

---

# *De novo* variants implicate chromatin modification, transcriptional regulation, and retinoic acid signaling in syndromic craniosynostosis

## Authors

Andrew T. Timberlake, Stephen McGee,  
Garrett Allington, ..., Richard P. Lifton,  
Kristopher T. Kahle, Paul Kruszka

## Correspondence

[rickl@rockefeller.edu](mailto:rickl@rockefeller.edu) (R.P.L.),  
[kahle.kristopher@mgh.harvard.edu](mailto:kahle.kristopher@mgh.harvard.edu) (K.T.K.)

**Craniosynostosis is the most frequent congenital cranial deformity. Analysis of exome sequence data from 526 probands with syndromic craniosynostosis identified 13 genes surpassing thresholds for genome-wide significance, implicating regulation of chromatin modification, gene transcription, and osteoblast differentiation in disease pathogenesis.**



# De novo variants implicate chromatin modification, transcriptional regulation, and retinoic acid signaling in syndromic craniosynostosis

Andrew T. Timberlake,<sup>1</sup> Stephen McGee,<sup>2</sup> Garrett Allington,<sup>3</sup> Emre Kiziltug,<sup>3</sup> Erin M. Wolfe,<sup>4</sup> Amy L. Stiegler,<sup>5</sup> Titus J. Boggon,<sup>6</sup> May Sanyoura,<sup>2</sup> Michelle Morrow,<sup>2</sup> Tara L. Wenger,<sup>7</sup> Erica M. Fernandes,<sup>8</sup> Oana Caluseriu,<sup>9</sup> John A. Persing,<sup>10</sup> Sheng Chih Jin,<sup>11</sup> Richard P. Lifton,<sup>12,\*</sup> Kristopher T. Kahle,<sup>13,14,15,\*</sup> and Paul Kruszka<sup>2</sup>

## Summary

Craniosynostosis (CS) is the most common congenital cranial anomaly. Several Mendelian forms of syndromic CS are well described, but a genetic etiology remains elusive in a substantial fraction of probands. Analysis of exome sequence data from 526 proband-parent trios with syndromic CS identified a marked excess (observed 98, expected 33,  $p = 4.83 \times 10^{-20}$ ) of damaging *de novo* variants (DNVs) in genes highly intolerant to loss-of-function variation (probability of LoF intolerance  $> 0.9$ ). 30 probands harbored damaging DNVs in 21 genes that were not previously implicated in CS but are involved in chromatin modification and remodeling (4.7-fold enrichment,  $p = 1.1 \times 10^{-11}$ ). 17 genes had multiple damaging DNVs, and 13 genes (*CDK13*, *NFIX*, *ADNP*, *KMT5B*, *SON*, *ARID1B*, *CASK*, *CHD7*, *MED13L*, *PSMD12*, *POLR2A*, *CHD3*, and *SETBP1*) surpassed thresholds for genome-wide significance. A recurrent gain-of-function DNV in the retinoic acid receptor alpha (*RARA*; c.865G>A [p.Gly289Arg]) was identified in two probands with similar CS phenotypes. CS risk genes overlap with those identified for autism and other neurodevelopmental disorders, are highly expressed in cranial neural crest cells, and converge in networks that regulate chromatin modification, gene transcription, and osteoblast differentiation. Our results identify several CS loci and have major implications for genetic testing and counseling.

## Introduction

Craniosynostosis (CS; [MIM: PS123100]) is the most frequent cranial birth defect in humans (1/2,000 births), requiring surgical intervention in infancy to decrease life-threatening increases in intracranial pressure and mitigate the risk of adverse neurodevelopmental outcomes. Approximately 15% of cases are syndromic and present with extracranial and developmental anomalies.<sup>1</sup> More than 40 known Mendelian forms of CS are caused by variants in genes that impact the FGF/MAPK, BMP, Wnt, hedgehog, retinoic acid, STAT, and ephrin signaling pathways.<sup>2</sup> These genetic variants have been speculated to cause premature suture fusion by altering the balance between suture stem cell proliferation and differentiation, but the molecular mechanisms of syndromic CS are still poorly understood.<sup>3</sup>

Despite many recent advances in CS genetics, a molecular diagnosis is still only achieved in about 70% of syndromic cases studied; nearly 30% of genetic evaluations remain unrevealing despite striking clinical presentations.<sup>4</sup> We previously reported exome sequencing of a

dozen proband-parent trios with unexplained syndromic CS and identified likely pathogenic variants in 9 of 12 cases studied.<sup>5</sup> These data suggest that syndromic CS is a particularly rich vein for the discovery of CS loci. We sought to build on these findings by studying a substantially larger CS cohort.

## Subjects and methods

### Enrollment

Individuals in this study were referred to GeneDx for clinical exome trio sequencing for diagnosis of a suspected Mendelian disorder. Clinical testing, including collection of structured phenotypic information as Human Phenotype Ontology (HPO) terms, was conducted as previously described.<sup>6</sup> 549 probands were identified with HPO terms under the parent term “craniosynostosis” (HP:0001363), and six additional parent-child trios from Yale were included for analysis in tandem with the GeneDx cohort; in total, we studied 526 probands to identify risk loci after excluding 29 probands with pathogenic variants at established loci. This study was conducted in accordance with all guidelines set forth by the Western institutional review board, Puyallup,

<sup>1</sup>Hansjörg Wyss Department of Plastic Surgery, NYU Langone Medical Center, New York, NY, USA; <sup>2</sup>GeneDx, Gaithersburg, MD, USA; <sup>3</sup>Department of Pathology, Yale University School of Medicine, New Haven, CT, USA; <sup>4</sup>Division of Plastic and Reconstructive Surgery, University of Miami Hospital, Miami, FL, USA; <sup>5</sup>Department of Pharmacology, Yale University, New Haven, CT, USA; <sup>6</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; <sup>7</sup>Division of Genetic Medicine, University of Washington, Seattle, WA, USA; <sup>8</sup>Nemours Children's Health, Wilmington, DE, USA; <sup>9</sup>Department of Medical Genetics, University of Alberta, AB, Canada; <sup>10</sup>Section of Plastic and Reconstructive Surgery, Yale University School of Medicine, New Haven, CT, USA; <sup>11</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; <sup>12</sup>Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY, USA; <sup>13</sup>Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; <sup>14</sup>Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, MA, USA; <sup>15</sup>Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

\*Correspondence: [rickl@rockefeller.edu](mailto:rickl@rockefeller.edu) (R.P.L.), [kahle.kristopher@mgh.harvard.edu](mailto:kahle.kristopher@mgh.harvard.edu) (K.T.K.)

<https://doi.org/10.1016/j.ajhg.2023.03.017>

© 2023 American Society of Human Genetics.

Washington (WIRB 20162523), and the inclusion of six trios from Yale was approved by the Yale Human Investigation Committee institutional review board. Informed consent for genetic testing was obtained from all individuals undergoing testing, and WIRB waived authorization for use of de-identified aggregate data for these purposes. Individuals or institutions who opted out of this type of data use were excluded. GeneDx exomes were captured with either SureSelect Human All Exon v.4 (Agilent Technologies, Santa Clara, CA), Clinical Research Exome (Agilent Technologies, Santa Clara, CA), or xGen Exome Research Panel v.1.0 (IDT, Coralville, IA) and sequenced with either 2 × 100 or 2 × 150 bp reads on the HiSeq 2000, 2500, or 4000 or NovaSeq 6000 (Illumina, San Diego, CA) platform. GATK joint genotyping variant calling was performed as previously described.<sup>7</sup> Exome sequencing of these samples was performed by exon capture with the same IDT xGen capture reagent, followed by 99-base paired-end sequencing on the Illumina HiSeq 2000. For identification of novel CS loci, de-identified data pertaining to DNVs in probands was abstracted from the combined 555 trios. For participating individuals with identifiable clinical images, explicit written consent was provided for publication of these images and associated clinical data.

### Variant annotation and imputation of ethnicities

Annotation of aggregated DNVs was performed with RefSeq transcripts and ANNOVAR.<sup>8</sup> The transcript with the most severe consequence was selected, and all associated annotations were based on the predicted effect of the variant on that specific transcript. Allele frequencies of identified variants were taken from the ExAC and GnomAD databases. Principal-component analysis, as previously described,<sup>7</sup> led to the following ancestry groups: 270 European, 95 Hispanic, 27 African American, 24 South Asian, 9 East Asian, and 22 Middle Eastern among the 526 individuals studied. The impact of nonsynonymous variants was predicted from the MetaSVM rank score; scores greater than 0.83357 served as a threshold for predicting that the variant was deleterious (MetaSVM “D,” D-mis).<sup>9</sup> Heterozygous variants (hemizygous for males in non-PAR X chromosome regions) with wildtype parental genotypes were considered. Variants were filtered on the basis of read depths of at least 10× in all family members, variant quality-score log odds (VQSLOD) of at least -10, Phred-scaled p value (Fisher’s exact test) < 30, with an alternate allele count of at least 4 in the proband, and MAF > 0.2 in the proband. Variants were further filtered on the basis of a population frequency not above 10<sup>-4</sup> in gnomAD. The number of observed *de novo* variants matched both expectation and prior experimental results. There were no differences in the rates of *de novo* variants identified among different ancestry groups.

### Burden of *de novo* variants

Statistical analysis of the burden of DNVs in individuals with CS and in autism controls was performed in R with the denovolyzeR package as previously described.<sup>10</sup> The expected number of DNVs in case and control cohorts across variant classes was calculated, and this value was compared with the observed number in each cohort via Poisson statistics.<sup>11</sup> For gene-set enrichment analyses, only variants observed or expected in genes within the specified gene set were included in each statistical test. Statistical analysis of the probability of observing multiple damaging DNVs in an individual gene in the syndromic-CS cohort was also performed in R with denovolyzeR.<sup>11,12</sup> p values for enrichment in damaging DNVs at the individual gene level represent the upper tail of the

Poisson probability density function. These values were corrected via the Benjamini-Hochberg method for all genes with probability of LoF intolerance (pLI) > 0.9 (n = 3,063) to give false discovery rates; q values < 0.05 were considered significant. Control trios were those sequenced from the Simons Foundation Autism Research Initiative Simplex Collection.<sup>13</sup> Simplex families, comprising two unaffected parents, one child with autism, and one unaffected sibling, underwent whole-exome sequencing; 1,789 trios of unaffected family members were sequenced on the same platform and served as controls for this study.

### Contribution of *de novo* variants to syndromic CS

We infer that the number of probands with *de novo* loss-of-function (LoF) variants in high-pLI genes in 526 trios who had sporadic syndromic CS but did not have variants in known genes (n = 97) more than those expected by chance (n = 33) represents the number of subjects in whom these variants confer CS risk (n = 64). Comparing this number to the total number of trios considered (64/526) yields an estimate of the fraction of individuals in whom these variants are expected to contribute to disease risk: ~12.2%.

### DisGeNET analysis

Overlap between the genes with damaging DNVs in our CS cohort and the DisGeNET Intellectual Disability, Neurodevelopmental Disorder, and Autism Spectrum Disorder curated gene lists were mapped with the disease2disease\_by\_gene function of the disgenet2r R package ([www.disgenet.org](http://www.disgenet.org)). The significance of each overlap was calculated with one-sided Fisher’s exact tests corrected for false discovery via the Benjamini-Hochberg method. The total background number of DisGeNET risk genes used for these calculations at the time of accession was 21,666.

### Probability of observing 17 genes with multiple damaging DNVs

We determined the probability of observing 17 genes with at least two damaging DNVs in our cohort by using a permutation function in denovolyzeR-denovolyzeMultiHits().<sup>11</sup> In total, 182 damaging (LoF + deleterious missense [D-mis]) DNVs were observed in our cohort. Given the individual probability of DNV in each gene, the probability of observing more than two damaging DNVs by chance in 17/182 genes was determined.<sup>11</sup> The number of times at least 17 genes had more than one hit in an iteration was counted and was found to be zero. The expected number of genes that would be mutated more than once, on the basis of individual mutation probabilities, was 2.2.

### Probability of observing two identical *de novo* HUWE1 (p.Arg110Trp) or RARA (p.Gly289Arg) variants

Calculating this probability is analogous to the “birthday paradox”; i.e., the chance that in a set of “n” randomly chosen people, at least one pair will have the same birthday. A complete description of this calculation is found in [Note S1](#).

### Determination of risk gene lists

The “cohort syndromic CS risk gene list” includes individual genes surpassing thresholds for genome-wide significance in our cohort; these are genes in which damaging DNVs were identified in the chromatin organization gene ontology term and genes in which DNVs were identified in known syndromic loci. We added syndromic genes well established in the literature but not identified

in our cohort to this list to generate the “all high-confidence syndromic CS risk genes” list. A complete list of risk genes in each group can be found in [Table S1](#). We compiled a list of genes with rare risk variation in ASD from two papers describing genes that contribute to ASD risk via both *de novo* and rare inherited variants ([Table S1](#)).<sup>14,15</sup> We compiled a list of developmental disorder (DD) risk genes from the Deciphering Developmental Disorders study,<sup>16</sup> which describes genes enriched in damaging DNVs ([Table S1](#)).

### Assessing for enrichment in cranial neural crest cell genes

We used processed RNA-seq data from six human cranial neural crest cell (CNCC) samples derived from embryonic stem cells or induced pluripotent stem cells (GEO: GSM1817212, GSM1817213, GSM1817214, GSM1817215, GSM1817216, and GSM1817217). FPKM values across each sample were averaged, and the genes included in our exome capture were ranked according to expression. Genes in the top 25% of expression were designated “highly expressed” in human CNCCs ([Table S2](#)). The number of protein-damaging DNVs was compared to the expected distribution in *de novo*lyzeR with the “includeGenes” function, which included those genes highly expressed in CNCCs with pLI > 0.9 (n = 1,593).

### Weighted gene co-expression network analysis

A processed bulk-mRNAseq expression dataset encompassing sixteen human brain regions across human development was used for robust consensus weighted gene co-expression network analysis (rWGCNA).<sup>17</sup> The analysis was limited to the timepoints between gestational week 9 and postnatal year 3. Samples that were more than 3 standard deviations above the mean sample network connectivity were removed. Network analysis was performed with WGCNA, and genes were assigned to specific modules on the basis of bi-weight mid-correlations among genes.<sup>18</sup> A soft threshold power of 10 was chosen to yield scale-free topology ( $r^2 > 0.9$ ). Then a signed co-expression network was generated. The topological overlap matrix was clustered hierarchically via average linkage hierarchical clustering (“1 – TOM” was used as a dis-similarity measure). The topological overlap dendrogram was used for defining modules with a minimum module size of 40, deep split of 4, and merge threshold of 0.1.

### Module enrichment analysis

Module gene lists were obtained via WGCNA as described above. In a background set of all genes categorized in co-expression modules, logistic regression was used for an indicator-based enrichment:  $\text{is.disease} \sim \text{is.module} + \text{gene covariates}$  (GC content, gene length, and mean expression in bulk RNA-seq atlas), as described previously.<sup>19</sup> Of the 88 WGCNA modules, the gray module, by WGCNA convention,<sup>17</sup> contains all genes that do not co-express and are consequently unassigned to a co-expression network. Thus, the gray module was excluded from enrichment testing, and enrichment significance was defined at the Bonferroni multiple-testing cutoff ( $\alpha = 0.05/88 = 5.68 \times 10^{-4}$ ).

### Gene ontology enrichment analysis

Using the the EnrichR R package<sup>20</sup> and identified gene modules significantly enriched for genes in our cohort’s CS gene list (“brown,” “greenyellow,” “royalblue,” and “salmon” modules), we performed enrichment analysis for gene ontologies (biological processes). The top 10 terms with lowest p values were reported in the order of highest to lowest combined Z-scores. p values

less than 0.05 and Z scores greater than 2 were considered significant.

### Cell-type enrichment analysis

Cell-type-enriched genes (cell-type markers) were obtained from a scRNA-seq atlas that maps the human midgestational cortex (17–18 gestational weeks).<sup>21</sup> In a background set of all genes expressed in at least three cells of the scRNA-seq atlas, we used a logistic regression for indicator-based enrichment:  $\text{is.cell.type} \sim \text{is.disease} + \text{gene covariates}$  (GC content, gene length). All p values were adjusted with Bonferroni correction. Enrichment significance was defined at the Bonferroni multiple-testing cutoff ( $\alpha = 0.05/32 = 1.56 \times 10^{-3}$ ).

### Estimating the number of risk genes in syndromic CS

We followed a previously described Monte Carlo simulation strategy to estimate the number of risk genes that are DNV targets.<sup>22</sup> We defined K to be the number of observed damaging DNVs in all genes (n = 19,347) among cases. R1 indicates the number of genes mutated exactly twice in affected individuals, and R2 indicates the number of genes mutated three times or more. Defined as  $E = (M1 - M2)/M1$ , where M1 and M2 are the observed and expected count of damaging DNVs per trio, respectively. We then simulated the likelihood function as follows: First, we randomly selected G risk genes from the gene set (n = 19,347). Next, we simulated the number of contributing damaging variants in risk genes, i.e., C, by sampling once from the binomial (K,E) distribution. Then, we simulated C contributing damaging variants in G risk genes and K – C non-contributing damaging variants in the complete gene set by using each gene’s damaging mutability score as probability weights. We performed 20,000 simulations for G from 5 to 100 and calculated the likelihood function L(G) as the proportion of simulations in which the number of genes with two damaging variants equals R1 and the number of genes with three or more damaging variants equals R2. We then estimated the number of risk genes by using the maximum-likelihood estimate.

## Results

### Exome sequencing of trios with syndromic CS

To systematically identify genetic causes of syndromic CS, we ascertained 555 proband-parent trios who had undergone genetic evaluation for a suspected Mendelian cause of CS but for whom no causal variant identified (see [subjects and methods](#)). Proband had syndromic sagittal, metopic, coronal, lambdoid, or multi-suture CS. Whole-exome sequencing was performed on the Illumina platform, and variants were called and analyzed as described in the [subjects and methods](#).<sup>10</sup> The impact of nonsynonymous variants on protein function was inferred from the MetaSVM rank score.<sup>9</sup> D-mis and LoF variants were designated as protein-damaging variants ([subjects and methods](#)).

### 29 individuals harbor pathogenic variants in genes with known syndromic CS associations

In 555 trios, we identified 29 individuals with DNVs in known CS-associated genes.<sup>2</sup> The list of syndromic CS-associated genes analyzed appears in [Table S3](#). DNVs

were identified in *ERF* (MIM: 611888), *ZIC1* (MIM: 600470), *SKI* (MIM: 164780), *TGFBR2* (MIM: 190182), *EFNB1* (n = 3; MIM: 300035), *KMT2D* (n = 3; MIM: 602113), *KAT6A* (n = 2; MIM: 601408), *FAM20C* (MIM: 611061), *GNAS* (MIM: 139320), *JAG1* (MIM: 601920), *KRAS* (MIM: 190070), *SH3PXD2B* (MIM: 613293), and *ZEB2* (MIM: 605802). Two individuals had a variant consistent with Pfeiffer syndrome (*FGFR2* [MIM: 101600]; c.1019A>G [p.Tyr340Cys] [GenBank: NM\_001320658]; c.1697A>G [p.Glu566Gly] [GenBank: NM\_022970]), two with Crouzon syndrome (*FGFR2* [MIM: 123500]; c.314 A>G [p.Tyr105Cys] [GenBank: NM\_022970]; c.1040C>G [p.Ser347Cys] [GenBank: NM\_001320658]), one with Muenke syndrome (*FGFR3* [MIM: 602849]; c.749C>G [p.Pro250Arg] [GenBank: NM\_022965]), two with Crouzon syndrome with acanthosis nigricans (*FGFR3* [MIM: 612247]; c.1172C>A [p.Ala391Glu] [GenBank: NM\_001354810]), and one with the synonymous variant causing either Crouzon syndrome or craniosynostosis with Axenfeld-Rieger anomaly (*FGFR2* [MIM: 123500]; c.1032G>A [p.Ala344=] [GenBank: NM\_001320658]). We also replicated the previous association of *HUWE1* (MIM: 300697) variants at codon 110 to craniosynostosis<sup>23</sup> and identified three individuals with *de novo* missense variants at this codon (two with c.328C>T [p.Arg110Trp] [GenBank: NM\_031407]; one with c.329G>A [p.Arg110Gln] [GenBank: NM\_031407]). Pathogenic variants identified in known CS risk loci appear in [Table S4](#). The probability of observing any identical DNV (p.Arg110Gln) in a cohort of this size was  $6.5 \times 10^{-3}$  (see [subjects and methods](#)), and the probability of observing another variant at this same codon is exceptionally small, confirming the role of this specific codon in syndromic CS. Phenotypes observed in individuals with pathogenic variants at codon 110 of *HUWE1* are described in [Table S5](#).

