size of 0.15 with a p-value < 5.0×10^{-7} was 91%, based on 10 000 simulations assuming a sample size of 4 300 and complete linkage disequilibrium to a marker with minor allele frequency of 0.3.

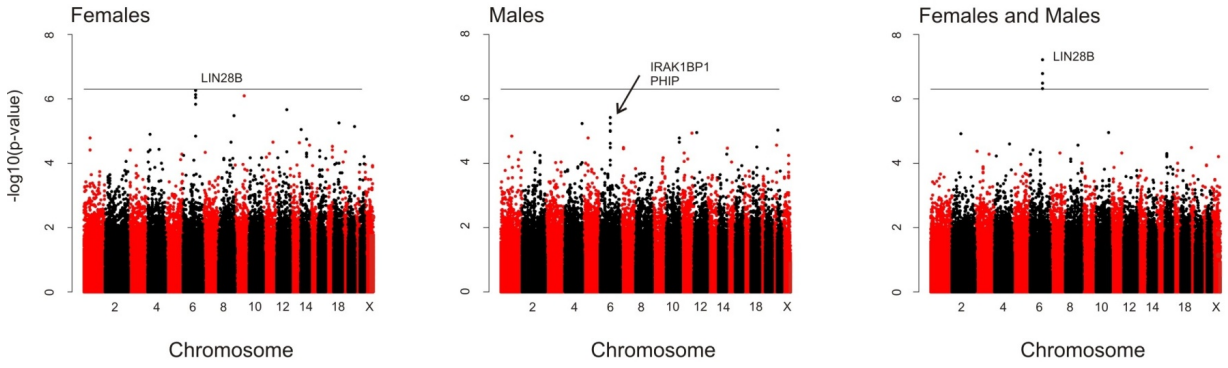

Figure S2. Manhattan Plots for Growth during Late Adolescence in Northern Finland Birth Cohort 1966

The analysis results from analyzing females are shown to the left, the results from analyzing males are shown in the middle panel and the results from the combined analysis of females and males are shown to the right. The horizontal line corresponds to a p-value threshold of 5×10^{-7}. The analysis procedure is described in detail in the legend of Supplementary Figure 1.

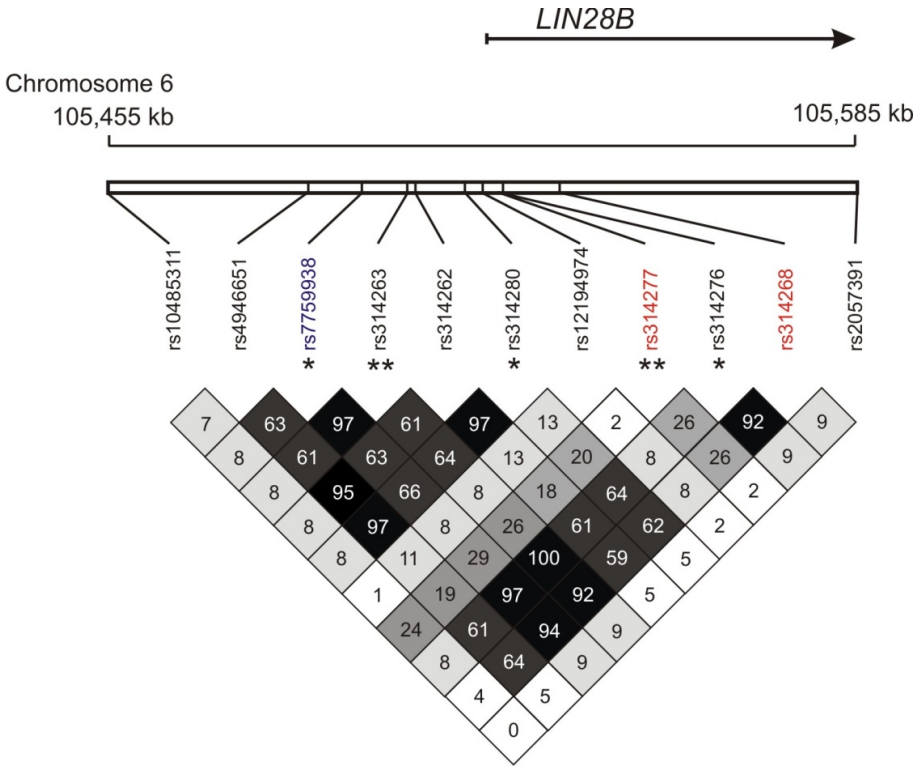

Figure S3. Pairwise Correlations between Markers in the *LIN28B* Region Present on the Illumina Infinium 370CNV Duo Arrays and Markers Previously Associated with Age of Menarche[3-6] (Indicated with an Asterisk)

