HGGA, Volume *3*

Supplemental information

Genome-wide analysis of copy number variants and normal facial variation in a large cohort of Bantu Africans

Megan Null, Feyza Yilmaz, David Astling, Hung-Chun Yu, Joanne B. Cole, Benedikt Hallgrímsson, Stephanie A. Santorico, Richard A. Spritz, Tamim H. Shaikh, and Audrey E. Hendricks

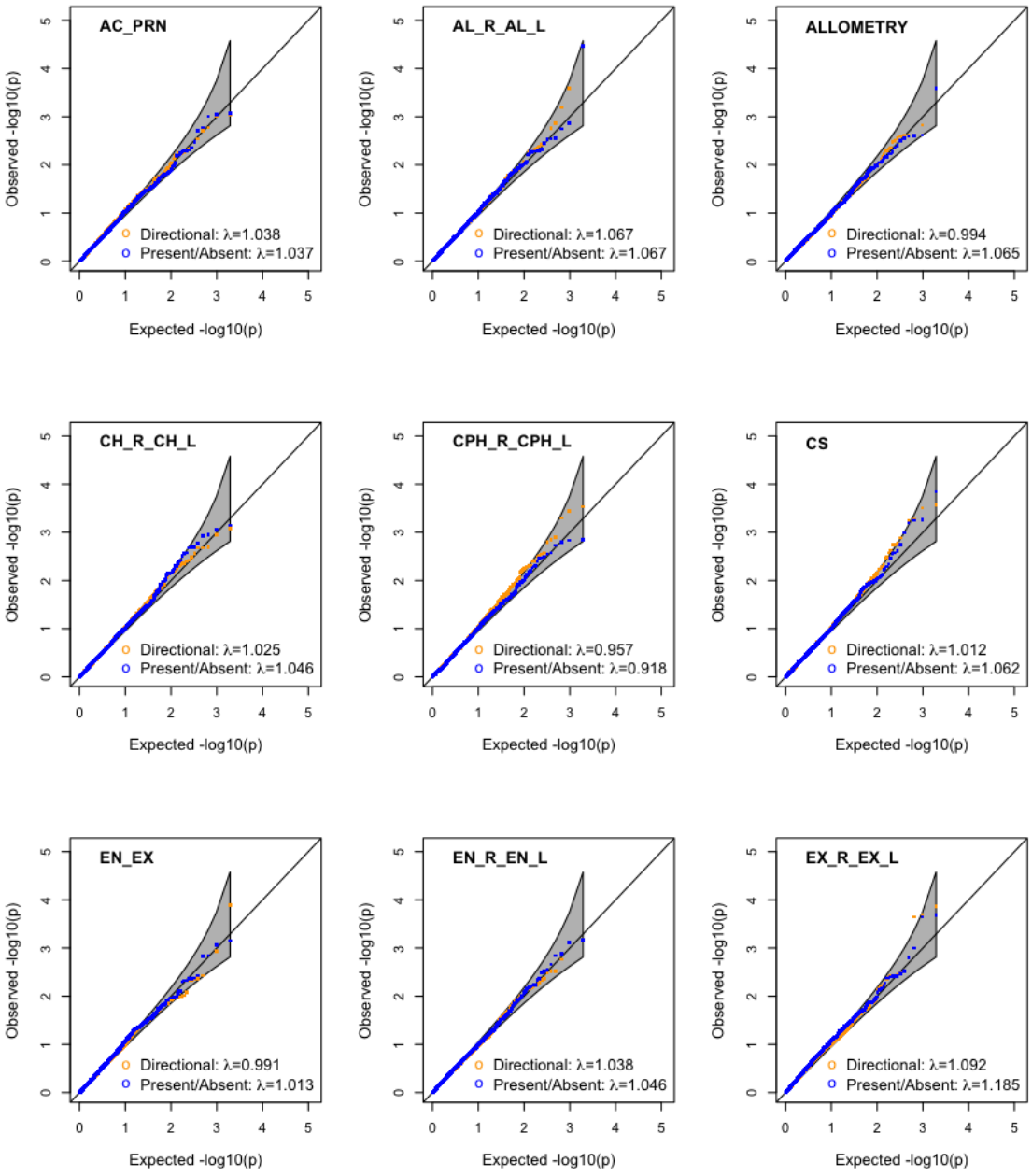

Figure S1. Quantile-Quantile Plots. QQ plots for nine phenotypes for the *absent/present* (blue) and *directional* (orange) models with respective λ values. Due to high correlation between overlapping analysis windows within a region, a random window from each region was selected for the QQ plot.