### Burden of *de novo* variants in 526 probands with unsolved syndromic CS

After exclusion of 29 individuals with the variants described above, we analyzed the remaining 526 trios to identify genetic causes of syndromic CS. On average, we identified 1.16 *de novo* variants per proband, matching both prior experimental results and expectation ([Table 1](#)). All DNVs identified are described in [Table S6](#). Although the overall number of DNVs matched expectation, and there was no enrichment of *de novo* synonymous or missense variants inferred to be tolerated (T-mis), we identified highly significant enrichment of protein-damaging variants (LoF + D-mis;  $p = 2.2 \times 10^{-7}$ ). Virtually all of this enrichment was among genes that are highly intolerant to LoF variation ( $pLI > 0.9$ ); 98 damaging DNVs were observed, in comparison to 33 expected ( $p = 4.83 \times 10^{-20}$ , [Table 1](#)). High- $pLI$  genes are of particular interest in this cohort because high scores suggest greater selection against individuals with damaging heterozygous mutations.<sup>24</sup> From the observed fraction of individuals with damaging variants, we infer that disease is caused by damaging

DNVs in high- $pLI$  genes in ~12.2% of individuals studied (see [subjects and methods](#)).

### Damaging *de novo* variants in chromatin and OMIM genes

We performed pathway analysis by using GOrilla on all genes harboring damaging DNVs.<sup>25</sup> We identified significant enrichment in several gene ontology (GO) terms, including chromatin organization (3.6-fold enrichment,  $q = 0.001$ ) and chromatin remodeling (GO: 0006338; 6.2-fold enrichment,  $q = 0.01$ ; [Table S7](#)). In sum, we identified 30 damaging variants in genes included in the chromatin remodeling GO term (GO: 0006325); these included 17 LoF and 13 D-mis variants (4.7-fold enrichment,  $p = 1.1 \times 10^{-11}$ ; [Tables 1](#) and [2](#); [Figure 1](#)). This pathway was not enriched among 1,789 healthy control trios ( $p = 0.48$ ; [Table S8](#)). 21 different genes are mutated in this term, and 19 of these have a  $pLI > 0.95$  ([Table 2](#)). Although it was not encompassed by the GO term, we also identified a nonsense DNV in *INO80D* (MIM: 619207), a core component of the INO80 ATP-dependent chromatin remodeling complex (c.1669C>T [p.Arg557\*] [GenBank: NM\_017759],  $pLI = 1$ ). This enrichment remained highly significant after we removed variants in 13 genes implicated at the individual gene level (3.6-fold enrichment in damaging DNVs,  $p = 4.3 \times 10^{-7}$ ; [Table S9](#)). The results implicate variants disrupting the epigenome as a frequent cause of syndromic CS, accounting for ~6% of individuals studied and half the signal from DNVs in our cohort.

Among syndromic CS probands there were 76 damaging DNVs in 50 different high- $pLI$  genes that had previously been implicated in Mendelian diseases whose phenotypic spectrum did not previously include CS (OMIM genes; expected number = 10.9, 7.0-fold enrichment,  $p = 8.98 \times 10^{-38}$ ; [Tables 1](#) and [S10](#)). We assessed overlap between damaging DNVs identified in syndromic CS probands and those in probands with other diseases by using DisGeNET ([subjects and methods](#)). Variants in CS were enriched in genes previously implicated in intellectual disability ( $p = 4.16 \times 10^{-22}$ ; Fisher's exact test), autism ( $p = 5.83 \times 10^{-11}$ ; Fisher's exact test), and other neurodevelopmental disorders (NDDs) ( $p = 6.55 \times 10^{-20}$ ; Fisher's exact test; [Figure S1](#)). Control trios demonstrated a paucity of damaging variants in intolerant OMIM genes, as expected ( $p = 0.97$ ; [Table S8](#)).

The 22 damaging DNVs in high- $pLI$  genes without dominant OMIM associations were distributed across 21 genes with a variety of biological functions, including chromatin modification, RNA splicing, Wnt signaling, and neural development ([Table 3](#)). Although these 21 genes are strong candidates for conferring CS risk, sequencing additional syndromic CS trios will be necessary to determine which are *bona fide* CS risk genes.

### Thirteen genome-wide-significant CS risk genes

Ten individual genes surpassed thresholds for genome-wide significance of damaging DNVs in our cohort,

**Table 1. Burden of *de novo* variants in 526 probands with syndromic craniosynostosis**

	Observed		Expected		Enrichment	p
	n	Rate	n	Rate		
<b>All genes</b>						
Total	609	1.16	587.9	1.12	1.04	0.20
Synonymous	144	0.27	166.9	0.32	0.86	0.97
T-mis	283	0.54	300.1	0.57	0.94	0.85
D-mis	113	0.21	69.2	0.13	1.63	<i>8.72 × 10<sup>-7</sup></i>
LoF	67	0.13	51.7	0.098	1.30	0.02
Damaging	182	0.35	121.0	0.23	1.50	<i>1.42 × 10<sup>-7</sup></i>
<b>High-pLI genes (&gt;0.9) (n = 3,063)</b>						
Total	214	0.41	148.2	0.28	1.44	<i>2.26 × 10<sup>-7</sup></i>
Synonymous	38	0.072	41.7	0.080	0.91	0.74
T-mis	78	0.15	73.5	0.14	1.06	0.31
D-mis	56	0.11	19.6	0.037	2.85	<i>1.66 × 10<sup>-11</sup></i>
LoF	42	0.080	13.3	0.025	3.14	<i>3.10 × 10<sup>-10</sup></i>
Damaging	98	0.19	33.0	0.063	2.97	<i>4.83 × 10<sup>-20</sup></i>
<b>High-pLI OMIM genes (n = 643)</b>						
Total	104	0.198	36.3	0.069	2.86	<i>4.73 × 10<sup>-20</sup></i>
Synonymous	5	0.010	10.3	0.020	0.49	0.98
T-mis	23	0.044	15.2	0.029	1.51	0.03
D-mis	41	0.078	7.7	0.015	5.32	<i>3.81 × 10<sup>-17</sup></i>
LoF	35	0.067	3.2	0.006	10.9	<i>2.33 × 10<sup>-24</sup></i>
Damaging	76	0.144	10.9	0.021	7.0	<i>8.98 × 10<sup>-38</sup></i>
<b>Chromatin modifiers and remodelers (n = 614)</b>						
Total	48	0.091	26.1	0.050	1.84	<i>7.95 × 10<sup>-5</sup></i>
Synonymous	6	0.011	7.1	0.014	0.84	0.72
T-mis	12	0.023	12.6	0.024	0.95	0.61
D-mis	13	0.025	3.9	0.0073	3.37	<i>1.93 × 10<sup>-4</sup></i>
LoF	17	0.032	2.5	0.0048	6.76	<i>1.69 × 10<sup>-9</sup></i>
Damaging	30	0.057	6.4	0.012	4.71	<i>1.07 × 10<sup>-11</sup></i>

n, number of *de novo* variants in 526 subjects; Rate, number of *de novo* variants per subject; damaging or tolerated missense variants as called by MetaSVM (D-mis, T-mis respectively); loss of function (LoF) denotes premature termination, frameshift, splice-site variant, start-loss, or stop-loss variants. "Damaging" includes LoF, D-mis, and inframe insertion or deletion variants. P values represent the upper tail of the Poisson probability density function; p values in italics represent significant values.

each with between 2 and 4 damaging DNVs. These included *CDK13* (MIM: 603309), *NFIX* (MIM: 164005), *ADNP* (MIM: 611386), *KMT5B* (MIM: 610881), *SON* (MIM: 182465), *ARID1B* (MIM: 614556), *CASK* (MIM: 300172), *MED13L* (MIM: 608771), *PSMD12* (MIM: 604450), and *POLR2A* (MIM: 180660) (Table 4; Figure S2). Three additional genes demonstrated significant enrichment for D-mis (*CHD3*; MIM: 602120) or LoF (*CHD7* [MIM: 608892], *SETBP1* [MIM: 611060]) DNVs; these remained significant after correction for multiple testing, each at a false discovery rate (FDR) < 0.05 (Table 4; Figure S2, subjects and methods).

A description of phenotypes associated with variation in each gene can be found in Table S5.

#### Chromatin remodelers

Among this set, 17 probands had DNVs in one of six genes implicated in chromatin remodeling, including *ARID1B*, *ADNP*, *KMT5B*, *CHD7*, *CHD3*, and *SETBP1*. Three LoF DNVs were identified in AT-rich interactive domain-containing protein 1B (*ARID1B*: c.3834delC [p.Asn1278Lysfs\*8] [GenBank: NM\_020732]; c.4063C>T [p.Gln1355\*]; and c.4105C>T [p.Gln1369\*]), a component of the SWI/SNF chromatin remodeling complex that causes an NDD resembling Coffin-Siris syndrome (Table S5).<sup>26</sup> Three damaging

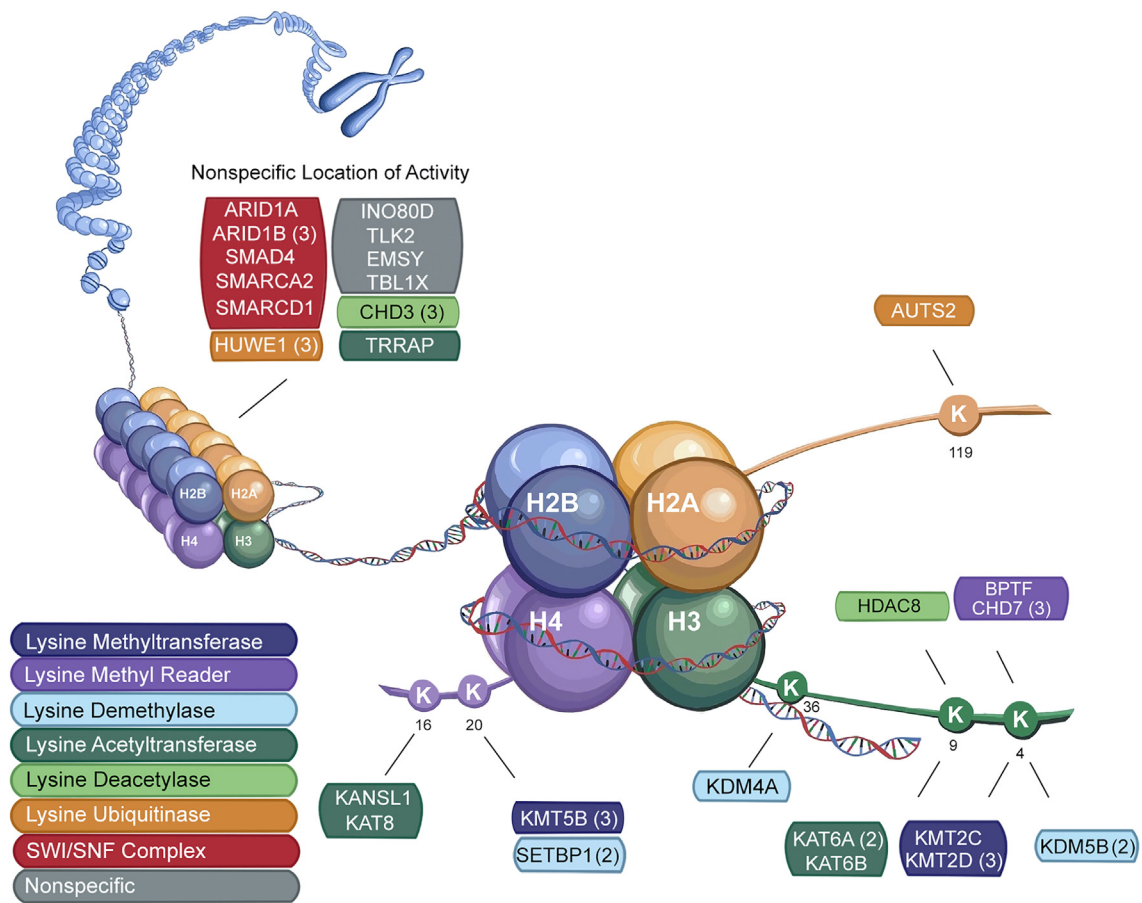
**Table 2. Thirty damaging *de novo* variants in chromatin modifiers and remodelers among 526 syndromic CS probands**

Gene	Impact	GnomAD frequency	pLI	OMIM
<i>ARID1A</i>	p.Gly276*	0	1	Y
<i>ARID1B</i>	p.Asn1278Lysfs*8	0	1	Y
<i>ARID1B</i>	p.Gln1355*	0	1	Y
<i>ARID1B</i>	p.Gln1369*	0	1	Y
<i>AUTS2</i>	p.Arg495*	0	1	Y
<i>BPTF</i>	p.Met1?	0	1	Y
<i>CHD3</i>	p.Asn1159Lys	0	1	Y
<i>CHD3</i>	p.Arg1169Trp	0	1	Y
<i>CHD3</i>	p.Arg1172Trp	0	1	Y
<i>CHD7</i>	c.2836-2A>T	0	1	Y
<i>CHD7</i>	c.5050+1G>A	0	1	Y
<i>CHD7</i>	p.Arg1820*	0	1	Y
<i>EMSY</i>	p.Gln477*	0	1	N
<i>HDAC8</i>	p.Thr311Met	$5.5 \times 10^{-6}$	1	Y
<i>KANSL1</i>	p.Leu270Valfs*11	$2.5 \times 10^{-5}$	1	Y
<i>KAT6B</i>	p.Glu1371*	0	1	Y
<i>KAT8</i>	p.Arg99Gln	0	0.22	Y
<i>KDM4A</i>	p.Ala286Thr	0	1	N
<i>KDMSB</i>	p.Pro67Leu	0	0	N
<i>KDMSB</i>	p.Lys517Arg	0	0	N
<i>KMT2C</i>	p.Tyr1594*	0	1	Y
<i>KMT5B</i>	p. Ala97Glyfs*3	0	1	Y
<i>KMT5B</i>	p.Gly156Asp	0	1	Y
<i>KMT5B</i>	p.Glu661Glyfs*26	0	1	Y
<i>SMAD4</i>	p.Arg496Cys	$8.0 \times 10^{-6}$	1	Y
<i>SMARCA2</i>	p.Glu852Lys	0	1	Y
<i>SMARCD1</i>	p.Gln508*	0	1	Y
<i>TBLIX</i>	p.Phe550Leu	0	0.98	Y
<i>TLK2</i>	p.Arg303*	0	1	Y
<i>TRRAP</i>	p.Glu104Lys	0	1	Y

*De novo* LoF and D-mis variants in genes associated with the gene ontology term for chromatin organization (GO: 0006325) in 526 trios with sporadic syndromic CS. The impact of each variant is provided at the protein level for missense, frameshift, and nonsense variants and at the DNA level for splice variants. Complete annotations for each variant are found in Table S6. pLI = probability of LoF intolerance. Allele frequencies and pLI scores were obtained from GnomAD v.2.1.1. For each of the three variants found in GnomAD, the allele was designated as “pathogenic” in ClinVar in at least one child with NDD. The OMIM (Online Mendelian Inheritance in Man) column indicates whether the gene has a known dominant Mendelian phenotype association (Y/N).

DNVs were identified in activity-dependent neuroprotector homeobox protein (*ADNP*: c.2189G>C [p.Arg730Pro] [GenBank: NM\_001282531]; c.1084C>T [p.Gln362\*]; and c.2213C>A [p.Ser738\*]), a transcription factor involved in the SWI/SNF chromatin remodeling complex known to cause syndromic autism spectrum disorder (ASD).<sup>27</sup> These individuals presented with either sagittal or coronal CS (Table S5). *KMT5B*, encoding histone-lysine N-methyltransferase 5B, which methylates H4K20, harbored three damaging DNVs (c.290delC [p.Ala97Glyfs\*3] [Gen-

Bank: NM\_001369432]; c.1982delA [p.Glu661Glyfs\*26]; and c.467G>A [p.Glu156Asp]) in individuals with metopic CS (Table S5). Variants in *KMT5B* cause both syndromic and non-syndromic NDDs.<sup>28</sup> Three LoF DNVs were identified in chromodomain-helicase-DNA-binding protein 7 (*CHD7*; [GenBank: NM\_017780] c.5050+1G>A; c.2836-2A>T; c.5458C>T [p.Arg1820\*]), a chromatin-associated transcription factor implicated in multiple developmental anomalies (CHARGE syndrome).<sup>29</sup> These individuals presented with metopic or lambdoid CS (Table S5). Three



**Figure 1. Variants in chromatin modifiers and remodelers in probands with CS**

Damaging *de novo* variants identified in probands with syndromic CS are noted; each color represents the corresponding modification or function noted in the legend. Numbers in black indicate the lysine (K) residue on each histone tail. Numbers in parentheses indicate the number of damaging DNVs identified in our cohort.

individuals have previously been reported as having CS in association with CHARGE syndrome.<sup>30</sup> Three damaging missense DNVs were identified in the helicase domain of *CHD3*, which encodes a component of the nucleosome remodeling and deacetylase complex (Figure S3A).<sup>31</sup> Two of these variants (c.3477C>G [p.Asn1159Lys] [GenBank: NM\_005852] and c.3505C>T [p.Arg1169Trp]) were previously identified in probands with syndromic NDD, and the third variant (c.3514C>T [p.Arg1172Trp] [GenBank: NM\_005852]) occurs at a codon known to be mutated in *CHD3*-related NDD.<sup>31</sup> Those harboring these variants had metopic +/- sagittal CS (Table S5). The observed distribution of variants suggests that CS might be specific to variants within the helicase domain of *CHD3*. Two LoF DNVs were identified in *SETBP1* (c.1264C>T [p.Gln422\*] [NM\_015559] and c.1633G>T [p.Glu545\*]), a gene whose protein product binds SET domains and functions as part of a complex that demethylates H4K20,<sup>32</sup> in individuals with coronal CS (Table S5). These findings suggest that CS is a frequent feature of chromatinopathies.