The figure shows the r^2 calculated with Haploview (http://www.broadinstitute.org/mpg/haploview) based on CEU genotypes from the Hapmap3 dataset. The marker locus rs7759938 is strongly associated with pubertal growth in the current study. Marker loci rs314277, previously associated with height with a p-value of 1.1×10^{-8} in roughly 25 000 individuals[7], and rs314268, previously associated with adult height with a p-value of 7.7×10^{-7} in 49 000 individuals[8], are shown in red font. The marker rs7759938 is well correlated with rs314268 but only partially correlated with rs314277. ** Two markers in incomplete LD best associated with age of menarche in the study by He et al[5] ($p<10^{-12}$).

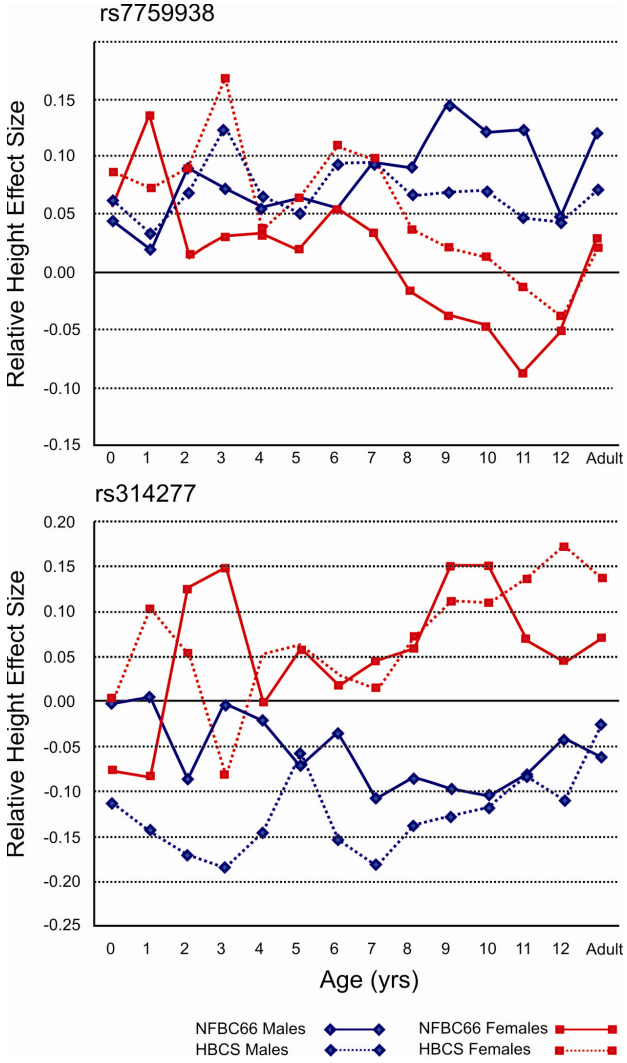

Figure S4. Linear Regression Analysis Birth Length, Childhood Height, and Adult Stature in Northern Finland Birth Cohort 1966 and Helsinki Birth Cohort Study

The sex specific effect sizes of birth length and height at 1 year intervals between age 1 and 12 and adult stature have been calculated including both rs7759938 and rs314277 in the regression model. The results at rs7759938 are shown in the upper panel and rs314277 in the lower panel. The effect alleles are G and A at rs7759938 and rs314277 respectively. Birth length in both cohorts has been adjusted for gestational age.