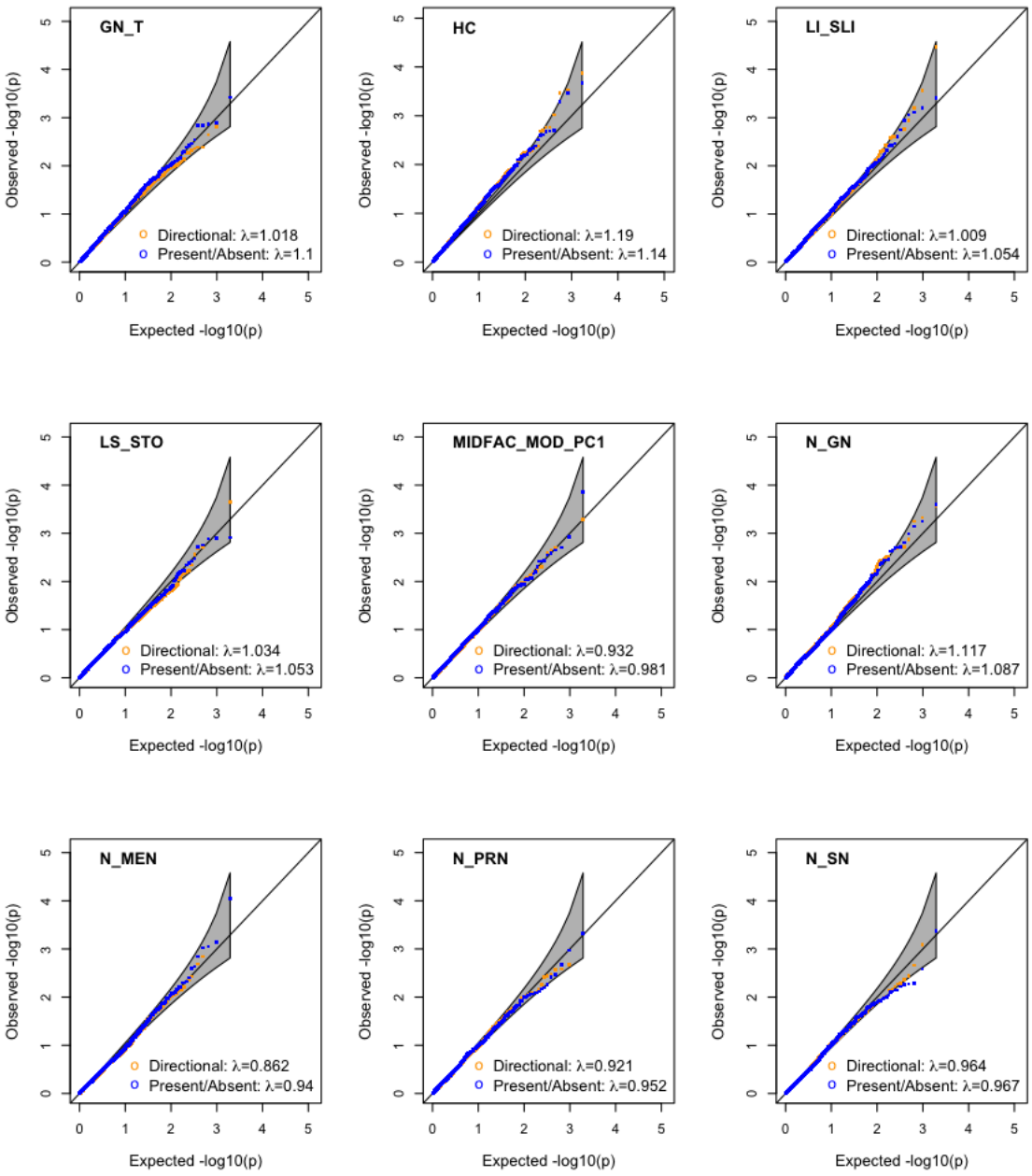

Figure S2. Quantile-Quantile Plots. QQ plots for nine phenotypes for the *absent/present* (blue) and *directional* (orange) models with respective λ values. Due to high correlation between overlapping analysis windows within a region, a random window from each region was selected for the QQ plot.