#### Transcriptional regulators

Ten probands had DNVs in three genes, *CDK13*, *MED13L*, and *POLR2A*, implicated in RNA-polymerase-II-mediated

transcription. Four damaging missense variants were identified within the kinase domain of *CDK13* (c.2149G>A [p.Gly717Arg] [GenBank: NM\_031267]; c.2509G>A [p.Asp837Asn]; c.2525A>G [p.Asn842Ser]; and c.2626A>G [p.Thr876Ala]), a cyclin-dependent kinase that phosphorylates RNA polymerase II to enable initiation of transcription (Figure S3B). Each of these four individuals had metopic craniosynostosis (Table S5). Variants in *CDK13* have previously been demonstrated to cause both syndromic congenital heart disease (CHD) and an NDD.<sup>16,33</sup> Three damaging missense DNVs were identified in mediator of RNA polymerase II transcription subunit 13-like (*MED13L*; c.2605C>T [p.Pro869Ser] [GenBank: NM\_015335]; c.3428C>T [p.Ala1143Val]; and c.5023C>G [p.Pro1675Ala]; Figure S3C), a gene whose protein product serves as a scaffold for assembly of the RNA polymerase II preinitiation complex. These individuals presented with sagittal or metopic CS (Table S5). DNVs in *MED13L* cause syndromic ID.<sup>34</sup> Three damaging DNVs were identified in *POLR2A* (c.163dupG [p.Leu57Alafs\*10] [GenBank: NM\_000937]; c.4252G>A [p.Gly1418Arg]; c.1671+1G>A), which encodes the largest component of RNA polymerase II.<sup>35</sup> Each had metopic craniosynostosis (Table S5). These



**Table 3. Damaging *de novo* variants in high-pLI genes in syndromic CS probands**

Gene	Impact	GnomAD frequency	pLI
<i>DLGAP1</i>	p.Arg648*	0	1
<i>DST</i>	p.Glu1585*	0	1
<i>EMSY</i>	p.Gln477*	0	1
<i>GFPT1</i>	p.Tyr158Ser	0	0.90
<i>GNAI1</i>	p.Cys224Tyr	0	0.91
<i>IGF2BP1</i>	p.Arg93*	0	1
<i>IKZF2</i>	p.Cys147Arg	0	0.99
<i>INO80D</i>	p.Arg557*	0	1
<i>KDM4A</i>	p.Ala286Thr	0	1
<i>LONP1</i>	p.Arg474Lys	0	1
<i>LRP1</i>	p.Asp1615Glu	0	1
<i>LRP2</i>	p.Arg3086Cys	$2.0 \times 10^{-5}$	1
<i>MGAT5</i>	p.Pro416Ser	$8.0 \times 10^{-6}$	1
<i>PHF23</i>	p.Arg183*	0	1
<i>PSMC5</i>	p.Arg325Trp	0	0.93
<i>RARA</i>	p.Gly289Arg x2	0	0.96
<i>SLIT1</i>	p.Leu52Pro	0	1
<i>TENM3</i>	p.Arg2111Trp	$6.8 \times 10^{-5}$	1
<i>TNIK</i>	p.Gly336Arg	0	1
<i>TTC28</i>	p.Phe1699Ser	0	1
<i>VEZF1</i>	p.Arg116*	0	0.99

Damaging *de novo* variants identified in high-pLI genes with no dominant OMIM association in probands with sporadic syndromic CS. The impact of each variant is described at the protein level; full annotation of each variant is provided in Table S6. pLI = probability of LoF intolerance. Allele frequencies and pLI scores were obtained from GnomAD v.2.1.1.

variants identify transcriptional dysregulation as a cause of syndromic CS.

#### Other genes with significant enrichment in DNVs

Eleven probands had DNVs in four genes involved in RNA splicing, transcriptional regulation, plasma-membrane maintenance, and proteasome function. Three LoF variants were identified in *SON* (c.3653delG [p.Ser1218-Metfs\*23] [GenBank: NM\_138927]; c.5751\_5754del [p.Val1918Glu fs\*87]; and c.5812C>T [p.Arg1938\*]), an RNA-binding protein that acts as an mRNA splicing cofactor. These individuals presented with sagittal (n = 2) or coronal (n = 1) CS (Table S5). Variants in this gene have been shown to cause syndromic intellectual disability (ID) via disruption of RNA splicing.<sup>36</sup> Three damaging variants were identified in nuclear factor IX (*NFIX*), a ubiquitously expressed transcription factor regulating neurogenesis and neural stem cell proliferation. These included two variants within the DNA-binding domain (c.355T>C [p.Cys119Arg] [GenBank: NM\_002501] and c.361C>T [p.Arg121Cys]) and a nonsense variant (c.772C>T [p.Gln258\*]). These individuals had lambdoid, metopic, or lambdoid and bicoronal CS (Table S5). LoF variants and missense variants in the DNA-binding domain of

*NFIX* are known to cause Sotos and Malan syndromes, which are characterized by overgrowth and neurodevelopmental anomalies.<sup>37</sup> We identified three damaging DNVs (c.1874\_1876delinsTTG [p.Pro625Leu] [GenBank: NM\_003688]; c.2259delT [p.Leu754Serfs\*38]; and c.2695A>T [p.Arg899\*]) in *CASK*, which encodes a scaffolding protein that anchors synaptic transmembrane proteins. Variants in *CASK* are known to cause a syndromic NDD with microcephaly.<sup>13</sup> These individuals presented with sagittal CS and microcephaly. It is possible that CS was secondary to microcephaly in these individuals; however, the complex craniofacial phenotype with concurrent limb anomalies was suggestive of a craniofacial syndrome (Table S5). Two damaging DNVs were identified in *PSMD12* (c.253C>T [p.Arg85\*] [GenBank: NM\_001316341] and c.1105T>C [p.Trp369Arg]), a subunit of the 26S proteasome.<sup>38</sup> These individuals had sagittal and metopic or lambdoid CS (Table S5).

Although these genes have previously been implicated in autism or other NDDs, CS is not an established feature of any of these syndromes, with the exception of a suggested relationship between CS and CHARGE syndrome (*CHD7*). In the most comprehensive expert-curated gene

**Table 4. 15 high-pLI genes with multiple damaging *de novo* variants in syndromic CS probands**

Gene	pLI	D-mis observed	D-mis p	LoF observed	LoF p	Damaging observed	Damaging p	Damaging FDR q
<i>CDK13</i>	0.91	4	<b><math>5.07 \times 10^{-12}</math></b>	0	1	4	<b><math>6.17 \times 10^{-10}</math></b>	<b><math>1.89 \times 10^{-6}</math></b>
<i>NFIX</i>	1	2	<b><math>7.44 \times 10^{-6}</math></b>	1	$1.66 \times 10^{-3}$	3	<b><math>2.81 \times 10^{-8}</math></b>	<b><math>4.30 \times 10^{-5}</math></b>
<i>ADNP</i>	1	1	$2.56 \times 10^{-3}$	2	<b><math>1.11 \times 10^{-5}</math></b>	3	<b><math>6.44 \times 10^{-8}</math></b>	<b><math>6.58 \times 10^{-5}</math></b>
<i>KMT5B</i>	1	1	$3.47 \times 10^{-3}$	2	<b><math>9.88 \times 10^{-6}</math></b>	3	<b><math>8.24 \times 10^{-8}</math></b>	<b><math>6.31 \times 10^{-5}</math></b>
<i>ARID1B</i>	1	0	1	3	<b><math>1.64 \times 10^{-7}</math></b>	3	<b><math>1.87 \times 10^{-7}</math></b>	<b><math>1.15 \times 10^{-4}</math></b>
<i>CASK</i>	1	1	$5.40 \times 10^{-3}$	2	<b><math>1.56 \times 10^{-5}</math></b>	3	<b><math>2.21 \times 10^{-7}</math></b>	<b><math>1.13 \times 10^{-4}</math></b>
<i>SON</i>	1	0	1	3	<b><math>1.61 \times 10^{-7}</math></b>	3	<b><math>2.95 \times 10^{-7}</math></b>	<b><math>1.29 \times 10^{-4}</math></b>
<i>MED13L</i>	1	3	<b><math>1.80 \times 10^{-6}</math></b>	0	1	3	<b><math>5.14 \times 10^{-6}</math></b>	<b><math>1.97 \times 10^{-3}</math></b>
<i>PSMD12</i>	1	1	$3.79 \times 10^{-4}$	1	$3.16 \times 10^{-3}$	2	<b><math>6.26 \times 10^{-6}</math></b>	<b><math>2.13 \times 10^{-3}</math></b>
<i>POLR2A</i>	1	1	0.035	2	<b><math>3.48 \times 10^{-5}</math></b>	3	<b><math>1.39 \times 10^{-5}</math></b>	<b><math>4.26 \times 10^{-3}</math></b>
<i>CHD7</i>	1	0	1	3	<b><math>6.25 \times 10^{-7}</math></b>	3	$2.54 \times 10^{-5}$	<b><math>7.07 \times 10^{-3}</math></b>
<i>CHD3</i>	1	3	<b><math>1.56 \times 10^{-5}</math></b>	0	1	3	$3.38 \times 10^{-5}$	<b><math>8.63 \times 10^{-3}</math></b>
<i>PTPN11</i>	1	2	$8.01 \times 10^{-5}$	0	1	2	$1.31 \times 10^{-4}$	<b>0.03</b>
<i>SETBP1</i>	1	0	1	2	<b><math>1.63 \times 10^{-5}</math></b>	2	$2.00 \times 10^{-4}$	<b>0.04</b>
<i>RARA</i>	0.96	2	$1.57 \times 10^{-4}$	0	1	2	$2.19 \times 10^{-4}$	<b>0.04</b>

15 high-pLI genes harbored multiple damaging *de novo* variants. p and q values surpassing genome-wide significance after correction for multiple tests appear in bold. No low-pLI genes (pLI < 0.9) had FDR < 0.05. pLI, probability of loss of function intolerance; D-mis, damaging missense as called by MetaSVM; LoF, loss of function; FDR, false discovery rate.

panel for CS,<sup>39</sup> eleven of these genes do not appear on the panel, one (*CHD7*) carries an “amber” rating indicating inadequate evidence for genome interpretation, and one (*NFIX*) carries a “red” rating for insufficient evidence for any involvement in CS.

When we consider individual genes in which we identify a genome-wide-significant enrichment in DNVs, 29 of 40 DNVs identified were reported by the clinical laboratory as PATH (pathogenic) or LPATH (likely pathogenic), whereas 11 of 40 DNVs were not reported as either PATH or LPATH (Table S6). These findings highlight the importance of large-scale gene-discovery efforts to provide evidence for novel gene associations and phenotypic expansions, both of which improve the diagnostic yield of clinical testing.

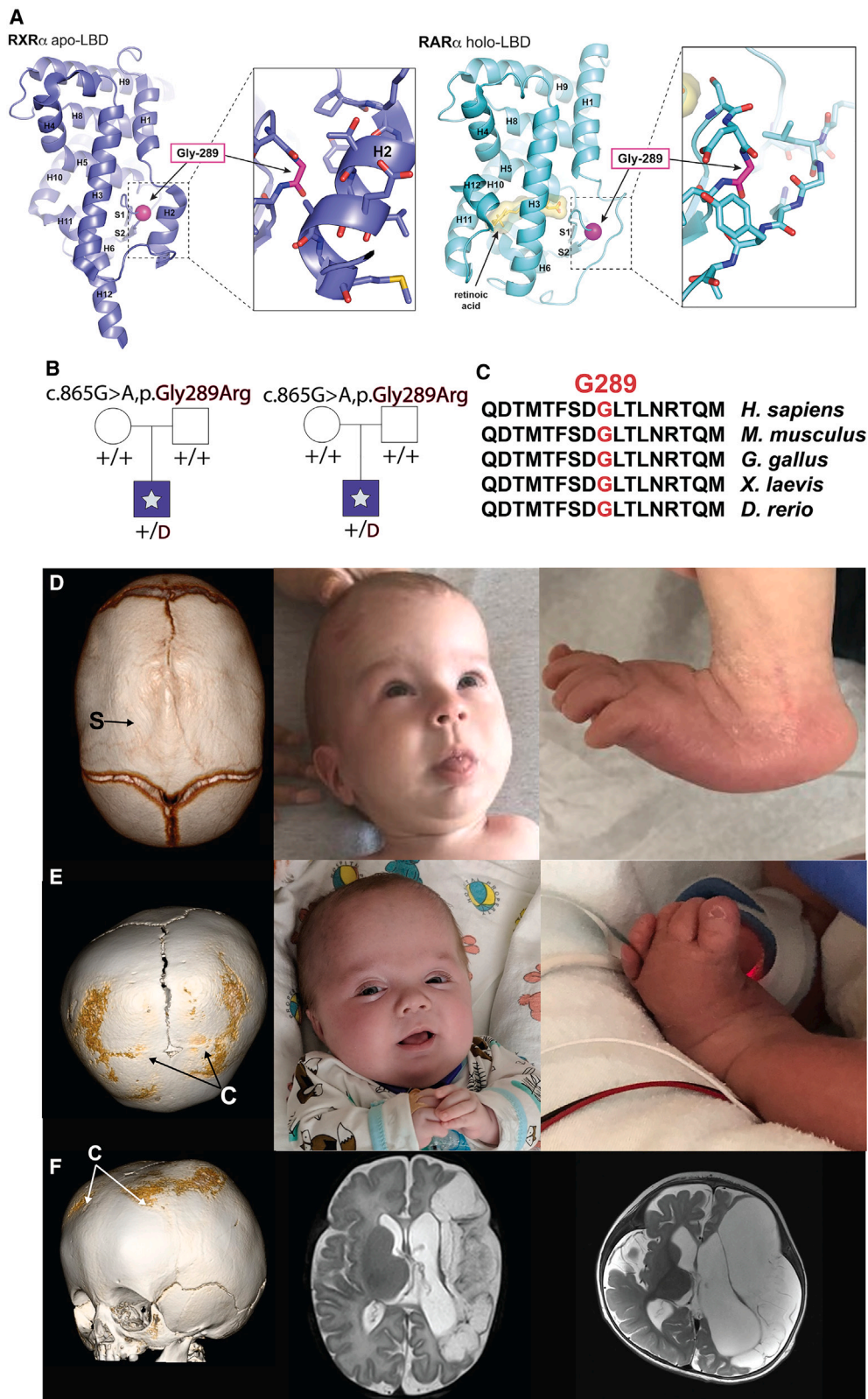
We performed phenomic analyses to attempt to identify whether individuals with DNVs in the same gene presented with a consistent constellation of phenotypic features, as described in Note S2. We were not able to make definitive conclusions regarding specific genotype-phenotype relationships by analyzing the data from the current cohort; more individuals with variants in each gene will be necessary for full characterization of the phenotypic impact of variants in each.

#### Identification of additional high-confidence CS risk genes with multiple damaging DNVs

Seventeen genes had at least two damaging DNVs, an event highly unlikely to occur by chance ( $p < 10^{-8}$ , 10M permutations; Figure S2, see subjects and methods). On

the basis of the expected number of genes mutated at least twice ( $n = 2.2$ ), we expect that at least 15 of these genes are likely to contribute to CS risk. In addition to the 13 genes described above, two individuals had damaging DNVs in *PTPN11* (MIM: 176876; c.172A>G [p.Asn58Asp] [GenBank: NM\_080601] and c.218C>T [p.Thr73Ile]). Both are designated as pathogenic variants in ClinVar, consistent with a diagnosis of Noonan syndrome, which occasionally involves CS<sup>40</sup> (Figure S2). Among genes without known dominant OMIM associations, we identified two damaging DNVs in *KDM5B* (MIM: 605393; c.200C>T [p.Pro67Leu] [GenBank: NM\_006618] and c.1550A>G [p.Lys517Arg]), which encodes an H3K4 demethylase, two damaging DNVs in the lysine methyltransferase *EEF1AKMT4* (c.222C>A [p.Tyr74\*] [GenBank: NM\_032331] and c.380T>G [p.Val127Gly]), and a recurrent damaging DNV in the retinoic acid receptor alpha (*RARA* [MIM: 180240]; c.865G>A [p.Gly289Arg in two individuals] [GenBank: NM\_001145301]).

Although not classified as “damaging” by MetaSVM, additional LoF-intolerant genes with no known OMIM association had multiple protein-altering variants; such genes included *CAMSAP1* (MIM: 613774); pLI = 1; c.1188C>A [p.His396Gln] [GenBank: NM\_015447] and c.2350C>A [p.Asp784Asn]) and *RDX* (MIM: 179410; pLI = 1; c.1411C>G [p.Pro471Ala] [GenBank: NM\_002906] and c.683-2A>G [GenBank: NM\_001260494]), which play roles in microtubule formation and actin filament membrane anchoring, respectively, and *PTBP1* (MIM: 600693; pLI = 1; c.142A>G [p.Lys48Glu] [GenBank: NM\_031991] and



**Figure 2. A recurrent *de novo* p.Gly289Arg variant in *RARA* defines a craniosynostosis syndrome**

(A) Left (purple): Ribbon diagram of the ligand binding domain of human retinoic acid receptor alpha. The diagram is modeled on the structure of the highly homologous retinoid receptor RXR $\alpha$  (PDB: 6HN6<sup>43</sup>). Right (cyan): Ribbon diagram of the ligand binding domain of human retinoic acid receptor alpha (RAR $\alpha$ ) bound to all-trans retinoic acid (PDB: 3A9E<sup>44</sup>). The retinoic acid receptor alpha LBD, which is responsible for ligand binding, dimerization, and recruitment of coregulatory factors, adopts the typical helical sandwich fold that is characteristic of the large superfamily of nuclear receptors.<sup>45</sup> It comprises twelve alpha-helices (H1–H12) and a beta turn (containing strands S1 and S2); retinoic acid ligands bind in a hydrophobic pocket involving helices H3, H5, and H11 and the beta turn. In response

(legend continued on next page)

c.1511C>T [p.Ala504Val]), which encodes a pre-mRNA splicing factor. The potential role of these genes in syndromic CS will require further study.

### A recurrent *de novo* variant in *RARA*

Among probands without variants in known CS-associated genes, we sought to identify recurrent DNVs at the same codon. We identified a single recurrent *de novo* missense variant (c.865G>A [p.Gly289Arg]) in the retinoic acid receptor alpha (*RARA*) in two probands with similar phenotypes. Observing a recurrent DNV at any position in a cohort of this size is highly unlikely to occur by chance ( $p = 6.5 \times 10^{-3}$ , see [subjects and methods](#)). Being severely affected with either sagittal or bicoronal CS, each of them also demonstrated limb anomalies including rocker-bottom feet, bowing of the legs, and short upper and lower limbs. These children had additional craniofacial manifestations of disease, including microtia, conductive hearing loss, ankyloglossia, esotropia, hypoplastic nasal bones, and oligodontia. Additional anomalies included renal dysplasia with cysts, tracheomalacia, pulmonary arterial hypertension, developmental delays, hypotonia, cryptorchidism, seizures, adrenal insufficiency, and other phenotypes detailed in [Table S5](#) and [Note S3](#).

Both children were heterozygous for the c.865G>A (p.Gly289Arg) variant and displayed no evidence of mosaicism. Retinoic acid receptors (RARs) and retinoid receptors (RXRs) are related members of the nuclear receptor superfamily and contain a DNA binding domain and a ligand binding domain (LBD) ([Figure 2](#)). Gly289 is located within the beta turn of the LBD, and in both the apo (free) and holo (ligand-bound) conformations it is inaccessible to the surface of the domain ([Figure 2A](#)). Thus, mutation to arginine, with its bulky and positively charged sidechain, is expected to destabilize this region of the LBD to alter either ligand binding or overall solubility and thus impact transcriptional activation. Somatic missense variants at codon Gly289 confer treatment resistance in acute promyelocytic leukemia via gain-of-function, supporting the functional significance of this residue.<sup>41</sup> *In utero* exposure to retinoid is an established cause of CS via neural-crest defects,<sup>42</sup> further supporting a gain-of-function mechanism for these variants.

### DNVs in individuals with syndromic CS in the DECIPHER cohort

To further support our findings, we queried the DECIPHER database for individuals who were associated with the phenotype term “craniosynostosis” and who had exonic DNVs outside of established CS loci, and we identified 43 individuals with syndromic CS.<sup>46</sup> Within these 43 individuals, DNVs of interest were identified in *ARID1A* (MIM: 603024), *AHDC1* (MIM: 615790), *FBXO11* (MIM: 607871), *SMARCA2* (MIM: 600014), *HNRNPK* (MIM: 600712), *KAT6B* ( $n = 2$ ), *MED13L*, *ANKRD11* ( $n = 2$ ; [MIM: 611192]), *CHD3*, *NF1* (MIM: 613113), *KANSL1* (MIM: 612452), *TLK2* (MIM: 608439), *SETBP1*, *NFIX*, and *SON* ( $n = 3$ ), which are each mutated in our cohort. Thus, 44% of individuals with unexplained syndromic CS with exonic DNVs in DECIPHER harbored protein-damaging variants in chromatin or OMIM genes found to be mutated in our cohort, providing compelling evidence that these and other genes identified are *bona fide* CS loci.