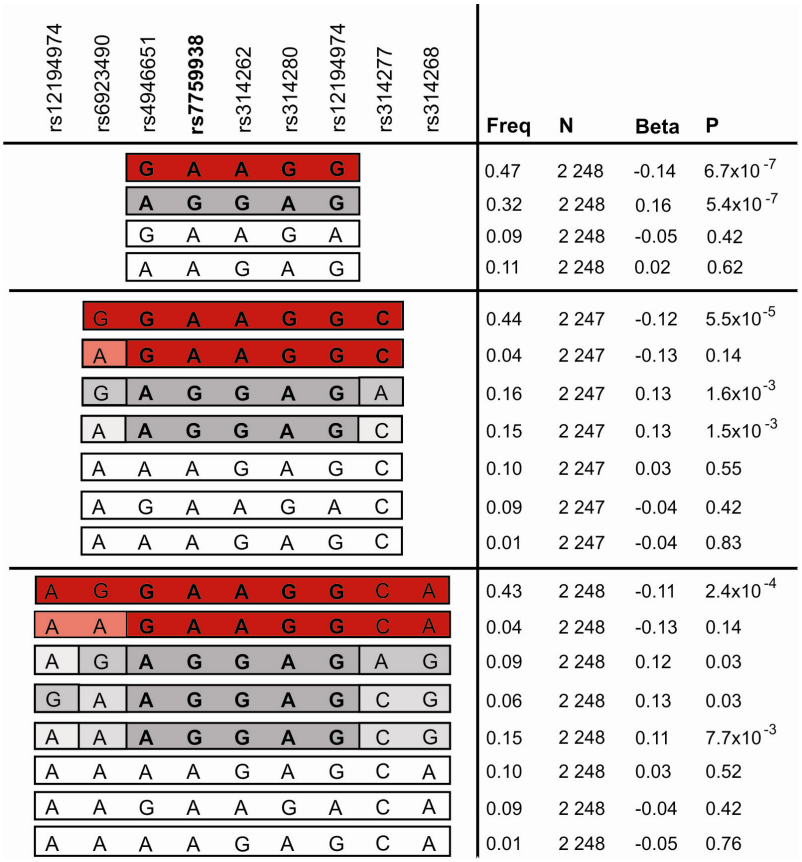

Figure S5. **Haplotypes Associated with Pubertal Growth and Maturation in the *LIN28B* Region**

Haplotypes affecting pubertal growth and maturation in female study subjects from the Northern Finland Birth Cohort 1966 (NFBC1966) are shown. All haplotypes in windows of a fixed number of SNPs and the corresponding haplotype-based association tests with growth after age 14 in NFBC1966 females were determined by Plink[1] using 5-marker, 7-marker and 9-marker windows surrounding rs7759938. The marker map corresponds to markers successfully genotyped using Illumina Infinium 370CNV Duo arrays. The haplotype which contains the G allele of rs7759938 and is associated with later adolescent growth and maturation (grey) in the uppermost window, breaks up into multiple haplotypes immediately upstream and downstream. The most common haplotype (red), associated with earlier adolescent growth and maturation, extends almost intact from approximately 105,455 – 105,599 kb (data not shown). Of note, the association results for the grey haplotype containing the G allele at rs7759938 and the red haplotype containing the A allele at rs7759938 are not independent, because they account for 79% of all haplotypes in the uppermost window.

