Predicting the functional effect of compound heterozygous genotypes from large scale variant effect maps

Michael Xie[1,2], Amy Sirr[1], Russell S. Lo[1], Gareth A. Cromie[1], Michelle Tang[1], Julee Ashmead[1], Katherine Owens[1,3], J. Nathan Kutz[3], Richard N. McLaughlin Jr.[1] and Aimée M. Dudley[1,2]*

[1] Pacific Northwest Research Institute, Seattle, Washington, USA

[2]Molecular Engineering Graduate Program, University of Washington, Seattle, Washington, USA

[3] Department of Applied Mathematics, University of Washington, Seattle, Washington, USA

Keywords: Yeast, Serine Deficiency, Neu-Laxova Syndrome, PSAT1, Inborn Errors of Metabolism

***Corresponding Author:**

Aimée M. Dudley, Ph.D.

Pacific Northwest Research Institute

720 Broadway

Seattle, WA 98122

aimee.dudley@gmail.com

Abstract

Loss of function mutations in *PHGDH*, *PSAT1*, and *PSPH* cause a set of rare, autosomal recessive diseases known as Neu-Laxova syndrome (NLS) or serine-deficiency disorders. The diseases present with a broad range of phenotypes, including lethality, severe neurological manifestations, seizures, and intellectual disability. However, because L-serine supplementation, especially if started early in pregnancy, can ameliorate and in some cases even prevent symptoms, knowledge of pathogenic variants is highly actionable. Recently, our laboratory established a yeast-based assay for human *PSAT1* function. We have now applied it at scale to assay the functional impact of 1,914 SNV-accessible amino acid substitutions. Our results agree well with clinical interpretations and protein structure-function relationships, supporting the use of our data as functional evidence under the ACMG interpretation guidelines. In addition to assaying the functional impact of individual variants in yeast haploid cells, we can assay pairwise combinations of *PSAT1* alleles that recapitulate human genotypes, including compound heterozygotes, in yeast diploids. Results from this diploid assay successfully distinguish patient genotypes from those of healthy carriers and agree well with disease severity. Finally, we present a linear model that can accurately predict biallelic function in diploids using information from the individual allele measurements in haploids. Taken together, our work provides an example of how large-scale functional assays in model systems can be powerfully applied to the study of a rare disease.

63 **Background:**

64 High throughput sequencing technologies have accelerated the discovery of human genetic
65 variation. Data from the Genome Aggregation Database (gnomAD) [1] suggest that the average
66 person harbors approximately 200 rare (allele frequency <0.1%) protein coding variants [2].
67 However, leveraging genomic information for the diagnosis and treatment of medically
68 actionable diseases is currently limited by the relatively small number of alleles for which
69 clinical interpretation is readily available. Most of the ~550K missense variants in ClinVar [3],
70 even those in well-studied disease genes, are variants of uncertain significance (VUS). The
71 difficulty of variant interpretation is compounded further in rare diseases, and patients often
72 face long diagnostic odysseys or remain undiagnosed.

73 Inborn errors of metabolism (IEM), genetic alterations that disrupt metabolic homeostasis, are
74 a relatively large class of rare diseases. Over 700 IEMs [4] have been described and new
75 disorders continue to be discovered [5]. Although they are rare individually, collectively these
76 diseases are common, with estimates suggesting that IEMs may affect 1 in 800-2600 live births
77 [6,7]. Many are severe and manifest early in life, but some are also medically actionable and
78 timely diagnosis can prevent the onset of irreversible damage. Disruptions of metabolic
79 pathways can lead to toxic levels of substrate accumulation or deficiency of an essential
80 product. Therapeutic strategies often focus on amending these imbalances. For example,
81 dietary restriction is prescribed for the treatment of urea cycle disorders, phenylketonuria, and
82 galactosemia, while dietary supplementation is prescribed for homocystinuria and pyridoxine-
83 dependent epilepsy [8–10].

84 Serine deficiency disorders are an emerging group of IEMs with promising case studies
85 describing successful treatment with dietary supplementation (**Figure 1**). Impairment of any of
86 the three L-serine biosynthesis pathway enzymes, phosphoglycerate dehydrogenase (PGDH;
87 encoded by *PHGDH*), phosphoserine aminotransferase (PSAT; encoded by *PSAT1*), and
88 phosphoserine phosphatase (PSP; encoded by *PSPH*), result in systemic serine deficiency [11].
89 Serine metabolism is central to numerous biological processes, including the synthesis of
90 proteins, nucleotides, and phospholipids, as well as the formation of the neuromodulators D-
91 serine and glycine. As a result, serine biosynthesis defects negatively impact nervous system
92 development and function, with clinical manifestations that include microcephaly, seizures,
93 intellectual disability, and neuropathies. Case studies have demonstrated that oral serine
94 supplementation can reduce and, in some cases, prevent the onset of these severe symptoms
95 that typically manifest very early in life [12–16] Prompt diagnosis is crucial, as the impact of
96 therapeutic intervention, including prenatal dietary supplementation, is greatest before
97 patients become symptomatic and irreversible neurological damage occurs.