### Transcriptomic analysis of syndromic CS-associated genes

Having identified striking overlap between genes mutated in autism, NDD, and syndromic CS, we sought to determine whether genes mutated in our cohort clustered in neurodevelopmental processes that might concomitantly impact brain and suture morphogenesis. We tested whether genes mutated in syndromic CS probands converged in gene co-expression networks comprising RNA extracted from 16 regions in each of 27 human brains sampled at different timepoints from fetal development into early childhood (see [subjects and methods](#); [Table S1](#)). We constructed 88 co-expression networks by using weighted gene co-expression network analysis (WGCNA) and conducted module enrichment analysis by using CS risk genes ([subjects and methods](#)). We identified significant enrichment of novel CS risk genes in four modules, namely the “brown,” “greenyellow,” “royalblue,” and “salmon” modules ([Figure 3A](#)). The gene lists included in each module were assessed for enrichment in biological pathways. The top enriched Gene Ontology process terms showed enrichment for mutated genes in the “brown” module in transcriptional regulation, the “greenyellow” module in RNA processing and splicing, the “royalblue” module in histone modification and chromatin remodeling, and the “salmon” module

---

to ligand binding, the LBD undergoes a conformational change in helices H11, H12, H3, and H2. This conformational change leads to transcriptional activation. In both diagrams, alpha helices (H) and beta strands (S) are labeled according to the standard nomenclature for nuclear hormone receptor superfamily.<sup>45</sup> Gly289, shown in magenta, is located within the beta turn loop connecting strands S1 and S2. Insets show the details of the sidechains of neighboring residues and demonstrate the surface inaccessibility of Gly289.

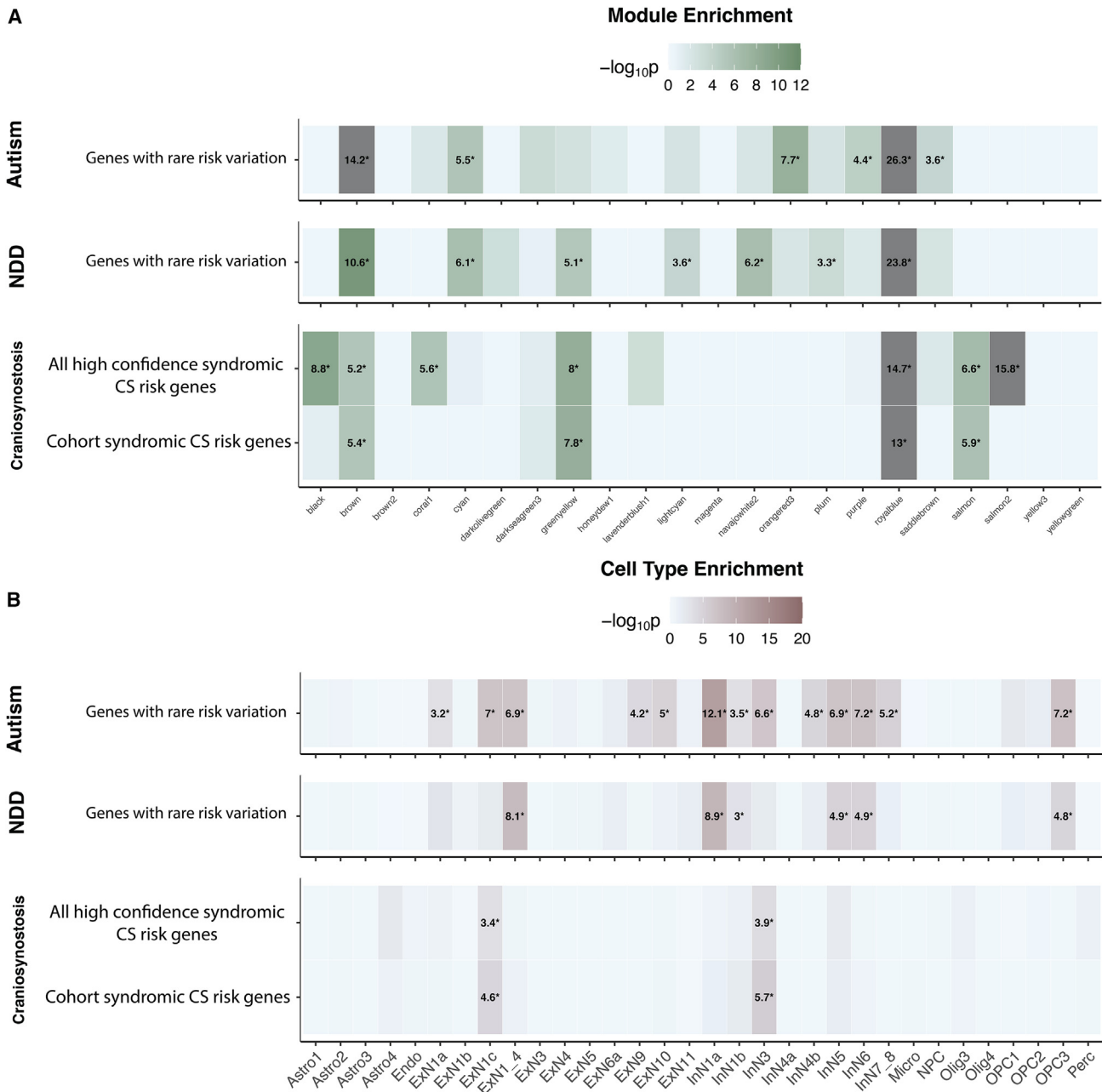
(B) Pedigrees of two families in which a recurrent *de novo* variant in *RARA* were detected. + represents a wild-type allele, and D represents the variant described above each pedigree.

(C) Protein sequence alignment of *RARA*, demonstrating evolutionary conservation of Gly289 and its flanking sequence across several recent orthologs.

(D) 3D CT scan demonstrating sagittal CS (S) in a proband, as well as associated head shape and rocker-bottom feet.

(E) CT scan demonstrating bilateral coronal CS (C), associated brachycephaly, and similar foot deformity present bilaterally.

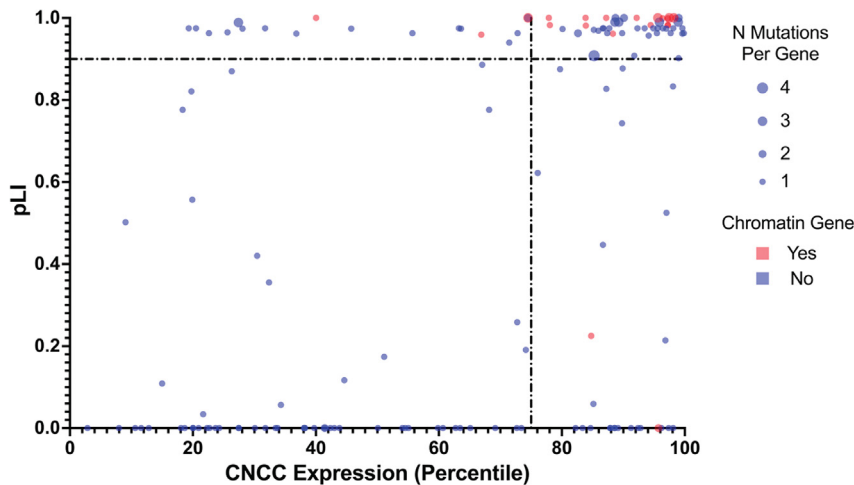
(F) Frontal view of 3D CT reconstruction demonstrating bilateral coronal CS, and MRI images demonstrating sequelae of perinatal middle cerebral artery stroke.



**Figure 3. CS risk genes are enriched in co-expression networks pertinent to autism and other neurodevelopmental disorders**  
 (A) Enrichment analysis across weighted gene co-expression network analysis modules of midgestational human brain for genes with rare risk variants in syndromic CS (craniosynostosis), autism, and neurodevelopmental disorders (NDDs). Tiles labeled with  $-\log_{10}(p\text{-value})$  and an asterisk represent those in which significant enrichment was identified at the Bonferroni-corrected cutoff ( $\alpha = 0.05/17 = 2.94 \times 10^{-3}$ ).  
 (B) Enrichment analysis across each cell cluster identified in the midgestational human brain for genes with rare risk variants in autism, NDD, and CS. Tiles are again labeled with  $-\log_{10}(p\text{-value})$  and an asterisk to represent significant enrichment at the Bonferroni multiple-testing cutoff ( $\alpha = 0.05/32 = 1.56 \times 10^{-3}$ ).

in regulation of osteoblast proliferation (Figure S4). Interestingly, the “brown” and “royalblue” modules are also significantly enriched in autism and NDD risk genes, and the “greenyellow” module is enriched in NDD risk genes (Figure 3A). Predictably, the “salmon” module, which contains genes that regulate osteoblast proliferation, was uniquely enriched in CS risk genes. Addition of all other known syndromic CS-associated genes to the input list

maintains or strengthens each of these associations (Figure 3A). The results implicate various biological processes with established roles in neurodevelopment in syndromic CS; such processes include chromatin modification, transcriptional regulation, and RNA splicing. The implicated modules demonstrate significantly enriched expression during fetal development (post-conception weeks 9–17; Figure S5), which includes the period spanning



**Figure 4. Syndromic CS risk genes are highly expressed in cranial neural crest cells**  
 Genes in which damaging *de novo* variants were identified are plotted on the basis of each gene's pLI and expression percentile in human cranial neural crest cells. The size of each circle represents the number of damaging DNVs identified in that gene; the genes involved in chromatin remodeling and modification as described in Table 2 are shown in red. This plot makes use of the “jitter” function of ggplot2, which adds a negligible amount of random variation to the location of each point to reduce overplotting caused by overlapping values. The results demonstrate considerable enrichment of damaging *de novo* variants in intolerant genes highly expressed in CNCCs in probands with syndromic CS ( $p = 6.51 \times 10^{-22}$ ; Poisson distribution).

from initiation of skull ossification (~7 weeks) through suture formation (~16 weeks).<sup>47</sup>

Finally, we assessed for potential enrichment of CS-associated genes in specific neural cell types by using a single-cell RNA transcriptomic atlas of midgestational brain development<sup>19</sup> (Figure S6). Genes associated with syndromic CS were enriched in an upper-layer excitatory neuron subcluster (ExN1c;  $p = 3.2 \times 10^{-4}$ ) and an inhibitory neuron subcluster (InN3;  $p = 1.26 \times 10^{-6}$ ) (Figure 3B). Genes mutated in ASD demonstrate significant enrichment in each of these two cell types, suggesting that the neurodevelopmental phenotypes observed in many children with syndromic CS could be due, at least in part, to intrinsic brain dysfunction.<sup>48,49</sup> Determining the potential role of these excitatory and inhibitory neuron populations in the aberrant neurodevelopmental seen in syndromic CS will require further study.

#### Enrichment of damaging variants in genes highly expressed in cranial neural crest cells

Craniofacial anomalies frequently result from aberrations in the proliferation, differentiation, or migration of neural crest cells destined for the skull. Perturbations in CNCCs can disrupt progenitor cell populations that contribute to cranial suture development and maintenance of suture patency.<sup>3,50</sup> We sought to determine whether variants identified in syndromic CS probands were enriched in genes highly expressed in human CNCCs. Using bulk RNA-seq data generated from six human CNCC samples<sup>51</sup> (see subjects and methods), we ranked genes on the basis of their CNCC expression and assessed them for enrichment in DNVs in intolerant genes ( $pLI > 0.9$ ) in the highest quartile of CNCC expression (Table S2). 73 of 182 damaging DNVs were in intolerant genes with high CNCC expression (4-fold enrichment;  $p = 6.51 \times 10^{-22}$ , Poisson distribution; Figure 4; Table S11). These results remain significant after exclusion of probands with variants in the above-described 13 genes that were significant at the individual gene level ( $p = 5.0 \times 10^{-7}$ , Poisson distribution).

The results suggest that variants in many of the genes identified disrupt CNCCs, potentially contributing to CS in these individuals.

#### Discussion

The results describe a sizable syndromic CS cohort, in which we identified many CS loci. Biological processes, including chromatin modification and remodeling, transcriptional regulation, RNA splicing, and regulation of osteoblast differentiation appear crucial to maintaining cranial suture patency. We also establish CS as a frequent feature of chromatinopathies. In our previous studies of non-syndromic CS, DNVs contributing to genetic risk were identified in several regulators of the Wnt, BMP, and FGF-Ras/ERK signaling pathways, and most identified risk genes had very low pLI scores.<sup>10</sup> Individuals included in this previous study had isolated fusion of a cranial suture, without other systemic manifestations of disease. The study of syndromic CS with a wide variety of extracranial anomalies resulted in the identification of >30 CS loci, nearly all of which are highly intolerant to LoF variation. These findings highlight the utility of pLI in the study of syndromic conditions likely to be caused by pleiotropic genes, particularly those that present with neurodevelopmental phenotypes or multisystem involvement. After exclusion of probands with DNVs in known genes, ~6% of probands had plausible pathogenic DNVs in chromatin genes, another ~5.3% had genome-wide-significant DNVs in other genes, ~2.7% had plausible pathogenic DNVs in other intolerant OMIM genes (from the excess of damaging DNVs in comparison to expectation), and another ~0.5% harbored a recurrent gain-of-function DNV in *RARA* (Figures 1, 2, and 3; Tables 1 and 2). Dozens of additional candidate genes that merit further study were also identified. In sum, these DNVs in CS loci explain disease in ~15% of previously undiagnosed individuals. Importantly, none of these probands harbor additional DNVs reported as PATH or LPATH, suggesting that the

possibility of composite phenotypes resulting from multi-locus genomic variation is unlikely in these individuals.<sup>52</sup>

Variants in chromatin modifiers and remodelers appear to contribute to disease risk in a substantial fraction of individuals with syndromic CS; such variants include those in many genes with notable roles in conferring risk of congenital heart disease,<sup>53,54</sup> autism,<sup>13</sup> and other NDDs.<sup>16</sup> The pleiotropic effects of variants in chromatin modifiers are increasingly evident; however, the phenotypic heterogeneity observed in individuals with similar spectra of variants remains enigmatic.<sup>12</sup> Whether this phenotypic variability is due to genetic modifiers, is stochastic in nature, or is the result of widespread modifications to the epigenomic landscape remains to be studied. Variants in chromatin genes have been shown to alter the delicate balance between proliferation and differentiation of neural progenitor cells, resulting in structural and neurodevelopmental anomalies.<sup>55,56</sup> Perturbations affecting the balance between the proliferation and differentiation of skeletal progenitor cells that reside in cranial sutures has been demonstrated to cause CS as well.<sup>3</sup> The results suggest that CNCCs, which give rise to suture progenitor cell populations, might be particularly vulnerable to impact from genetic disruption in chromatin modifiers. The identification of variants in both distinct and overlapping chromatin modifiers in cohorts ascertained for unique birth defects continues to shed light on which chromatin genes might play more significant roles in the function of specific cell types, perhaps contributing to the diverse phenotypic presentations associated with their mutation. Although aberrant programming of suture progenitor cells is a plausible explanation for these findings, further work will be necessary to elucidate the specific mechanisms by which variants affecting the epigenome cause CS.

A recurrent gain-of-function variant in *RARA* provides further evidence that augmented retinoic-acid signaling is an important cause of CS. Recessive variants in *POR* (MIM: 124015), which encodes a P450 electron donor, cause syndromic CS via abnormal steroidogenesis with consequent elevations in circulating levels of retinoic acid.<sup>57</sup> Recessive variants in *CYP26B1* (MIM: 605207), which encodes a P450 enzyme responsible for retinoic-acid degradation, also cause syndromic CS via accelerated differentiation of osteoblasts.<sup>58</sup> Fetal retinoid syndrome, resulting from excessive retinoid exposure *in utero*, results in craniofacial malformations that include CS.<sup>59</sup> This germline gain-of-function variant in *RARA* merits further study, and additional individuals harboring the mutation will be necessary to fully characterize its phenotypic effects.

The identification of DNVs conferring strong risk of both NDD and CS has clinical implications. CS is usually present at birth, and syndromic features often prompt referral to a clinical geneticist before discharge from the hospital. Because the presence of CS is noted remarkably earlier than many of the behavioral or developmental traits associated with NDD, early evaluation of children with appar-

ently syndromic CS is likely to result in earlier molecular diagnoses for many affected children, allowing for prospective enrollment in early-intervention programs to potentially mitigate neurodevelopmental delays and optimize clinical outcomes. The identification of DNVs in genes not previously associated with CS suggests that broader exome or genome testing will provide additional clinical utility over targeted gene panels. Previous work established a role for DNVs in high pLI genes in neurodevelopmental outcomes in non-syndromic CS<sup>48</sup>; our results imply that genotypes identifiable at birth might have similar prognostic value in syndromic CS. Although surgery has a clear role in mitigating elevations in intracranial pressure in syndromic CS,<sup>49,60</sup> intrinsic brain dysfunction attributable to genetic lesions affecting neurodevelopment appears to explain a component of adverse neurodevelopmental outcomes.

These results increase the number of known syndromic CS loci by >25%, which stands to significantly augment the diagnostic utility of both exome and targeted sequencing approaches in this population. From the observed excess of damaging DNVs and the number of genes mutated multiple times, we estimate that approximately 38 novel genes contribute to syndromic CS risk in this cohort (Figure S7; see [subjects and methods](#)). Although the current analysis implicates 15 individual genes with significant enrichment, and an additional 18 chromatin genes with damaging DNVs, these findings support recruitment of a substantially larger cohort to identify additional CS loci.

## Data and code availability

RNA-seq data are available in the NIH Gene Expression Omnibus platform under accession numbers GEO: GSM1817212, GSM1817213, GSM1817214, GSM1817215, GSM1817216, and GSM1817217. This study makes use of data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from <https://deciphergenomics.org/about/stats> and via e-mail from [contact@deciphergenomics.org](mailto:contact@deciphergenomics.org).

## Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2023.03.017>.

## Acknowledgments

Funding for the DECIPHER project was provided by the Wellcome Trust.

## Author contributions

A.T.T., R.P.L., and K.T.K. performed genetic analyses and wrote the manuscript. S.M., G.A., E.K., E.M.W., S.C.J., and P.K. performed genetic analyses, A.L.S. and T.J.B. generated structural models of

genes studied, and O.C., M.M., T.L.W., M.S., E.M.F., and J.A.P. contributed clinical evaluations. All authors reviewed and edited the manuscript.

## Declaration of interests

S.M., M.M., M.S., and P.K. are employees of GeneDx.

