Description of Additional Supplementary Files

File Name: Supplementary Data 1

Description: SNPs reaching at least genome-wide significance ($P < 5 \times 10^{-8}$) in the GWAS meta-analysis. For each lead SNP, this table contains chromosome (Chr), position (Pos, GRCh37/hg19), cytogenetic band (Locus), allele information, the frequency of the A1 allele in the US and UK sample ($A1_{freq}$ US & $A1_{freq}$ UK), the canonical correlation values (CC_{US} & CC_{UK}), one-sided, right-tailed identification P-values ($P_{CCA\text{-}US}$ & $P_{CCA\text{-}UK}$), one-sided, right-tailed verification P-values ($P_{UniVar\text{-}UK}$ & $P_{UniVar\text{-}US}$), one-sided meta-analysis *P*-value using either subsample as identification dataset ($P_{Meta\text{-}US}$ & $P_{Meta\text{-}UK}$), followed by the lowest meta-analysis P-value (PMeta_lowest) and whether this lowest *P* passed the study-wide significance threshold. Next, the table contains the candidate gene to which this signal was annotated, a list of SNPs in LD that were previously associated with facial shape, the R^2 and D' between the lead SNP and previously reported SNPs, a reference to the original report (Ref.) and whether this association was established in a cohort that is independent to the currently used cohort. The next column provides DOI references to previous literature associating additional SNPs in a genomic region of ± 250kb but not in LD, that were previously associated with facial morphology. The last three columns contain the RSID of the SNP with the minimum p-value for that locus in the Tanzanian cohort ($RSID_{Tanzanian}$) and the corresponding minimum adjusted ($P_{Tanzanian\_adjusted}$) and FDR-adjusted right-tailed identification p-values ($FDR_{Tanzanian}$).

File Name: Supplementary Data 2

Description: Gene-association scores for predicted high confidence interactions with FGFR3. For each gene, this table contains chromosome (Chr), transcription start and end sites (Tx start, Tx end, GRCh37/hg19), STRING confidence score of the predicted interaction (only interactions with score >0.7 are included), number of SNPs mapped to each gene (N_{SNPs}), aggregated gene association score (P_{gene}; one-sided F-test), MAGMA regression coefficient ($BETA_{gene\text{-}set}$), standard error of the regression coefficient ($SE_{gene\text{-}set}$), MAGMA gene-set p-value ($P_{gene\text{-}set}$; one-sided single sample t-test).

File Name: Supplementary Data 3

Description: Gene ontology (GO) biological process enrichment for the 19 achondroplasia (ACH)-derived lead SNPs (sheets 1. ACH-informed) and for the 203 lead SNPs of the White et al. GWAS (sheet 2. Face uninformed), generated using GREAT v.4.0.4. For each biological process, this table contains the hypergeometric fold enrichment ($Fold_{hyper}$), the corresponding FDR q-value (FDR Q_{hyper}), the hypergeometric rank ($Rank_{hyper}$), the number of observed hits in genes related to the process (Obs_{hyper}) and the total number of genes related to that process ($Total_{hyper}$), the hypergeometric gene set coverage (Cov_{hyper}), the binomial fold enrichment ($Fold_{binom}$), the corresponding FDR q-value (FDR Q_{binom}), the binomial rank ($Rank_{binom}$), the number of observed hits in regions related to the process (Obs_{binom}) and the binomial region set coverage (Cov_{binom}). Processes are displayed in ascending order of hypergeometric-based enrichment ranking.

File Name: Supplementary Data 4

Description: Comparison of the Gene Ontology (GO) biological process enrichment between different gene sets using a foreground-background approach. The different sheets contain the results of different combinations of foreground and background, as indicated in the file. Each sheet contains the biological processes with a nominally significant enrichment (right-tailed hypergeometric p-value < 0.05, p_up) in the foreground versus the background. The GO identifier (GO_ID) and name of the biological process are provided, as well as the fold enrichment between foreground and background set (fold), the observed and expected number of genes related to the biological process (n_obs, n_exp). For biological processes that showed significant biological process enrichment in the foreground (right-tailed hypergeometric FDR-adjusted p < 0.05, cfr. Supplementary Data 3), false discovery rate (FDR)-adjusted p-values (q_up) were calculated using an adapted Benjamini-Hochberg procedure with a 5% FDR.

File Name: Supplementary Data 5

Description: Description of the five achondroplasia (ACH)-associated traits (body height, infant head circumference, lung volume, obstructive sleep apnea, sitting height ratio) and two negative control traits (hormone-sensitive cancer, cigarettes per day measurement) that were used to compute genetic correlations with the ACH-derived facial trait. For each trait, this table contains the corresponding EFO ID, sample size (N) and ancestry of the discovery cohort, corresponding publication, and URL to the open-source GWAS summary statistics.

File Name: Supplementary Data 6

Description: This table details the closeness of the matching between individuals with ACH and controls when testing which facial segments differ significantly between ACH and controls. Each row represents an individual from the ACH sample with their sex (ACH_sex) and age (ACH_age) provided. For each facial segment (63 iterations), in a random order, each ACH individual is matched to the control of the same sex (Matched_control_sex) that was closest in age (Matched_control_age). Per individual, the age of the matched control per iterations is provided, as well as the standard deviation (SD) of these ages, the mean age of the matched control across the iterations and the difference in age between the individuals with ACH and the mean of their respective matched controls. These age differences are represented in the histogram below the chart, and summary statistics on a group level are provided.

File Name: Supplementary Data 7

Description: Table containing the Malahanobis distance to the control mean (mahal_dist) for each individual of the control sample and each individual of the achondroplasia sample, and the right-tailed p-values associated with these distances calculated using the chi-squared cumulative distribution function (p_value).