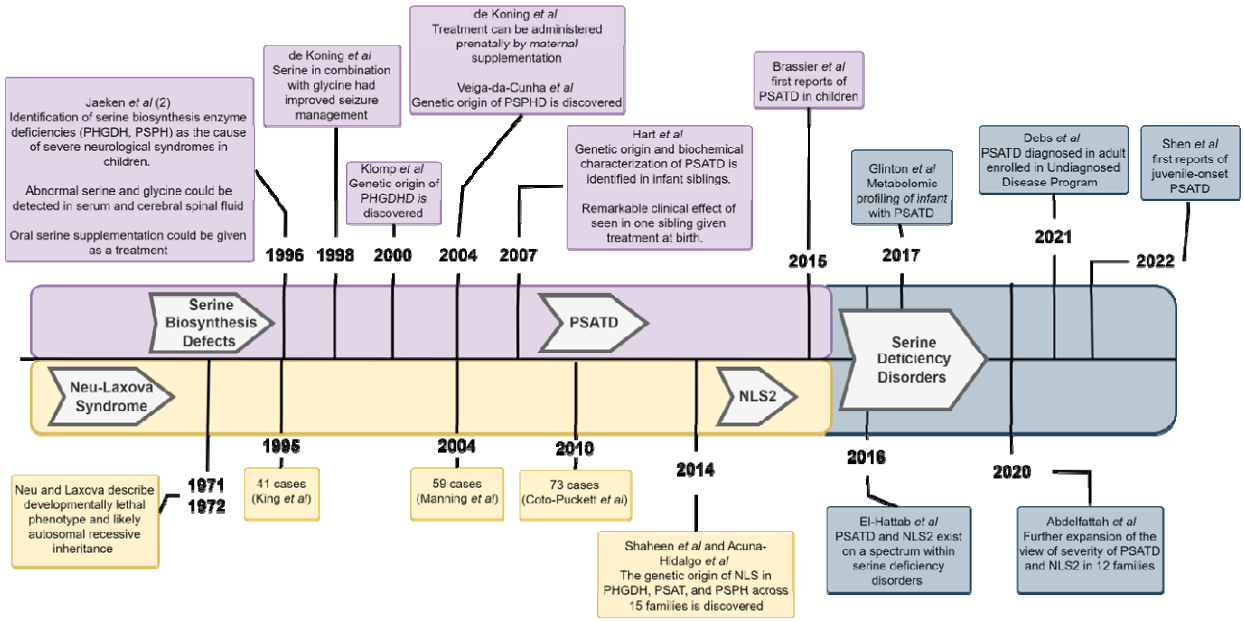


Figure 1: Key developments in the study of serine deficiency disorders, a set rare but treatable inborn errors of metabolism.

Unfortunately, like many rare diseases, diagnosing serine deficiency disorders is difficult in the absence of a family history of disease. One challenge is that the disorders display a broad phenotypic spectrum. At one extreme, Neu-Laxova syndrome (NLS) is a severe developmentally lethal congenital disease, while serine deficiency ranges from growth defects and severe neurological manifestations in infants to milder forms with intellectual disability, developmental delay, or progressive neuropathies that can manifest in children or adults [11,17]. Although it has been hypothesized that residual enzyme activity may explain the phenotypic variability [18–20] within serine deficiency disorders, functional studies regarding this subject are sparse and the natural history of the disorders are largely unknown. A second challenge is the difficulty of biochemical diagnosis, which requires measurement of amino acid levels in patient serum or cerebral spinal fluid compared to closely age-matched controls [17,21].

Identifying the presence of pathogenic sequence variants through whole exome/genome sequencing represents an attractive alternative for diagnosing serine deficiency disorders. However, this approach relies on having clinical interpretations for rare variants, which are generally not available. Instead, within the current disease literature, clinical sequencing has primarily been used as a tool by experts to provide molecular confirmation once a patient has become symptomatic. High-throughput approaches to variant interpretation that can be applied to rare variants are needed to increase the success rate of sequencing-based diagnostics. While computational prediction methods can be applied at scale, their relatively high error rates limit their utility (examples in [22]). Functional assays that can quantitatively

122 measure variant effects on protein function [23] are an alternative approach that can be used
123 as part of the criteria for variant interpretation established by the American College of Medical
124 Genetics (ACMG) [24]. Improvements in DNA synthesis technology have enabled time-and-cost
125 effective methods for building large variant libraries. When combined with high-throughput
126 phenotyping methods, functional assays can comprehensively assess variants that have been
127 identified in the human population as well as those that may arise in the future [25,26].

128 Recently, our laboratory established a yeast-based assay for human *PSAT1* function [27]. Here,
129 we apply it at scale to quantify the functional impact of 1,914 amino acid substitutions that are
130 accessible via a single nucleotide variant (SNV) in the human *PSAT1* coding sequence. Our
131 results agree well with clinically interpreted alleles and with protein structure-function
132 relationships, supporting the use of our data as functional evidence under the ACMG
133 interpretation guidelines. In addition to assaying the functional impact of single variants in
134 yeast haploids, we demonstrate that we can apply our assay to yeast diploids with pairwise
135 combinations of *PSAT1* alleles that recapitulate human genotypes. The results of the diploid
136 functional assay distinguish patient genotypes from those of healthy carriers and agree with the
137 stratification of patient genotypes by disease severity. Finally, we developed a mathematical
138 model that can accurately predict biallelic function in diploids from pairwise combinations of
139 individual allele activity measured experimentally in haploids. Taken together, our work
140 provides an example of how large-scale functional assays in model systems can be powerfully
141 applied in the study of rare disease and to inform future diagnostic efforts.

142 **Methods**

143 **Strain library construction**

144 All *Saccharomyces cerevisiae* strains used in this study (**Supplementary Table S1** and **S2**) were
145 derived from the isogenic lab strains FY4 (MAT**a**) and FY5 (MATα) [28]. Unless otherwise noted,
146 strains were grown in rich medium (YPD, 1% yeast extract, 2% peptone, and 2% glucose) or
147 minimal medium (SD, without amino acids, 2% glucose) using standard media conditions and
148 methods for yeast genetic manipulation [29].

149 Methodology concerning the design and construction of the yeast codon-optimized version
150 (*yPSAT1*) of human *PSAT1* isoform 1, the wild type (*yPSAT1*) strain, the deletion (*ser1Δ0)* strain,
151 and individual variant strains were previously described in Sirr *et al* [27]. Here, we describe the
152 design and construction of a variant library ~10 fold larger and modifications of the growth
153 assay for high throughput as follows:

154 The variant library was designed to capture the amino acid substitutions resulting from all SNV-
155 accessible missense mutations across 369 codons, excluding the start and stop, in the human
156 *PSAT1* isoform 1 cDNA sequence (CCDS6660.1; Consensus Coding Sequence database [30]. The

157 9 possible single nucleotide variants at each codon resulted in 4-7 unique amino acid
158 substitutions. *yPSAT1* variants encoding the complete set of these unique amino acid
159 substitutions (n = 2,182) were synthesized (Twist Biosciences) in *yPSAT1* as an oligonucleotide
160 library in which each well of a 96 well plate contained all amino acid substitutions (n=4-7) at a
161 given amino acid.

162 Individual wells of these plates were amplified to approximately 500 ng of DNA by 15-cycles of
163 high-fidelity PCR (Phusion High-Fidelity DNA Polymerase; Thermo Scientific) and transformed
164 into a MAT**a** haploid deletion strain (*ser1Δ0*) using standard methods. Single colonies from the
165 yeast transformations were isolated such that 6,335 individual transformants, each encoding a
166 single amino acid substitution, were arrayed into 96-well plates containing rich medium.
167 Because individual transformants are isolated and maintained as separate stocks (one strain per
168 well in a 96-well plate), each strain is an independently constructed biological isolate of the
169 variant it contains. For downstream phenotype normalization, each library plate also contained
170 replicates of the same control strains: 2 deletion (*ser1Δ0)* and 4 wild type (*yPSAT1*) strains.