Received: October 3, 2022

Accepted: March 24, 2023

Published: May 4, 2023

## References

1. Timberlake, A.T., and Persing, J.A. (2018). Genetics of nonsyndromic craniosynostosis. *Plast. Reconstr. Surg.* *141*, 1508–1516. <https://doi.org/10.1097/PRS.0000000000004374>.
2. Twigg, S.R.F., and Wilkie, A.O.M. (2015). A genetic-pathophysiological framework for craniosynostosis. *Am. J. Hum. Genet.* *97*, 359–377. <https://doi.org/10.1016/j.ajhg.2015.07.006>.
3. Menon, S., Salhotra, A., Shailendra, S., Tevlin, R., Ransom, R.C., Januszky, M., Chan, C.K.F., Behr, B., Wan, D.C., Longaker, M.T., and Quarto, N. (2021). Skeletal stem and progenitor cells maintain cranial suture patency and prevent craniosynostosis. *Nat. Commun.* *12*, 4640. <https://doi.org/10.1038/s41467-021-24801-6>.
4. Wilkie, A.O.M., Johnson, D., and Wall, S.A. (2017). Clinical genetics of craniosynostosis. *Curr. Opin. Pediatr.* *29*, 622–628. <https://doi.org/10.1097/MOP.0000000000000542>.
5. Timberlake, A.T., Jin, S.C., Nelson-Williams, C., Wu, R., Furey, C.G., Islam, B., Haider, S., Loring, E., Galm, A., et al.; Yale Center for Genome Analysis (2019). Mutations in TFAP2B and previously unimplicated genes of the BMP, Wnt, and Hedgehog pathways in syndromic craniosynostosis. *Proc. Natl. Acad. Sci. USA* *116*, 15116–15121. <https://doi.org/10.1073/pnas.1902041116>.
6. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibelini, E., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* *18*, 696–704. <https://doi.org/10.1038/gim.2015.148>.
7. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F., et al. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* *586*, 757–762. <https://doi.org/10.1038/s41586-020-2832-5>.
8. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* *38*, e164. <https://doi.org/10.1093/nar/gkq603>.
9. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* *24*, 2125–2137. <https://doi.org/10.1093/hmg/ddu733>.
10. Timberlake, A.T., Furey, C.G., Choi, J., Nelson-Williams, C., Yale Center for Genome Analysis, Loring, E., Galm, A., Kahle, K.T., Steinbacher, D.M., Larysz, D., et al. (2017). De novo mutations in inhibitors of Wnt, BMP, and Ras/ERK signaling pathways in non-syndromic midline craniosynostosis. *Proc. Natl. Acad. Sci. USA* *114*, E7341–E7347. <https://doi.org/10.1073/pnas.1709255114>.
11. Ware, J.S., Samocha, K.E., Homsy, J., and Daly, M.J. (2015). Interpreting de novo Variation in Human Disease Using denovolyzeR. *Curr. Protoc. Hum. Genet.* *87*, 7.25.1–7.25.15. <https://doi.org/10.1002/0471142905.hg0725s87>.
12. Timberlake, A.T., Choi, J., Zaidi, S., Lu, Q., Nelson-Williams, C., Brooks, E.D., Bilguvar, K., Tikhonova, I., Mane, S., Yang, J.F., et al. (2016). Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife* *5*, e20125. <https://doi.org/10.7554/eLife.20125>.
13. Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* *515*, 216–221. <https://doi.org/10.1038/nature13908>.
14. Ruzzo, E.K., Pérez-Cano, L., Jung, J.Y., Wang, L.K., Kashef-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J.N., Leventhal, O., et al. (2019). Inherited and de novo genetic risk for autism impacts shared networks. *Cell* *178*, 850–866.e26. <https://doi.org/10.1016/j.cell.2019.07.015>.
15. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* *180*, 568–584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>.
16. Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* *542*, 433–438. <https://doi.org/10.1038/nature21062>.
17. Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* *362*, eaat7615. <https://doi.org/10.1126/science.aat7615>.
18. Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* *4*, Article17. <https://doi.org/10.2202/1544-6115.1128>.
19. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell* *179*, 750–771.e22. <https://doi.org/10.1016/j.cell.2019.09.021>.
20. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* *44*, W90–W97. <https://doi.org/10.1093/nar/gkw377>.
21. Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* *103*, 785–801.e8. <https://doi.org/10.1016/j.neuron.2019.06.011>.
22. Jin, S.C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S.R., Zeng, X., Qi, H., Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* *49*, 1593–1601. <https://doi.org/10.1038/ng.3970>.



23. Moortgat, S., Berland, S., Aukrust, I., Maystadt, I., Baker, L., Benoit, V., Caro-Llopis, A., Cooper, N.S., Debray, F.G., Faivre, L., et al. (2018). HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. *Eur. J. Hum. Genet.* *26*, 64–74. <https://doi.org/10.1038/s41431-017-0038-6>.
24. Fuller, Z.L., Berg, J.J., Mostafavi, H., Sella, G., and Przeworski, M. (2019). Measuring intolerance to mutation in human genetics. *Nat. Genet.* *51*, 772–776. <https://doi.org/10.1038/s41588-019-0383-1>.
25. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf.* *10*, 48. <https://doi.org/10.1186/1471-2105-10-48>.
26. Hoyer, J., Ekici, A.B., Ende, S., Popp, B., Zweier, C., Wiesener, A., Wohlleber, E., Dufke, A., Rossier, E., Petsch, C., et al. (2012). Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* *90*, 565–572. <https://doi.org/10.1016/j.ajhg.2012.02.007>.
27. Helsmoortel, C., Vulto-van Silfhout, A.T., Coe, B.P., Vandeweyer, G., Rooms, L., van den Ende, J., Schuurs-Hoeijmakers, J.H.M., Marcelis, C.L., Willemsen, M.H., Vissers, L.E.L.M., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat. Genet.* *46*, 380–384. <https://doi.org/10.1038/ng.2899>.
28. Stessman, H.A.F., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerds, J., Trinh, S., Coemans, N., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* *49*, 515–526. <https://doi.org/10.1038/ng.3792>.
29. Vissers, L.E., van Ravenswaaij, C.M.A., Admiraal, R., Hurst, J.A., de Vries, B.B.A., Janssen, I.M., van der Vliet, W.A., Huys, E.H., de Jong, P.J., Hamel, B.C.J., et al. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat. Genet.* *36*, 955–957. <https://doi.org/10.1038/ng1407>.
30. De Luca, C., Picone, S., Cassina, M., Marziali, S., Morlino, S., Camerota, L., Tamburrini, G., Castori, M., Paolillo, P., Salviati, L., and Brancati, F. (2021). Craniosynostosis is a feature of CHD7-related CHARGE syndrome. *Am. J. Med. Genet.* *185*, 2160–2163. <https://doi.org/10.1002/ajmg.a.62208>.
31. Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L.H., Nowak, C.B., Douglas, J., Swoboda, K.J., et al. (2018). CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. *Nat. Commun.* *9*, 4619. <https://doi.org/10.1038/s41467-018-06014-6>.
32. Hoischen, A., van Bon, B.W.M., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* *42*, 483–485. <https://doi.org/10.1038/ng.581>.
33. Sifrim, A., Hitz, M.P., Wilsdon, A., Breckpot, J., Turki, S.H.A., Thienpont, B., McRae, J., Fitzgerald, T.W., Singh, T., Swaminathan, G.J., et al. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* *48*, 1060–1065. <https://doi.org/10.1038/ng.3627>.
34. Asadollahi, R., Oneda, B., Sheth, F., Azzarello-Burri, S., Bal-dinger, R., Joset, P., Latal, B., Knirsch, W., Desai, S., Baumer, A., et al. (2013). Dosage changes of MED13L further delineate its role in congenital heart defects and intellectual disability. *Eur. J. Hum. Genet.* *21*, 1100–1104. <https://doi.org/10.1038/ejhg.2013.17>.
35. Haijes, H.A., Koster, M.J.E., Rehmann, H., Li, D., Hakonarson, H., Cappuccio, G., Hancarova, M., Lehalle, D., Reardon, W., Schaefer, G.B., et al. (2019). De novo heterozygous POLR2A variants cause a neurodevelopmental syndrome with profound infantile-onset hypotonia. *Am. J. Hum. Genet.* *105*, 283–301. <https://doi.org/10.1016/j.ajhg.2019.06.016>.
36. Kim, J.H., Shinde, D.N., Reijnders, M.R.F., Hauser, N.S., Belmonte, R.L., Wilson, G.R., Bosch, D.G.M., Bubulya, P.A., Shashi, V., Petrovski, S., et al. (2016). De novo mutations in SON disrupt RNA splicing of genes essential for brain development and metabolism, causing an intellectual-disability syndrome. *Am. J. Hum. Genet.* *99*, 711–719. <https://doi.org/10.1016/j.ajhg.2016.06.029>.
37. Yoneda, Y., Saito, H., Touyama, M., Makita, Y., Miyamoto, A., Hamada, K., Kurotaki, N., Tomita, H., Nishiyama, K., Tsurusaki, Y., et al. (2012). Missense mutations in the DNA-binding/dimerization domain of NFIX cause Sotos-like features. *J. Hum. Genet.* *57*, 207–211. <https://doi.org/10.1038/jhg.2012.7>.
38. Küry, S., Besnard, T., Ebstein, F., Khan, T.N., Gambin, T., Douglas, J., Bacino, C.A., Craigen, W.J., Sanders, S.J., Lehmann, A., et al. (2017). De novo disruption of the proteasome regulatory subunit PSMD12 causes a syndromic neurodevelopmental disorder. *Am. J. Hum. Genet.* *100*, 352–363. <https://doi.org/10.1016/j.ajhg.2017.01.003>.
39. Stark, Z., Foulger, R.E., Williams, E., Thompson, B.A., Patel, C., Lunke, S., Snow, C., Leong, I.U.S., Puziakova, A., Daugherty, L.C., et al. (2021). Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am. J. Hum. Genet.* *108*, 1551–1557. <https://doi.org/10.1016/j.ajhg.2021.06.020>.
40. Ueda, K., Yaoita, M., Niihori, T., Aoki, Y., and Okamoto, N. (2017). Craniosynostosis in patients with RASopathies: accumulating clinical evidence for expanding the phenotype. *Am. J. Med. Genet.* *173*, 2346–2352. <https://doi.org/10.1002/ajmg.a.38337>.
41. Gallagher, R.E., Moser, B.K., Racevskis, J., Poiré, X., Bloomfield, C.D., Carroll, A.J., Ketterling, R.P., Roulston, D., Schachter-Tokarz, E., Zhou, D.C., et al. (2012). Treatment-influenced associations of PML-RARalpha mutations, FLT3 mutations, and additional chromosome abnormalities in relapsed acute promyelocytic leukemia. *Blood* *120*, 2098–2108. <https://doi.org/10.1182/blood-2012-01-407601>.
42. James, A.W., Levi, B., Xu, Y., Carre, A.L., and Longaker, M.T. (2010). Retinoic acid enhances osteogenesis in cranial suture-derived mesenchymal cells: potential mechanisms of retinoid-induced craniosynostosis. *Plast. Reconstr. Surg.* *125*, 1352–1361. <https://doi.org/10.1097/PRS.0b013e3181d62980>.
43. Eberhardt, J., McEwen, A.G., Bourguet, W., Moras, D., and Dejaegere, A. (2019). A revisited version of the apo structure of the ligand-binding domain of the human nuclear receptor retinoic X receptor alpha. *Acta Crystallogr. F Struct. Biol. Commun.* *75*, 98–104. <https://doi.org/10.1107/S2053230X18018022>.
44. Sato, Y., Ramalanjaona, N., Huet, T., Potier, N., Osz, J., Antony, P., Peluso-Iltis, C., Poussin-Courmontagne, P., Ennifar, E., Mély, Y., et al. (2010). The "Phantom Effect" of the Rixinoid LG100754: structural and functional insights. *PLoS One* *5*, e15119. <https://doi.org/10.1371/journal.pone.0015119>.

45. Wurtz, J.M., Bourguet, W., Renaud, J.P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* 3, 87–94. <https://doi.org/10.1038/nsb0196-87>.
46. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84, 524–533. <https://doi.org/10.1016/j.ajhg.2009.03.010>.
47. Jin, S.W., Sim, K.B., and Kim, S.D. (2016). Development and growth of the normal cranial vault : an embryologic review. *J. Korean Neurosurg. Soc.* 59, 192–196. <https://doi.org/10.3340/jkns.2016.59.3.192>.
48. Timberlake, A.T., Junn, A., Flores, R., Staffenberg, D.A., Lifton, R.P., and Persing, J.A. (2022). Genetic influence on neurodevelopment in nonsyndromic craniosynostosis. *Plast. Reconstr. Surg.* 149, 1157–1165. <https://doi.org/10.1097/PRS.0000000000008976>.
49. Brooks, E.D., Beckett, J.S., Yang, J., Timberlake, A.T., Sun, A.H., Chuang, C., and Persing, J.A. (2018). The etiology of neuronal development in craniosynostosis: a working hypothesis. *J. Craniofac. Surg.* 29, 49–55. <https://doi.org/10.1097/SCS.0000000000004040>.
50. Maruyama, T., Jeong, J., Sheu, T.J., and Hsu, W. (2016). Stem cells of the suture mesenchyme in craniofacial bone development, repair and regeneration. *Nat. Commun.* 7, 10526. <https://doi.org/10.1038/ncomms10526>.
51. Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 163, 68–83. <https://doi.org/10.1016/j.cell.2015.08.036>.
52. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. *N. Engl. J. Med.* 376, 21–31. <https://doi.org/10.1056/NEJMoa1516767>.
53. Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K., et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 498, 220–223. <https://doi.org/10.1038/nature12141>.
54. Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J., DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* 350, 1262–1266. <https://doi.org/10.1126/science.aac9396>.
55. Jin, S.C., Dong, W., Kundishora, A.J., Panchagnula, S., Moreno-De-Luca, A., Furey, C.G., Allocco, A.A., Walker, R.L., Nelson-Williams, C., Smith, H., et al. (2020). Exome sequencing implicates genetic disruption of prenatal neuro-gliogenesis in sporadic congenital hydrocephalus. *Nat. Med.* 26, 1754–1765. <https://doi.org/10.1038/s41591-020-1090-2>.
56. Sood, S., Weber, C.M., Hodges, H.C., Krokhotin, A., Shalizi, A., and Crabtree, G.R. (2020). CHD8 dosage regulates transcription in pluripotency and early murine neural differentiation. *Proc. Natl. Acad. Sci. USA* 117, 22331–22340. <https://doi.org/10.1073/pnas.1921963117>.
57. Flück, C.E., Tajima, T., Pandey, A.V., Arlt, W., Okuhara, K., Verge, C.F., Jabs, E.W., Mendonça, B.B., Fujieda, K., and Miller, W.L. (2004). Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat. Genet.* 36, 228–230. <https://doi.org/10.1038/ng1300>.
58. Laue, K., Pogoda, H.M., Daniel, P.B., van Haeringen, A., Alanay, Y., von Ameln, S., Rachwalski, M., Morgan, T., Gray, M.J., Breuning, M.H., et al. (2011). Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. *Am. J. Hum. Genet.* 89, 595–606. <https://doi.org/10.1016/j.ajhg.2011.09.015>.
59. Johnston, M.C., and Bronsky, P.T. (1995). Prenatal craniofacial development: new insights on normal and abnormal mechanisms. *Crit. Rev. Oral Biol. Med.* 6, 368–422. <https://doi.org/10.1177/10454411950060040601>.
60. Still, M.E.H., Chidarala, S., Alvarado, A., Neal, D., Governale, L., and Ching, J. (2022). Craniosynostosis surgery for increased intracranial pressure. *J. Craniofac. Surg.* 33, 1454–1457. <https://doi.org/10.1097/SCS.0000000000008357>.

**The American Journal of Human Genetics, Volume 110**

**Supplemental information**

***De novo* variants implicate chromatin modification,  
transcriptional regulation, and retinoic acid  
signaling in syndromic craniosynostosis**

**Andrew T. Timberlake, Stephen McGee, Garrett Allington, Emre Kiziltug, Erin M. Wolfe, Amy L. Stiegler, Titus J. Boggon, May Sanyoura, Michelle Morrow, Tara L. Wenger, Erica M. Fernandes, Oana Caluseriu, John A. Persing, Sheng Chih Jin, Richard P. Lifton, Kristopher T. Kahle, and Paul Kruszka**

## Table of Contents:

**Figure S1. Venn diagram demonstrating overlap between syndromic CS risk genes and genes implicated in autism, intellectual disability, and other NDDs**

**Figure S2. High pLI genes with multiple damaging *de novo* variants in syndromic CS probands.**

**Figure S3. Domain annotations for CHD3, CDK13, MED13L**

**Figure S4. Pathway analysis of genes in modules enriched in CS**

**Figure S5. Temporal expression of coexpression networks implicated in syndromic CS**

**Figure S6. UMAP plot of scRNA-seq data demonstrating 32 distinct clusters**

**Figure S7. Estimation of the number of syndromic craniosynostosis risk genes by simulation**

**Table S1. Curated List of Neurodevelopmental Disorder, Autism, and Craniosynostosis Risk Genes**

**Table S2. Neural Crest Cell Gene Expression Rank**

**Table S3. Known syndromic CS genes**

**Table S4. *De novo* variants in known genes in syndromic CS probands**

**Table S5. Phenotypes of patients with mutations in novel CS risk loci**

**Table S6. All DNVs identified in 526 probands with unsolved syndromic CS**

**Table S7. GOrilla pathway analysis of all genes harboring damaging DNVs**

**Table S8. Burden of *de novo* variants in 1,789 autism control trios**

**Table S9. Enrichment in chromatin genes in 498 trios with syndromic CS after exclusion of 38 probands with variants in 13 genes identified via *de novo* variant burden at the individual gene level**

**Table S10. *De novo* variants in high pLI OMIM genes in syndromic CS probands**

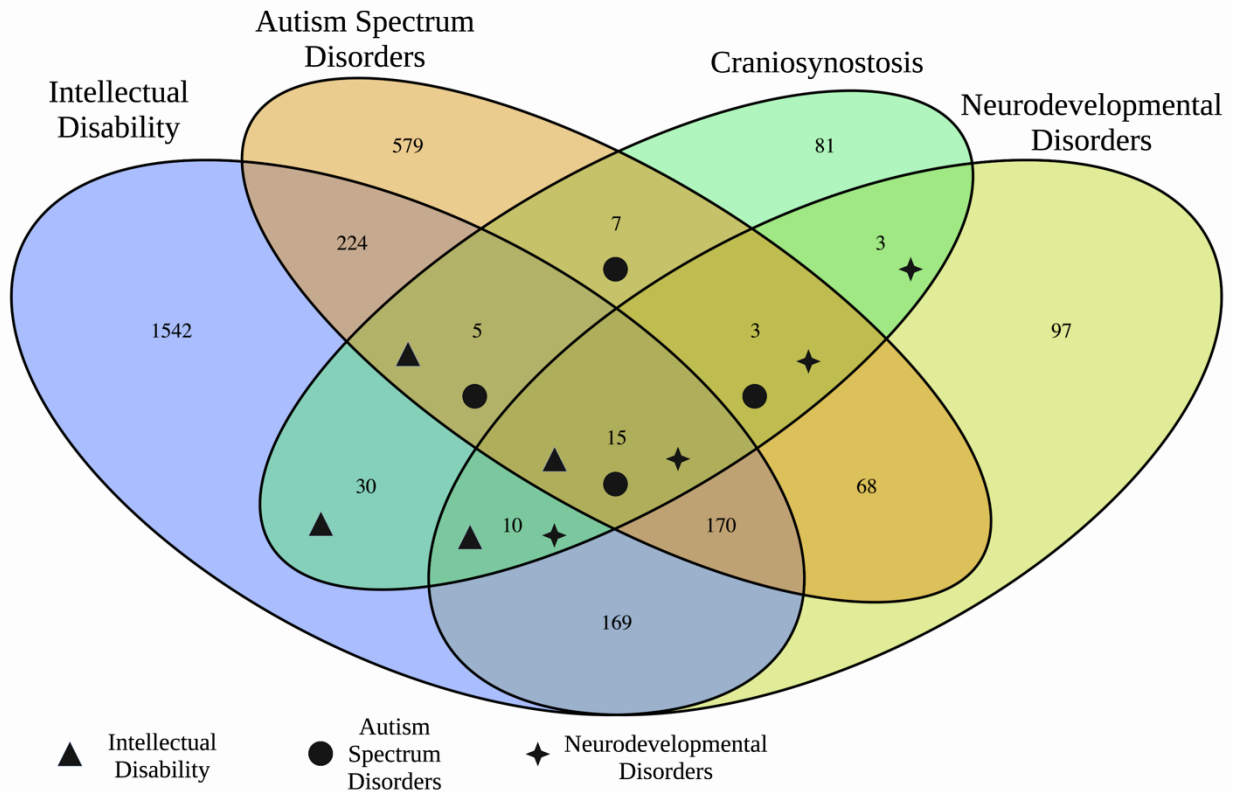
**Table S11. Enrichment of LOF intolerant genes with high expression in cranial neural crest cells in 526 trios with syndromic CS**