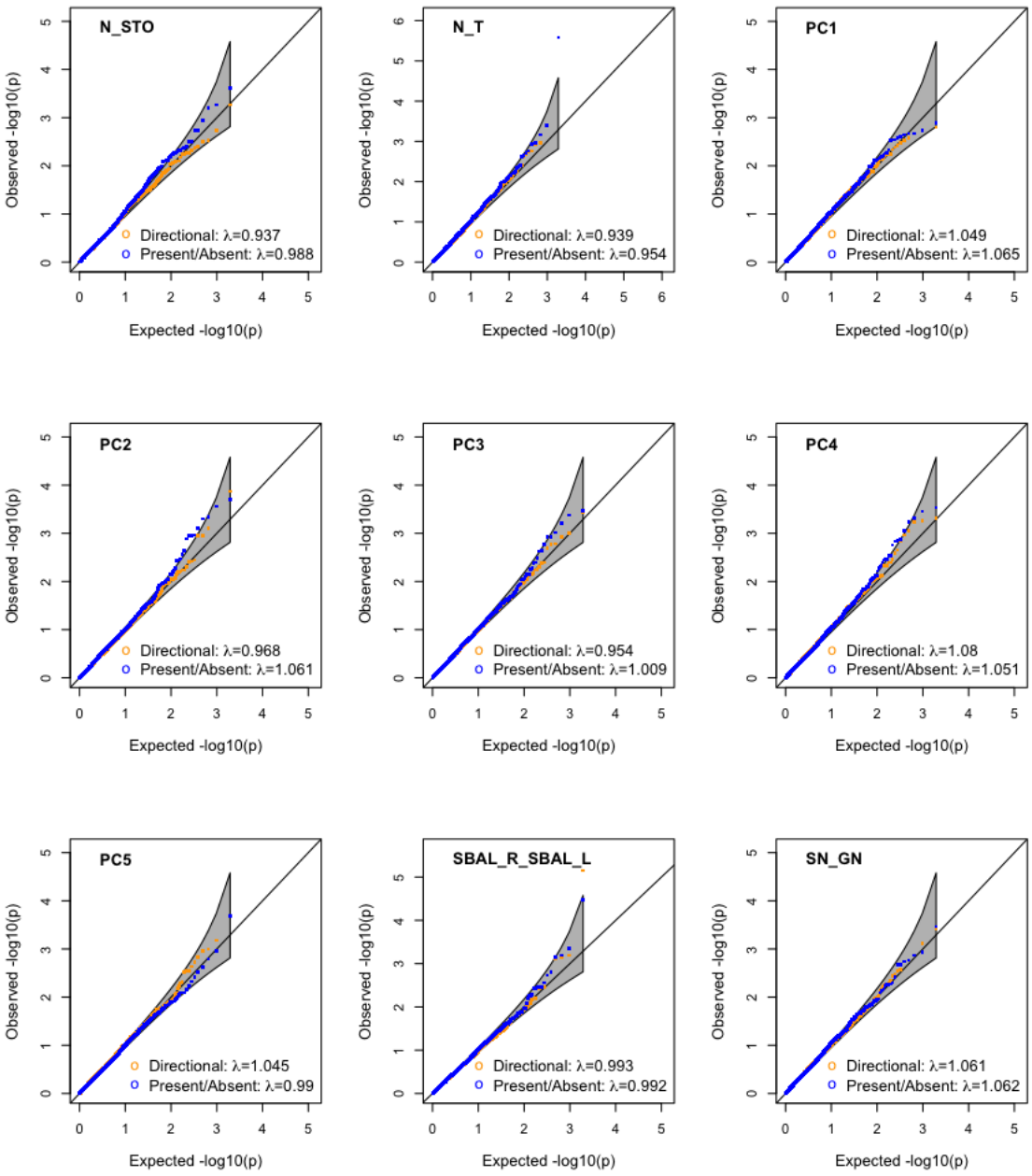

Figure S3. Quantile-Quantile Plots. QQ plots for nine phenotypes for the *absent/present* (blue) and *directional* (orange) models with respective λ values. Due to high correlation between overlapping analysis windows within a region, a random window from each region was selected for the QQ plot.

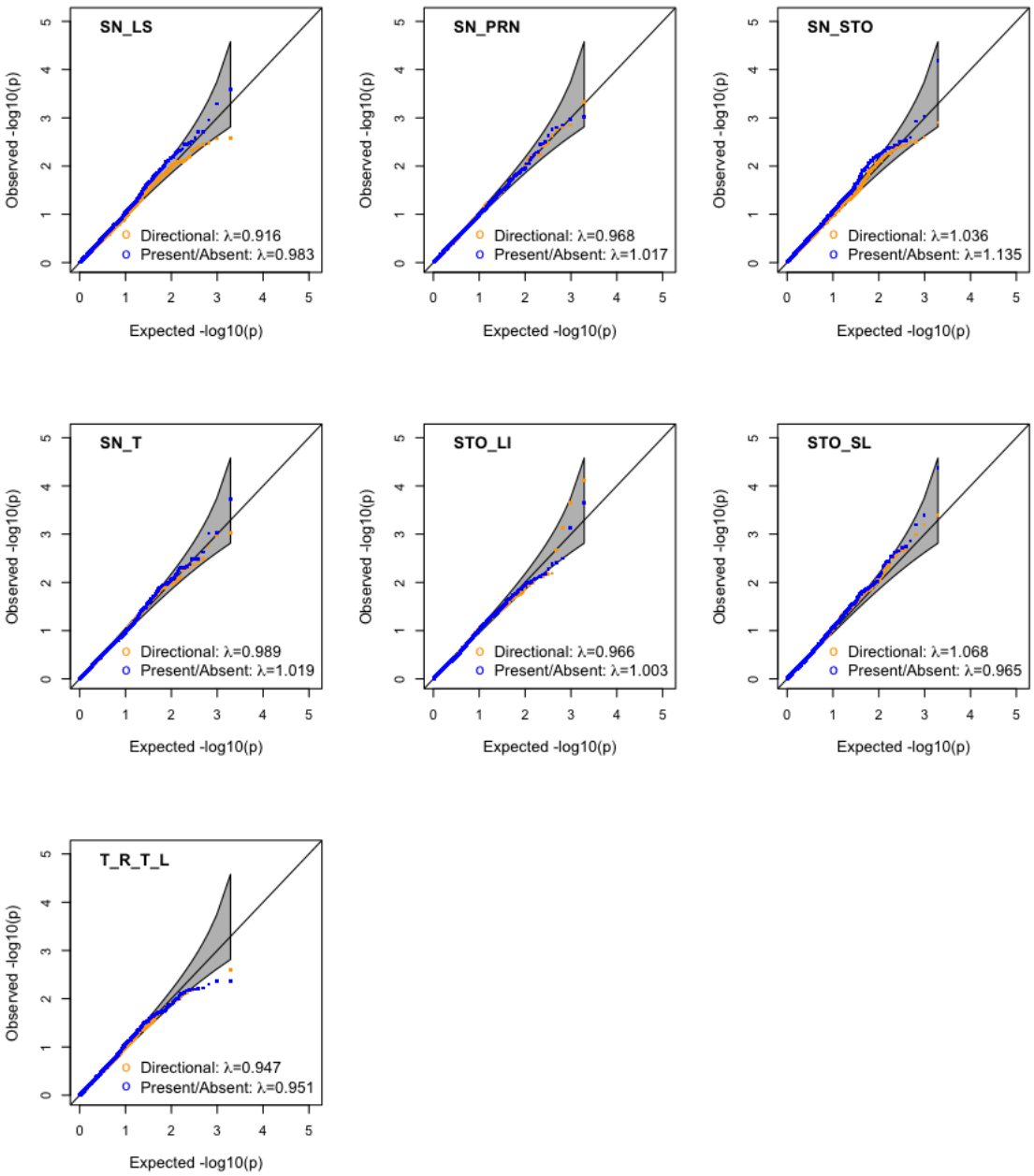

Figure S4. Quantile-Quantile Plots. QQ plots for seven phenotypes for the *absent/present* (blue) and *directional* (orange) models with respective λ values. Due to high correlation between overlapping analysis windows within a region, a random window from each region was selected for the QQ plot.