171 All of the original *yPSAT1* variant strains in the large-scale library are of a single mating type
172 (MAT**a**) (**Supplementary Table S1**). To generate diploid strains harboring combinations of
173 *yPSAT1* variants were first introduced into a strain of the opposite mating type and the resulting
174 *yPSAT1* MATα strains were mated in a pairwise manner to target *yPSAT1* MAT**a** strains using
175 standard methods [31]. The resulting diploid strains (**Supplementary Table S2**) harboring
176 combinations of *yPSAT1* alleles (missense variant, wild type, or null) were then arrayed in an
177 alternating checkerboard pattern that minimizes the influence of nutrient competition from
178 neighboring colonies during phenotype measurement.

179 **Variant library sequence confirmation**

180 For each *yPSAT1* transformant, we know which codon is mutated (target codon), but not which
181 specific variant is present. To determine this, we used a custom MinION (Oxford Nanopore
182 Technologies) sequencing pipeline, as described in Lo *et al* [32]. Briefly, individual
183 transformants were pooled in groups of 12, so that no target codon was represented more than
184 once in a single pool. The pool was then sequenced and at each target codon the most frequent
185 variant codon was identified (candidate variant) as well as the second most frequent variant. An
186 error model was then developed for each variant.

187 Candidate variants then underwent quality control. Variants passed if they were supported by
188 at least 15 reads, if the observed frequency of the variant was at least 3.3x the estimated error
189 frequency for that variant and if no missense or nonsense secondary mutations were present.
190 In addition, candidates were rejected when the second most frequent variant was both
191 enriched (>=10-fold) relative to its error frequency and was observed at >30% of the frequency
192 of the candidate variant.

193 **Growth assays**

194 Before phenotyping, the set of variant strains generated in this study were extended to include
195 the haploid strains carrying missense variants constructed and sequence confirmed by Sirr *et al*
196 [27]. These additional strains were re-arrayed to match the phenotyping layout used in this
197 study, with at least 2 isolates of each variant. Additional control plates included a plate with
198 wild type at every position, and plates consisting entirely of D100A or A99V (with control
199 positions) *yPSAT1* strains, which are ClinVar pathogenic alleles encoding missense variants.

200 A Biomek i7 robot outfitted with a V&P 96-pin head was utilized to pin strain plate libraries
201 between different culture media. Strains were initially grown to saturation in rich medium and
202 then pinned onto solid medium utilizing glycerol as the central carbon source (YPG,1% yeast
203 extract, 2% peptone, 2% glycerol) to remove any yeast cells lacking mitochondria (petite).
204 Strains were then pinned back into rich medium and grown to saturation. Each plate was then
205 pinned in replicate (n=3-6) onto solid minimal medium, which lacks serine, and grown for 3
206 days at 30°C. A mounted Canon PowerShot SX10 IS compact digital camera was used to take
207 images (ISO200, f4.5, 1/40s exposure) every 24 hours for three days under consistent lighting,
208 camera to subject distance, and zoom. Each plate was labeled with a custom code39 barcode
209 that was included in the frame of view. Images were acquired as jpg files.

210 **Image-based growth quantification**

211 Each plate image was processed using PyPl8 (https://github.com/lacyk3/PyPl8) to extract
212 features from each strain patch in each barcoded agar plate as follows: First, the barcode within
213 each image was detected and decoded to rename each file using the corresponding plate
214 name, replicate number, condition, and timepoint. Next, each image was cropped into 96
215 square tiles, segmented, and each replica pinned patch was identified using Otsu's method or
216 circle detection. Finally, the sum of the gray scale pixel intensities within each strain patch
217 (pixelsum), was extracted and used as the metric for growth estimation.

218 **Growth data fitting and normalization**

219 For haploid strains, raw phenotypic values were normalized, quality control filters were applied
220 to each isolate, and a final relative growth estimate for each variant was determined, as
221 described in Lo *et al* [32]. Briefly, normalization steps were carried out to account for the
222 effects of plate-to-plate variation, relative growth of neighboring patches, and plate edge
223 effects. Pin effect normalization did not reduce noise and was omitted. Isolates with
224 (nonsynonymous) secondary mutations were removed from the dataset as were all isolates of
225 variants that showed a high degree of variation in isolate-to-isolate growth values. This left a
226 final filtered dataset of 5,164 independent isolates. Finally, a linear model was used to estimate

227 the relative growth of each genotype, on a scale with growth of null controls set to 0 and
228 growth of wild type set to 1.

229 A similar approach was applied to phenotypic values extracted from the diploid growth assay,
230 although neighbor effects were assumed negligible because of the checkerboard pinning
231 arrangement. For normalization, a linear model was used to simultaneously estimate the
232 effects of genotype, plate edge positioning and plate-to-plate variation on growth. Genotype
233 effects were rescaled to set homozygous null to 0 and homozygous wild type to 1.

234 **Predicting diploid growth**

235 We developed a linear model to predict the growth of diploid strains based on which pair of
236 *yPSAT1* variants is present, using individual growth estimates of each allele in haploids. In this
237 model, for each diploid, the (haploid) growth estimates of the lower and higher growth alleles
238 were labelled as minimum and maximum (j, k), respectively. In cases of homozygous
239 combinations, the minimum and maximum were equal. Strains carrying a single copy of the
240 wild type allele (*yPSAT1*) or null (*ser1*Δ0) had their respective haploid estimates set equal to 1
241 and 0. To predict diploid growth (d) from the more impaired (x_j) and less impaired (x_k) alleles,
242 we performed an ordinary least squares regression to fit an additive pairwise model (d = a + bx_j
243 + cx_k). We also compared this model to two simpler regression models. In the first model,
244 diploid growth was predicted from the mean of the haploid estimates (x_j = x_k). In the second,
245 diploid growth was predicted from the higher growing haploid allele only, i.e. complete
246 dominance (x_j = 0). Leave one out cross validation was used to assess model performance in
247 the best model.

248 Next, we evaluated the performance of experimental versus predicted growth as a binary
249 classifier for identifying genotypes matching those of patients diagnosed with *PSAT1*-related
250 serine deficiency disorders vs carrier parents. A logistic regression model indicated a threshold
251 (p=0.5) decision value of 72% and 73% for experimental and predicted diploid growth
252 respectively.

