The American Journal of Human Genetics, Volume *89*

Supplemental Data

Genome-wide Association of Copy-Number Variation Reveals an Association between Short Stature and the Presence of Low-Frequency Genomic Deletions

Andrew Dauber, Yongguo Yu, Michael C. Turchin, Charleston W. Chiang, Yan A. Meng, Ellen W. Demerath, Sanjay R. Patel, Stephen S. Rich, Jerome I. Rotter, Pamela J. Schreiner, James G. Wilson, Yiping Shen, Bai-Lin Wu, Joel N. Hirschhorn

Figure S1. Histogram of length of deletion CNVs in the clinical cohort

Figure S2. Histogram of length of duplication CNVs in the clinical cohort

Figure S3. Scatter Plot of Total CNV Burden Versus Height in the Clinical Cohort

Figure S4. Scatter Plot of Total Lower Frequency (<5%) Deletion Burden Versus Height in the Clinical Cohort

Figure S5. 11q11 candidate region CNV association data

Figure S6. 14q11.2 candidate region CNV association data

Figure S7. 17q21.31 candidate region CNV association data

Figure S8. MLPA validation of candidate regions

Table S1. Custom designed MLPA probe sets for candidate regions

Table S2. Subjects with deletion or duplication syndromes associated with short stature

Table S3. Genic CNV burden association analysis in short cases versus controls

Table S4. Quantitative Trait CNV Association Analysis in Clinical Cohort

Table S5. Quantitative Trait CNV Association Analysis in Clinical Cohort Divided By Frequency and Type of CNV

Table S6. Height distribution in 3 candidate regions divided by copy number status