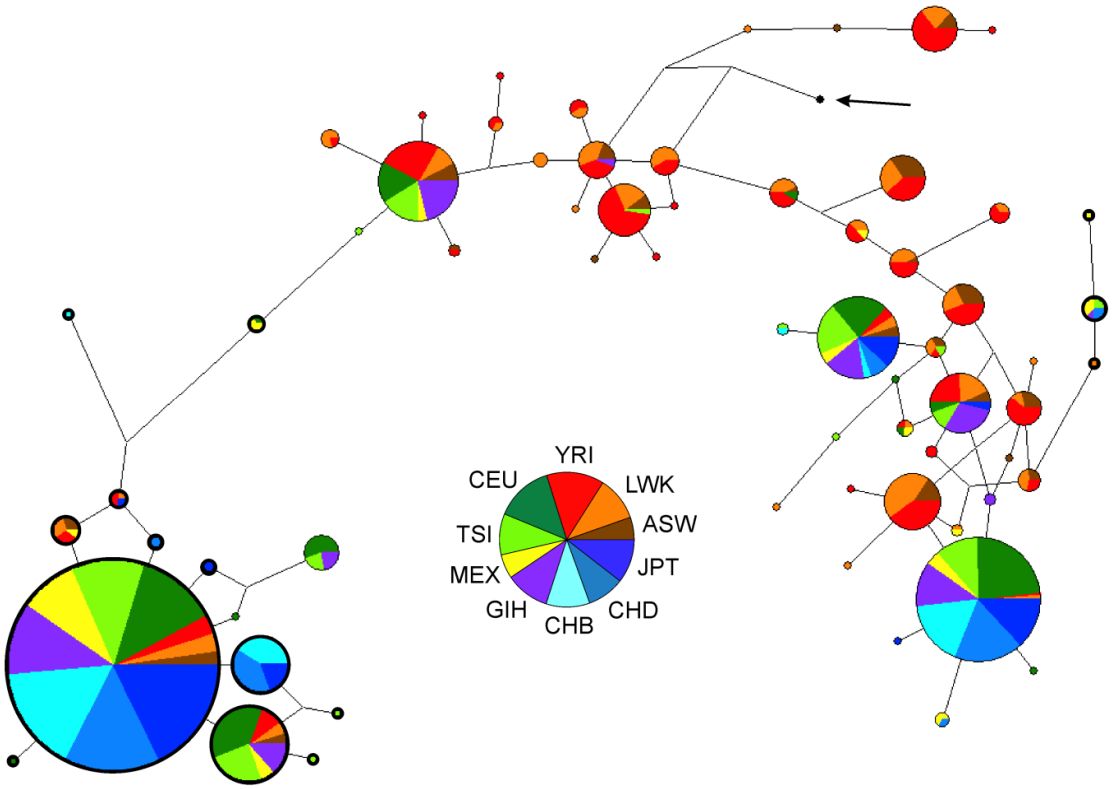

Figure S6. Median-Joining Network of Haplotypes of 22 SNPs in the Region chr6:105,460,000-105,570,000 Constructed with Network 4.5.0.2 (fluxus-engineering.com[9])
Positions are according to NCBI B36 assembly. Ten populations of the HapMap3 dataset were included in the analysis (http://www.sanger.ac.uk/humgen/hapmap3/). The Maasai in Kinyawa Kenya, were excluded from the analysis due to complex family structures. Networks were constructed from several haplotype blocks in the *LIN28B* region, and one of the networks displaying the least recombination is shown. Individual haplotypes each are represented by nodes, the area of which corresponds to the overall haplotype frequency in the whole dataset. The branches connecting the haplotypes denote the SNPs differing between haplotypes. The length of each branch corresponds to the number of SNPs differentiating a haplotype from its neighboring haplotypes. The ancestral haplotype (carrying the chimpanzee allele at each position) is marked by an arrow, and haplotypes surrounded by a bold line carry the haplotype rs7759938A-rs314262A-rs314280G that predisposes to early pubertal growth and maturation. The dataset used in the analysis includes 10 populations of the HapMap3 dataset, with SNPs genotyped using Affymetrix 6.0 and Illumina 1M arrays in samples from Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), Toscani in Italia (TSI), Mexican ancestry in Los Angeles, California (MEX), Gujarati Indians in Houston, Texas (GIH), Han Chinese in Beijing, China (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Japanese in Tokyo, Japan (JPT), Yoruba in Ibadan, Nigeria (YRI), Luhya in Webuye, Kenya (LWK), African ancestry in Southwest USA (ASW). The population key with the relative population frequencies is shown by a pie-chart located in the middle of the figure.