253 **Protein structure and conservation analysis**

254 Structural features were derived from the crystal structure of human phosphoserine
255 aminotransferase isoform 1 complexed with pyridoxal phosphate cofactor (PDB: 3e77; subunit
256 L17-L370, homodimer biological assembly). Secondary structure features were extracted from
257 this crystal structure using DSSP software [33,34]. AlphaFold [35,36] was used to visualize the
258 predicted structure of the missing 16 N-terminal residues. Molecular illustrations were created
259 by using the PyMol Molecular Graphics System Version 2.3.2 (Schrödinger). Tools from the
260 publicly available PyMol script and plugin repository were used to determine interfacial
261 residues and charge distribution. Dimer interface residues were defined as residues at which

262 their subunit-only based surface areas are larger than their complex-based surface areas (cutoff
263 value = 0.5). Macromolecular electrostatics were estimated using the Adaptive Poisson-
264 Boltzmann Solver plugin. Precomputed evolutionary conservation scores and grades, derived
265 from multiple sequence alignments of amino acid homologs with known structure, for subunit
266 A of PDB 3e77 were extracted from the ConSurf Database [38,39].

267 **Results**

268 **Surveying the functional impact of large-scale missense variation in *PSAT1***

269 Our laboratory had previously established a yeast-based complementation assay that leveraged
270 the ability of the human *PSAT1* coding sequence to functionally replace its yeast ortholog, *SER1*
271 [27]. In this assay, growth on minimal medium lacking serine provides a quantitative readout of
272 PSAT1 activity, allowing the functional impact of protein coding variants to be assessed. Variant
273 impact is expressed on a relatively intuitive scale of activity between the level associated with
274 no activity (that of a complete gene deletion) and that conferred by the wild type human
275 protein coding sequence (*yPSAT1*). Variants causing a reduction in PSAT activity could have
276 their effect via decreases in enzymatic activity, protein stability, or a combination of both.

277 In this study, we applied the assay at scale to measure the effect of thousands of amino acid
278 substitutions as follows. First, with the exception of the translational start and stop codons, we
279 identified all unique amino acid substitutions (n = 2,182) that were accessible via a SNV across
280 the full length of human *PSAT1* isoform 1 cDNA (1,113 bp). Variant codons encoding each
281 amino acid substitution were then introduced into the yeast codon-optimized version of *PSAT1*
282 (*yPSAT1*). The resulting variant library was transformed into a haploid *SER1* deletion strain
283 (*ser1Δ0*) and integrated in single copy at the *SER1* locus of the yeast genome, under the control
284 of the endogenous *SER1* transcriptional promoter and terminator. Next, transformants were
285 individually arrayed in 96-well plate format, and the identity of the variant codon present in
286 each transformant was determined by Oxford Nanopore MinION sequencing (**Methods**). Strains
287 that harbored secondary mutations in the protein coding sequence were removed from further
288 consideration. The arrayed strain library was then grown, in triplicate, on solid medium lacking
289 serine and imaged after three days at 30°C. Data from these images were used to measure the
290 growth of each strain, relative to wild type (*yPSAT1*) and null (*ser1Δ0*), using a custom
291 automated image analysis pipeline and normalization software (**Methods**).

292 Here, we report the functional impact of 1,914 amino acid substitutions in *PSAT1*
293 (**Supplementary Table S3**), corresponding to ~88% of all unique SNV-accessible amino acid
294 substitutions (n=2,182) in the human *PSAT1* cDNA sequence. A subset of these substitutions
295 (n=196) was also assayed in our previous study [27] which sought to quantify the functional
296 impact of the full set of *PSAT1* missense variants described in ClinVar, gnomAD, or the clinical
297 literature at that time. Reanalysis in the current study allowed us to assess these amino acid

substitutions using the same data processing pipeline as the 1,718 new substitutions, thereby placing them on a common scale to facilitate direct comparison. Despite the use of an improved analysis pipeline, there was excellent agreement (R^2=0.85) between the normalized growth estimates of the 196 substitutions in the two studies.

Among the full set of 1,914 amino acid substitutions, the distribution of growth values is bimodal, with a large peak centered approximately on the value of the wild type *yPSAT1* strain (normalized growth=1) and a smaller peak centered on the value of the null (deletion) strain (normalized growth=0) (**Figure 2**). These results are consistent with a large group of protein coding variants showing little or no functional impairment relative to wild type, and a smaller group behaving as complete loss of function alleles, comparable to the null control. The remaining protein coding variants showed varying levels of less severe functional impairment (**Figure 2**).

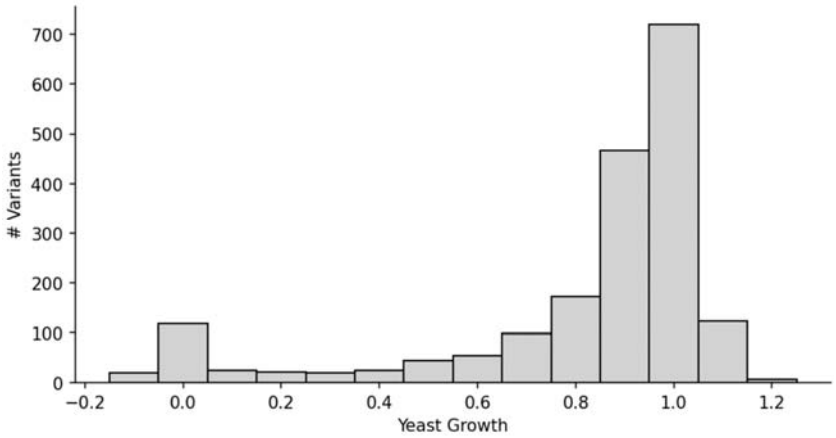


Figure 2. Distribution of variant effect for 1,914 PSAT1 unique SNV-accessible missense amino acid substitutions. Frequency of experimentally measured yeast growth values scaled relative to wild type yPSAT1 (normalized growth=1) and null (normalized growth=0).