Supplemental Acknowledgments

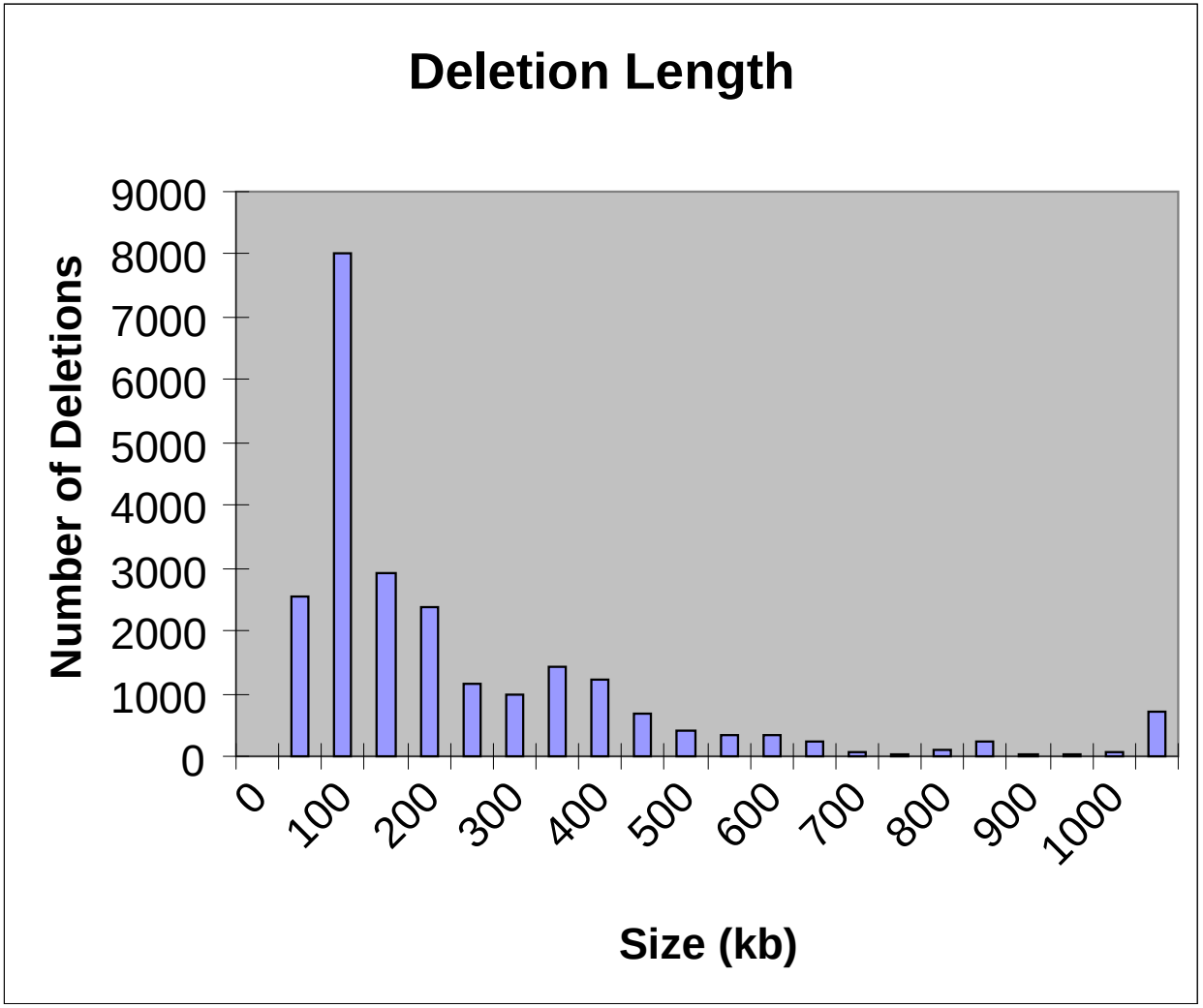

Figure S1. Histogram of length of deletion CNVs in the clinical cohort

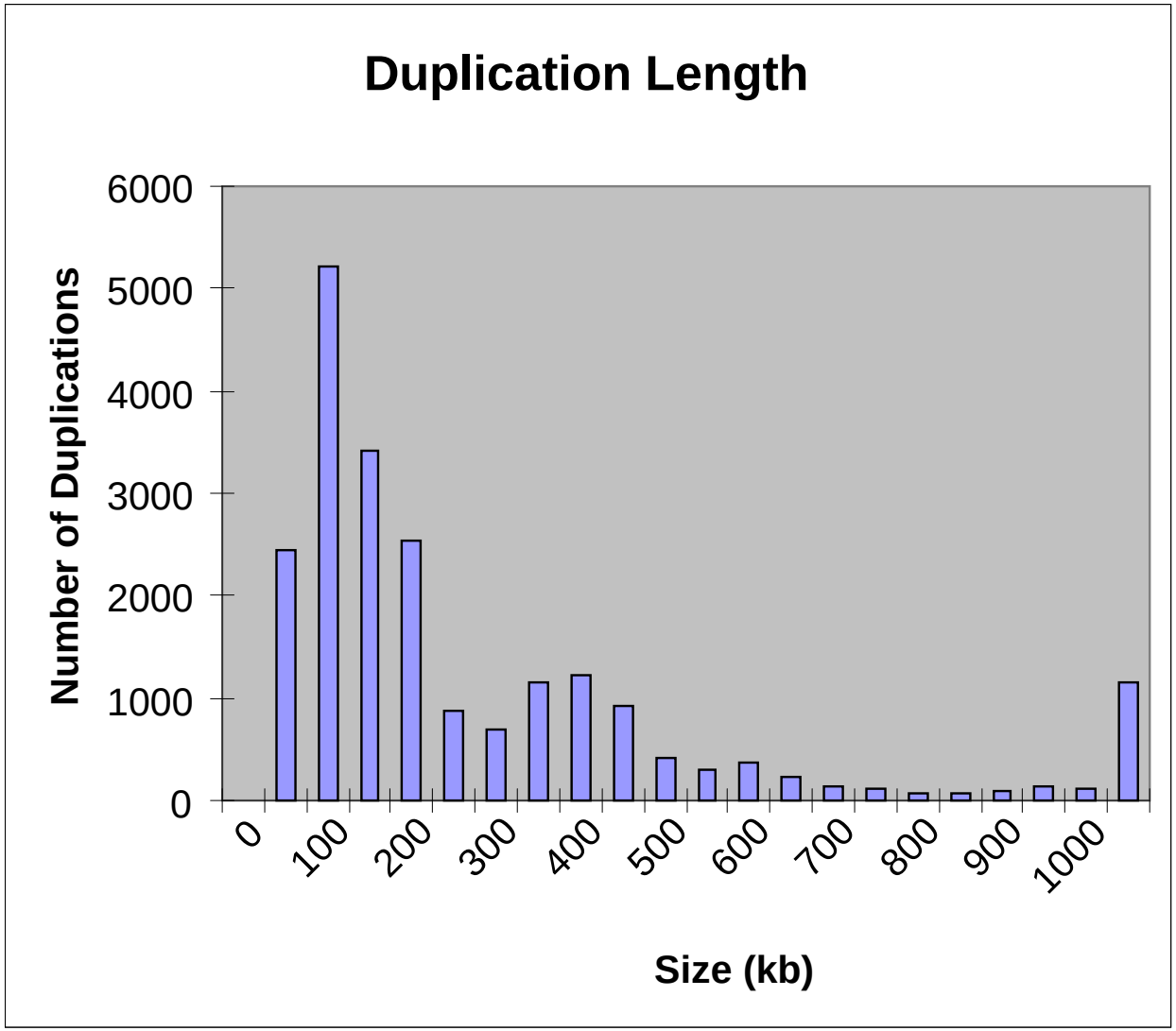

Figure S2. Histogram of length of duplication CNVs in the clinical cohort

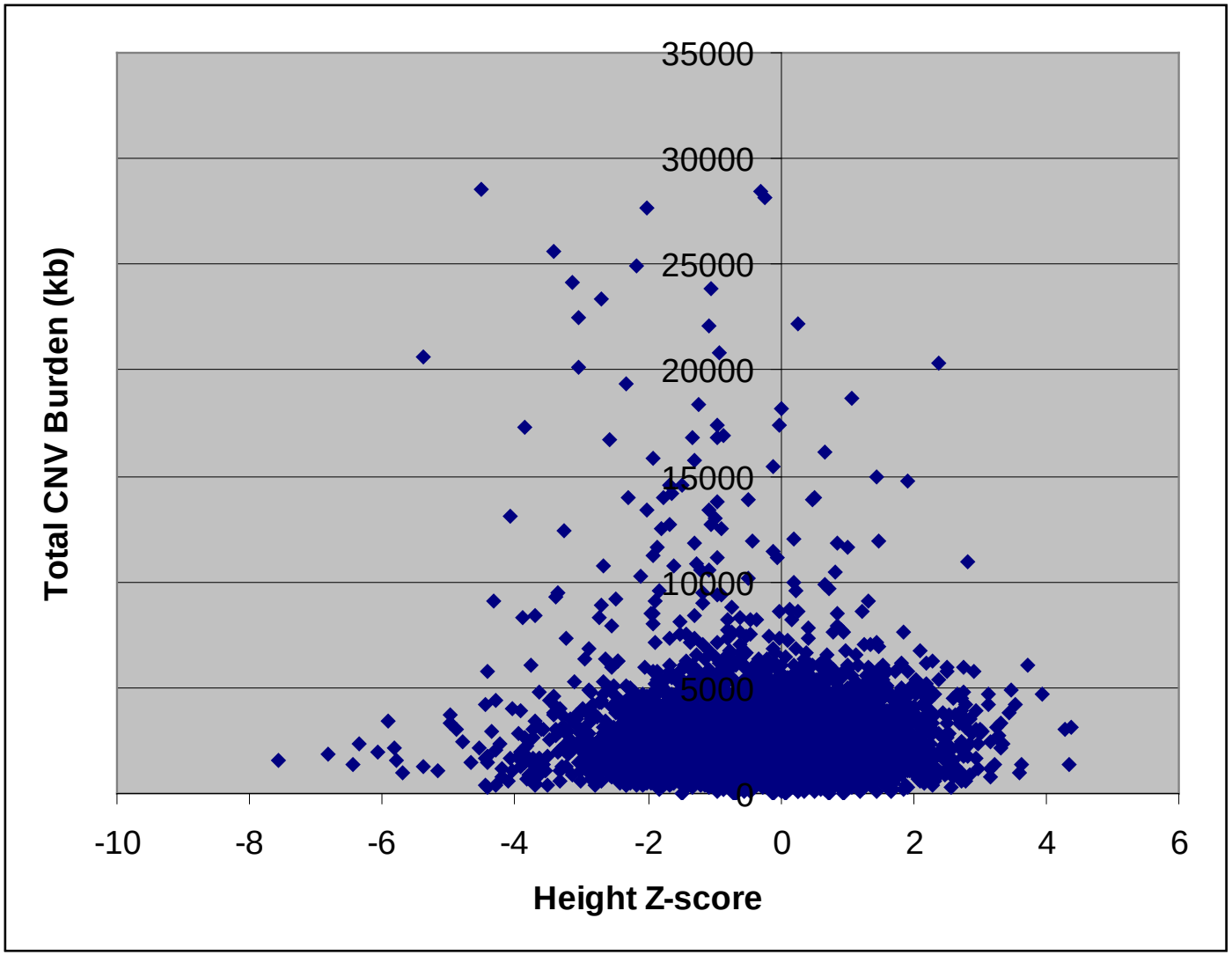

Figure S3. Scatter Plot of Total CNV Burden Versus Height in the Clinical Cohort

The top 12 outliers for CNV burden were removed to allow for adequate graphical representation of the majority of the cohort. These 12 outliers