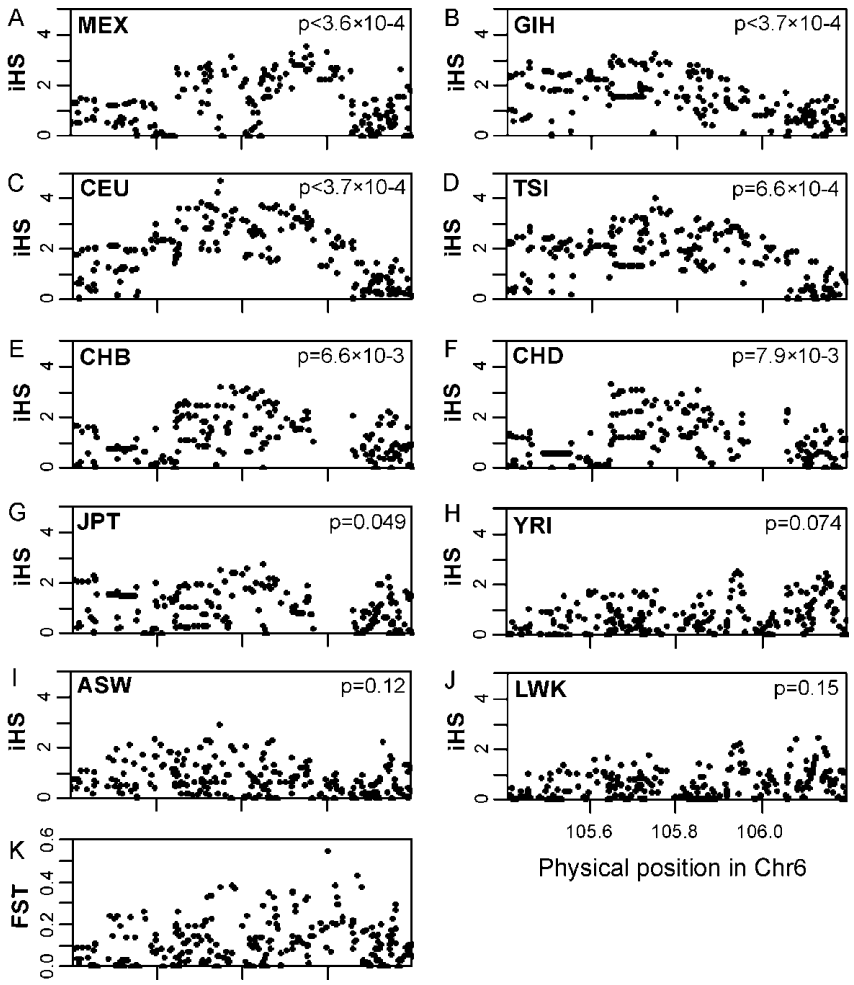

Figure S7. The Landscape of the iHS Statistic (Panels A-J) in the Region chr6:105,400,000-106,200,000

The positions are according to NCBI B36 assembly. Ten populations of the HapMap3 dataset, with SNPs genotyped using Affymetrix 6.0 and Illumina 1M arrays, are included in the analyses (http://www.sanger.ac.uk/humgen/hapmap3/). The physical position is shown on the x-axis and the iHS scores are shown on the y-axis. *LIN28B* is located at 105,511,616 – 105,637,899 bp. The iHS analysis results, obtained from the different populations, are ordered based on the statistical significance of the iHS-score: Residents of Los Angeles, California with Mexican ancestry in (MEX) in panel A, Gujarati Indians in Houston, Texas (GIH) in panel B, Utah residents of Northern and Western European ancestry from the CEPH collection (CEU) in panel C, Toscans in Italy (TSI) in panel D, Han Chinese in Beijing, China (CHB) in panel E, Chinese in Metropolitan Denver, Colorado (CHD) in panel F, Japanese in Tokyo, Japan (JPT) in panel G, Yoruba in Ibadan, Nigeria (YRI) in panel H, African ancestry in Southwest USA (ASW) in panel I, and Luhya in Webuye, Kenya (LWK) in panel J. Haplotypes on chromosome 6 were phased in each population using Beagle[10] and iHS was calculated for all the common SNPs (minor allele frequency >5%) using Sweep[11]. The iHS scores were standardized in frequency bins[11,12]. The F_{ST} statistic (panel K) was calculated across all the populations according to Akey et al[13]. The empirical p-values for the iHS were calculated as by Pickrell et al[14] using the proportion of SNPs in the 100 kb window with |iHS|>2.5 or F_{ST}>0.211 (5th percentile) as the test statistic, a conservative measure when there may be a long recently-selected haplotype.