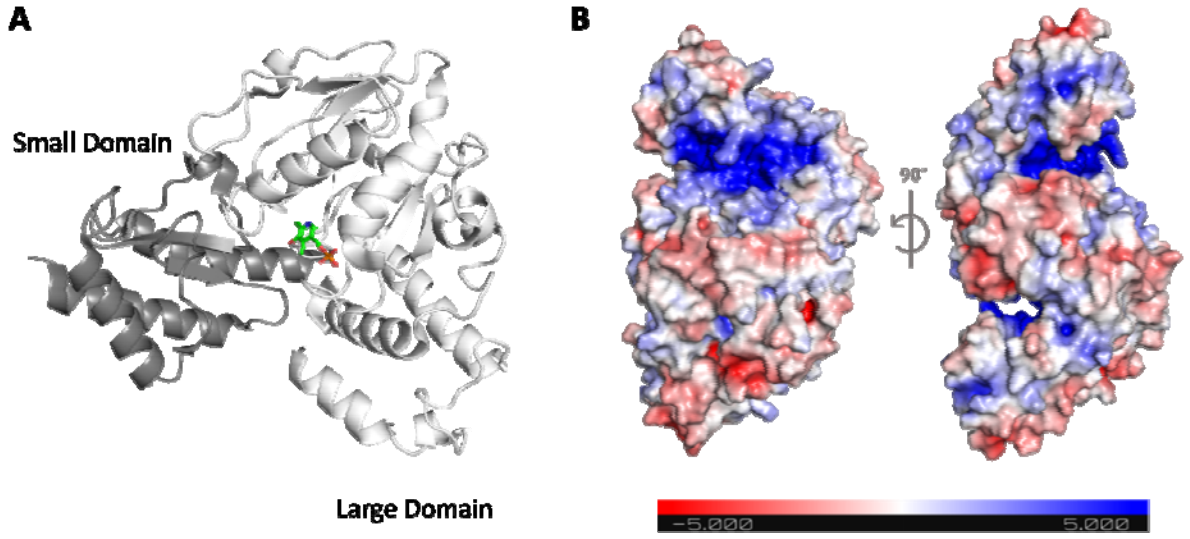
Mapping functional effects to protein structure and conservation

To provide a better understanding of the impact of missense substitutions on protein function, we examined the results of our assay in the context of mechanistic, structural, and evolutionary features of the human phosphoserine aminotransferase isoform 1 (PSAT) protein. PSAT belongs to the large family of fold type I pyridoxal phosphate (PLP; active form of vitamin b6) cofactor-dependent enzymes, and more specifically, the class IV group of aminotransferases [40,41] In PLP-dependent enzymes the cofactor serves two ubiquitous functions in catalysis: to form a reactive Schiff base covalent linkage with the substrate amino group and to act as an electron sink to stabilize intermediate states [42–44]. Reaction specificity is determined by the conformation of the active site as well as how residues interact with PLP and the substrates. For

325 aminotransferases, these enzymes typically form functional homomultimers with interfacial
326 PLP-containing active sites, where transamination reactions proceed via substrate binding,
327 amino group transfer, and product release [41].

328 PSAT catalyzes the reversible conversion of glutamate to α-ketoglutarate and 3-
329 phosphohydroxypyruvate (3-PHP) to phosphoserine [45]. The crystal structure PSAT complexed
330 with PLP (PDB: 3e77; residues L17-L370) is characterized by a s-shaped homodimer assembly in
331 which two separate PLP-containing active sites reside along the interface. An overall positive
332 charge distribution is also localized near these active sites (**Figure 3**.**B**). Each subunit consists of
333 two domains with mixed α/β fold; a large N-terminal catalytic domain (L17-G263), which forms
334 the dimer interface and binds PLP, and a small C-terminal domain (A264-L370) (**Figure 3**.**A**).
335 Although the 16 N-terminal amino acids (M1-K16) are not included in the crystal structure,
336 AlphaFold [35,36] predicts these residues to be placed between the two domains, proximal to
337 the active site (**Supplemental Figure S1**).

338 The overall structural organization and active site architecture of PSAT are conserved with
339 other members of the class IV family of aminotransferases [41], and phosphoserine
340 aminotransferases from eukaryotic and prokaryotic organisms [46]. Evolutionary conservation
341 scores for PSAT (pdb 3e77; subunit A) from the ConSurf DataBase [38,39], which are derived
342 from sequence alignments of amino acid homologs with known structure, illustrate that more
343 conserved (negative ConSurf score) regions localize near the active sites (**Figure 4**.**B**).
344 Furthermore, structural analysis of related phosphoserine aminotransferases [47–50] with
345 additional crystallographic experiments in substrate (or analog) bound states (not currently
346 available for human PSAT), have identified substrate binding residues that are highly conserved.



347

348 **Figure 3.** Overall subunit organization and homodimer assembly of human PSAT. **A)** Ribbon diagram representation
349 of the secondary structure of a subunit of huPSAT (PDB 3e77; L17-L370) complexed with PLP (green) in stick

350 representation. The small C-terminal domain is colored as dark grey and large N-terminal domain as white. **B)** Local
351 charge distribution on the surface of the huPSAT homodimer assembly, on a scale of negative charge (-5 kT; red) to
352 positive charge (+5 kT; blue) as calculated by the Adaptive Poisson–Boltzmann Solver program.

353 Some of our knowledge of human PSAT protein structure-function is currently based on studies
354 of non-human phosphoserine aminotransferases. Our dataset of 1,914 PSAT variants
355 characterizes the functional impact of a distribution of 4-7 amino acid substitutions at each
356 position across the entire length of the human PSAT protein (370 aa) (**Figure 4**.**A**). These
357 substitutions display a range of activity from complete loss of function (scores ≤ 5%) to not
358 impaired relative to wild type (scores ≥ 95%). Impaired variants were non-randomly distributed
359 across the length of the protein. We identified several positions where the majority of SNV-
360 accessible variants exhibited some degree of functional impairment. In particular and
361 consistent with expectation, substitutions in conserved active site residues (**Figure 4**.**C**) were
362 often associated with more severe loss of activity (median scores ≤ 35%). However, not all
363 substitutions at highly conserved residues exhibited loss of function in our assay, and not all
364 substitutions at less conserved residues had activities close to wild type.