all had height Z-scores <0 and 3 had Z-scores <-2. Their total CNV burden ranged from 31-151 Mb.

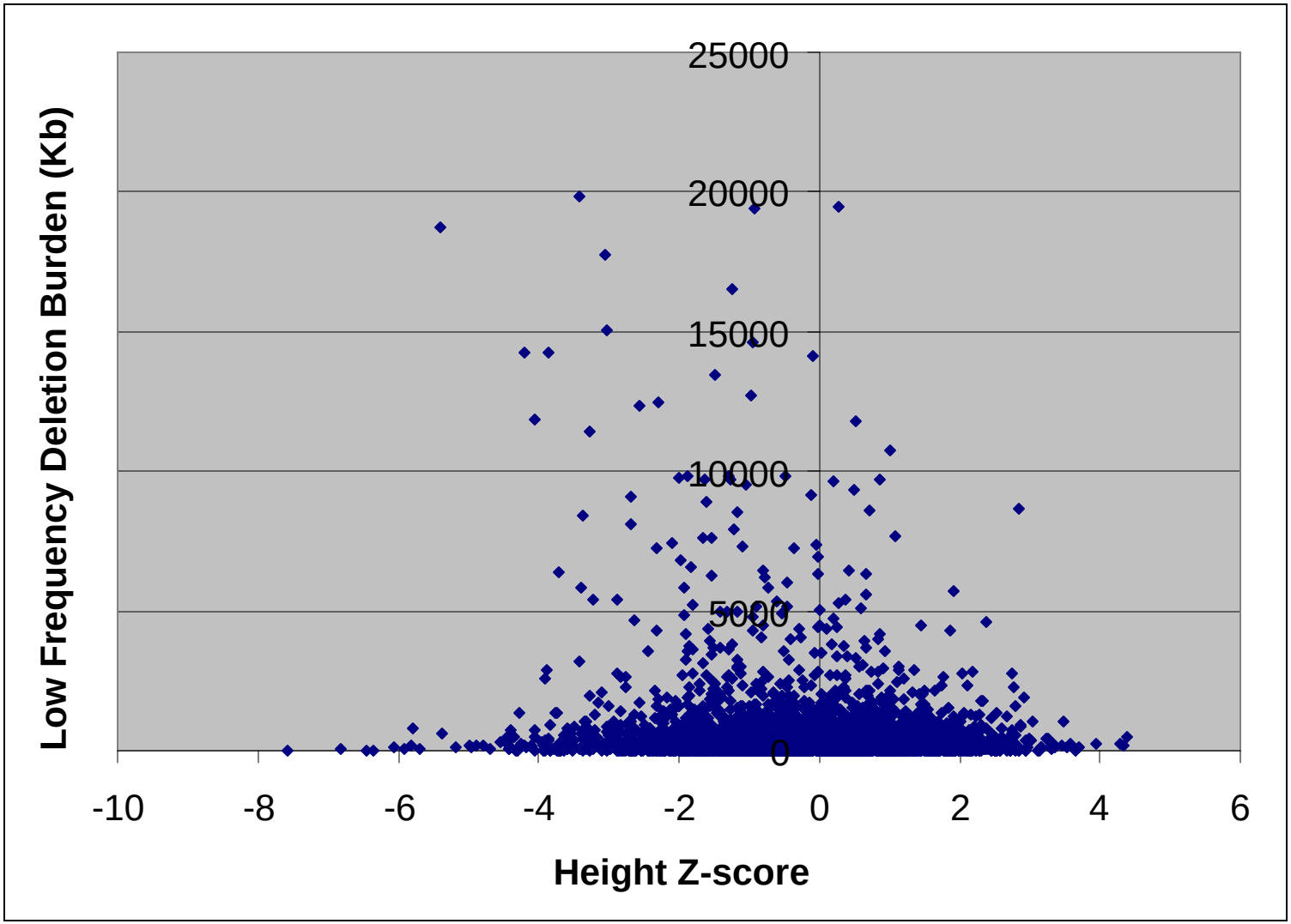

Figure S4. Scatter Plot of Total Lower Frequency (<5%) Deletion Burden Versus Height in the Clinical Cohort

The top 9 outliers for low frequency deletion burden were removed to allow for adequate graphical representation of the majority of the cohort. These 9 outliers all had height Z-scores <0 and 4 had Z-scores <-2. Their total low frequency deletion burden ranged from 20-149 Mb.