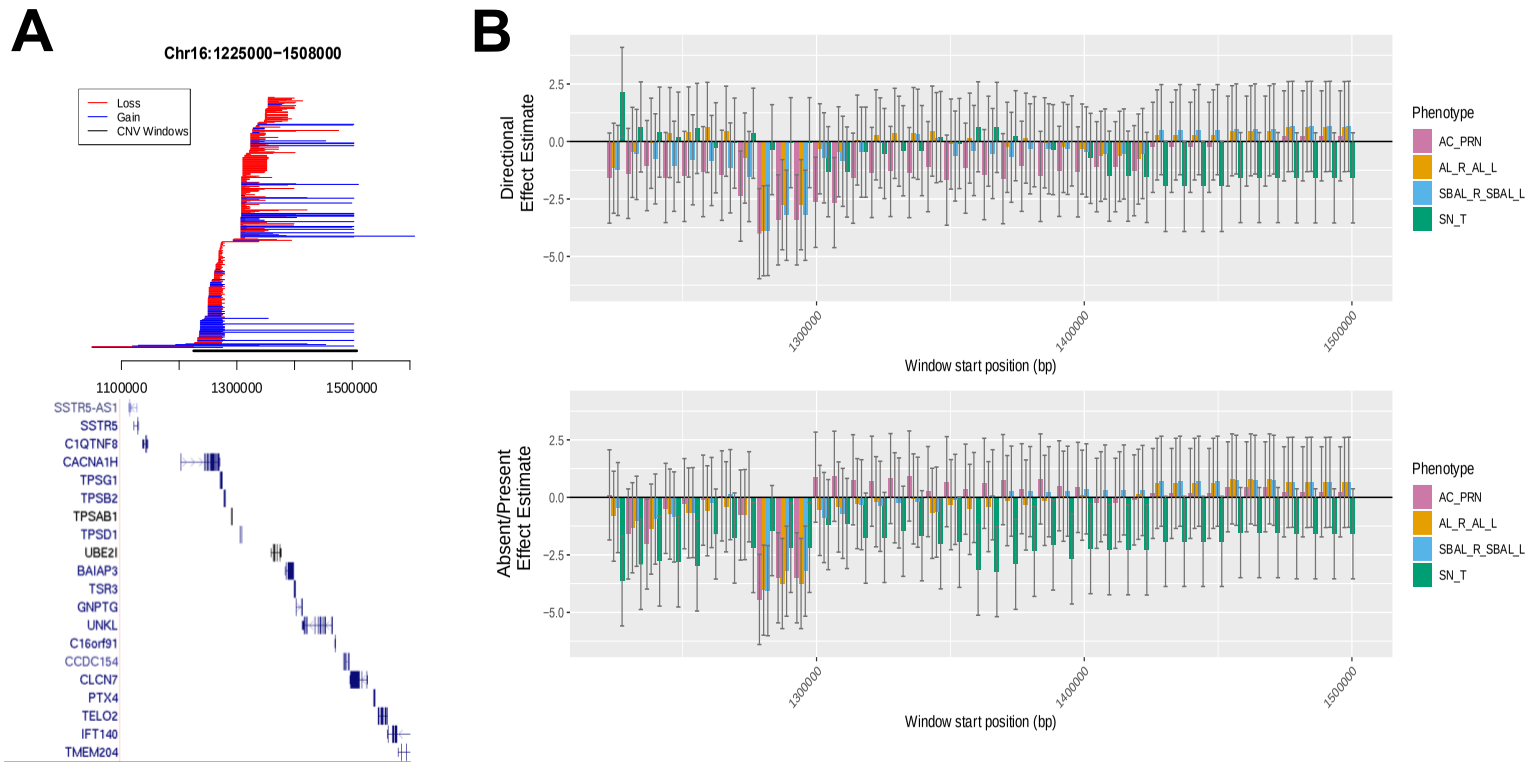
A
Chr16:1225000−1508000
Loss
Gain
CNV Windows
1100000
1300000
1500000
SSTR5-AS1
SSTR5
C1QTNF8
CACNA1H
TPSG1
TPSB2
TPSAB1
TPSD1
UBE2I
BAIAP3
TSR3
GNPTG
UNKL
C16orf91
CCDC154
CLCN7
PTX4
TELO2
IFT140
TMEM204
B
Directional Effect Estimate
2.5
0.0
−2.5
−5.0
1300000
1400000
1500000
Window start position (bp)
Phenotype
AC_PRN
AL_R_AL_L
SBAL_R_SBAL_L
SN_T
Absent/Present Effect Estimate
2.5
0.0
−2.5
−5.0
1300000
1400000
1500000
Window start position (bp)
Phenotype
AC_PRN
AL_R_AL_L
SBAL_R_SBAL_L
SN_T

Figure S5 A). Region Plot, Chromosome 16. Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black. **B). Region Plot, Chromosome 16.** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < $5x10^{-4}$ are shown: nasal ala length (AC_PRN), nasal width (AL_R_AL_L), subnasal width (SBAL_R_SBAL_L), and midfacial depth (SN_T).