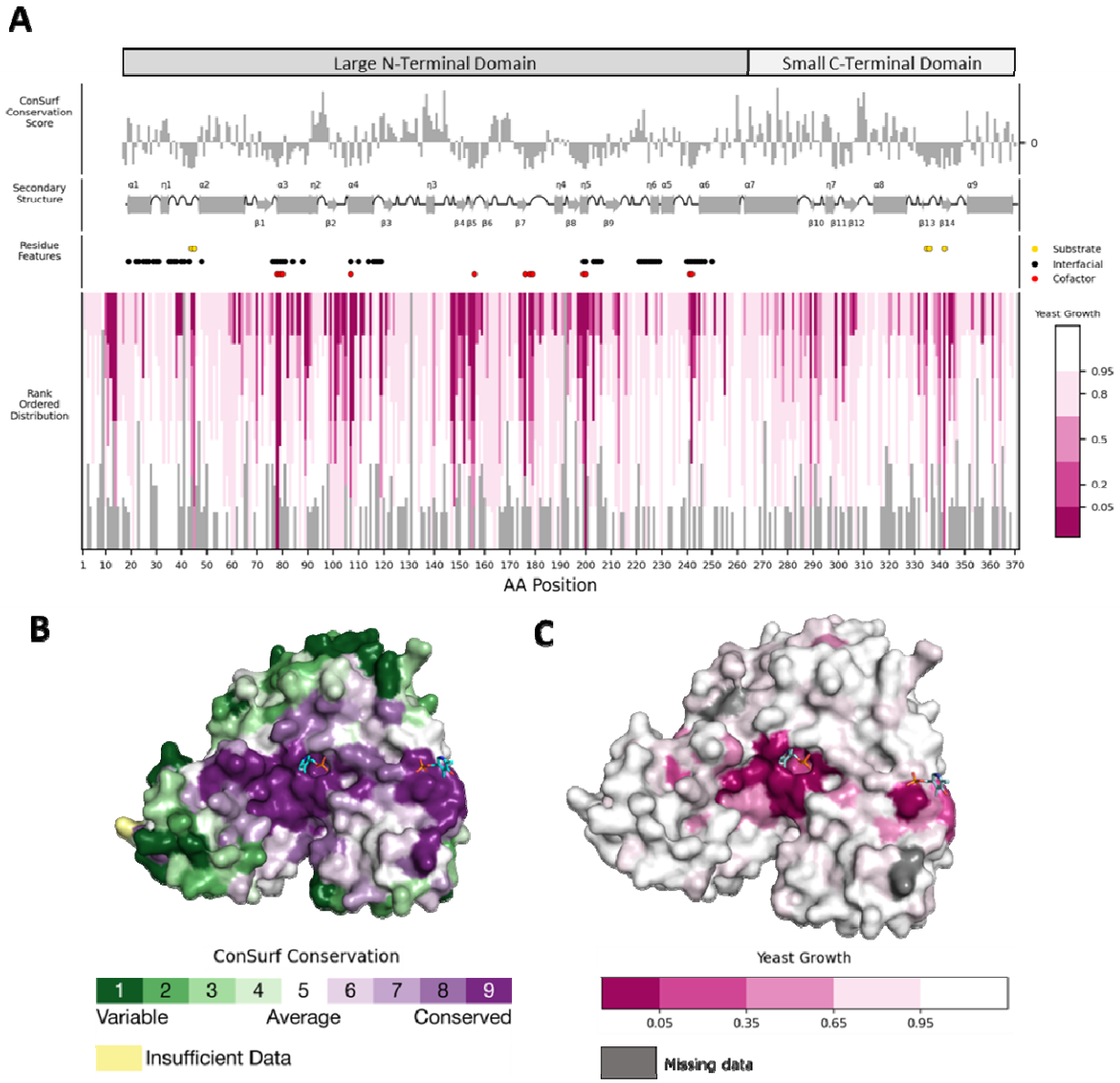


Figure 4. Missense variant effect map across the length of human PSAT (PDB: 3e77; L17-L370) compared to structural features and conservation. **A)** The bottom most plot represents the rank ordered distribution of variant growth at each amino acid position. Each stacked bar depicts the effect size of missense variants, grouped into normalized mean haploid growth values bins of 30% (righthand colorbar), and ordered from most impaired (top) to least impaired (bottom) growth bins. Variants that are not impaired relative to wild type (≥ 95%) are colored as white. Variants with growth values that were impaired relative to wild type (≤ 95%) are colored in shades of pink, with darker shades indicating more impaired growth. Grey fields correspond to missing data. The height of each color in each stacked bar represents the number of variants of a given growth bin, over the total number of possible variants at that position. Overlayed above are the secondary structure of PSAT, ConSurf evolutionary conservation scores [38,39], subunit domains and functional residues. Residues implicated in substrate binding (yellow), formation of the dimer interface (black), or cofactor binding (red) are depicted by circles. Alpha-helices (α; cylinder), 3_{10}-helices (η; cylinder), beta-sheets (β; arrows) are colored in gray and labeled, with turns shown as upward half-coils. **B-C)** Subunit surface colored by **B)** the ConSurf evolutionary conservation [38,39] grade and **C)** median yeast growth estimate per position. Two PLP cofactors are shown in each surface representation, to indicate the other interfacial active site bound to the opposite subunit (not depicted).