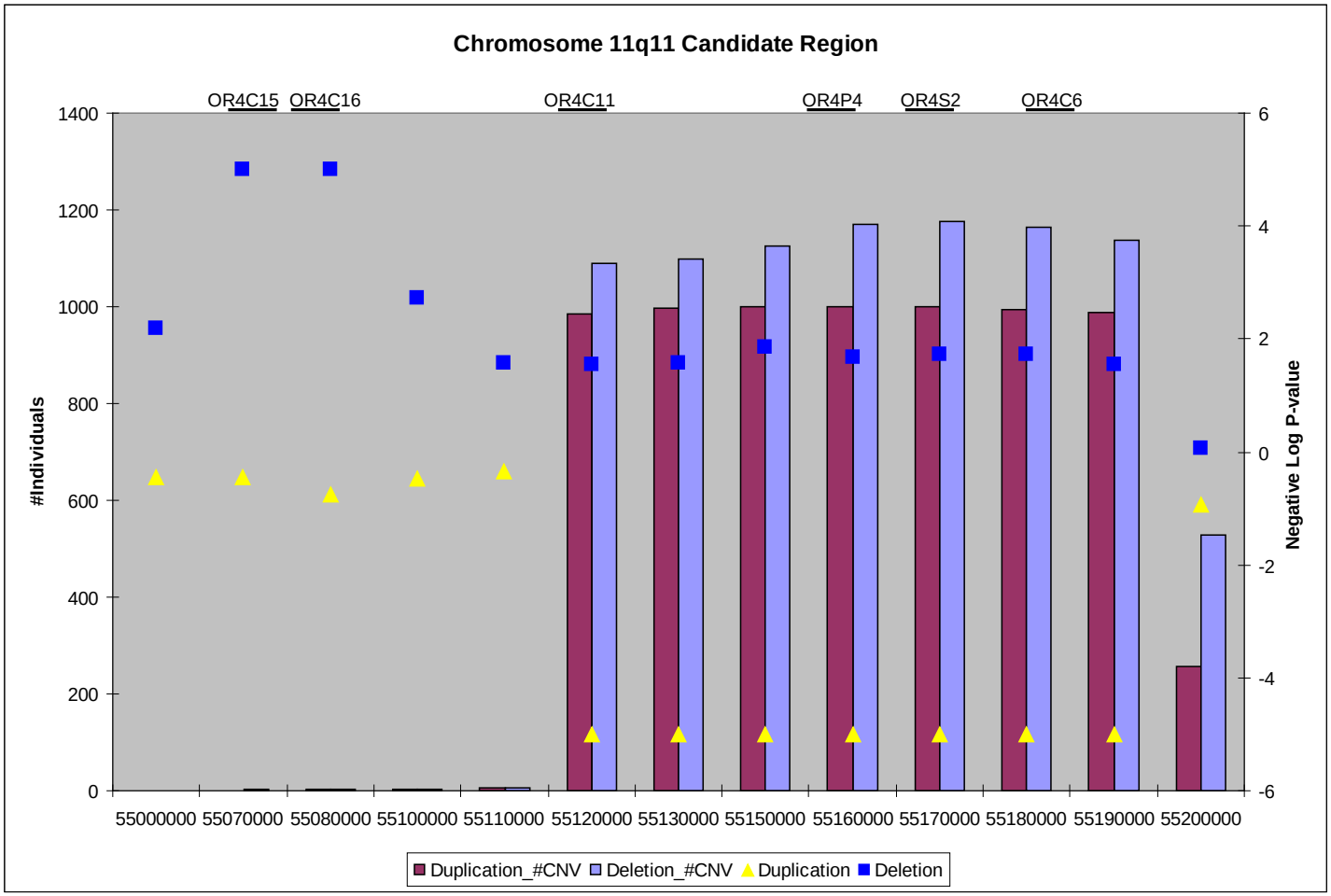

Figure S5. 11q11 candidate region CNV association data

Bars represent number of individuals with the CNV in our cohort (out of 4411 individuals). Triangles and squares represent the negative log p-value for the regional height association for duplications and deletions respectively. Positive values indicate that the CNV is height increasing and negative values indicated that the CNV is height decreasing. Gene names for genes overlapping region are placed above their approximate positions.