**Supplementary Note 1: Probability of observing two identical *de novo* *HUWE1* (p.Arg110Trp) or *RARA* (p.Gly289Arg) variants**

**Supplementary Note 2: Phenomic analyses**

**Supplementary Note 3: Case Reports**

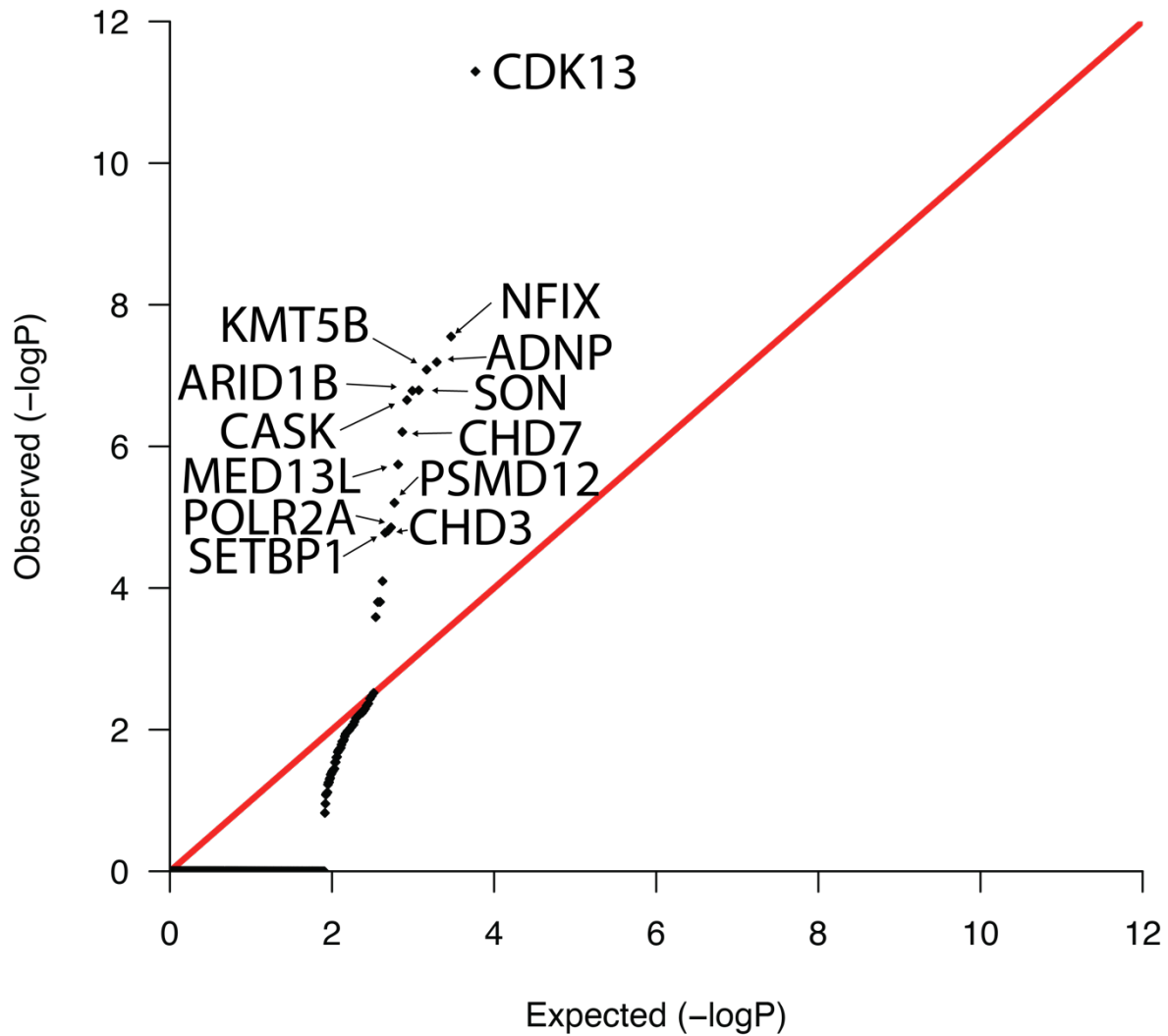
**Figure S1. Venn diagram demonstrating overlap between syndromic CS risk genes and genes implicated in autism, intellectual disability, and other NDDs.**



Disease	Genes in DisGeNET	CS Candidate Genes Overlapping with DisGeNET	P-Value
Neurodevelopmental Disorders	535	31	✦ $6.55 \times 10^{-20}$
Autism Spectrum Disorders	1071	30	● $5.83 \times 10^{-11}$
Intellectual Disability	2165	60	▲ $4.16 \times 10^{-22}$

**Figure S1. Venn diagram demonstrating overlap between syndromic CS risk genes and genes implicated in autism, intellectual disability, and other NDDs. Overlap between genes implicated in CS and each phenotype was significant by Fisher's exact test (Methods).**

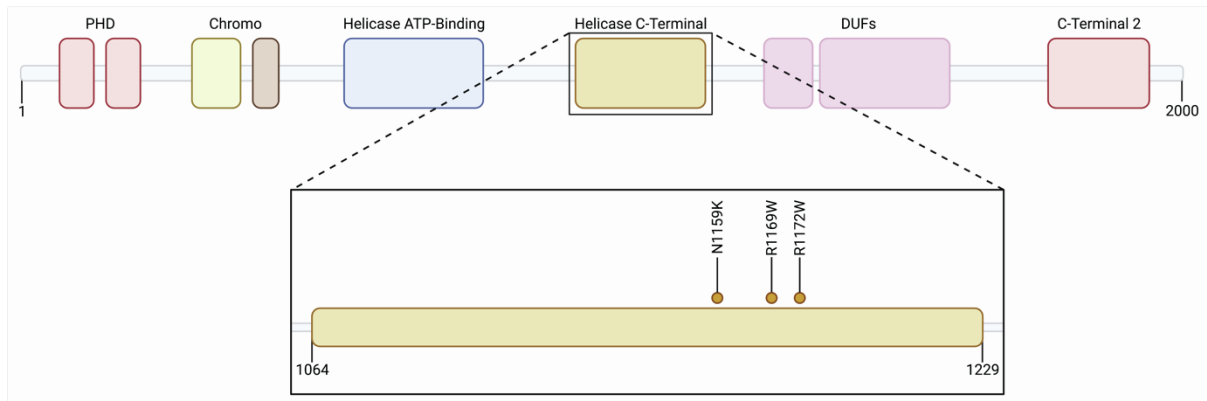
**Figure S2. High pLI genes with multiple damaging *de novo* variants in syndromic CS probands.**



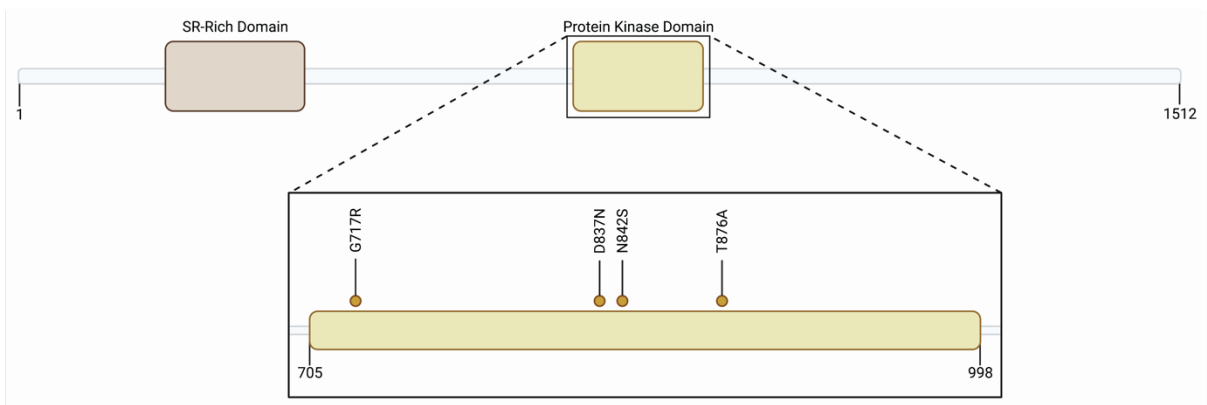
**Figure S2. High pLI genes with multiple damaging *de novo* variants in syndromic CS probands.** Quantile-quantile plot of log transformed observed vs expected P-values for genes with damaging *de novo* variants in syndromic CS probands (Poisson distribution, see Methods). Thirteen genes surpassing the threshold for genome-wide significance after Bonferroni correction for high pLI genes ( $0.05/3,063 = 1.63 \times 10^{-5}$ ) are labeled.

**Figure S3. Domain annotations for CHD3, CDK13, MED13L**

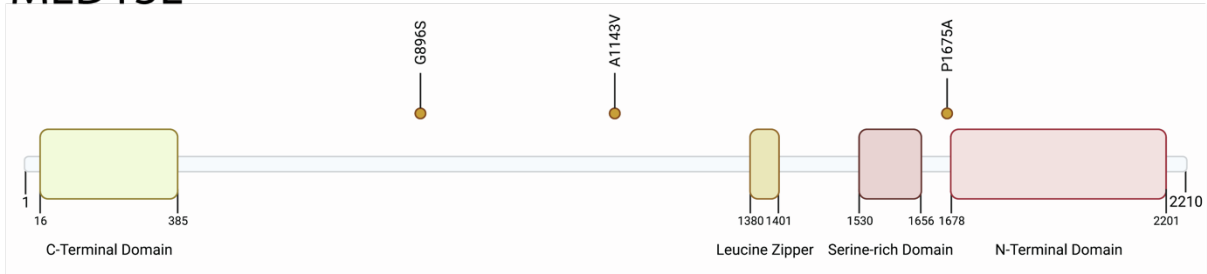
### A. CHD3



### B. CDK13

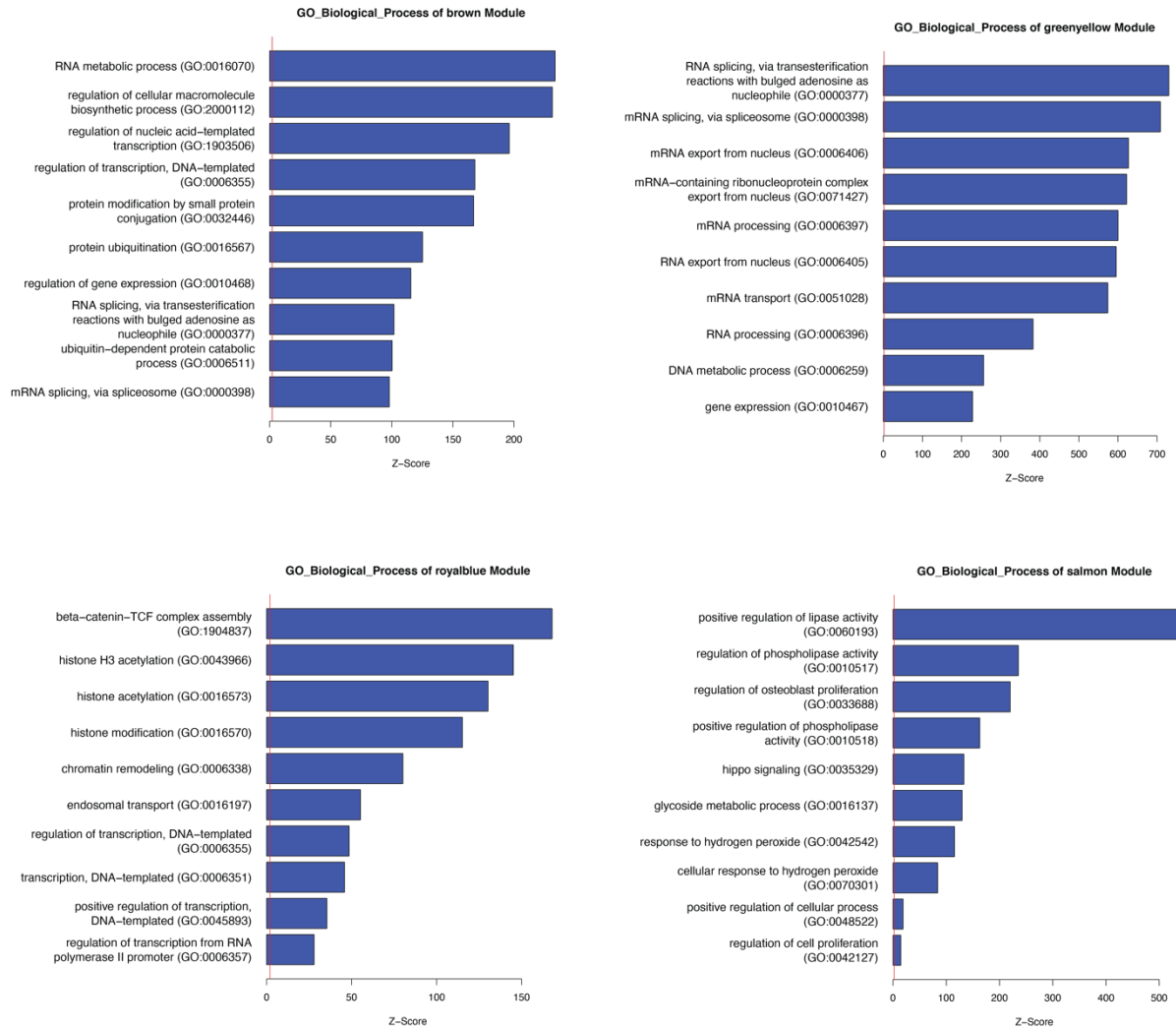


### C. MED13L



**Figure S3. Domain annotations for CHD3, CDK13, MED13L.** Domain annotations appear for genes in which damaging *de novo* variants identified in syndromic CS probands were uniquely missense variants. **A.** In CHD3, missense variants are all in the helicase C-terminal domain. **B.** In CDK13, all 4 variants appear within the kinase domain. **C.** In MED13L, variants are scattered throughout the protein, supporting haploinsufficiency as a possible mechanism of contribution to CS risk.

**Figure S4. Pathway analysis of genes in modules enriched in CS.**



**Figure S4. Pathway analysis of genes in modules enriched in CS.** Pathway analysis of genes encompassed by each of the four transcriptional networks enriched for genes mutated in syndromic CS. The brown module genes are implicated in transcriptional regulation, the greenyellow module genes are implicated in RNA splicing, the royalblue module genes are implicated in histone modification, and the salmon module genes are involved in osteoblast differentiation.



Figure S5. Temporal expression of coexpression networks implicated in syndromic CS.

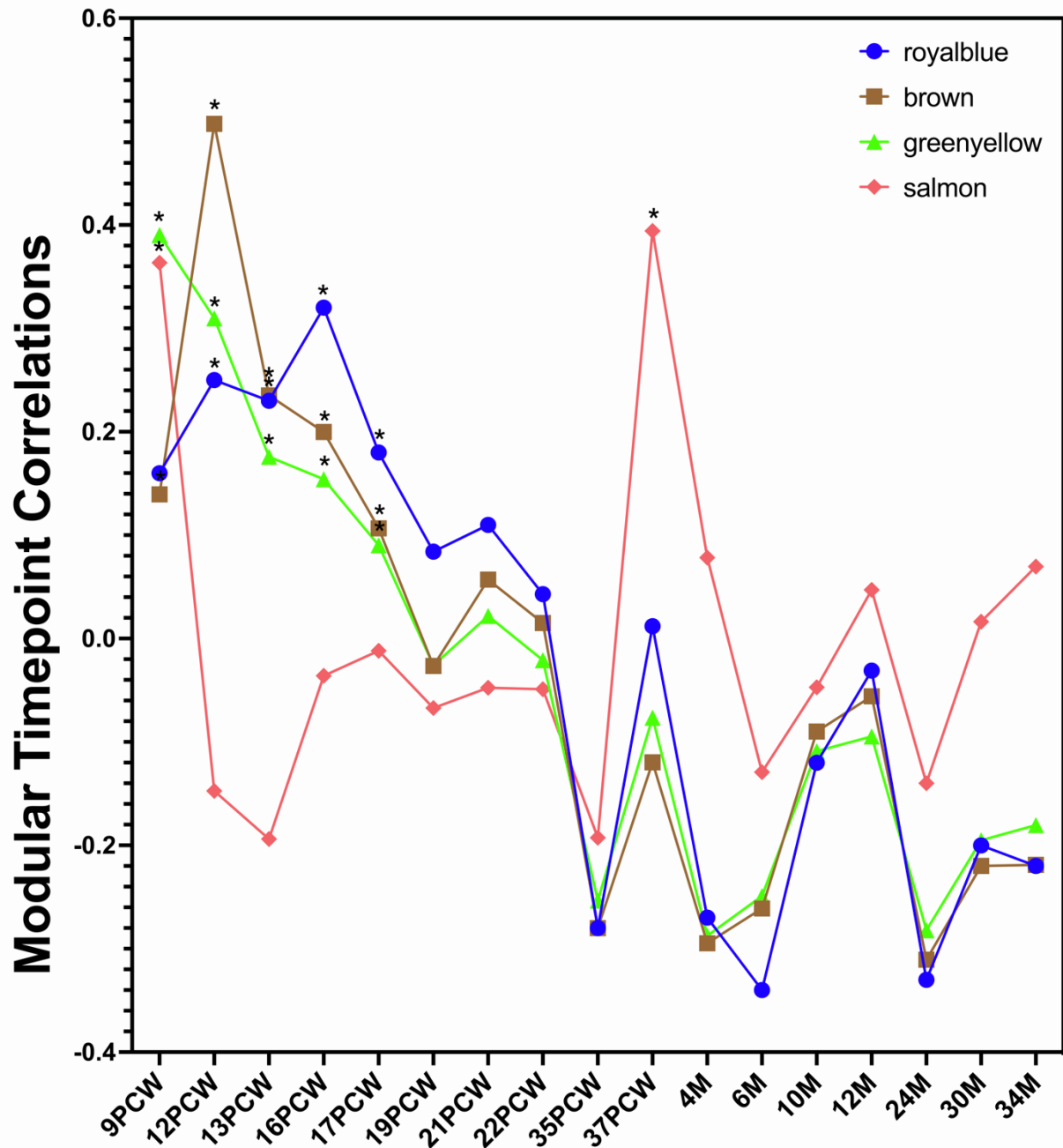
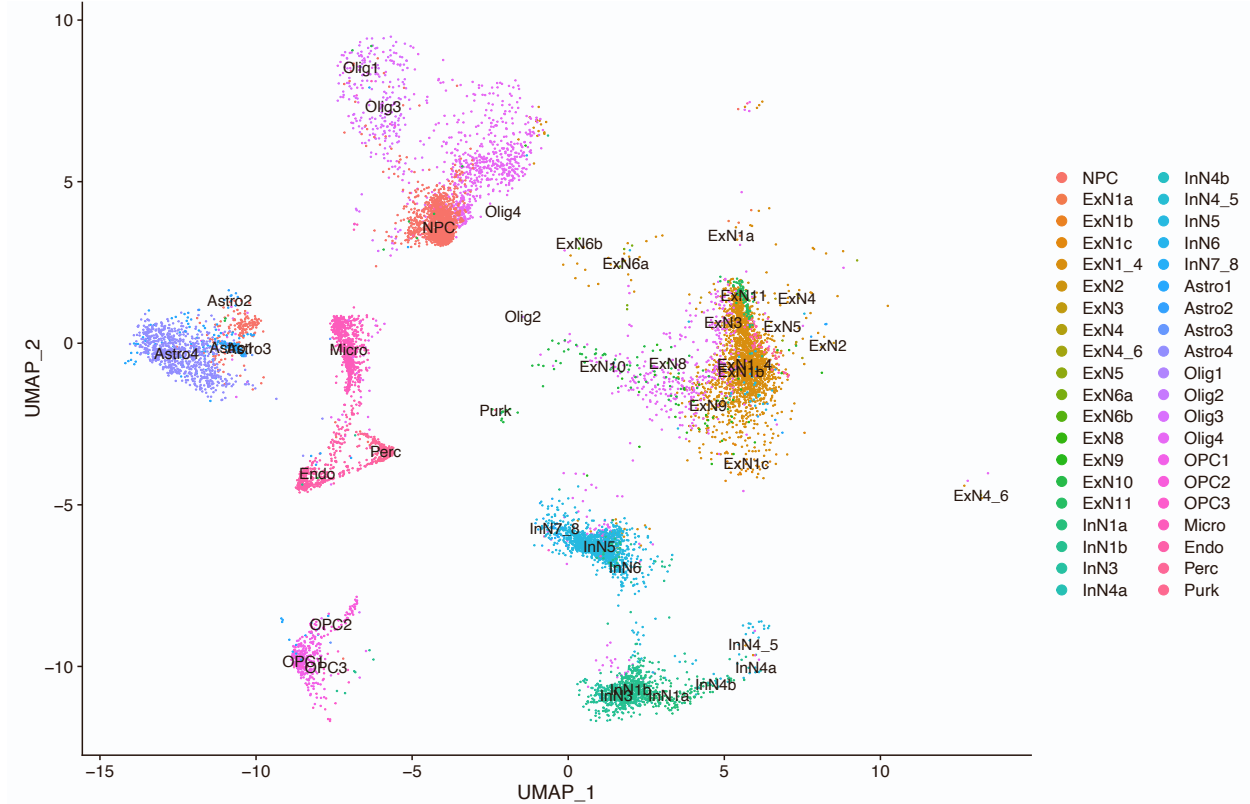


Figure S5. Temporal expression of coexpression networks implicated in syndromic CS.

