**Recognition of a critical functional domain and improved *PHOX2B* missense variant interpretation using *in silico* prediction tools**

Andy Drackley†1,2

Email: [adrackley@luriechildrens.org](mailto:adrackley@luriechildrens.org)

ORCID iD: 0000-0003-1822-1728

Andrew D. Skol†1,3

Email: [askol@luriechildrens.org](mailto:askol@luriechildrens.org)

ORCID iD: 0009-0005-7530-8127

Casey M. Rand3,4

Email: [crand@luriechildrens.org](mailto:crand@luriechildrens.org)

ORCID iD: 0000-0003-1077-8978

Debra E. Weese-Mayer3,4,5

Email: [dweese-mayer@luriechildrens.org](mailto:dweese-mayer@luriechildrens.org)

ORCID iD: 0000-0001-6061-220X

Kai Lee Yap\*1,6

Email: [klyap@luriechildrens.org](mailto:klyap@luriechildrens.org)

ORCID iD: 0000-0003-4496-5464

† *Authors contributed equally*

*\* Corresponding author*

**Affiliations**

1. Department of Pathology & Laboratory Medicine, Molecular Diagnostics Laboratory, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL, USA
2. Division of Genetics, Genomics and Metabolism, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL, USA
3. Stanley Manne Children’s Research Institute, Chicago, IL, USA
4. Department of Pediatrics, Division of Autonomic Medicine, Center for Autonomic Medicine in Pediatrics (CAMP), Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL, USA
5. Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
6. Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

**List of abbreviations:**

CCHS: congenital central hypoventilation syndrome

PARM: polyalanine repeat mutation

NPARM: non-polyalanine repeat mutation

ACMG: American College of Medical Genetics and Genomics

AMP: Association for Molecular Pathology

CI: confidence interval

LRT: likelihood ratio test

AUC: area under the curve

ROC: receiver operating curve

Delins: deletion-insertion

VUS: variant of uncertain significance

**DECLARATIONS**

**Ethics approval and consent to participate**

The study was approved by Ann & Robert H. Lurie Children’s Hospital of Chicago’s IRB (IRB #2013-15273).

**Consent for publication**

Not applicable

**Availability of data and materials**

All non-patient data generated or analyzed during this study are included in this published article and its supplementary information files. Patient-specific data may be available from the authors upon request.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

CMR receives funding from the Chicago Community Trust Foundation *PHOX2B* Patent Fund.

NIH R03TR003869 (Weese-Mayer PI)

**Contributions**

Conceptualization: A.D., K.L.Y.; Methodology: A.D., K.L.Y., A.D.S.; software: A.D.S.; validation: A.D., A.D.S.; formal analysis: A.D.S.; investigation: A.D., A.D.S.; resources: A.D., A.D.S., C.M.R., D.E.W.-M., K.L.Y.; data curation: A.D., A.D.S., C.M.R., D.E.W.-M., K.L.Y.; writing – original draft: A.D., A.D.S.; writing – review & editing: A.D., A.D.S., C.M.R., D.E.W.-M., K.L.Y.; supervision: K.L.Y.; project administration: A.D. All authors read and approved the final manuscript.

**Keywords**

*PHOX2B*;Congenital central hypoventilation syndrome; CCHS; *In silico* predictions; REVEL; CADD; BayesDel; AlphaMissense; homeodomain

**ABSTRACT** (350-word limit – currently at 337)

**Background (65 words)**

Pathogenic heterozygous variants in *PHOX2B* are associated with congenital central hypoventilation syndrome (CCHS). Most *PHOX2B* missense variation is difficult to interpret due to the lack of clinical and experimental evidence for individual variants and limited understanding of their effect on protein function. An improved ability to distinguish pathogenic missense variants from benign ones could dramatically aid in the diagnosis and management of individuals with CCHS.

**Methods (105 words)**

All *PHOX2B* missense variants in the literature and public and private databases were extracted and given consensus pathogenicity classifications. CADD, REVEL, BayesDel, and AlphaMissense scores were queried for all variants. We performed a weighted logistic regression in a multiple imputation framework to determine if and at what score thresholds these prediction tools can be used for application of the ACMG-AMP guidelines’ PP3/BP4 criteria at various strength levels and calculated the area under the curve for each tool. We then analyzed the positional distribution of Pathogenic/Benign variants as well as the four tools’ gene-wide pathogenicity predictions to assess potential hotspots or critical functional domains in *PHOX2B*.

**Results (102 words)**

All four tools can reach “strong” level for BP4, whereas CADD, REVEL, and BayesDel can reach “strong” level for PP3, consistent with ClinGen recommendations for regular use. By a small margin, BayesDel was the strongest prediction tool, with the greatest area under the curve (0.992) and second fewest “Uncertain” variants with a prediction score in the indeterminate range. Additionally, there is a clustering of pathogenic, and lack of benign, variants in the region encoding the PHOX2B homeodomain, with all predictiontools supporting the putative functional importance of the homeodomain. The PM1 criterion should therefore be considered for in-frame variants altering the homeodomain.

**Conclusions (65 words)**

BayesDel, with CADD and REVEL only slightly behind, is the highest performing pathogenicity prediction tool for *PHOX2B* missense variants and may be preferentially utilized for their assessment. In combination with the support for application of the PM1 criterion for variants in the PHOX2B homeodomain, our improved ability to differentiate pathogenic from benign variants will allow for fewer VUS classifications in favor of more conclusive results.

**BACKGROUND**

Congenital central hypoventilation syndrome (CCHS, MIM #209880) is a rare autonomic disorder characterized primarily by alveolar hypoventilation, abnormal to absent control of breathing, and autonomic nervous system dysregulation (Weese-Mayer et al., 2010). Heterozygous pathogenic variation involving the highly conserved Paired-Like Homeobox 2B (*PHOX2B,* MIM #603851) gene is the dominant known genetic etiology of CCHS. *PHOX2B* pathogenic variants are disease-defining in the CCHS phenotype including additional anomalies such as cardiac arrhythmias, diminished pupillary response to light, and abnormalities associated with maldevelopment of neural crest-derived structures including Hirschsprung disease or severe constipation, neural crest tumors, and esophageal dysmotility (Weese-Mayer et al., 2010; Weese-Mayer et al., 2021).