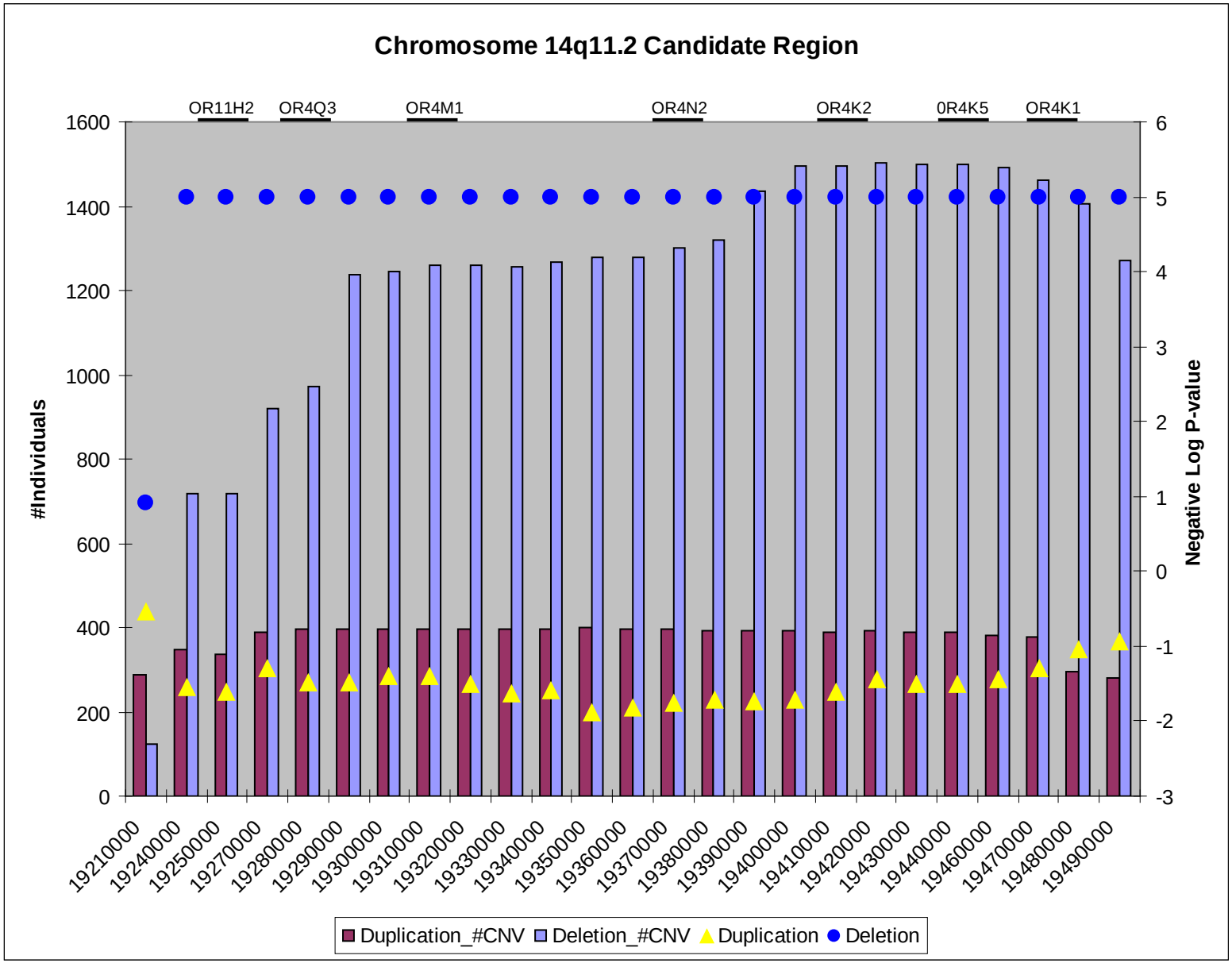

Figure S6. 14q11.2 candidate region CNV association data

See Figure S5 for legend.