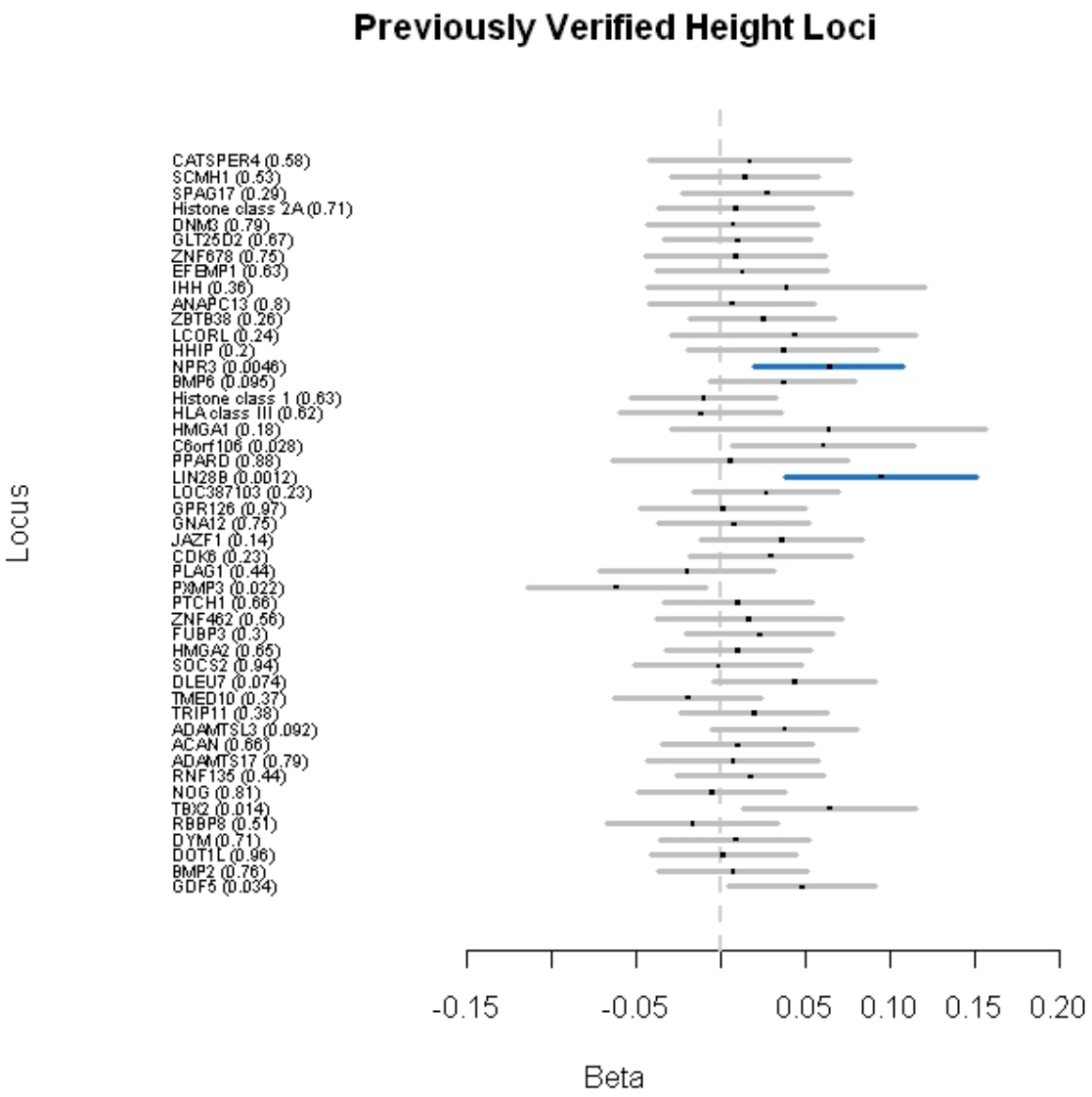

Figure S8. Forest Plot Showing the Association Results (Beta and S.E.) for Growth during Late Adolescence in the Combined Analysis of Males and Females from Northern Finland Birth Cohort Study 1966

Loci with genome-wide significant ($p<5x10^{-8}$) association with adult stature were selected from the review by Lettre et al[15]. A nearby gene is shown to the left and the p-value of the corresponding analysis of adolescent growth is shown in parenthesis. Proxy snps (r2 >0.8) for markers not found on the Illumina chips were determined using the SNP Annotation and Proxy Search from the Broad Institute (http://www.broadinstitute.org). Markers with an association p-value of <0.01 are in blue.

Table S1. Association between Height Growth during Adolescence and Marker rs7759938