Identification of a *PHOX2B*pathogenic variant is an essential component of the CCHS diagnostic criteria and therefore the ability to adequately evaluate the clinical significance of identified variants is critical for the diagnosis and management of this patient population (Weese-Mayer et al., 2010). The predominant and best characterized type of disease-causing variants are polyalanine repeat mutations (PARMs), expansions of the C-terminal polyalanine tract in exon 3 of *PHOX2B*. Less common are the non-polyalanine repeat mutations (NPARMs), estimated to represent approximately 10% of CCHS cases (Zhou et al., 2021).Pathogenicity and genotype-phenotype correlations are well characterized for both PARMs and the truncating and frameshift NPARMs (Weese-Mayer et al., 2003; Matera et al., 2004; Weese-Mayer et al., 2017; Zhou et al., 2021). Accurate interpretation of the well characterized *PHOX2B* genotypes has enabled disease prognostication and prediction of the associated phenotypic features, which has improved and personalized management and disease surveillance of CCHS patients.

Despite these advances, there is a striking lack of information to support interpretation of *PHOX2B* missense variants due to the unpredictable impact of missense variants on gene function. The assessment of the clinical significance of *PHOX2B* missense variants remains very challenging due to their rarity, the scarcity of relevant clinical and experimental functional data, and that a significant proportion of variants are unique to a single individual or family (Zhou et al., 2021). Consequently, the vast majority of detected *PHOX2B* missense variants are classified as variants of uncertain significance (VUS) due to insufficient evidence for or against pathogenicity (Richards et al., 2015; Landrum et al., 2018), leading clinical providers to delay diagnosing CCHS and introducing appropriate management, despite the clinical phenotype of CCHS.

Numerous computational pathogenicity prediction algorithms, typically focusing on structural and/or biochemical properties of the reference and alternative amino acids, the degree of evolutionary conservation at the nucleotide and amino acid level, among other components, have been developed to aid in the interpretation of missense variants. Ensemble or meta-prediction tools that combine several algorithms to produce a single numerical summary have been shown to be more effective at predicting variant pathogenicity compared to a single algorithm or *ad hoc* combinations of individual algorithms (Tian et al., 2019; Pejaver et al., 2022). Impressively, and consequentially, Pejaver et al. (2022) found that several prediction tools are sufficiently accurate, that in many instances they are able to increase the strength of evidence for the PP3/BP4 criteria (*in silico algorithms support a pathogenic or benign impact,* respectively) from the “supporting” level assigned in the original ACMG/AMP guidelines (Richards et al., 2015; e.g., from “supporting” level to “moderate” or “strong”). Tian et al. and Pejaver et al. evaluated numerous tools’ performances for variants across thousands of genes and found that all tools are not equally efficacious for all genes and diseases. Indeed, Pejaver et al. (2022) endorsed completion of similar evaluations to assess and calibrate the different prediction tools for individual genes (and diseases) of interest.

With scant clinical and experimental evidence for evaluation of *PHOX2B* missense variants, we sought to determine the most appropriate *in silico* pathogenicity prediction tool. We therefore endeavored to calibrate *PHOX2B-*specific score thresholds for these tools and to identify and recommend an individual tool for variants in *PHOX2B.* Our goal is to improve interpretation of *PHOX2B* missense variants’ pathogenicity potential, and ultimately aid in the clinical management of patients potentially affected with CCHS.

**METHODS**

**Variant Search and Curation of the Dataset**

Queries of public databases (2023-07-07) for all reported *PHOX2B* missense variants were performed including ClinVar, *PHOX2B* Leiden Open Variation Database, DECIPHER, the biomedical literature, and the large private internal patient database managed by authors D.E.W.-M. and C.M.R.; These queries also extracted all available clinical information (i.e., phenotype, segregation data) for each variant. Variants in gnomAD v4.1.0 and their filtering allele frequency data were obtained following its release. Deletion-insertion (delins) variants resulting in amino acid substitutions that are impossible via a single nucleotide variant (SNV), and thus do not have pre-calculated prediction scores, were excluded from our assessment (e.g., c.679\_680delinsTT:p.Ala227Leu); delins variants resulting in amino acid substitutions also possible via SNV (e.g., c.432\_433delinsTC:p.Trp145Arg compared to c.433T>A:p.Trp145Arg) and thus possessing pre-calculated scores were retained. Different SNVs resulting in identical amino acid substitutions (e.g., c.564G>T:p.Lys188Asn and c.564G>C:p.Lys188Asn) and thus identical prediction scores were considered equivalent, with only one retained. Variants in gnomAD annotated as non-coding on the MANE Select (Morales et al., 2022) transcript (NM\_003924.4/ENST00000226382.4) but impacting a minor transcript were also excluded from analyses.

**Assigning Variant Pathogenicity**

*Predictive value of in silico variant effect prediction tools.*

We sought to assess the predictive value of four well regarded *in silico* variant effect prediction tools. Nearly all *PHOX2B* missense variants have insufficient evidence to reach a (likely) pathogenic classification according to the ACMG-AMP guidelines due to insufficient literature specific to a unique variant. To create truth sets and enable critical evaluation of the performance of *in silico* predictive algorithms, we developed a framework for consensus classifications of variants’ presumed clinical significance by a group of clinicians and laboratorians with considerable experience in *PHOX2B* and CCHS; this framework is heavily influenced by the ACMG-AMP guidelines and considers the comprehensive phenotypic features and family testing of the patients for variant classification. Tamana et al. (2022) established a similar scheme for assessing variants in hemoglobinopathy genes in which, like *PHOX2B*, a huge proportion of variants are unique to a single family.

*Benign variants*

To define variants as Benign, we first sought to determine the population maximum tolerated allele frequency for CCHS (Whiffin et al, 2017). The factors used in this calculation, which were all conservatively set in order to minimize the risk of inappropriately mis-classifying variants as Benign, include the following: a CCHS prevalence of 1/100,000, more common than the highest current estimate of its prevalence (Shimokaze et al., 2015); a penetrance of 0.40 (40%) to allow for incomplete penetrance and variable expressivity or somatic mosaicism resulting in subclinical disease (Weese-Mayer et al., 2003; Bachetti et al., 2011); a genetic heterogeneity of 0.80 to account for the CCHS cases without diagnostic *PHOX2B* variants identified; and an allelic heterogeneity of 0.10, which represents the approximate proportion of CCHS cases caused by all NPARMs and is therefore highly conservative for consideration of individual missense variants. This yielded a maximum tolerated allele frequency of 0.0001% (0.000001). Variants’ gnomAD GroupMax filtering allele frequency (FAF) (95% CI) estimates (the combined exome and genome [joint] datasets) were then compared to the calculated disease-specific frequency threshold as recommended by Gudmundsson et al. (2022). All variants with GroupMax FAF that exceed the disease-specific frequency threshold frequency (0.0001%) were thus classified as Benign.