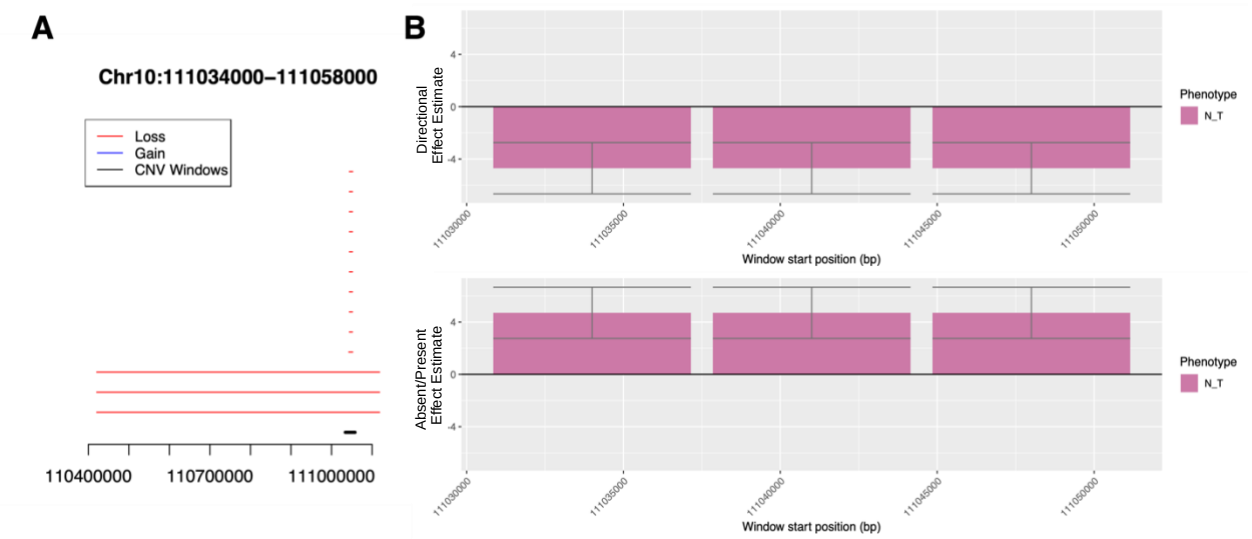

Figure S6. Region Plot, Chromosome 10. A) Loss (red) and gain (blue, none present) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black. This region does not have any genes. **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals the *directional* model (top) and *absent/present* model (bottom). Upper facial depth (N_T), the only phenotype with p-value < $5x10^{-4}$, is shown.

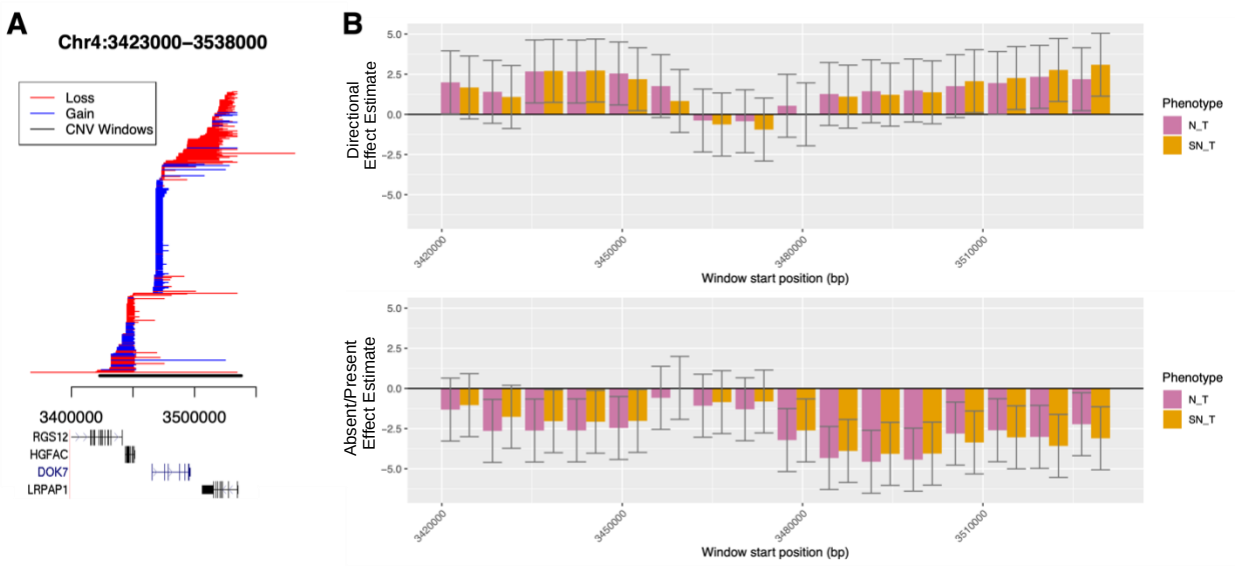

Figure S7. Region Plot, Chromosome 4. A) Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < $5x10^{-4}$ are shown: upper facial depth (N_T) and midfacial depth (SN_T).

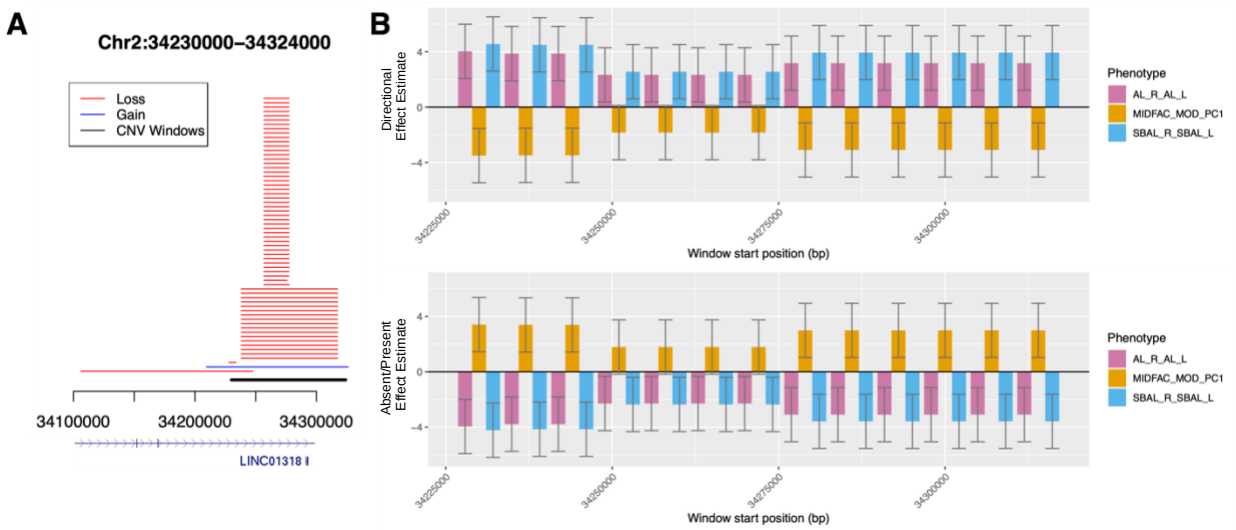

Figure S8. Region Plot, Chromosome 2. A) Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < $5x10^{-4}$ are shown: nasal width (AL_R_AL_L), mid-face principal component 1 (MIDFAC_MOD_PC1), and subnasal width (SBAL_R_SBAL_L).