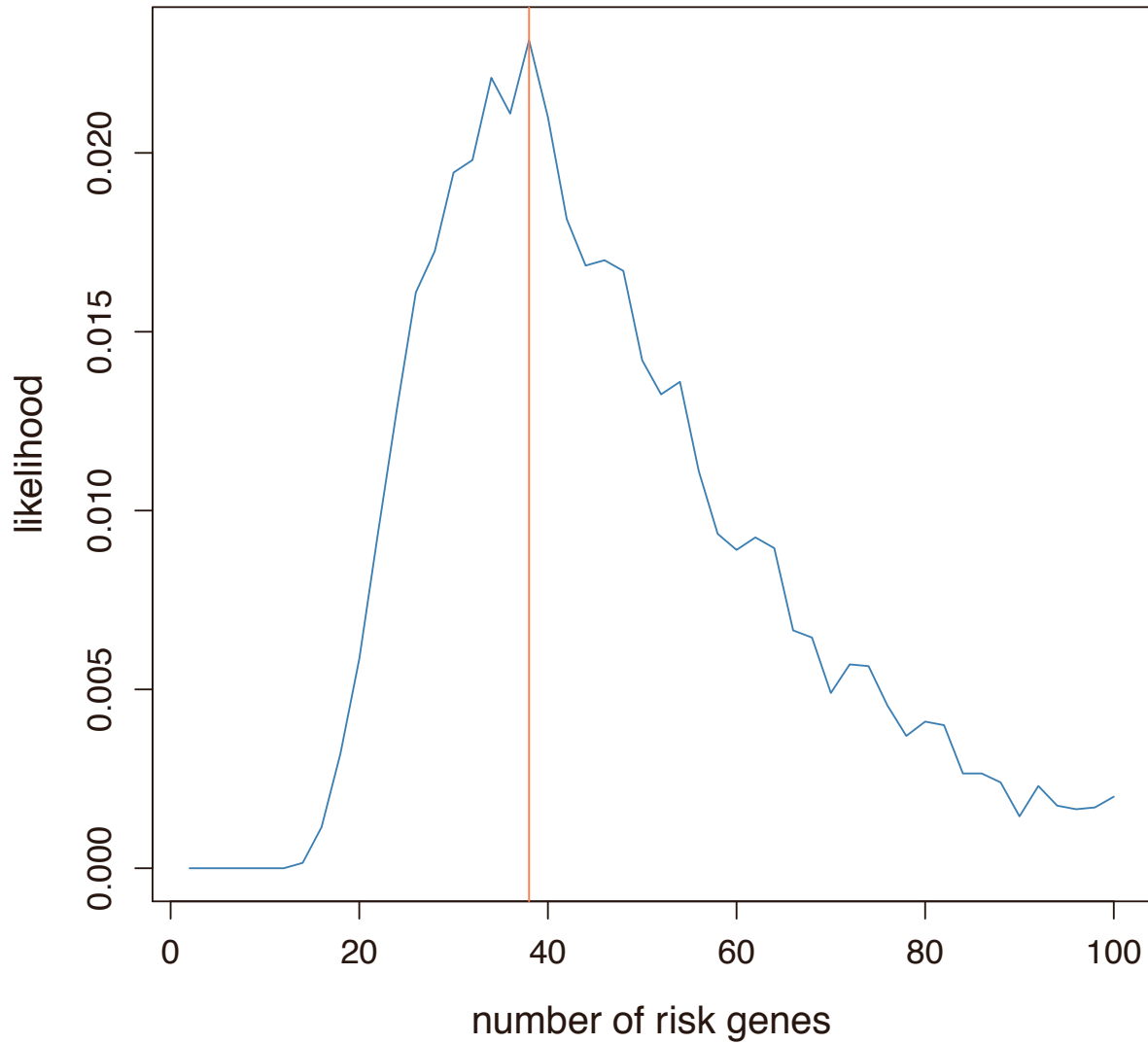
Temporal expression of the brown, greenyellow, royalblue, and salmon modules are plotted, with periods with significantly enriched expression indicated with asterisks (Methods). Implicated modules are highly expressed during the mid-gestational period during which the skull ossifies and sutures form, namely post-conception weeks 9-17.

**Figure S6. UMAP plot of scRNA-seq data demonstrating 32 distinct clusters.**



**Figure S6. UMAP plot of scRNA-seq data demonstrating 32 distinct clusters.** Clusters identified using over 40,000 cells were used to identify specific neural cell populations, which were assessed for enrichment in CS risk genes (Methods). NPC, neural progenitor cells; ExN, excitatory neurons; InN, inhibitory neurons; Olig, oligodendrocytes; OPC, oligodendrocyte progenitor cells; Micro, microglia; Astro, astrocytes; Endo, endothelial cells; Perc, pericytes; Purk, purkinjee cells.

**Figure S7. Estimation of the number of syndromic craniosynostosis risk genes by simulation.**



**Figure S7. Estimation of the number of syndromic craniosynostosis risk genes by simulation.** The likelihood of the model under the assumption of there being 0-100 risk genes is plotted, with maximum likelihood observed at 38 genes (Methods).

**Table S1. Curated List of Neurodevelopmental Disorder, Autism, and Craniosynostosis Risk Genes (provided as Excel File Table S1)**

**Table S2. Neural Crest Cell Gene Expression Rank (provided as Excel File Table S2)**

**Table S3. Known syndromic CS genes**

Genes	
<i>ADAMTSL4</i>	<i>IL11RA</i>
<i>ALPL</i>	<i>IRX5</i>
<i>ALX4</i>	<i>JAG1</i>
<i>ASXL1</i>	<i>KAT6A</i>
<i>ATR</i>	<i>KMT2D</i>
<i>CDC45</i>	<i>KRAS</i>
<i>COLEC11</i>	<i>LMX1B</i>
<i>CTSK</i>	<i>LRP5</i>
<i>CYP26B1</i>	<i>MEGF8</i>
<i>EFNA4</i>	<i>MSX2</i>
<i>EFNB1</i>	<i>PHEX</i>
<i>ERF</i>	<i>POR</i>
<i>ESCO2</i>	<i>RAB23</i>
<i>FAM20C</i>	<i>RECQL4</i>
<i>FBN1</i>	<i>RUNX2</i>
<i>FGFR1</i>	<i>SCARF2</i>
<i>FGFR2</i>	<i>SH3PXD2B</i>
<i>FGFR3</i>	<i>SKI</i>
<i>FLNA</i>	<i>SPECC1L</i>
<i>GLI3</i>	<i>STAT3</i>
<i>GNAS</i>	<i>TCF12</i>
<i>GNPTAB</i>	<i>TGFBR1</i>
<i>GPC3</i>	<i>TGFBR2</i>
<i>HUWE1</i>	<i>TMCO1</i>
<i>IDS</i>	<i>TWIST1</i>
<i>IDUA</i>	<i>WDR35</i>
<i>IFT122</i>	<i>ZEB2</i>
<i>IHH</i>	<i>ZIC1</i>

Known syndromic CS genes, derived from Twigg et al[2]. Probandes were screened for variants in these genes, and those with identified variants were removed from further assessment.

**Table S4. De novo variants in known genes in syndromic CS probands**

Gene	Impact	GnomAD Frequency	pLI
<i>EFNB1</i>	p.Arg109Gly	0	0.93
<i>EFNB1</i>	p.Gln214*	0	0.93
<i>EFNB1</i>	p.Arg66*	0	0.93
<i>ERF</i>	p.Ser297fs*9	0	0.99
<i>FAM20C</i>	p.Gly280Arg	4.9x10 <sup>-6</sup>	0.31
<i>FGFR2</i>	p.Glu566Gly	0	1
<i>FGFR2</i>	p.Try105Cys	4.0x10 <sup>-6</sup>	1
<i>FGFR2</i>	p.Tyr340Cys	0	1
<i>FGFR2</i>	p.Ser347Cys	0	1
<i>FGFR2</i>	c.G1032A; p.=	0	1
<i>FGFR3</i>	p.Pro250Arg	7.4x10 <sup>-6</sup>	0
<i>FGFR3</i>	p.Ala391Glu	0	0
<i>FGFR3</i>	p.Ala391Glu	0	0
<i>GNAS</i>	p.Asp799Asn	0	0.68
<i>HUWE1</i>	p.Arg110Trp	0	1
<i>HUWE1</i>	p.Arg110Trp	0	1
<i>HUWE1</i>	p.Arg110Gln	0	1
<i>JAG1</i>	p.Ile819Leufs*2	0	1
<i>KAT6A</i>	p.Arg1024*	0	1
<i>KAT6A</i>	p.Ser1880*	0	1
<i>KMT2D</i>	c.15922-1G>A	0	1
<i>KMT2D</i>	p.Ile3420Val	0	1
<i>KMT2D</i>	p.Asn4027His	0	1
<i>KRAS</i>	p.Gly60Ser	4.0x10 <sup>-6</sup>	0
<i>SH3PXD2B</i>	p.Ile633Val	0	0
<i>SKI</i>	p.Asn128Ser	0	1
<i>TGFBR2</i>	p.Arg553Cys	0	0.13
<i>ZEB2</i>	p.Leu349*	0	1
<i>ZIC1</i>	p.Asp348Asn	0	0.94

Variants identified in known syndromic CS genes in 555 trios with sporadic CS. The impact of each variant is provided at the protein level for missense, frameshift, and nonsense variants, and at the DNA level for splice variants. Complete annotations for each variant are found in Table S6. pLI= probability of LOF intolerance. Allele frequencies and pLI scores were obtained from GnomAD v2.1.1. In the case of *FGFR2*, *FGFR3*, and *KRAS* variants found once in GnomAD, each is also annotated as 'pathogenic' by ClinVar. Variants in known genes explained 5.2% of cases studied.

**Table S5. Phenotypes of patients with mutations in novel CS risk loci (provided as Excel File Table S5)**

**Table S6. All DNVs identified in 526 probands with unsolved syndromic CS (provided as Excel File Table S6)**

**Table S7. GOrilla pathway analysis of all genes harboring damaging DNVs (provided as Excel File Table S7)**

**Table S8. Burden of *de novo* variants in 1,789 autism control trios**

	Observed		Expected		Enrichment	P
	n	Rate	n	Rate		
<b>High pLI Genes (&gt;0.9) (n=3,063)</b>						
Total	453	0.25	491.3	0.27	0.92	0.96
Synonymous	113	0.06	137.1	0.077	0.83	0.98
Total missense	305	0.17	309.2	0.18	0.99	0.60
D-mis	61	0.034	66.2	0.037	0.92	0.76
LoF	35	0.020	45.0	0.025	0.78	0.95
Damaging	96	0.054	111.2	0.062	0.86	0.93
<b>High pLI OMIM Genes (n=643)</b>						
Total	111	0.062	120.3	0.067	0.92	0.81
Synonymous	25	0.014	33.7	0.019	0.74	0.95
Total missense	80	0.044	75.9	0.042	1.05	0.33
D-mis	20	0.011	25.9	0.014	0.77	0.90
LoF	6	0.0034	10.8	0.0060	0.56	0.96
Damaging	26	0.015	36.7	0.021	0.71	0.97
<b>Chromatin Modifiers and Remodelers (n=614)</b>						
Total	77	0.043	87.4	0.049	0.88	0.88
Synonymous	17	0.0095	23.7	0.013	0.72	0.94
Total missense	54	0.031	55.2	0.031	0.98	0.58
D-mis	16	0.0089	13.0	0.0073	1.23	0.23
LoF	6	0.0034	8.5	0.0048	0.71	0.85
Damaging	22	0.012	21.5	0.012	1.02	0.48

n, number of *de novo* variants in 1,789 control trios; Rate, number of *de novo* variants per subject; Damaging missense called by MetaSVM (D-mis); Loss of function denotes premature termination, frameshift, splice site variant, startloss, or stoploss variants. P-values represent the upper tail of the Poisson probability density function.



**Table S9. Enrichment in chromatin genes in 498 trios with syndromic CS after exclusion of 38 probands with variants in 13 genes identified via *de novo* variant burden at the individual gene level**

Class	Observed		Expected		Enrichment	P-value
	#	#/subject	#	#/subject		
All variants	40	0.08	24.7	0.05	1.62	<i>0.003</i>
Synonymous	6	0.01	6.76	0.01	0.88	0.67
T-mis	12	0.02	11.9	0.02	1.00	0.53
D-mis	13	0.03	3.65	0.01	3.56	<i>1.14x10<sup>-4</sup></i>
Loss of function (LOF)	9	0.02	2.38	0.005	3.78	<i>8.13x10<sup>-4</sup></i>
Damaging	22	0.04	6.03	0.01	3.65	<i>4.23x10<sup>-7</sup></i>

D-mis as called by Meta-SVM. Damaging refers to D-mis and LOF alleles. P-values represent the upper tail of the Poisson distribution; significant P-values appear in italics.

**Table S10. De novo variants in high pLI OMIM genes in syndromic CS probands**

Gene	Impact	GnomAD Frequency	pLI
<i>AHDC1</i>	p.Gln1156*	0	1
<i>CACNA1E</i>	p.le1451Met	0	1
<i>COL11A1</i>	c.3468+1G>A	0	1
<i>COL2A1</i>	p.Arg904Cys	0	1
<i>COL3A1</i>	p.Arg1358Gln	1.2x10 <sup>-5</sup> *VUS	1
<i>COL4A1</i>	p.Gly888Arg	0	1
<i>CTNNA1</i>	p.Val374_375insSerTrpLysMetLys	0	0.97
<i>CTNND1</i>	p.Arg458*	0	1
<i>FBN2</i>	p.Cys1156Phe	0	1
<i>HCN4</i>	p.Leu438Val	0	1
<i>KCND3</i>	p.Val401Met	0	0.99
<i>MAST1</i>	p.Arg496His	0	1
<i>MFN2</i>	p.Arg95Thr	0	0.99
<i>NAA15</i>	p.Leu89Pro	0	1
<i>NF1</i>	p.Thr2135Ile	0	0.90
<i>PDHA1</i>	p.Arg340Cys	0	0.99
<i>PRKG1</i>	p.Pro537fs	0	1
<i>SCN1A</i>	p.Val846Ile	0	1
<i>SLC20A2</i>	p.Pro568Leu	1.2x10 <sup>-5</sup> *P	0.97
<i>SOX4</i>	p.Gly94Asp	0	0.93
<i>TRIO</i>	p.Gln1427Arg	0	1
<i>USP9X</i>	c.7219-2A>G	0	1
<i>ANKRD11</i>	p.Ala1751Ser	0	1
<i>BICD2</i>	p.Lys734Asn	0	0.98
<i>CLTC</i>	p.Pro890Leu	0	1
<i>CSNK2A1</i>	p.Arg47Gln	0	0.99
<i>FBXO11</i>	p.Pro109Leu	0	1
<i>FOXF1</i>	p.Asn219Ser	0	0.96
<i>HNRNPK</i>	p.Glu85Lys	0	1
<i>HNRNPK</i>	p.Asp190Ala	0	1
<i>HTT</i>	p.Thr341Ser	0	1
<i>HTT</i>	p.Met1623Val	0	1
<i>KANSL1</i>	p.Arg942Trp	0	1
<i>MCM6</i>	p.Cys158Tyr	0	0.98
<i>KIF11</i>	p.Leu517Val	0	1
<i>MED12</i>	p.Val1964Leu	0	1
<i>MTOR</i>	p.Arg2018Pro	0	1
<i>NOTCH1</i>	p.Ala16Thr	0	1
<i>PKD1</i>	p.Ser322Leu	1.7x10 <sup>-5</sup>	1
<i>PPP2CA</i>	p.Ser30Thr	0	0.99
<i>ROBO2</i>	p.Ser545Thr	0	1
<i>SPOP</i>	p.Thr25Ala	0	1
<i>YAP1</i>	p.Ala351Thr	0	1
<i>ZBTB20</i>	p.Gly63Arg	5.6x10 <sup>-5</sup>	0.97
<i>ZNF462</i>	p.Ala1171Thr	1.1x10 <sup>-5</sup>	1

Variants identified in known high pLI OMIM genes in 526 trios with sporadic syndromic CS not identified at the individual gene level. The impact of each variant is provided at the protein level for missense, frameshift, and nonsense variants, and at the DNA level for splice variants. pLI= probability of LOF intolerance. Allele frequencies and pLI scores were obtained from GnomAD v2.1.1. Gene names in bold represent LOF or D-mis alleles, while those not in bold represent missense variants not predicted to be damaging by MetaSVM, as these also demonstrated significant enrichment in probands. \*VUS represents a variant in ClinVar described as a variant of unknown significance, and \*P represents a ClinVar variant designated pathogenic.

**Table S11. Enrichment of LOF intolerant genes with high expression in cranial neural crest cells in 526 trios with syndromic CS**

Class	Observed		Expected		Enrichment	P-value
	#	#/subject	#	#/subject		
All variants	148	0.28	82.7	0.16	1.79	<i>6.75x10<sup>-11</sup></i>
Synonymous	22	0.042	23.0	0.044	0.96	0.61
T-mis	53	0.10	41.3	0.079	1.27	<i>0.04</i>
D-mis	39	0.07	10.8	0.02	3.62	<i>2.55x10<sup>-11</sup></i>
Loss of function (LOF)	34	0.06	7.63	0.015	4.46	<i>2.11x10<sup>-12</sup></i>
Damaging	73	0.14	18.4	0.035	3.97	<i>6.51x10<sup>-22</sup></i>

D-mis as called by Meta-SVM. Damaging refers to D-mis and LOF alleles. P-values represent the upper tail of the Poisson distribution; significant P-values appear in italics. Genes in the top 25% of CNCC expression with pLI>0.9 (n=1,593) were included for gene set enrichment analysis.

**Supplementary Note 1: Probability of observing two identical *de novo* *HUWE1* (p.Arg110Trp) or *RARA* (p.Gly289Arg) variants.**

Calculating this probability is analogous to the “birthday paradox”; i.e., the chance that in a set of “n” randomly chosen people, at least one pair will have the same birthday. In this scenario, we let  $R$  denote the total number of two identical DNVs in the cohort;  $R_j$  denotes the number of two identical DNVs in the  $j$ th tri-nucleotide category (e.g., TCC → TTC; based on flanking base context, there are  $4 \times 4 \times 4 \times 3 = 192$  categories in total);  $M$  denotes the total number of DNVs;  $M_j$  is the number of DNVs in the  $j$ th tri-nucleotide category.