*Pathogenic variants*

Variants were classified as Pathogenic through evaluation of the following factors: the variant is present in one or more affected probands, with more weight given if the proband(s) are documented to have a formal CCHS diagnosis or a multi-systemic *PHOX2B*-related phenotype (i.e., CCHS, Hirschsprung disease, neural crest tumors, etc.); family testing demonstrates co-segregation of the variant with disease in family members, or the variant is found to be *de novo* in the absence of a relevant family history; the variant is extremely rare or absent in gnomAD v4.1.0; and if there is experimental evidence supporting a deleterious impact of the variant.

*Uncertain variants*

Variants that did not meet criteria for Benign or Pathogenic, had insufficient or conflicting evidence, or consensus among evaluators could not be reached, were classified as Uncertain.

**Computational Prediction Tools**

CADD (Kircher et al., 2014), REVEL (Ioannidis et al., 2016), BayesDel (Feng, 2014) (noAF, without incorporation of population allele frequency data to avoid double-counting of this evidence), and AlphaMissense (Cheng et al., 2023) scores were collected for all variants. The former three have been shown to perform well in multiple large-scale evaluations, with REVEL and BayesDel considered particularly effective (Tian et al., 2019;Pejaver et al., 2022), and AlphaMissense, the most recently developed pathogenicity prediction methodology based on AlphaFold’s (Jumper et al., 2021) predicted protein structures, is a promising tool. Importantly, all tools are free to use for clinical application with varying levels of bioinformatic integration required.

**Statistical Analyses**

*Identifying prediction tool score thresholds*

The method proposed by Pejaver et al. (2022) uses an empirical approach to estimate the posterior probability of pathogenicity and benignity conditional on prediction score to set thresholds for the strength of evidence that a variant is pathogenic or benign. These thresholds are based on finding the prediction score values at which the estimated “positive likelihood ratio” (lr+) achieves a set of predetermined values (posterior probabilities of pathogenicity (or benignity) of 0.98 (very strong), 0.61 (strong), 0.21 (moderate) and 0.10 (supporting) (see Table 1 of Pejaver et al.). Identifying empirical estimates of these thresholds, as Pejaver et al. did, protects against biases that can be introduced through parametric modeling. However, in our study, we are focusing on a single gene and disease (*PHOX2B* and CCHS) and variant type (missense) which greatly limits the number of pathogenic and benign variants we can use to estimate pathogenic/benign thresholds. Because using such a small sample of variants would result in extremely conservative boundaries, as an alternative we used weighted logistic regression in a multiple imputation framework. Logistic regression innately models odds ratios, an important component of lr+, and multiple imputation allows us to extract information from the uncertain variants to improve the accuracy and precision of our parameter estimates. The initial logistic regression weights are used to account for the unbalanced sampling of variants in our dataset. Weights for pathogenic and benign variants were set to PPath / P­Path\_samp and (1 – PPath) / (1 – PPath\_samp), respectively, where PPath and PPath\_samp are the estimated proportion of pathogenic missense variants in the population and our sample (omitting unclassified and synonymous variants). Twenty-five imputation datasets were generated by assigning Pathogenic/Benign status to Uncertain variants using the probability estimates from the initial weighted logistic regression parameter estimates. For each imputed dataset, weighted logistic regression was preformed where weight = 1 for the variants with known pathogenicity/benignity, and *2 x |0.5* – pi,path| for the ith Uncertain variant. We use Rubin’s rule to incorporate the results from the multiple analyses into the parameter estimates and their covariance matrix (Rubin, 2009). Using these results, we estimate the odds of pathogenicity conditional on prediction score, and the odds of a variant being pathogenic, both of which are used to calculate the positive likelihood ratio. Specifically,

As suggested by Pejaver et al., we used the 95% lower bound of the estimate of the odds in lieu of the odds directly. This lr+ lower bound was used to identify the prediction score (*s*) satisfying the pathogenicity class values in Table 1 of Pejaver et al. is:

The same formula is used to determine benign prediction score thresholds, except that we invert the odds and use the upper rather than lower confidence bound of the odds (e.g., change the sign in front of 1.64). We used an estimate of 0.044 x (1+Pmissense) for PPath based on the Pejaver et al. (2022) estimate of PPath and adjusted for our exclusion of non-missense variants. We performed one-dimensional root finding using the *uniroot* function in R (Brent, 2013) to find the prediction score (*s*) that solved for *lr+*­­*­­­LB* for the likelihood ratios given in Table 1 of Pejaver et al.

*Exploring the predictive value of pathogenicity predictions*

We calculated the area under the curve (AUC) of the receiver operator curve (ROC) using 80% of the data as a training set. We accounted for our multiple imputation method and the modest number of variants available to build testing and training sets by taking 200 random training and testing samples, generating a single imputed dataset from each, and calculating the AUC based on the testing set. AUC was calculated using the *ROCR* package in R.

An ROC was calculated and plotted using a single imputed dataset from a round of our multiple imputation procedure. This is a good approximation to the true ROC curves based on the similarity of the AUC estimates derived from these data and those from our more rigorous procedure described above.

We used the prediction tool score thresholds described earlier to demonstrate the number of variants in the Uncertain set that could be assigned different PP3/BP4 strength levels.