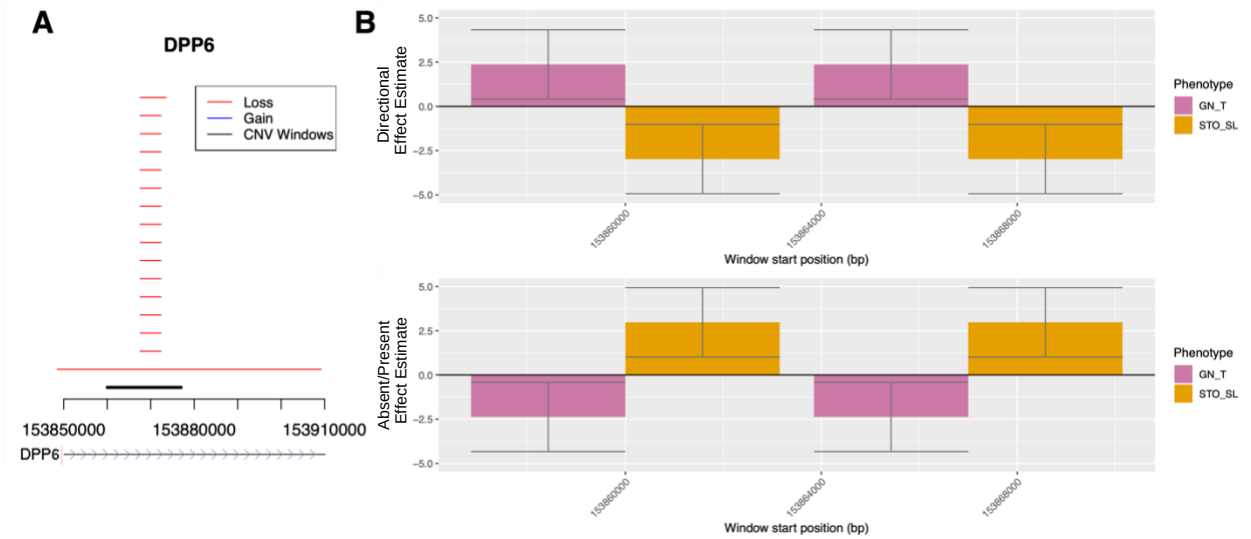

Figure S9. Region Plot, *DPP6*. A) Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < $5x10^{-4}$ are shown: lower facial depth (GN_T) and lower lip height (STO_SL).

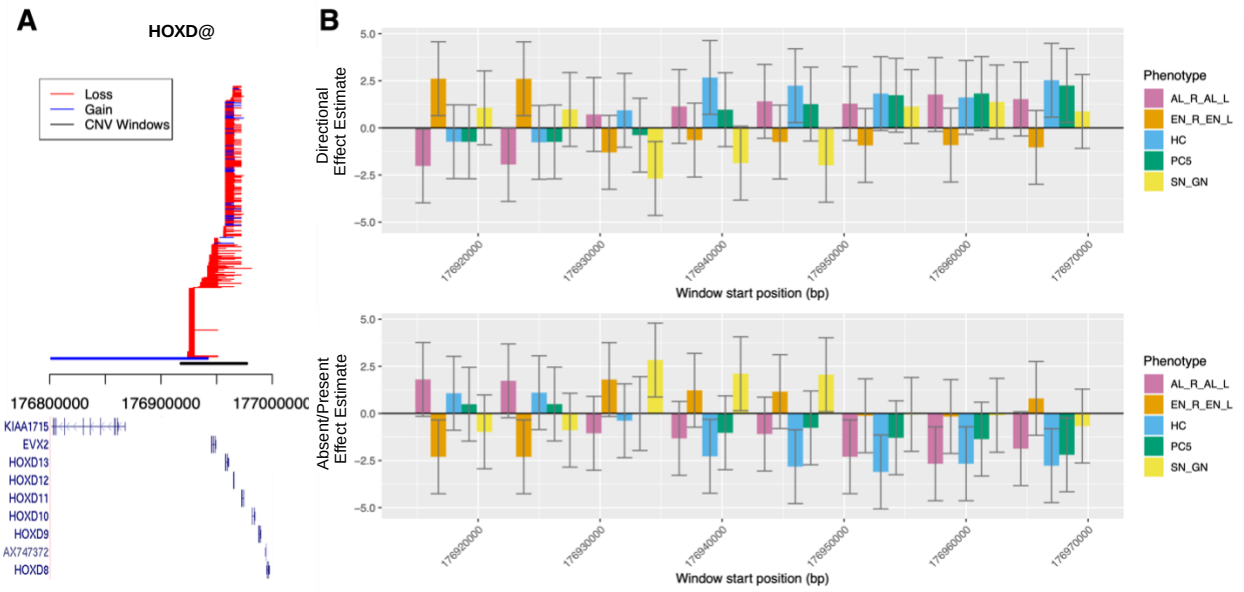

Figure S10. Region Plot, *HOXD@*. A) Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < $5x10^{-4}$ are shown: nasal width (AL_R_AL_L), inner canthal width (EN_R_EN_L), head circumference (HC), principal component 5: nose shape, height of mouth (PC5), and lower facial height (SN_GN).