Additionally, denote the number of potential DNVs in the  $j$ th tri-nucleotide category as  $L_j$ ; per-base variant probability of the  $j$ th category is  $\mu_j$ . The quantity of interest is then:

$$\begin{aligned}
 & \mathbb{P}(R > 0 \mid M = m) \\
 & \approx \sum_{j=1}^{192} \mathbb{P}(R_j > 0 \mid M = m) \\
 & = \sum_{j=1}^{192} \sum_{k=2}^m \mathbb{P}(R_j > 0, M_j = k \mid M = m) \\
 & = \sum_{j=1}^{192} \sum_{k=2}^m \mathbb{P}(R_j > 0 \mid M_j = k) \times \mathbb{P}(M_j = k \mid M = m)
 \end{aligned}$$

To calculate  $\mathbb{P}(R_j > 0 \mid M = m)$ :

$$\begin{aligned}
 \mathbb{P}(M_j = k \mid M = m) & = \binom{m}{k} \times [\mathbb{E}(M_j \mid M = 1)]^k \times [1 - \mathbb{E}(M_j \mid M = 1)]^{m-k} \\
 & = \binom{m}{k} \times \left( \frac{\mu_j L_j}{\sum_{i=1}^{192} \mu_i L_i} \right)^k \times \left( 1 - \frac{\mu_j L_j}{\sum_{i=1}^{192} \mu_i L_i} \right)^{m-k}
 \end{aligned}$$

We use the Poisson tail probability to estimate  $\mathbb{P}(R_j > 0 \mid M_j = k)$ :

$$\mathbb{P}(R_j > 0 \mid M_j = k) \approx 1 - \exp\left(-\binom{k}{2}/L_j^2\right)$$

Taken together, we have

$$\mathbb{P}(R > 0 \mid M = m) \approx \sum_{j=1}^{192} \sum_{k=2}^m (1 - \exp\left(-\binom{k}{2}/L_j^2\right)) \times \binom{m}{k} \times \left(\frac{\mu_j L_j}{\sum_{i=1}^{192} \mu_i L_i}\right)^k \times \left(1 - \frac{\mu_j L_j}{\sum_{i=1}^{192} \mu_i L_i}\right)^{m-k}$$

We use this equation to estimate the probability of having two identical DNVs given a total of 609

DNVs ( $P = 6.5 \times 10^{-3}$ ).

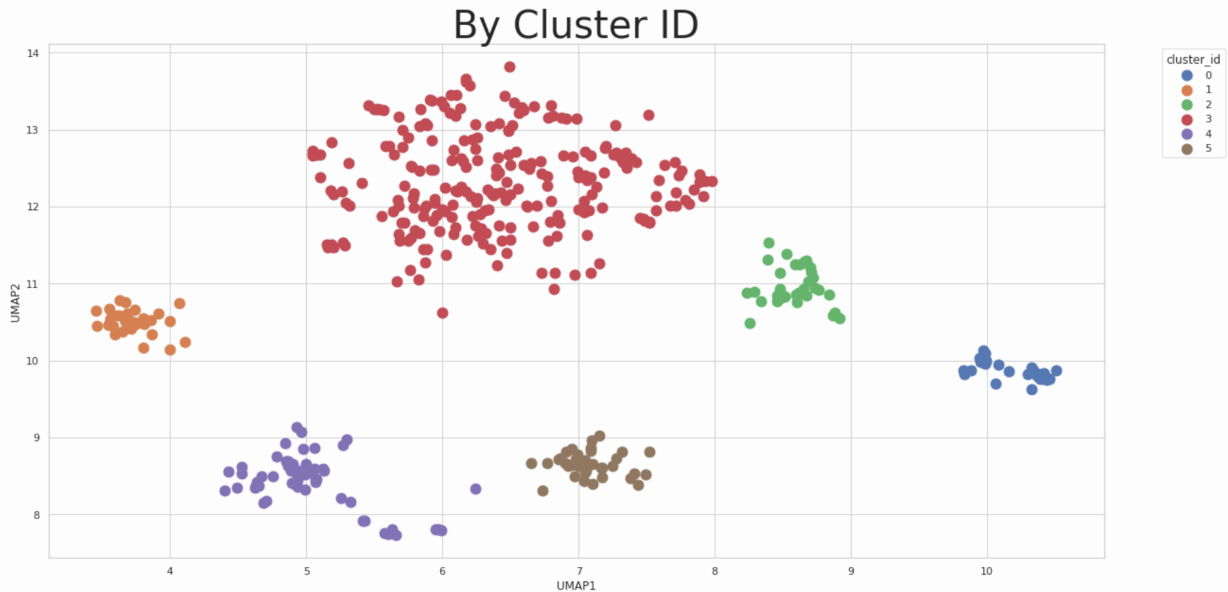
## Supplementary Note 2: Phenomic Analyses

### Compressing the HPO tree for dimensionality reduction

There are currently >16,000 disease-related terms in the Human Phenotype Ontology (HPO). To perform our phenomics analyses, we needed to reduce the number of features to a more manageable number by first group similar phenotype terms together. Terms in the ontology that are more similar based on Hybrid Relative Specificity Similarity (HRSS) are grouped together using a bottom-up approach to hierarchical clustering. Initially, all HPO terms begin as separate clusters. Then, cluster pairs are joined together with the most similar going first. Using this process, the optimal number of clusters (phenotype groups) can be adjusted based on a desired specificity. In this analysis, we compressed the HPO tree down to 250 phenotype groups. Each patient is then represented by a 250-length vector, where the  $i^{th}$  number represents the count of HPO terms that this patient exhibits within the  $i^{th}$  phenotype group.

### Running UMAP/HDBSCAN

Using these vectors, we can then perform Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, enabling us to visualize our patients in two-dimensional space. In this two-dimensional space, we then perform Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to assign cluster IDs to each of the patients. Of the 526 patients, we were able to successfully cluster 498 samples together across 6 separate phenotype clusters (shown below). The resulting 98 samples which did not cluster with the other groups were omitted from the remainder of the visualizations.



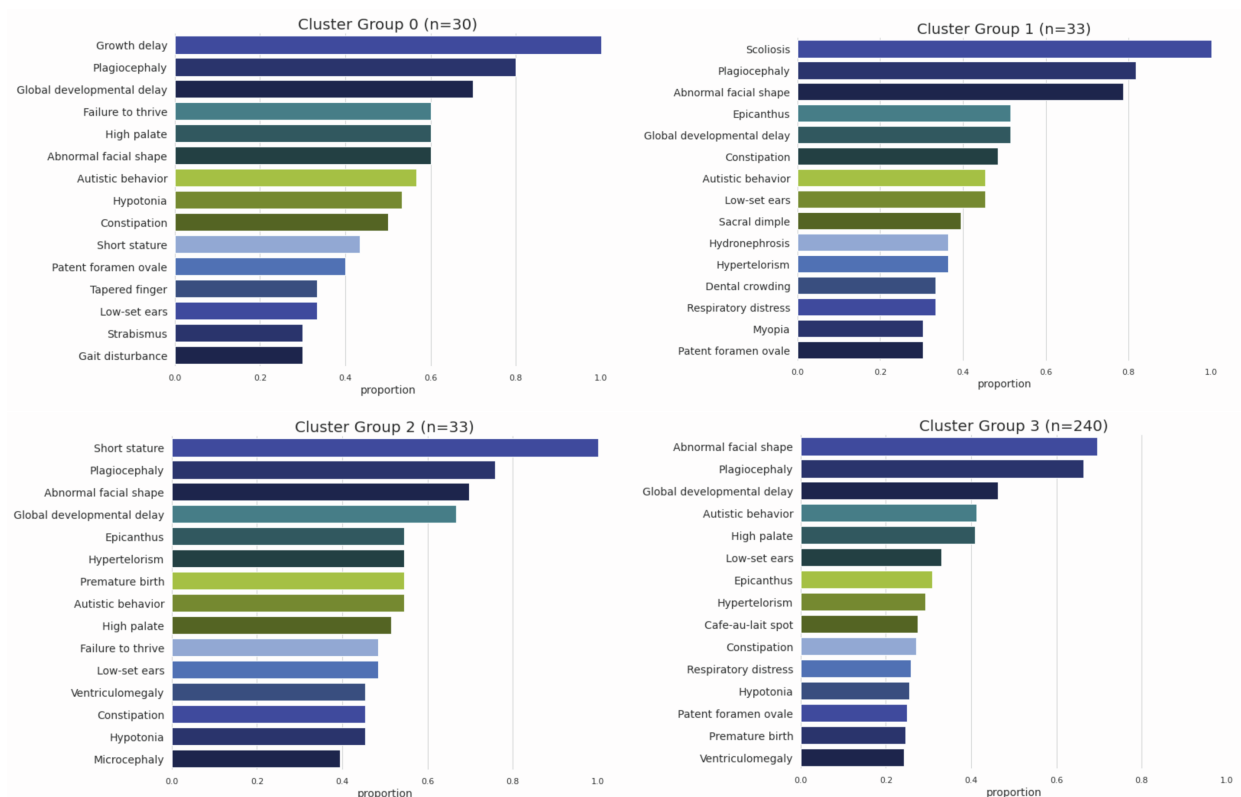
### Basic cluster stats

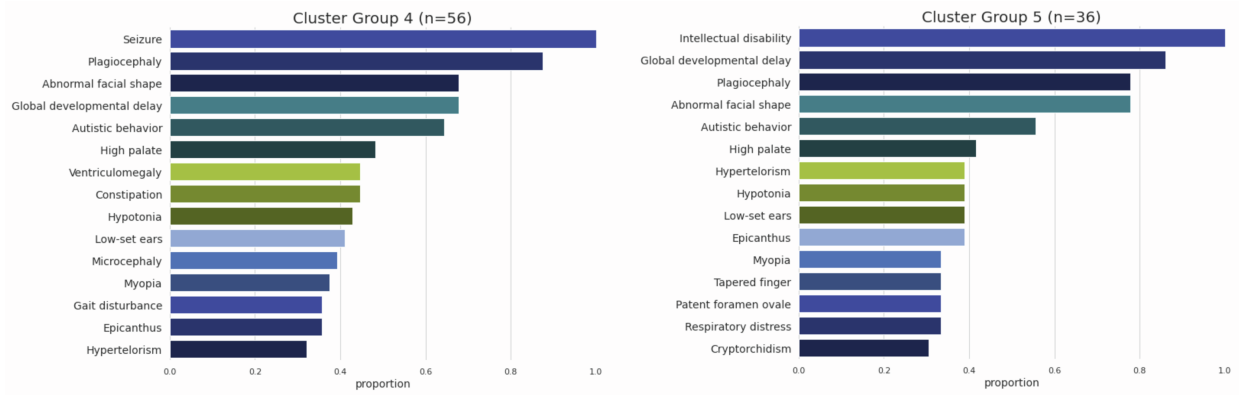
In looking at each cluster, we show the proportion of females, mean and standard deviation of age. As a reminder, cluster -1 pertains to patients which were not assigned to a cluster using our HDBSCAN algorithm.

Cluster	N	prop_female	mean_Age	std_age
-1	98	0.45	4.98	6.09
0	30	0.40	3.54	3.45
1	33	0.39	4.93	4.94
2	33	0.36	6.75	8.13
3	240	0.39	5.57	7.24
4	56	0.32	6.51	6.64
5	36	0.25	7.51	8.20

### Phenotype group frequencies per cluster

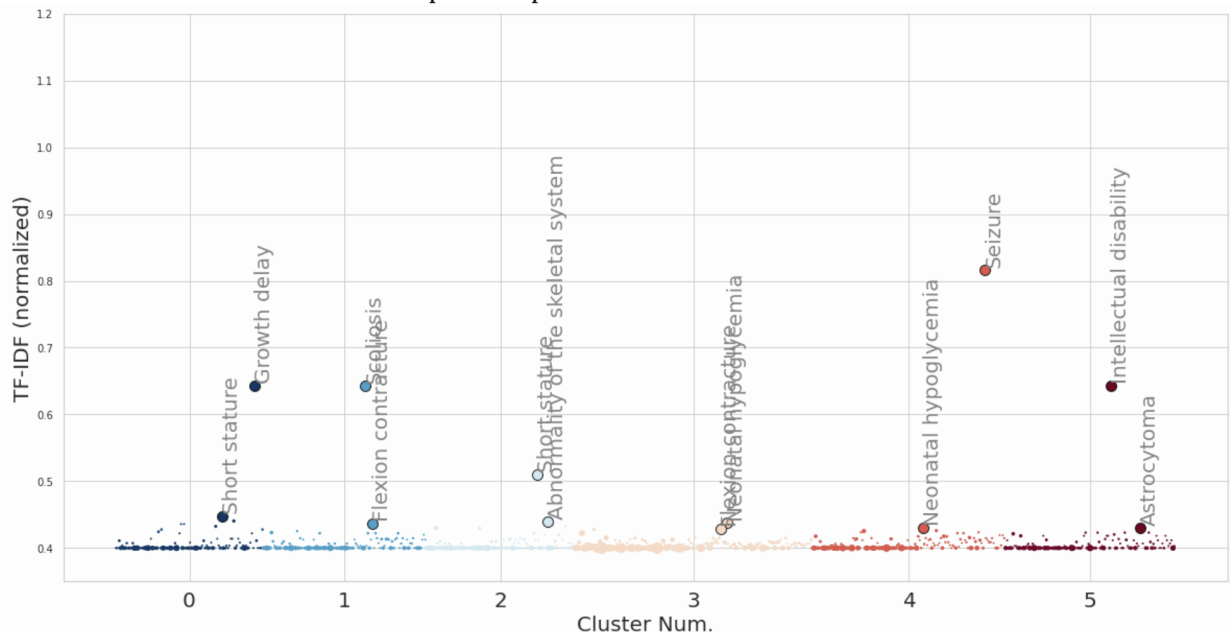
To further characterize each cluster, we generated a series of waterfall plots showing the frequencies of phenotype groups that each set of patients exhibit. Each term in the waterfall plot represents the most-used term from the phenotype group in our in-house dataset. For instance, in group 1, 100% of patients exhibit a phenotype term that resides in the same phenotype group (and is therefore most similar to) “Scoliosis”. This allows us to quickly evaluate the defining feature for several groups (e.g. – Group 4 appears to predominantly exhibit “seizure”-related phenotypes).





### Can we better characterize each cluster?

In looking at the waterfall plots above, we see that “plagiocephaly” appears in every cluster, indicating that it is not a particularly differentiating factor. To define the most differentiating phenotypes within each phenotype cluster, we relied on a statistical technique, Term Frequency-Inverse Document Frequency (TF-IDF). This method is commonly used in document analysis to define the most important words within a document, while down weighting terms that occur in many documents. In our case, TF-IDF would essentially ignore “plagiocephaly” as a differentiating phenotype given that it exists in all clusters. The image below shows the normalized TF-IDF values for all terms in each cluster. The top terms per cluster have been annotated.



While we were able to identify the above clusters, these patients did not share similar genotypes in most all cases, thus we can't make any conclusions about specific genotype/phenotype relationships with our current number of samples. Instead, these results suggest that mutations in these pleiotropic genes affect various organ systems as expected, but a larger number of patients with mutations in each will be necessary to fully characterize the phenotypic impact of mutations in each.



## Supplementary Note 3: Case Reports

### Clinical synopses of probands with a recurrent *de novo* variant in *RARA*

#### Proband 1:

Patient is 21-month-old male

Birth: full term, 8 lbs 7 oz, 20 inches long, head size “larger”

Prenatal: no concerns. Negative noninvasive prenatal testing.

Newborn: Difficulty feeding due to ankyloglossia, poor weight gain. Noted to have bilateral rocker bottom feet (found to be vertical tali) and abnormal head shape (scaphocephaly). Diagnosed with sagittal craniosynostosis.

- CT head: Fusion of the anterior sagittal suture. The remainder of the sutures are patent.

#### Diagnoses:

- Bilateral vertical tali with calcaneovalgus deformities
- Partial sagittal craniosynostosis (HC at 17 months 48 cm, 78<sup>th</sup>%)
- Poor weight gain
- Failure to thrive (weight at 19 months 7.79 kg, 0.05 percentile, z score -3.27)
- Short stature (length at 19 months was 74 cm, 0.02%, z score -3.49)
- Hypotonia
- Unilateral cryptorchidism with hernia
- Ankyloglossia

Infancy: poor weight gain necessitating NG tube feedings. Hypotonia and motor delay. Hospitalized at 3m of age due to poor weight gain. Sagittal suturectomy performed in early infancy. NG tube was feeding was stopped at 16 months of age. He continues to have very slow weight gain despite very high calorie diet and maintains a thin body habitus despite this caloric intake.

Ortho: vertical tali bilaterally. Followed by orthopedics and has had casting and surgical repair. Also initially had some overlapping fingers and clenched fists, which has now resolved, though fingers are somewhat long in appearance.

Exam: thin body habitus, elongated head, full cheeks, accessory nipple on right, rocker bottom feet, hypotonia

Development: motor delay due to casting and surgeries on feet. Seems on track socially and verbally. Feeding is improving and takes all feeds orally.

Testing: SNP microarray, normal male

WES: *de novo* *RARA* variant; c.865G>A, p.Gly289Arg (classified as variant of unknown significance)

## Proband 2:

This child was born at 35-4/7 weeks' gestation to a 27-year-old G2P1-2 via urgent C-section. During the pregnancy, his mother was noted to have pelvic kidneys and she underwent nonstress tests twice per week throughout the third trimester, which were all reassuring. She went into labor at 35 weeks and 4 days. When she arrived at the hospital, the baby continued to be in breech presentation with oligohydramnios and therefore a C-section was performed. She noted that fetal monitoring was normal during the C-section itself and during her monitoring time after arrival at the hospital. At birth, he did not cry. He had initial Apgars of 1 at 1 minute, 4 at 5 minutes and 8 at 10 minutes. During this time, he was intubated and placed on an oscillator. He weighed 4 pounds 13 ounces at birth. He was noted at birth to have several anomalies, including clubfeet, overlapping toes and a head shape consistent with persistent breech presentation and an occipital shelf. He was noted to have small ears at that time. He had difficulty weaning off of the ventilator and it was ultimately determined that he needed repeat head imaging. Initial head ultrasound at birth was normal; however, repeat imaging demonstrated a perinatal left MCA stroke that was quite extensive. Since that time, he has had a significant volume loss as well as development of seizures, initially spasms followed by tonic seizures, dysphagia, developmental delay and dependence on tracheostomy with ventilator. His ventilator requirement is likely multifactorial, including chronic respiratory failure secondary to upper airway obstruction, pulmonary hypoplasia, history of tracheal stenosis status post dilation, and severe tracheobronchomalacia. At the age of 3 years he was using the ventilator overnight but it was not needed during the day. He has a diagnosis of spastic hemiplegia, and has global delays. His primary method of ambulation at age 3 years was scooting, and was working on learning to walk independently. He could babble and his primary method of communication was signs and pointing. He was very social and emotionally bonded with his family.

His head shape became progressively brachycephalic in infancy, and CT scan demonstrated bilateral coronal craniosynostosis and bilateral intraoccipital synchondroses (Figure 3). He has not needed cranial vault surgery because his MCA stroke resulted in volume loss (Figure 3). Other medical issues include renal dysplasia with cysts and chronic kidney disease, as well as undescended testes and persistent asymmetric overlapping toes. He was also noted to have esotropia. Moderate to severe conductive hearing loss was noted. He also had an abnormal pattern of tooth eruption.

SNP array was completed and was normal. Trio exome sequencing was subsequently done through GeneDx. He was found to have a *de novo* variant, c.865G>A, p.Gly289Arg, in exon 7 of *RARA*. At the time of the clinical testing, this variant was classified as a variant of uncertain significance.