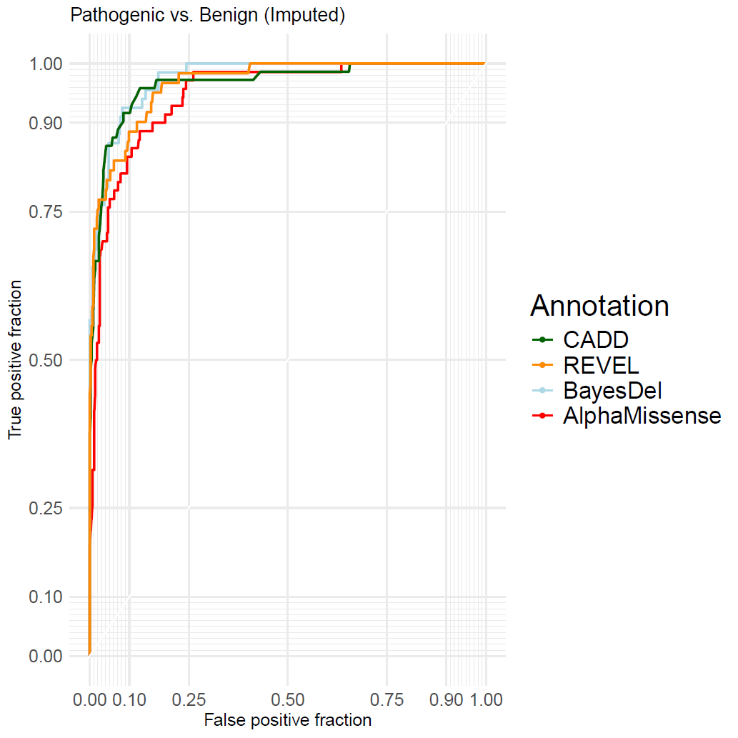
**RESULTS**

The dataset comprised 506 missense variants. Eighty-one variants (16.0%) were classified as Benign, 33 variants (7.7%) were classified as Pathogenic, and 392 variants (65.2%) were classified as Uncertain. Each tool’s average score values for the Pathogenic and Benign sets differed significantly (Table 1).The AUC for CADD, REVEL, BayesDel, and AlphaMissense (Figure 1) were 0.963, 0.982, 0.9992, and 0.961, respectively.

Table 1: Mean pathogenicity prediction scores and areas under the receiver operating curve for Benign and Pathogenic variants in *PHOX2B*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool** | **Benign** | **Pathogenic** | **p-value (difference)** | **AUC** |
| CADD | 19.00 | 29.00 | 2.0x10-21 | 0.963 |
| REVEL | 0.36 | 0.90 | 3.2x10-30 | 0.982 |
| BayesDel | -0.14 | 0.49 | 2.9x10-29 | 0.992 |
| AlphaMissense | 0.30 | 0.96 | 4.7x10-29 | 0.961 |

**Figure 1:** ROC curves showing the ability of the four pathogenicity prediction tools to differentiate between Benign and Pathogenic *PHOX2B* variants (based on all variants from an imputed dataset).



Pejaver et al. (2022) suggest using the posterior odds of pathogenicity (or benignity) to categorize a variant’s strength of evidence for pathogenicity (or benignity). Logistic regression is well parameterized to determine the strength levels, as its regression parameter represents the change in log odds of a variant’s pathogenicity resulting from a one unit increase in prediction score. The change in the log odds of pathogenicity per prediction score unit is given Table 2. The prediction score thresholds for all PP3 and BP4 strength levels are provided in Table 3.

The large AUC values, highly significant regression parameters (Table 2), and level of separation between Pathogenic and Benign variants based on prediction scores (Figure 2), all point to these prediction tools performing very well as classifiers.

After determining the classification thresholds, we next assigned each variant in the Uncertain set into a BP4/PP3 evidence bin. Table 4 shows the number of Uncertain variants falling into different strength levels by each of the tools. In general, classifications were very similar across prediction tools. Of the 392 variants of uncertain significance, 22 had the same classification across all four tools, 185 shared a classification across 3 tools (for 135 of these variants the disagreeing tool’s prediction strength was one level away from consensus (e.g. supporting versus moderate)). Even among the remaining 165 variants that shared only a single strength level across tools, most (124) had strength levels across adjacent levels (e.g. supporting, moderate, strong). There was, however, variation across tools in the proportion of Uncertain variants predicted to be pathogenic (any level); CADD predicted 17.9% (70/392) to be pathogenic, REVEL 12.8% (50/392), BayesDel 11.7% (46/392), and AlphaMissense 26.3% (103/392).

**Table 2:** Estimated change in the log(OR) as a function of pathogenicity prediction tool scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Prediction Tool** | **Coef** | **SE (Coef)** | **p-value** | **AUC** |
| CADD | 1.10 | 0.17 | 1.1x10-6 | 0.963 |
| REVEL | 15.44 | 2.45 | 1.6x10-6 | 0.982 |
| BayesDel | 15.82 | 2.71 | 5.0x10-6 | 0.992 |
| Alpha  Missense | 11.31 | 2.08 | 1.4x10-5 | 0.961 |

**Table 3:** Estimated score threshold intervals for the four pathogenicity prediction tools evaluated in this study as they relate to the different pathogenic and benign strength levels