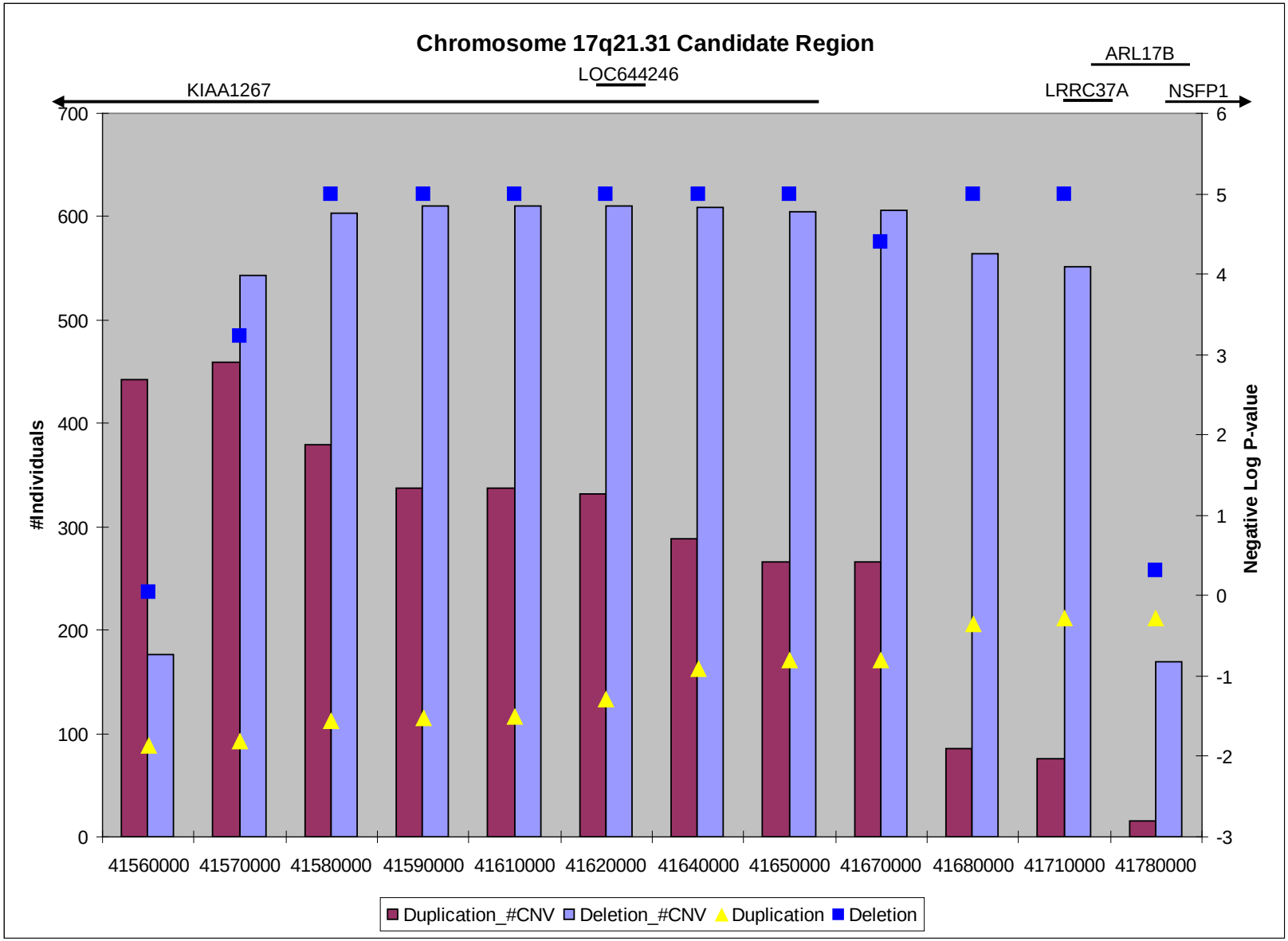

Figure S7. 17q21.31 candidate region CNV association data

See Figure S5 for legend.

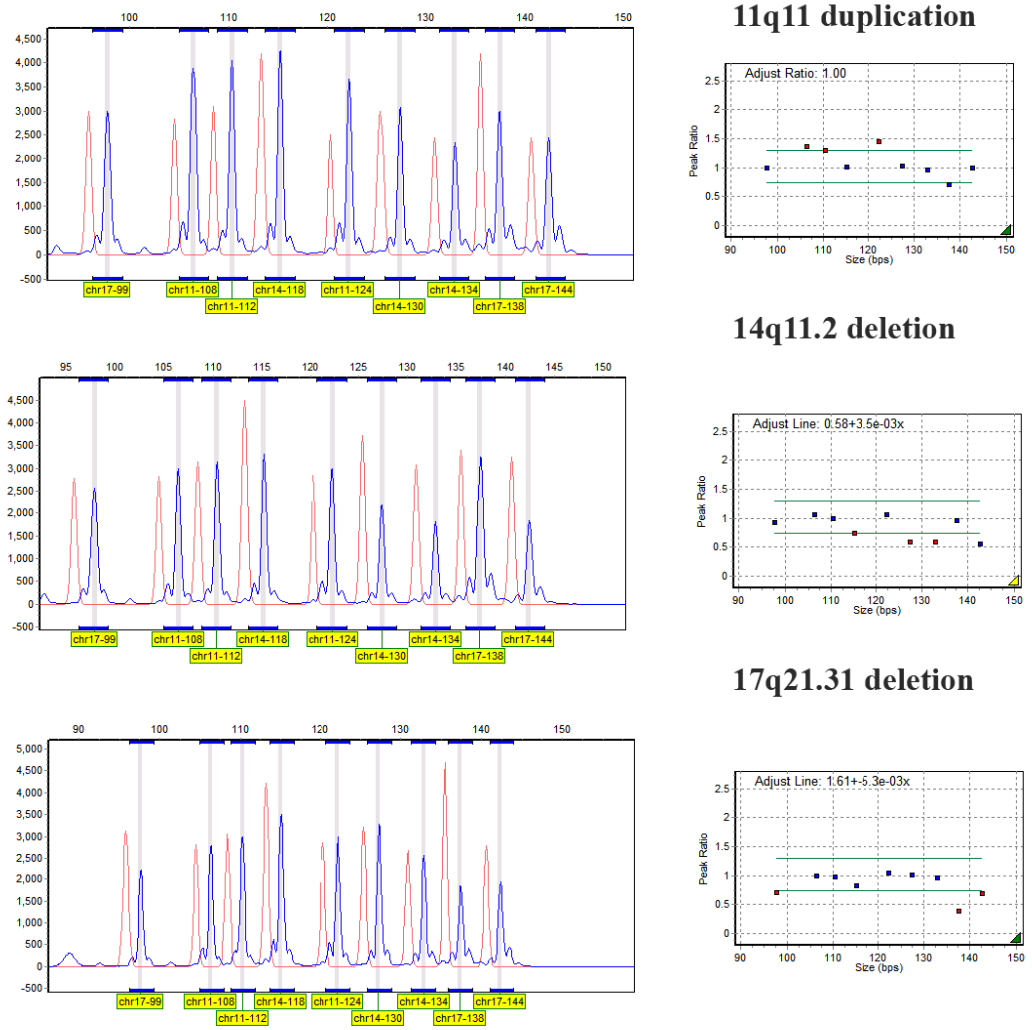

Figure S8. MLPA validation of candidate regions

Representative MLPA results were shown for three loci (from top to bottom: 11q11 duplication, 14q11.2 deletion and 17q21.31 deletion). Images on the left side are peak profiles (red traces were peak profiles from normal control samples and blue traces were peak profiles of test samples) and images on the right are ratio profiles (red dots above the upper bar indicate duplication: case/control ratio>1.25, red dots below the lower bar indicate deletion: case/control ratio<0.75)