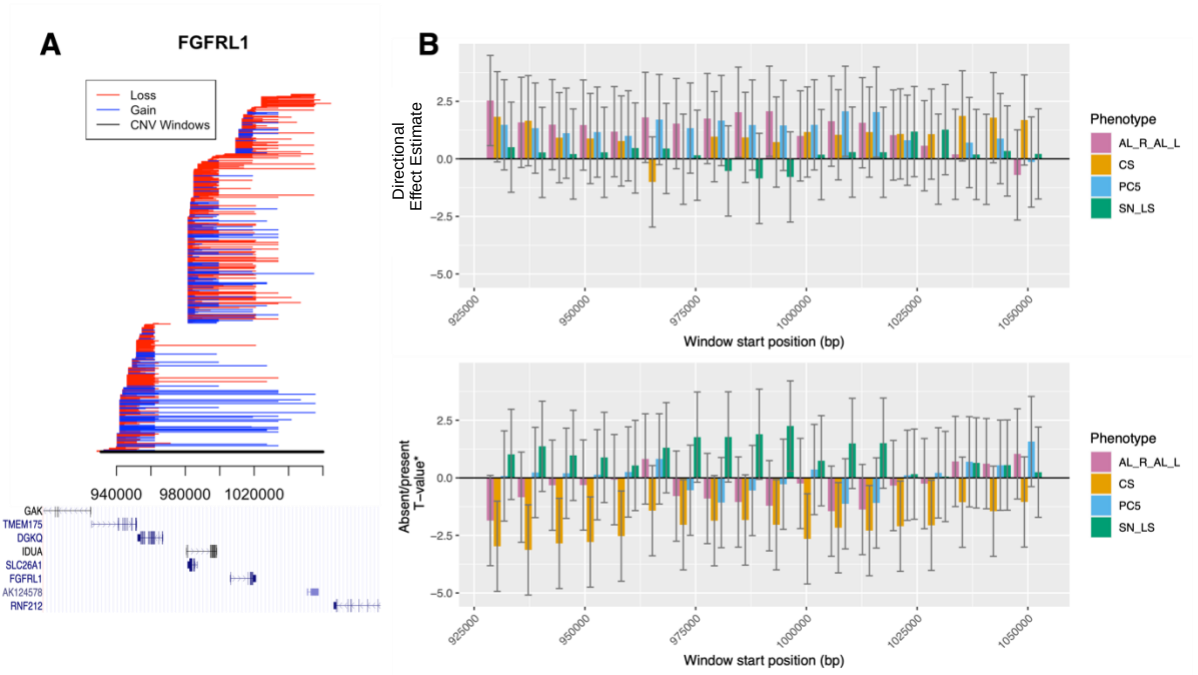

Figure S11. Region Plot, *FGFRL1*. A) Loss (red) and gain (blue) CNVs. Each line represents a unique CNV allele from one individual with the genes in the region shown below. The CNV analysis region is shown in black **B)** Test statistic t-values (effect estimate / standard error of effect estimate) across the region with 95% confidence intervals in the *directional* model (top) and *absent/present* model (bottom). Phenotypes with at least one window with p-value < 0.05 are shown: nasal width (AL_R_AL_L), centroid size (CS), principal component 5: nose shape, height of mouth (PC5), and philtrum length (SN_LS).

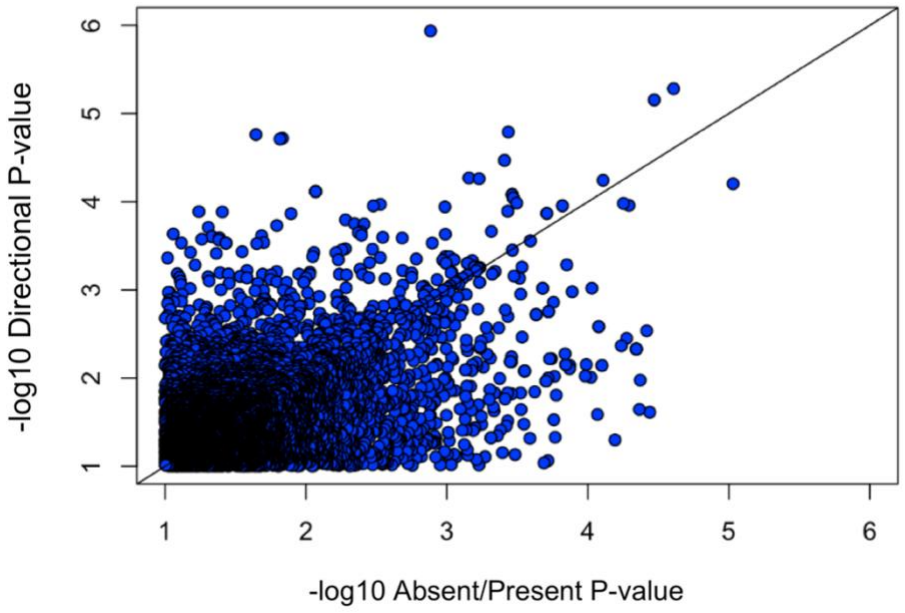

Figure S12. Scatterplot Comparing *Absent/Present* and *Directional* models. Scatterplot of $-\log_{10} pvalue$ from each model. Each point represents a window with both gains and losses.

Table S1. Facial Phenotypes

Measurement	Description
Total Facial Measurement	
Centroid Size	Face size
Allometry	Variation in shape due to size
Mid-Face PC1	The first principal component from a midfacial landmark network around the nose and mouth
Principal Component 1	Upper facial height, mid facial width
Principal Component 2	Overall facial height, lower facial height
Principal Component 3	Upper and middle facial width
Principal Component 4	Width of nose, mandible height
Principal Component 5	Nose shape, height of mouth
Linear Distances	
AL_R_AL_L	Nasal Width
AC_PRN	Nasal Ala Length (average)
CH_R_CH_L	Mouth Width
CPH_R_CPH_L	Philtrum Width
EN_EX	Palpebral Fissure Length (average)
EN_R_EN_L	Inner canthal Width
EX_R_EX_L	Outer canthal Width
GN_T	Lower Facial Depth (average)
LI_SL	Cutaneous Lower Lip Height
LS_STO	Upper Vermillion Height
N_GN	Morphological Facial Height
N_MEN	Nasion to Midendocanthion
N_PRN	Nasal Bridge Length
N_SN	Nasal Height
N_STO	Upper Facial Height
N_T	Upper Facial Depth (average)
SBAL_R_SBAL_L	Subnasal Width
SN_GN	Lower Facial Height
SN_LS	Philtrum Length
SN_PRN	Nasal Protrusion
SN_STO	Upper Lip Height
SN_T	Midfacial Depth (average)
STO_LI	Lower Vermillion Height
STO_SL	Lower lip height
T_R_T_L	Facial Width
Non-landmark defined	
Head Circumference	Direct occipital frontal circumference