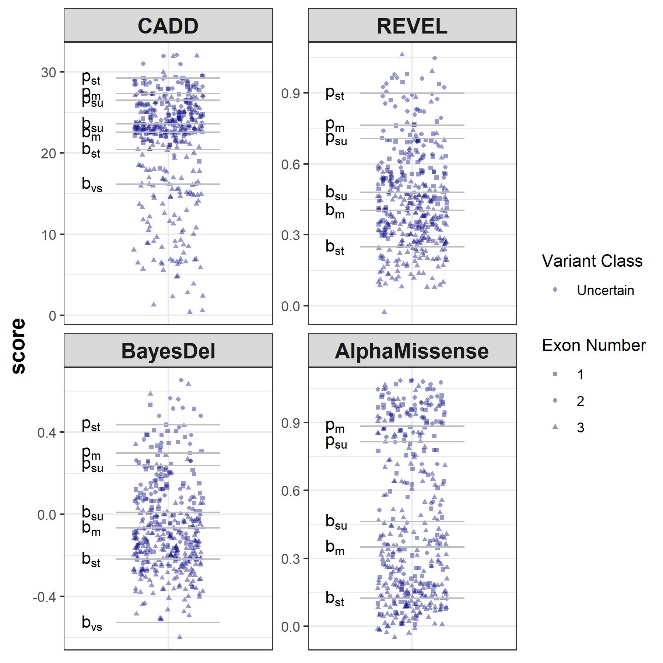
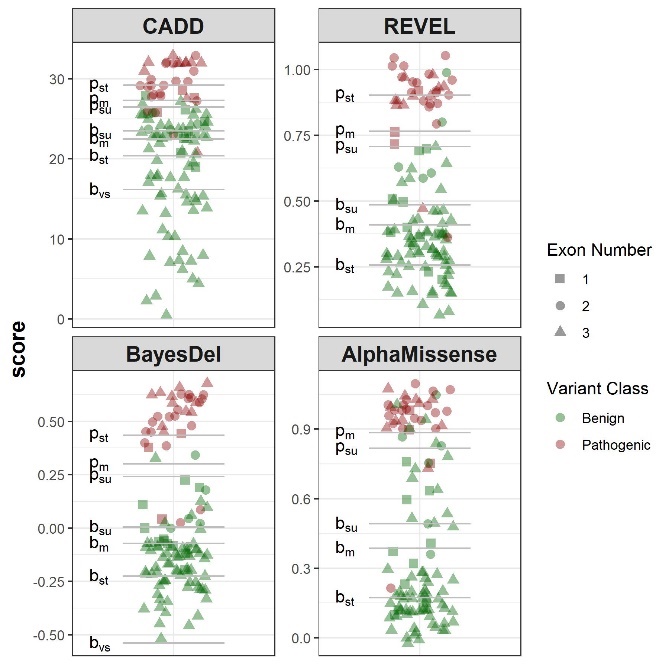
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prediction Tool** | **Benign (BP4)** | | | | **Pathogenic (PP3)** | | | |
| Very Strong | Strong | Moderate | Supporting | Supporting | Moderate | Strong | Very Strong |
| CADD | ≤ 16.17 | (16.17, 20.41] | (20.41, 22.51] | (22.51, 23.54] | [26.49, 27.33) | [27.33, 29.24) | ≥ 29.24 | — |
| REVEL | — | ≤ 0.26 | (0.26, 0.41] | (0.41, 0.48] | [0.71, 0.77) | [0.77, 0.90) | ≥ 0.90 | — |
| BayesDel | ≤ -0.52 | (-0.52, 0.22] | (0.22, -0.07] | (-0.07, 0.01] | [0.25, 0.30) | [0.30, 0.43) | ≥ 0.43 | — |
| Alpha  Missense | — | ≤ 0.15 | (0.15, 0.37] | (0.37, 0.48] | [0.81, 0.88) | ≥ 0.88 | — | — |

Table 3: ‘—' implies that the given tool did not meet the posterior probability (likelihood ratio) threshold. Parentheses represent exclusion of the end value; brackets represent inclusion of the end value.

**Table 4:** Number of predicted pathogenic and predicted benign variants at different strengths among the set of Uncertain variants (n = 392) based on predicted probability

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prediction Tool** | **Benign (BP4)** | | | | **Pathogenic (PP3)** | | | | **Indeterminate** |
| Very Strong | Strong | Moderate | Supporting | Supporting | Moderate | Strong | Very Strong |
| CADD | 79 | 28 | 51 | 63 | 23 | 34 | 13 | — | 101 |
| REVEL | — | 67 | 131 | 50 | 16 | 22 | 12 | — | 94 |
| BayesDel | 1 | 91 | 132 | 51 | 12 | 21 | 13 | — | 71 |
| Alpha Missense | — | 104 | 96 | 28 | 22 | 81 | — | — | 61 |

**Figure 2:** Prediction scores with pathogenic and benign evidence levels for Benign and Pathogenic (Left) and Uncertain (Right) variants. (p/b: pathogenic / benign; su:supporting, m:moderate, st: strong, vs: very strong)



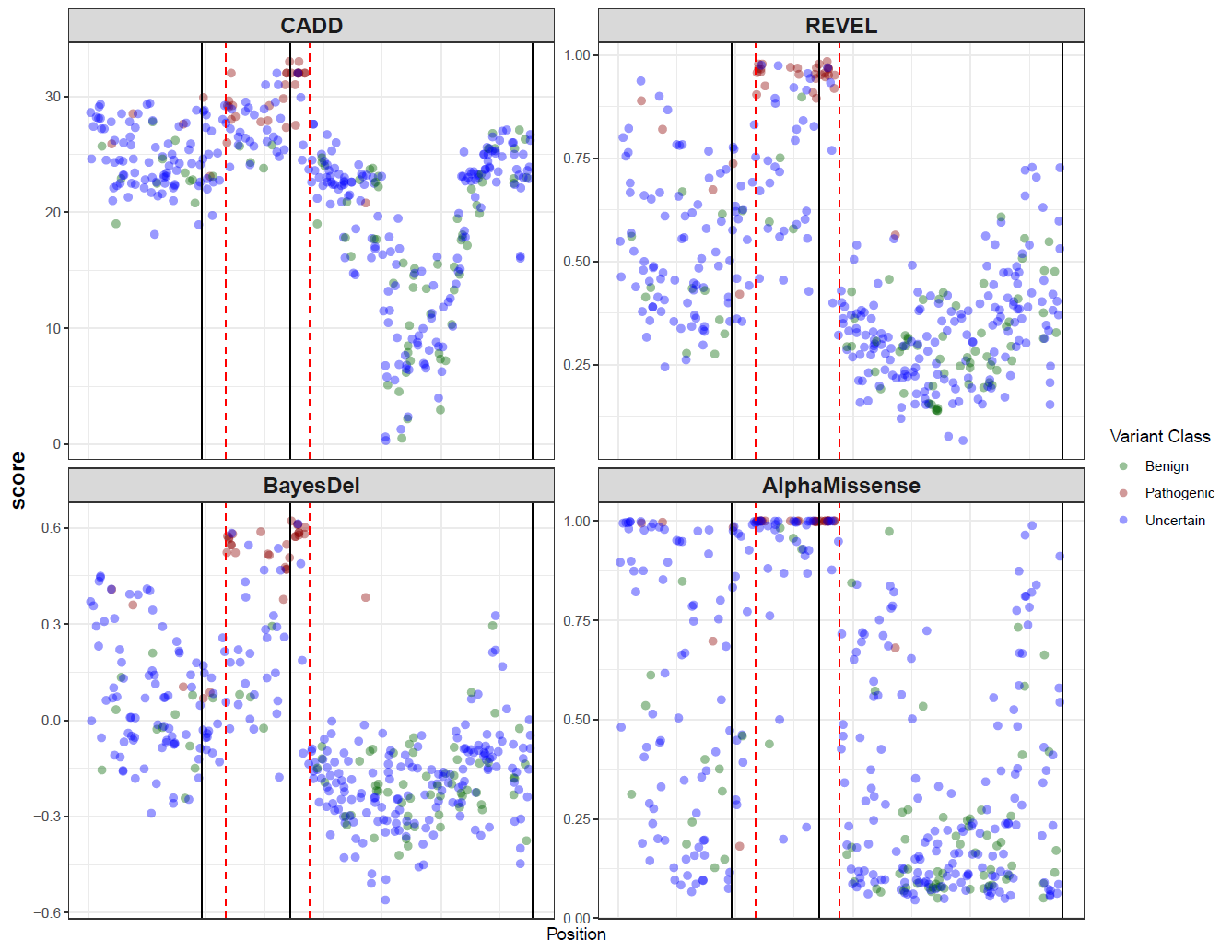
**Clustering of Pathogenic Variants in the Homeodomain**