Table S1. Custom designed MLPA probe sets

Loci	Probe coordinates (hg18)	Probe sequences	GC/Tm	amplicom size
14q11	chr14:19463852-19463925	Chr14Ex16L,**GGGTTCCCTAAGGGTTGGA**TGCAAGTTAGTCACAGGACAAGGAGGAAAGTATCAGAGG	50/76.78	118
		Chr14Ex16R,TACACTTTCCAGGAATGCAGAGTAACAGAGTCCTCCA**TCTAGATTGGATCTTGCTGGCAC**	46.67/75.70	
	chr14:19465044-19465103	Chr14Ex17L,**GGGTTCCCTAAGGGTTGGA**taaaaaactaccgtCAGCAGCAAGCCATCACACATTGGAAATGG	47.62/76.49	130
		Chr14Ex17R,CATGTGTAGGACCACGCTGAAGCAGGTATAgaaaagtcggtgga**TCTAGATTGGATCTTGCTGGCAC**	49.25/77.63	
	chr14:19468930-19468989	Chr14Ex19L,**GGGTTCCCTAAGGGTTGGA**tctggacccgtgatggCTCACTACATCAGGACAGGACCTGAGGACC	58.46/78.18	134
		Chr14Ex19R,TGAGAGGCTTCCTTCCATACGGCATAGTGTcattctctggttttcg**TCTAGATTGGATCTTGCTGGCAC**	47.83/77.26	
11q11	chr11:55078474-55078558	Chr11Ex4L,**GGGTTCCCTAAGGGTTGGA**CTATACATGATCCCTGTTGGAGCTTTCATCTTTTCCTTGG	47.46/75.88	124
		Chr11Ex4R,GAAACATGCAAAACCAAAGCTTTGTAACTGAGTTTGTCCTCC**TCTAGATTGGATCTTGCTGGCAC**	43.08/74.87	
	chr11:55190115-55190184	Chr11Ex9L,**GGGTTCCCTAAGGGTTGGA**CTGGATGAAATGGGAGGCTTTGGCTGGGAAATA	51.92/76.57	112
		Chr11Ex9R,ACTGCAATGCTGAGAACATCATGTATTTCCCAAAAGG**TCTAGATTGGATCTTGCTGGCAC**	43.33/74.33	
	chr11:55344586-55344651	Chr11Ex13L,**GGGTTCCCTAAGGGTTGGA**TCACCGAGATACTGGACACCAAAGTCTTCTCTT	50/75.78	108
		Chr11Ex13R,ACTGAGCCTGTTACTTTCATGGAGTTTGTCACA**TCTAGATTGGATCTTGCTGGCAC**	44.64/74.28	
17q21.31	chr17:41629594-41629659	Chr17Ex4L,**GGGTTCCCTAAGGGTTGGA**tcatccggtgaagagattGATGGAGGAAAAAACAGTCCCTCTTCCTCAGAC	50/78.26	144
		Chr17Ex4R,AGTGACACCTCAAACTGCTACAGTCACCTTTGCgagccacctgacagtgtg**TCTAGATTGGATCTTGCTGGCAC**	51.35/79.20	
	chr17:41839514-41839593	Chr17Ex12L,**GGGTTCCCTAAGGGTTGGA**tcggcgttGGATTTCCAGTCTGGCCAGTGAGTATCTGACTTTGTTTTC	50.75/78.24	138
		Chr17Ex12R,TTTTAACCTGCTAAGTGGCATTCGGGAAACTTCCAGAGAGttccggaa**TCTAGATTGGATCTTGCTGGCAC**	46.48/76.91	

(Bold capital letters are universal primer sequences. Small letters are stuffer sequences, capital letters are target specific sequences)

Table S2. Subjects with deletion or duplication syndromes associated with short stature

Name of Syndrome	**# of Subjects**
22q11 deletion	11
1q44 deletion	1
1p36 deletion	4
16p13 deletion	9
Smith-Magenis	4
Potocki-Lupski	1
Mowat-Wilson	3
Wolf-Hirschhorn	2
Williams	3
SHOX/Xq deletion	6

Table S3. Genic CNV Burden Association Analysis in 415 Short Cases versus 3800 Controls

	All Frequencies				Common (>5%)				Lower Frequency (<5%)				Rare (<1%)		
	Case	Control	P-value		Case	Control	P-value		Case	Control	P-value		Case	Control	P-value
Deletions and Duplications															
Total Number of CNV	3178	28785			875	7959			581	5352			364	3357	
Number of CNV per individual	7.7	7.6	0.31		2.1	2.1	0.43		1.4	1.4	0.53		0.88	0.88	0.49
Total CNV burden per individual (kb)	3090	2259	**0.001**		137	137	0.51		221	202	0.10		195	175	0.12
Average CNV size per individual (kb)	390	298	**0.002**		60	59	0.12		114	105	0.05		124	108	**0.02**
Deletions Only															
Total Number of CNV	1674	15237			407	3901			329	2867			199	1712	
Number of CNV per individual	4.0	4.0	0.43		1.0	1.0	0.83		0.79	0.75	0.22		0.48	0.45	0.23
Total CNV burden per individual (kb)	1781	1121	**0.002**		92	95	0.77		142	137	0.32		136	125	0.19
Average CNV size per individual (kb)	521	264	**0.0002**		60	60	0.52		97	93	0.27		104	98	0.23
Duplications Only															
Total Number of CNV	1504	13548			413	3636			308	2911			231	2145	
Number of CNV per individual	3.6	3.6	0.30		1.0	0.96	0.2		0.74	0.77	0.61		0.56	0.56	0.51
Total CNV burden per individual (kb)	1398	1226	0.08		85	83	0.27		181	166	0.16		183	162	0.13
Average CNV size per individual (kb)	354	324	0.15		58	56	0.13		123	111	0.07		133	115	**0.04**

P-values showing significant associations are in bold.

Table S4. Quantitative Trait CNV Association Analysis in Clinical Cohort

	All Subjects N=4411				Subjects with height Z<0 N=2334				Subjects with height Z>0 N=2067		
	BETA	STD ERR	P		BETA	STD ERR	P		BETA	STD ERR	P
Global CNV Burden											
Number of CNV per individual	0.022	0.011	0.045		-0.003	0.021	0.88		0.037	0.030	0.22
Total CNV burden per individual	-0.027	0.011	0.013		-0.040	0.022	0.070		0.017	0.029	0.54
Average CNV size per individual	-0.045	0.011	3.1E-05		-0.042	0.022	0.055		-0.005	0.028	0.87
Genic CNV Burden											
Number of CNV per individual	0.006	0.011	0.55		-0.016	0.021	0.45		0.041	0.030	0.17
Total CNV burden per individual	-0.045	0.011	3.4E-05		-0.053	0.022	0.016		0.008	0.028	0.77
Average CNV size per individual	-0.054	0.011	5.0E-07		-0.052	0.021	0.017		-0.010	0.028	0.73

All CNV measures were inverse normalized due to the non-normal distribution in the clinical cohort. Linear regression was then performed looking at each measure of inverse normalized CNV burden as a function of height Z-score. The beta value indicates the magnitude of increase in Z-score of CNV measure for every 1 standard deviation increase in stature. The analyses of subjects with height Z<0 or Z>0 indicate that the linear regression was repeated just including subjects with height Z-scores less than or greater than 0. These analyses demonstrate that the associations between increased burden of CNV and larger average size of CNV with decreased height are driven by subjects with height Z-scores less than 0. These associations are not present in subjects with height Z-scores greater than 0. This is consistent with the findings in the case control association analyses in which these associations were only present for short cases compared to controls but not tall cases compared to controls.