Table S2. Non-missing Phenotypes

Phenotype	Non-missing Subjects
CS	3388
ALLOMETRY	3388
MIDFAC_MOD_PC1	3387
PC1	3388
PC2	3388
PC3	3388
PC4	3388
PC5	3388
T_R_T_L	3388
N_T	3386
SN_T	3388
GN_T	3385
N_GN	3388
N_STO	3387
SN_GN	3387
EN_R_EN_L	3386
EX_R_EX_L	3387
EN_EX	3384
AL_R_AL_L	3388
SBAL_R_SBAL_L	3384
SN_PRN	3387
AC_PRN	3378
N_SN	3385
N_PRN	3386
CH_R_CH_L	3387
CPH_R_CPH_L	3387
SN_LS	3388
SN_STO	3386
STO_SL	3385
LS_STO	3387
STO_LI	3388
LI_SLI	3382
N_MEN	3387
HC	2589

Table S5. Family Wise Error Rate Significance Thresholds

Analysis	Effective Number of Phenotypes	Effective Number of Tests*	Study wide FWER significance threshold
Window Analysis (Primary)	23	6913	3.14×10^{-7}
Window Analysis (Secondary)	23	1519	1.433×10^{-6}
Common CNV Analysis	23	166	1.31×10^{-5}
Common Facial Variation SNP GWAS gene set (Primary)	22	17	1.34×10^{-4}
Common Facial Variation SNP GWAS gene set (Secondary)	22	11	2.01×10^{-4}
Phenotypic GWAS gene set (Primary)	22	26	8.74×10^{-5}
Phenotypic GWAS gene set (Secondary)	22	26	8.74×10^{-5}

*For one phenotype

Table S7. CNV summary statistics

	CNV Length	Loss Length	Gain Length	Total Number of CNVs per Person	Total Number of Losses per Person	Total Number of Gains per Person	Gain/Loss Ratio per Person*
Minimum	1001	1001	1001	18	15	0	0.006536
Quantile 1	4375	4088	8251	43	35	5	0.1111
Median	8904	7946	18680	50	41	8	0.1818
Mean	21600	16960	44310	63.88	53.05	10.83	0.2591
Quantile 3	20950	17100	45940	63	48	12	0.2745
Maximum	2741000	1753000	2741000	342	413	154	4.667
Standard Deviation	45564.48	31301.84	82605.6	45.705	44.22	11.359	0.3122

*Subset of subjects with at least one gain

Table S8. Top regions with CNVs <10 kb

Region	Associated Phenotype	Absent/ Present All CNVs	Absent/Present CNVs >10kb	Directional All CNVs	Directional CNVs >10kb	All CNVs N loss; n gain (CNVs >10kb: N loss; n gain)
Chr18: 77147000-77283000	Head Circumference	1.31×10^{-3}	1.34×10^{-3}	1.16×10^{-6}	2.41×10^{-6}	73;12 (67; 12)
	Lower Facial Depth (average)	1.03×10^{-4}	1.67×10^{-4}	7.03×10^{-3}	4.29×10^{-3}	55;7 (53; 4)
	Upper Lip Height	3.47×10^{-4}	1.14×10^{-4}	5.80×10^{-3}	1.14×10^{-3}	55;7 (53; 4)
	PC1	3.71×10^{-4}	6.45×10^{-4}	3.05×10^{-2}	3.00×10^{-3}	55;7 (53; 4)
Chr10: 111034000-111058000	Upper Facial Depth (average)	2.64×10^{-6}	Too few CNVs	2.64×10^{-6}	Too few CNVs	13;0 (3;0)
Chr4: 3423000-3538000	Upper Facial Depth – average	5.20×10^{-6}	No small CNVs	1.51×10^{-1}	No small CNVs	41; 7 (41;7)
	Midfacial Depth (average)	4.79×10^{-5}		2.21×10^{-1}		41; 7 (41;7)
Chr2: 34230000-34324000	Subnasal Width	2.47×10^{-5}	3.38×10^{-5}	5.23×10^{-6}	7.03×10^{-6}	19; 1 (18;1)
	Nasal Width	7.82×10^{-5}	1.52×10^{-4}	5.71×10^{-5}	1.11×10^{-4}	19; 1 (18;1)
	Mid-Face PC1	6.63×10^{-4}	7.16×10^{-4}	4.65×10^{-4}	4.99×10^{-4}	19; 1 (18;1)
Chr16: 1225000-1508000	Nasal Ala Length – average	9.35×10^{-6}	No small CNVs	6.26×10^{-5}	No small CNVs	1; 9 (1;9)
	Subnasal Width	5.12×10^{-5}		1.10×10^{-4}		1; 9 (1;9)
	Nasal Width	5.60×10^{-5}		1.05×10^{-4}		1; 9 (1;9)
	Midfacial Depth (average)	2.85×10^{-4}		3.32×10^{-2}		12; 7 (12;7)

Table S12. Primary analysis windows below p-value thresholds

P-value Threshold	Total	Directional model only		Both models		Absent/Present model only	
	N Windows	N Windows	Percent	N Windows	Percent	N Windows	Percent
$\mathbf{5.0 \times 10^{-2}}$	20758	7494	36.1%	3761	18.1%	9503	45.8%
$\mathbf{1.0 \times 10^{-2}}$	4726	1797	38.0%	656	13.9%	2273	48.1%
$\mathbf{1.0 \times 10^{-3}}$	580	215	37.1%	76	13.1%	289	49.8%
$\mathbf{1.0 \times 10^{-4}}$	77	31	40.3%	5	6.5%	41	53.2%
$\mathbf{1.0 \times 10^{-5}}$	7	4	57.1%	0	0.0%	3	42.9%