		MALES Mean Growth (cm)			FEMALES Mean Growth (cm)		
		9-12	**12-15†**	**15-adult‡**	**9-12**	**12-15†**	**15-adult‡**
NFBC1966	AA	15.7 (2.9)	15.0 (4.1)	15.0 (6.7)	18.1 (3.2)	11.3 (4.0)	3.7 (2.7)
	AG	15.4 (2.5)	14.8 (4.2)	15.7 (6.7)	17.8 (3.1)	11.7 (4.0)	4.3 (2.9)
	GG	15.6 (2.3)	14.4 (4.3)	16.2 (6.9)	17.4 (3.1)	11.9 (4.2)	4.5 (3.2)
	N	863	1253	2068	850	1 205	2 242
YF	AA	16.5 (3.2)	20.1 (3.9)	6.6 (5.0)	19.3 (2.9)	9.8 (4.5)	1.8 (1.6)
	AG	15.8 (2.8)	20.3 (3.6)	7.8 (5.5)	18.7 (3.1)	9.9 (4.3)	2.1 (1.7)
	GG	15.6 (2.3)	19.0 (3.3)	8.0 (6.6)	18.4 (3.3)	11.3 (4.4)	2.4 (2.1)
	N	362	357	570	413	424	664
HBCS	AA	14.1 (2.2)	NA	NA	16.8 (2.8)	NA	NA
	AG	14.1 (2.1)	NA	NA	16.5 (2.9)	NA	NA
	GG	13.9 (2.1)	NA	NA	16.7 (2.7)	NA	NA
	N	658	NA	NA	865	NA	NA
		MALES			**FEMALES**		
Beta (SE)		-0.082 (0.038)	0.064 (0.044)	0.103 (0.025)	-0.099 (0.032)	0.101 (0.037)	0.161 (0.028)
P		0.03	0.14	$4.4x10^{-4}$	0.002	0.006	$8.2x10^{-9}$
N		1 883	1 610	2 638	2 127	1 629	2 906
		MALES AND FEMALES COMBINED					
Beta (S.E.)		-0.092 (0.025)	-0.032 (0.028)	0.128 (0.018)			
P		$1.7x10^{-4}$	0.26	$5x10^{-11}$			
N		4 010	3 239	5 544			

Growth is estimated as mean height increase (cm) in age windows from 9 to 12 years, from 12 to 14/15 years and from 14/15 years to adulthood in Northern Finland Birth Cohort 1966 (NFBC1966), Cardiovascular Risk in Young Finns Study (YF) and Helsinki Birth Cohort Study (HBCS). The height data are presented as mean relative height and standard deviation. Height was obtained at 14 years in NFBC1966 and at 15 years in YF. The cohort-specific growth distributions have been logarithm-transformed (if necessary) and standardized prior to within-cohort linear

regression analysis. The within-cohort association analysis results have been combined with meta-analysis first in females and males separately and in males and females combined within each age window. NFBC1966 study subjects were genotyped on Illumina Infinium 370CNV Duo arrays. Genotyping in YF was carried out by TaqMan 5' nuclease assays (further details available on request), in which fluorescence was detected post-polymerase chain reaction using the ABI Prism 7900HT Sequence Detector System. Genotypes were called using ABI Prism SDS software version 2.1. The study samples of HBCS were genotyped using the Sequenom MassArray genotyping platform.

Table S2. Previously Published Genome-wide Significant Association Results between Age of Menarche and Markers in the *LIN28B* Region

	Effect on Age of Menarche				**Effect on Growth during Late Adolescence**			
Reference	**Published SNP**	**Ref SNP Alleles**	**Effect Allele***	**P value**	**Genotyped proxy**	**r2**	**Effect Allele****	**P value*****
Ong et al[3]	rs314276	A/C	A	$3.6x10^{-16}$	rs7759938#	0.97	C (proxy)	$5.9x10^{-8}$
Sulem et al[4]	rs314280	C/T	T	$1.8x10^{-14}$	-		T	$3.3x10^{-7}$
[a)]He et al[5]	rs314277	A/C	A	$2.7x10^{-13}$	-		A	0.001
[b)]He et al[5]	rs314263	C/T	C	$3.2x10^{-13}$	rs7759938#	0.97	C (proxy)	$5.9x10^{-8}$
Perry et al[6]	rs7759939	C/T	C	$7.0x10^{-9}$	-		C	$5.9x10^{-8}$

the pair-wise correlation with the proxy=0.97

*associated with later age of menarche

** associated with later pubertal growth

*** Northern Finland Birth Cohort 1966; females and males combined, n=4 321

[a)] best associated SNP

[b)] next best associated SNP

Data on one or two markers yielding the strongest evidence for association in each published study is shown. The corresponding results testing the same marker locus for association with growth during late adolescence in Northern Finland Birth Cohort 1966 is shown in the right-hand columns of the table. In case the originally published SNP was not genotyped in the current study, association data for a proxy is given instead.