Table S5. Quantitative Trait CNV Association Analysis in Clinical Cohort Divided By Frequency and Type of CNV

CNV CATEGORY		ALL FREQUENCIES				COMMON (>5%)				LOWER FREQUENCY (< 5%)				RARE (< 1%)		
	ANALYSIS	BETA	STD ERR	P		BETA	STD ERR	P		BETA	STD ERR	P		BETA	STD ERR	P
Global Deletions	Number of CNV per individual	0.026	0.011	0.014		0.041	0.010	**0.0001**		-0.009	0.010	0.39		-0.011	0.009	0.25
	Total CNV burden per individual	-0.011	0.011	0.32		0.040	0.011	**0.0001**		-0.038	0.010	**0.0002**		-0.034	0.010	**0.0003**
	Average CNV size per individual	-0.022	0.011	0.038		0.027	0.011	0.013		-0.037	0.010	**0.0003**		-0.034	0.010	**0.0003**
Global Duplications	Number of CNV per individual	0.001	0.011	0.92		-0.003	0.010	0.75		0.005	0.010	0.61		0.002	0.009	0.82
	Total CNV burden per individual	-0.017	0.011	0.11		-0.006	0.011	0.58		-0.014	0.010	0.19		-0.013	0.010	0.18
	Average CNV size per individual	-0.021	0.011	0.047		-0.006	0.011	0.57		-0.017	0.010	0.10		-0.014	0.010	0.14
Genic Deletions	Number of CNV per individual	0.020	0.011	0.058		0.034	0.010	**0.0008**		0.001	0.010	0.94		-0.004	0.009	0.68
	Total CNV burden per individual	-0.019	0.011	0.083		0.040	0.010	**0.0001**		-0.010	0.010	0.31		-0.010	0.009	0.29
	Average CNV size per individual	-0.029	0.011	0.007		0.028	0.010	0.007		-0.009	0.009	0.34		-0.009	0.009	0.32
Genic Duplications	Number of CNV per individual	-0.012	0.011	0.25		0.004	0.010	0.70		-0.0004	0.010	0.97		-0.002	0.009	0.84
	Total CNV burden per individual	-0.030	0.011	0.005		0.010	0.010	0.31		-0.014	0.010	0.15		-0.015	0.009	0.12
	Average CNV size per individual	-0.028	0.011	0.009		0.009	0.010	0.38		-0.014	0.010	0.18		-0.014	0.009	0.13

P-values less than 0.001 are in bold (Bonferroni corrected threshold for this table). All CNV measures were inverse normalized due to the non-normal distribution in the clinical cohort. The beta value indicates the magnitude of increase in Z-score of CNV measure for every 1 standard deviation increase in stature based on the linear regression.

Table S6. Height distribution in 3 candidate regions divided by copy number status

	Deletions	Duplications	Copy Number 2
11q11			
#CNV	1125	999	2279
Mean Height Z-score	-0.105	-0.388	-0.150
Median Height Z-score	0.027	-0.353	-0.083
Variance	1.974	2.019	1.861
Interquartile Range	-0.97 to 0.84	-1.30 to 0.54	-0.95 to 0.75
14q11.2			
#CNV	1502	392	2509
Mean Height Z-score	-0.058	-0.334	-0.251
Median Height Z-score	0.037	-0.335	-0.180
Variance	1.878	1.969	1.950
Interquartile Range	-0.83 to 0.86	-1.29 to 0.71	-1.10 to 0.67
17q21.31			
#CNV	609	288	3506
Mean Height Z-score	0.031	-0.314	-0.221
Median Height Z-score	0.055	-0.323	-0.124
Variance	1.824	1.654	1.969
Interquartile Range	-0.77 to 0.93	-1.17 to 0.55	-1.08 to 0.72

Funding Acknowledgments for the CARe Cohorts

Atherosclerotic Risk in Communities (ARIC): The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367, R01HL086694 and RC2 HL102419; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

Coronary Artery Risk in Young Adults (CARDIA): University of Alabama at Birmingham, Coordinating Center, N01-HC-95095; University of Alabama at Birmingham, Field Center, N01-HC-48047; University of Minnesota, Field Center and Diet Reading Center (Year 20 Exam), N01-HC-48048; Northwestern University, Field Center, N01-HC-48049; Kaiser Foundation Research Institute, N01-HC-48050; University of California, Irvine, Echocardiography Reading Center (Year 5 & 10), N01-HC-45134; Harbor-UCLA Research Education Institute, Computed Tomography Reading Center (Year 15 Exam), N01-HC-05187; Wake Forest University (Year 20 Exam), N01-HC-45205; New England Medical Center (Year 20 Exam), N01-HC-45204 from the National Heart, Lung and Blood Institute.

Cleveland Family Study (CFS): Case Western Reserve University (RO1 HL46380-01-16).

Jackson Heart Study (JHS): Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), Tougaloo College (N01-HC-95172).

Multi-Ethnic Study of Atherosclerosis (MESA): MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts N01 HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and RR-024156.