Notably, 27/34 (79.4%) of the variants classified as Pathogenic in this study were at positions between p.Arg99 and p.Arg154 in *PHOX2B*. Codons 98-157 are responsible for encoding the PHOX2B homeodomain (Pattyn et al., 1997). Examination of the distribution of the 81 Benign variants demonstrated that only four (4.9%) are in the homeodomain (p.Ala108Thr, p.Arg115Lys, p.Ile125Val, p. Leu131Val), none of which occur at positions at which there also exists a known Pathogenic variant. Upon noticing pathogenic variant clustering in the homeodomain, we sought to determine whether these four tools would support this evidence that the homeodomain is a critical functional domain characterized by an enrichment of pathogenic variation. All four tools have significantly higher mean prediction scores (p<10-8) for all unique theoretical missense variants at positions in the homeodomain compared to the remaining theoretical variants in the rest of *PHOX2B* (Table 5)*.*

***Table 5:*** Average pathogenicity prediction scores for all unique theoretical missense variants within the homeodomain compared to those for the rest of *PHOX2B*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Prediction Tool** | | | |
| **Region** | CADD | REVEL | BayesDel | AlphaMissense |
| Homeodomain (codons 98-157) | 29.89 | 0.823 | 0.341 | 0.965 |
| Non-homeodomain  (codons 2-97, 158-314 | 23.85 | 0.433 | -0.070 | 0.465 |

**Figure 3:** Distribution of pathogenicity prediction scores for all variants in our study (black line denotes exon boundaries (exon 1 on left); red dotted lines demarcate the homeodomain encoded by amino acid positions 98-157)



**DISCUSSION**

As the usage of the ACMG/AMP variant classification guidelines have evolved, it is sufficiently inadequate to consider variant classifications as arbitrary categorization based on rule sets. The usage of a likelihood ratio (LR) to define a continuum of overall variant ‘pathogenicity’ is likely necessary (Tavtigian et al., 2018). With respect to *in silico* predictions and the PP3/BP4 criteria, trichotomized (benign vs. pathogenic vs. indeterminate) use of predictive algorithms is likely preferred to their use as binary predictions (benign vs. pathogenic) (Pejaver et al., 2022). This trend has been noted by other groups, including assessments of *in silico* prediction algorithms for *RYR1* (Johnston et al., 2021)as well as for theglobin genes, for which Tamana et al. (2022) found that use of the algorithms as binary predictors resulted in too low specificity such that the predictions could not meet the requisite LR for use in the Bayesian classification scheme (Tavtigian et al., 2018, 2020). Our evaluation of these algorithms’ performance in a trichotomized manner enabled us to define discrete thresholds in which the PP3/BP4 criteria are applicable at different strengths, as well as locating the indeterminate range in which the predictions support neither deleterious nor benign effect.

Overall, the four *in silico* tools all have significant predictive ability for *PHOX2B* missense variants. However, Pejaver et al. (2022) recommend that laboratories elect to use tools that can reach a “strong” level of evidence for pathogenicity and “moderate” for benignity. While all four tools evaluated in this study satisfied the recommendation for benign strength predictions, AlphaMissense was the sole tool evaluated in this study that was unable to reach a “strong” strength level for pathogenic predictions, and therefore should not be depended on for routine clinical assessment. Of the three tools meeting Pejaver et al.’s recommendations, BayesDel had far fewer “indeterminate” predictions among the study’s Uncertain set (BayesDel: n=71, CADD: n=101, REVEL: n=94). BayesDel also had the highest AUC (0.992) and the lowest proportion (11.7%) of predicted pathogenic variants from the Uncertain set, with REVEL at a similar level (12.8%) and a slightly lower AUC (0.982). We expect the proportion of true pathogenic variants to be closer to that predicted by these two tools than that predicted by CADD (17.9%) or AlphaMissense (26.3%). While related to AUC, tools’ proportions of predicted pathogenic variants in the Uncertain set are important consideration for their use; tools should not dramatically over-predict pathogenicity so as not to lead to false positives.

Given BayesDel’s ability to reach the requisite PP3/BP4 strength thresholds, its low rate of indeterminate predictions and highest AUC, and that its predicted proportion of Uncertain variants is closest to the presumed true burden of pathogenic variants in that cohort, we suggest that BayesDel is the most appropriate prediction tool for routine *PHOX2B* clinical variant interpretation. REVEL, the overall performance of which lags only slightly behind BayesDel, should be utilized if implementation of BayesDel scores is not possible or practical. Both Tian et al. (2019) and Pejaver et al. (2022) found that BayesDel and REVEL outperform other tools in their large-scale evaluations, and our results support their findings for this much narrower single-gene use. CADD, though not quite as strong a performer for *PHOX2B* in this study, is very much still adequate for routine use.

Additional research is necessary to determine the precise mechanism by which missense variation in *PHOX2B* leads to the isolated or syndromic CCHS phenotypes. Nevertheless, our review of all reported pathogenic missense variants shows there is a clear pattern of pathogenic variation clustering within the homeodomain, encoded by codons 98-157, with a striking paucity of benign variants in the same region. The four tools’ pathogenicitypredictions for all possible missense variants in the homeodomain support the functional importance of this region. We found that the PHOX2B homeodomain has an enrichment of pathogenic missense variation and is largely depleted of benign variation. This is unsurprising given the PHOX2B homeodomain’s involvement in dimerization, DNA-binding, and nuclear localization (Di Lascio et al., 2016), and the recognition of homeodomains’ importance in many other disease-associated genes including *SHOX* (MIM #312865), *PAX3* (MIM #193500), *PITX2* (MIM #601542), and *ARX* (MIM #300382). Like in *PHOX2B*, pathogenic missense variants have been found to cluster in the homeodomain-encoding regions of these genes (D'Elia et al., 2001; Schneider et al., 2005; Holland et al., 2007; Thai et al., 2020; Zhou et al., 2023). Missense variants affecting the PHOX2Bhomeodomain thus begin at a baseline increased likelihood of being pathogenic compared to variants outside of this domain, and we propose that the PM1 criterion (*variant is located within a functionally critical domain or in a mutation hotspot*) be considered for variants altering one of these 60 residues.