381 **Functional impact of missense variation at the active site**

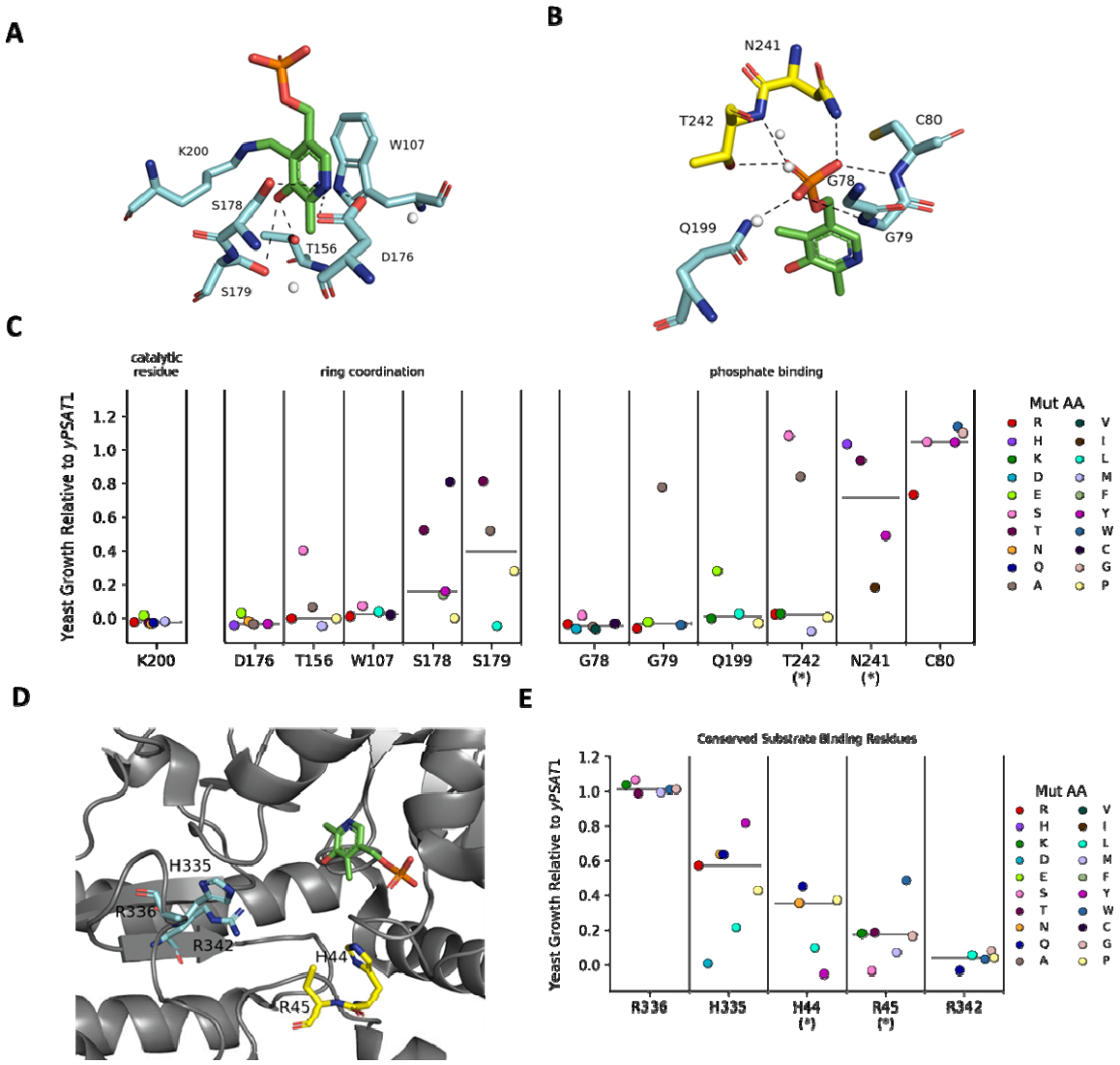
382 To compare our results to known structural features of human PSAT, we examined the impact
383 of amino acid substitutions at active site residues known bind the PLP cofactor, and a subset of
384 conserved residues that have been found to bind substrates in other phosphoserine
385 aminotransferases. First, we considered 12 residues that directly interacted with functional
386 groups of PLP: the aldehyde group, the pyridoxal ring, and the phosphate group (**Figure 5.A** and
387 **Figure 5.B**). The role of the aldehyde group in forming the initial Schiff base linkage with the
388 catalytic lysine, and the role of the pyridoxal ring in stabilizing reaction intermediates though
389 delocalization, are well established for PLP-dependent enzymes [41,43,44]. Each of the six SNV-
390 accessible missense substitutions at the catalytic lysine (K200) resulted in null activity (scores ≤
391 5%) (**Figure 5.C**). For ring coordination residues involved in pi-stacking (W107), or hydrogen
392 bonding to the phenol (T156, S179) or nitrogen (D176, S178) groups of PLP, the majority of
393 tested substitutions (15/23) displayed null-like activity (**Figure 5.C**). Notably, substitutions
394 between the biochemically similar amino acids of threonine and serine at these positions
395 resulted in less-impaired growth scores (T156S = 40%, S178T = 52%, S179T = 81%). Only two
396 positions, S178 and S179, displayed median variant growth above the null level but most
397 substitutions were still strongly impaired (S178 median score = 16%, S179 median score = 40%).

398 The role of the phosphate group of PLP is less well characterized for PLP-dependent enzymes. It
399 may act as anchor point for the protein [51], or even directly interact with nearby substrates in
400 [44]. In PSAT, six interfacial residues, G78, G79, C80, Q199, N241*, and T242*(asterisk indicates
401 residues from opposite subunit), bind to the phosphate group (**Figure 5.B**). All substitutions at
402 G78, G79, and Q199 displayed functional impairment (≤ 95%), with Q199E (score = 28%) and
403 G79A (score = 78%) being the only two variants with activity above null (>5%) (**Figure 5.C**).
404 Broader distributions of variant effects are observed at N241* and T242*, with two variants
405 N241H (score = 103%), T242S (score = 108%) displaying no functional impairment relative to
406 wild type (**Figure 5.C**). Interestingly, most substitutions at the C80 residue display wild type like
407 activity (C80Y|S|G|W) with the only exception being C80R (growth score = 80%) (**Figure 5.C**).
408 This residue (C80) is typically found as a serine, threonine (not SNV-accessible in human), or
409 arginine in other organisms (ConSurf DataBase [38,39] profile for 3e77 subunit A). It has been
410 hypothesized that C80 contributes to the higher phosphoserine substrate affinity (Km = 5 µM)
411 observed in PSAT relative to other phosphoserine aminotransferases [46], and possibly the
412 positive charge introduction represented by C80R negatively impacts this binding.

413 However, like most aminotransferases, the specific mechanisms of substrate binding are less
414 understood for PSAT. As an alternative, we considered 5 substrate binding residues that were
415 identified in *Escherichia coli* [47], *Bacillus alcalophilus* [48], and *Arabidopsis thaliana* [49]
416 phosphoserine aminotransferases that had crystal structures of phosphoserine or α-methyl-L-
417 glutamate (analog) bound states (not available for human PSAT). Amino acid substitutions at all

418 but one of these substrate binding residues, corresponding to H335, R336, R342, H44* and
419 R45* (asterisk indicates residues from opposite subunit) in human PSAT, show strong loss of
420 function (**Figure 5.D and Figure 5.E**). The exception is R336, which displayed activity similar to
421 wild type for all tested substitutions. Thus, R336 may be a conserved residue that does not
422 directly participate in substrate stabilization in PSAT but does in other orthologs.