Table S3. Linear Regression Analysis of the Change in Relative Height between age 6 and 9 in Females of Northern Finland Birth Cohort 1966 and Helsinki Birth Cohort Study Including Markers rs7759938, rs314277 and Their Interaction in the Regression Model

FEMALES **Independent Variable**	**Beta**	**S.E**	**P-value**
rs7759938	-0.214	0.054	6.65×10^{-5}
rs314277	-0.023	0.135	0.867
rs759938 x rs314277 interaction	0.145	0.093	0.122

MALES **Independent Variable**	**Beta**	**S.E**	**P-value**
rs7759938	-0.039	0.058	0.500
rs314277	0.078	0.155	0.616
rs759938 x rs314277 interaction	-0.027	0.102	0.794

The regression analysis was run using cohort- and sex-specific Z-scores as the dependent variable The AA, AG and GG genotypes at marker locus rs7759938 were coded as 0,1,2, the CC, CA and AA genotypes at rs314277 as 0,1,2. N = 1 460

Table S4. Linear Regression Analysis of Growth during Infancy (age 0-2 yrs), during Childhood (2-9 yrs) and Puberty (9 yrs – adulthood) in Northern Finland Birth Cohort 1966 and Helsinki Birth Cohort Study Including Markers rs7759938 and rs314277 in the Regression Model

		MALES		
		0-2 yrs	**2-9 yrs**	**9 yrs - adulthood**
rs7759938	**Beta (S.E)**	0.021(0.049)	0.053(0.053)	0.029(0.047)
	p	0.66	0.32	0.54
rs314277	**Beta (S.E)**	-0.062(0.064)	-0.041(0.069)	0.086(0.061)
	p	0.33	0.56	0.16
	n	1 552	1 346	1 731

		FEMALES		
		0-2 yrs	**2-9 yrs**	**9 yrs - adulthood**
rs7759938	**Beta (S.E)**	0.001(0.048)	-0.085(0.050)	0.031(0.045)
	p	0.98	0.09	0.49
rs314277	**Beta (S.E)**	0.080(0.062)	0.097(0.064)	0.076(0.057)
	p	0.20	0.13	0.18
	n	1 747	1 549	1 917

The AA, AG and GG genotypes at marker locus rs7759938 were coded as 0,1,2, the CC, CA and AA genotypes at rs314277 as 0,1,2.

Supplemental References

1. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. *81*, 559-575.

2. Sabatti, C., Service, S., Hartikainen, A.-L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., et al. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat. Genet. *41,* 35-46.

3. Ong, K.K., Elks, C.E., Li, S., Zhao, J.H., Luan, J., Andersen, L.B., Bingham, S.A., Brage, S., Davey Smith, G., Ekelund, U., et al. (2009). Genetic variation in LIN28B is associated with the timing of puberty. Nat. Genet. *41*, 729-733.

4. Sulem, P., Gudbjartsson, D.F., Rafnar, T., Holm, H., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Alexandersen, P., Feenstra, B., Boyd, H.A., et al. (2009). Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. Nat. Genet. *41*, 734-738.

5. He, C., Kraft, P., Chen, C., Buring, J.E., Paré, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D. J., Chasman, D.I. (2009). Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat. Genet. *41*, 724-728.

6. Perry, J.R.B., Stolk, L., Franceschini, N., Lunetta, K.L., Zhai, G., McArdle, P.F., Smith, A. V., Aspelund, T., Bandinelli, S., Boerwinkle, E., et al. (2009). Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat. Genet. *41*, 648-650.

7. Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guiducci, C., et al. (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. Nat. Genet. *40*, 584-591.

8. Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., et al. (2008). Many sequence variants affecting diversity of adult human height. Nat. Genet. *40*, 609-615.

9. Bandelt, H.J., Forster, A., Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol.Biol.Evol* 1, 37-48.

10. Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., et al. (2005). Natural selection on protein-coding genes in the human genome. Nature. *7062,* 1153-1157.

11. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S. and The International HapMap Consortium. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature. *7164*, 913–918.

12. Lappalainen, T., Salmela, E., Andersen, P.M., Dahlman-Wright, K., Sistonen, P., Savontaus, M.-L., Schreiber, S., Lahermo, P., Kere, J. Genomic landscape of positive natural selection in Northern European populations. (2009). Eur J Hum Genet. Oct 21, [Epub ahead of print]

13. Akey, J.M., Zhang, G., Zhang, K., Jin, L., Shriver, M.D. (2002). Interrogating a high-density SNP map for signatures of natural selection. Genome Res. *12,* 1805-1814.

14. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., Pritchard, J.K. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res. *19*, 826-837.

15. Lettre, G. (2009). Genetic regulation of adult stature. Curr Opin Pediatr. *21*, 515-22.