Combining PP3 and PM1 for variants in the homeodomain predicted to be pathogenic should be performed as recommended by Pejaver et al. (2022) to avoid inappropriately double counting overlapping evidence used to both develop *in silico* algorithms and to recognize a functionally critical domain. Specifically, the maximum strength of their combined application should not exceed a single “strong” criterion per Tavtigian et al.’s Bayesian framework (2020); i.e., the maximum awarded score should be either PP3\_moderate + PM1 or PP3\_strong, both equivalent to 4 points in the suggested framework.

Even with the increased strength at which the PP3/BP4 criteria may be applied and the availability of the PM1 criterion for some variants, these improvements in interpretation may have only a modest impact on the reduction of VUSs in favor of (likely) benign or (likely) pathogenic classifications until more clinical and experimental data are available. Yet, *in silico* pathogenicity prediction and recognition of the homeodomain’s importance may help guide clinicians about the level of pathogenic evidence in instances when the testing laboratory reports it as a VUS. Several groups have recently advocated for diagnostic genetic testing laboratories to provide additional context beyond the relative suspicion level when reporting VUSs. This may include incorporating the variant’s “score” corresponding to Tavtigian et al.’s (2020) Bayesian points-based classification scheme or including descriptors such as “hot” or “cold” to relate the magnitude of evidence for VUSs in the report for increased transparency and clinical utility (Ellard et al., 2020; Joynt et al., 2021; Loong et al., 2022). Though this will not resolve the issue surrounding the importance of identifying a (likely) pathogenic *PHOX2B* variant in the formal establishment of CCHS diagnoses in all patients with missense variants, an improved clinician understanding of the suspicion of certain VUSs can still play a key role in informing medical management to expedite life-saving management, genetic counseling, and expanded genetic testing decisions.

**Case Vignette**

Following completion of this study’s analyses, an external provider requested a consultation by our group on a young patient with hypoxemia and hypercarbia who had not yet undergone sleep studies nor further evaluation for CCHS. Proband-only exome sequencing at a reference laboratory identified the previously unreported *PHOX2B* variant c.296G>C (p.Arg99Pro), reported by the laboratory as a VUS. Utilizing the results of our analyses, a reassessment of this variant by our group determined that this variant is better classified as Likely Pathogenic: it is located in the homeodomain and its BayesDel, CADD, and REVEL scores of 0.572, 30.0, and 0.972, respectively, are therefore sufficient for either PP3\_strong or PM1 + PP3\_moderate; it occurs at the same amino acid position as another (likely) pathogenic variant (c.295C>T:p.Arg99Trp; PM5); and it is absent from gnomAD v4.1.0 (PM2\_supporting). An improved understanding of this variant’s presumed clinical significance was critical to the patient receiving a CCHS diagnosis with the accompanying medical management recommendations and accurate counseling of the family.

**Limitations**

Our dataset and analyses do not include functional data except for a small number of variants. Pairing the clinical and prediction data with experimentally derived functional data, such as that from a multiplex assay of variant effect, or MAVE, would further strengthen the confidence in the established score thresholds (Starita et al., 2017). This type of study has not been performed for *PHOX2B* but could be of significant value to the interpretation of missense variants.

The classifications of the variants in our dataset are based on a set of criteria agreed upon by the authors of this study. The variant classifications made in the study were agreed upon by experts involved in the clinical and molecular diagnoses of *PHOX2B*-related disease and reflect the current understanding of the clinical significance of variation in this gene. These criteria are related to several aspects of, but not explicitly based on, the ACMG-AMP guidelines; there is currently no available standard for classification of variants in *PHOX2B*, and the paucity of clinical and experimental data severely limits the ability to reach pathogenic/likely pathogenic and benign/likely benign classifications when adhering to the ACMG/AMP guidelines as written. Despite the low prevalence of CCHS, and the small proportion of cases associated with *PHOX2B* missense variation, we were able to leverage the considerable patient database maintained by two of our authors (CMR and DEW-M) and utilize the wealth of their clinical experience and expertise to obtain and classify variant test sets. The results from these analyses are reported with the important caveat that they reflect the currently available information about *PHOX2B* missense variation, and that this type of analysis should be performed periodically as our datasets, both clinical and population, continue to grow.

**CONCLUSION**

*In silico* prediction tools can distinguish between pathogenic and benign missense variants in *PHOX2B* with relatively high accuracy and support the functional importance of the protein’s homeodomain. This improved understanding for interpretation of *PHOX2B* missense variation will increase the number of conclusive classifications of missense variants previously classified as VUS and will provide the relative degree of suspicion for those remaining VUS, all of which will assist in the management of these fragile and underserved ventilator-dependent patients.

