Description of Additional Supplementary Files

Supplementary Data 1: FinnGen discovery cohort GWAS results of variants with P values below 1E-4 in combined cleft lip with or without cleft palate (CL/P) and cleft palate only (CP) cohorts (n = 355). Twosided P values are obtained from a likelihood ratio test in regression analysis and are not corrected for multiple comparisons.

Supplementary Data 2: FinnGen discovery cohort GWAS results of variants with P values below 1E-4 in cleft palate only (CP) (n = 228). Two-sided P values are obtained from a likelihood ratio test in regression analysis and are not corrected for multiple comparisons.

Supplementary Data 3: FinnGen discovery cohort GWAS results of variants with P values below 1E-4 cleft lip with or without cleft palate (CL/P) (n = 151). Two-sided P values are obtained from a likelihood ratio test in regression analysis and are not corrected for multiple comparisons.

Supplementary Data 4: Frequencies of rs570516915 (T>G) and rs7546014 (G>A) genotypes in the combined Finnish cohorts. Significance was evaluated by pairwise two-sided chi-squared test between distributions of genotypes with a one allele dosage difference. Thus, each genotype distribution was compared with that on the previous line, expect for TT/AA which was compared with TT/GA.

Supplementary Data 5: FinnGen replication cohort GWAS results of variants with P values below 1E-4 in cleft palate only (CP) (n = 165). Two-sided P values are obtained from a likelihood ratio test in regression analysis and are not corrected for multiple comparisons.

Supplementary Data 6: Estonian Biobank cohort GWAS results of variants with P values below 0.05 in cleft palate only (CP) (n = 71). Two-sided P values are obtained from a likelihood ratio test in regression analysis and are not corrected for multiple comparisons.

Supplementary Data 7: GWAS meta-analysis results of variants with P values below 1E-4 in cleft palate only (CP) (FinnGen Discovery, FinGen Replication and Estonian Biobank cohorts, n = 464, direction of effect shown in the same order). P values of the meta-analysis are two-sided P values derived from a fixed-effect meta-analysis and were not adjusted for multiple testing

Supplementary Data 8: Number of CP cases in FinRegistry and rs570516915 allele frequencies in 19 administrative regions of Finland. Minor allele frequency of rs570516915 was estimated based on birthplace data of 306,678 FinnGen study participants. The code column shows region codes from Statistics Finland (www.tilastokeskus.fi/en/luokitukset/maakunta).

Supplementary Data 9: Number of transgenic mouse embryos expressing LacZ under MCS-9.7 with the non-risk (T) and risk (G) alleles of rs570516915 and results from blinded assessment of the X-gal staining. S, single insertions; T, tandem insertions.

Supplementary Data 10: Primers used for CRISPR screening and qRT-PCR.

Supplementary Data 11: Primers used for ChIP-qPCR.