**REFERENCES**

1. Weese-Mayer DE, Berry-Kravis EM, Ceccherini I, Keens TG, Loghmanee DA, Trang H, et al. An official ATS clinical policy statement: Congenital central hypoventilation syndrome: genetic basis, diagnosis, and management. Am J Respir Crit Care Med. 2010;181(6):626-44. doi: 10.1164/rccm.200807-1069ST
2. Weese-Mayer DE, Rand CM, Khaytin I, Slattery SM, Yap KL, Marazita ML, Bery-Kravis EM. Congenital Central Hypoventilation Syndrome. In MP Adam et al. (Eds). GeneReviews. 2021.
3. Zhou A, Rand CM, Hockney SM, Niewijk G, Reineke P, Speare V, et al. Paired-like homeobox gene (PHOX2B) nonpolyalanine repeat expansion mutations (NPARMs): genotype-phenotype correlation in congenital central hypoventilation syndrome (CCHS). Genet Med. 2021;23(9):1656-63. doi: 10.1038/s41436-021-01178-x
4. Weese-Mayer DE, Berry-Kravis EM, Zhou L, Maher BS, Silvestri JM, Curran ME, et al. Idiopathic congenital central hypoventilation syndrome: analysis of genes pertinent to early autonomic nervous system embryologic development and identification of mutations in PHOX2b. Am J Med Genet A. 2003;123A(3):267-78. doi: 10.1002/ajmg.a.20527
5. Matera I, Bachetti T, Puppo F, Di Duca M, Morandi F, Casiraghi GM, et al. PHOX2B mutations and polyalanine expansions correlate with the severity of the respiratory phenotype and associated symptoms in both congenital and late onset Central Hypoventilation syndrome. J Med Genet. 2004;41(5):373-80. doi: 10.1136/jmg.2003.015412
6. Weese-Mayer DE, Rand CM, Zhou A, Carroll MS, Hunt CE. Congenital central hypoventilation syndrome: a bedside-to-bench success story for advancing early diagnosis and treatment and improved survival and quality of life. Pediatr Res. 2017;81(1-2):192-201. doi: 10.1038/pr.2016.196
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24. doi: 10.1038/gim.2015.30
8. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-7. doi: 10.1093/nar/gkx1153
9. Tian Y, Pesaran T, Chamberlin A, Fenwick RB, Li S, Gau CL, et al. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. Sci Rep. 2019;9(1):12752. doi: 10.1038/s41598-019-49224-8
10. Pejaver V, Byrne AB, Feng BJ, Pagel KA, Mooney SD, Karchin R, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet. 2022;109(12):2163-77. doi: 10.1016/j.ajhg.2022.10.013
11. Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604(7905):310-5. doi: 10.1038/s41586-022-04558-8
12. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5. doi: 10.1038/ng.2892
13. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99(4):877-85. doi: 10.1016/j.ajhg.2016.08.016
14. Feng BJ. PERCH: A Unified Framework for Disease Gene Prioritization. Hum Mutat. 2017;38(3):243-51. doi: 10.1002/humu.23158
15. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381(6364):eadg7492. doi: 10.1126/science.adg7492
16. Tamana S, Xenophontos M, Minaidou A, Stephanou C, Harteveld CL, Bento C, et al. Evaluation of in silico predictors on short nucleotide variants in HBA1, HBA2, and HBB associated with haemoglobinopathies. Elife. 2022;11:e79713. doi: 10.7554/eLife.79713
17. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med. 2017;19(10):1151-8. doi: 10.1038/gim.2017.26
18. Shimokaze T, Sasaki A, Meguro T, Hasegawa H, Hiraku Y, Yoshikawa T, et al. Genotype-phenotype relationship in Japanese patients with congenital central hypoventilation syndrome. J Hum Genet. 2015;60(9):473-7. doi: 10.1038/jhg.2015.65
19. Bachetti T, Parodi S, Di Duca M, Santamaria G, Ravazzolo R, Ceccherini I. Low amounts of PHOX2B expanded alleles in asymptomatic parents suggest unsuspected recurrence risk in congenital central hypoventilation syndrome. J Mol Med (Berl). 2011;89(5):505-13. doi: 10.1007/s00109-010-0718-y
20. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: Lessons from gnomAD. Hum Mutat. 2022;43(8):1012-30. doi: 10.1002/humu.24309
21. Preston CG, Wright MW, Madhavrao R, Harrison SM, Goldstein JL, Luo X, et al. ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. Genome Med. 2022;14
22. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi: 10.1038/s41586-021-03819-2
23. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons; 2009.
24. Brent RP. Algorithms for Minimization Without Derivatives. Dover Publications; 2013.
25. Pattyn A, Morin X, Cremer H, Goridis C, Brunet JF. Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. Development. 1997;124(20):4065-75.
26. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med. 2018;20(9):1054-60. doi: 10.1038/gim.2017.210
27. Johnston JJ, Dirksen RT, Girard T, Gonsalves SG, Hopkins PM, Riazi S, et al. Variant curation expert panel recommendations for RYR1 pathogenicity classifications in malignant hyperthermia susceptibility. Genet Med. 2021;23(7):1288-95. doi: 10.1038/s41436-021-01125-w
28. Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. Hum Mutat. 2020;41(10):1734-7. doi: 10.1002/humu.24088
29. Di Lascio S, Belperio D, Benfante R, Fornasari D. Alanine Expansions Associated with Congenital Central Hypoventilation Syndrome Impair PHOX2B Homeodomain-mediated Dimerization and Nuclear Import. J Biol Chem. 2016;291(25):13375-93. doi: 10.1074/jbc.M115.679027
30. D'Elia AV, Tell G, Paron I, Pellizzari L, Lonigro R, Damante G. Missense mutations of human homeoboxes: A review. Hum Mutat. 2001;18(5):361-74. doi: 10.1002/humu.1207
31. Holland PW, Booth HA, Bruford EA. Classification and nomenclature of all human homeobox genes. BMC Biol. 2007;5:47. doi: 10.1186/1741-7007-5-47
32. Schneider KU, Marchini A, Nitin Sabherwal, Röth R, Beate Niesler, Marttila T, et al. Alteration of DNA binding, dimerization, and nuclear translocation of SHOX homeodomain mutations identified in idiopathic short stature and Leri-Weill dyschondrosteosis. Human mutation. 2005 Jan 1;26(1):44–52. doi: 10.1002/humu.20187
33. Thai MHN, Gardner A, Redpath L, Mattiske T, Dearsley O, Shaw M, et al. Constraint and conservation of paired-type homeodomains predicts the clinical outcome of missense variants of uncertain significance. Hum Mutat. 2020;41(8):1407-24. doi: 10.1002/humu.24034
34. Zhou L, Wang X, An J, Zhang Y, He M, Tang L. Genotype-phenotype association of PITX2 and FOXC1 in Axenfeld-Rieger syndrome. Exp Eye Res. 2023;226:109307. doi: 10.1016/j.exer.2022.109307
35. Ellard S, Baple EL, Callaway A, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020. Association for Clinical Genomic Science; 2020. https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf. Accessed November 9, 2023.
36. Joynt ACM, Axford MM, Chad L, Costain G. Understanding genetic variants of uncertain significance. Paediatr Child Health. 2021;27(1):10-1. doi: 10.1093/pch/pxab070
37. Loong L, Garrett A, Allen S, Choi S, Durkie M, Callaway A, et al. Reclassification of clinically-detected sequence variants: Framework for genetic clinicians and clinical scientists by CanVIG-UK (Cancer Variant Interpretation Group UK). Genet Med. 2022;24(9):1867-77. doi: 10.1016/j.gim.2022.05.002
38. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 2017;101(3):315-25. doi: 10.1016/j.ajhg.2017.07.014