423



424

425 **Figure 5**. Functional impact of amino acid substitutions at active site residues. Stick representations for residues in
426 one subunit are colored in cyan, residues from the opposite subunit in the homodimer assembly are in yellow, and
427 the PLP cofactor in green in all panels where the crystal structure of human PSAT complexed with PLP (PDB: 3e77)
428 is visualized. **A-B)** Organization of PLP cofactor binding residues. Hydrogen bonding is depicted by dashed black
429 lines and nearby water molecules are represented by white spheres. **A)** Residues that bind or coordinate PLP's
430 heterocyclic ring, with the catalytic lysine shown covalently bound to PLP. **B)** Residues that bind to the phosphate
431 group of PLP. **C)** The distribution of variant growth scores for cofactor binding residues. Each circle is colored
432 according to the amino acid that it is substituted as shown (Mut AA). Asterisks here denote amino acids from the

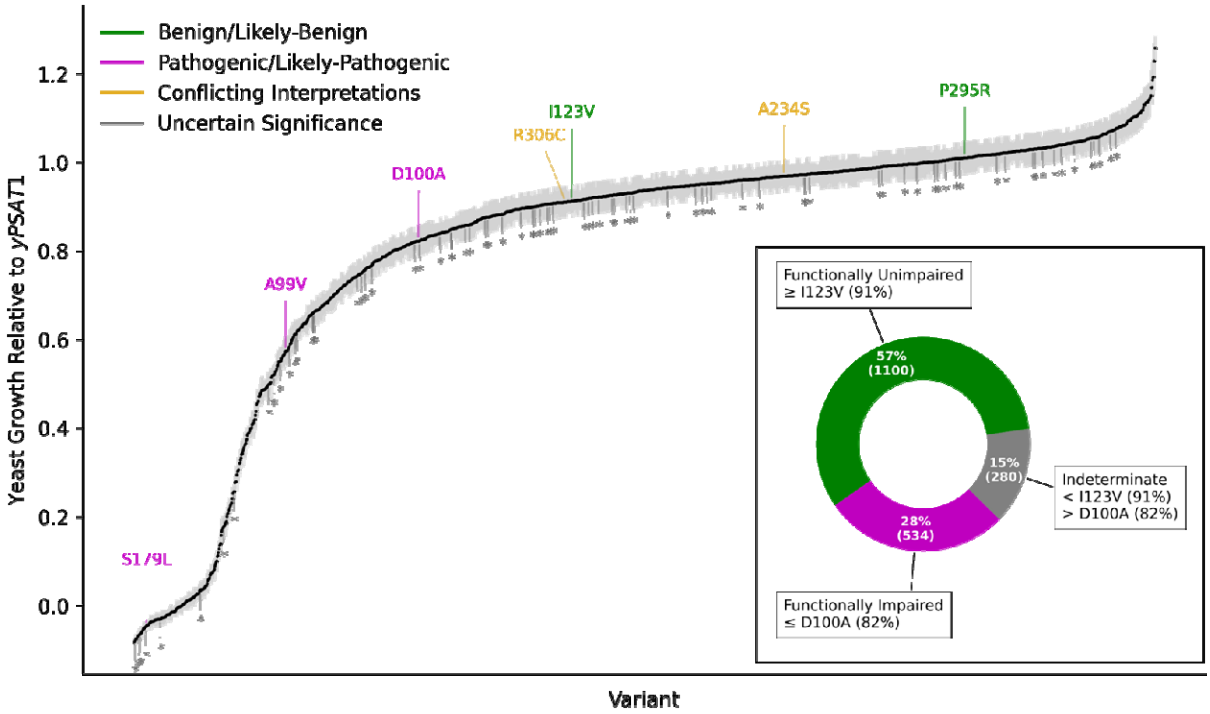
433 other subunit. The median growth score for each position is shown as a horizontal grey bar. **D)** Relative position of
434 highly conserved substrate binding residues to PLP, in the unbound (no substrate) crystal structure of human PSAT.
435 **E)** The distribution of functional effects for these conserved substrate binding residues, with the same coloring
436 scheme and annotation as in **C**.

437 **Mapping functional effects to clinical interpretations**

438 Although as a rare disease there are a limited number of pathogenic alleles, a number of those
439 alleles reside at these functionally important residues. We next compared our results to clinical
440 classifications for all corresponding missense variants in ClinVar. We expected that
441 substitutions derived from known pathogenic variants would produce strong reductions in PSAT
442 activity in our assay, while those derived from known benign variants would have activities
443 comparable to wild type or display only weak reductions in activity. Of the 90 amino acid
444 substitutions derived from variants currently in ClinVar, three (S179L, A99V, and D100A) are
445 derived from variants annotated as pathogenic/likely pathogenic, two (I123V and P295R) from
446 variants annotated as benign/likely-benign, and three (V149M, A234S, and R306C) from
447 variants with conflicting interpretations [3] (November 2022). Since our previous study [27],
448 while the small number of missense alleles with definitive clinical significance calls remained
449 largely unchanged, the number of missense VUS increased nearly 12-fold (from 7 to 82).
450 Consistent with expectation and our previous results [27], the pathogenic/likely pathogenic
451 substitutions all resulted in substantial functional impairment in our assay, with normalized
452 growth values of 82% or below. Similarly, the benign/likely-benign substitutions demonstrated
453 little impairment, with normalized growth values above 91% (**Figure 6**).

454 To facilitate the use of our functional data as supporting evidence for human sequence variant
455 annotation [24], we used the activities of the substitutions with associated clinical significance
456 calls in a simple thresholding approach, similar that that used in our previous study [27]. We
457 defined a functionally impaired range for scores ≤82% growth in our assay, which is less than or
458 equal to the scores of substitutions derived from known pathogenic variants (**Figure 6**). We also
459 defined a functionally unimpaired range for scores ≥91%, which is greater than or equal to the
460 scores of substitutions derived from known benign variants (**Figure 6**). Because gain of function
461 mutations have not been associated with serine deficiency disorders, we included values
462 greater than wild type in the unimpaired range. Finally, because there are no clinically
463 annotated variants associated with substitutions having scores between 82% and 91%, this
464 range of the assay is of unknown clinical significance, and we have labeled this range
465 indeterminate. Within these ranges of our assay, 534 (28%) of all amino acid substitutions are
466 functionally impaired, 1,100 (57%) are functionally unimpaired, and 280 (15%) are
467 indeterminant (**Figure 6**). Of the 80 ClinVar missense variants currently annotated as VUS that
468 were tested in our assay, 22 result in substitutions that fall in our impaired range, 41 in our
469 unimpaired range and 17 in the intermediate range. Thus, our dataset provides functional

470 information for a large number of variants that lack clinical interpretation, including 77% of the
471 missense VUS currently listed in ClinVar.



472

473 **Figure 6**: Mutational scan of 1,914 PSAT1 unique SNV-accessible missense amino acid substitutions. Rank ordered
474 (lowest to highest) normalized variant growth. Black dots represent the mean normalized growth estimate for each
475 substitution, with light grey bars indicating standard errors. All current ClinVar annotations associated with tested
476 substitutions are labelled above their growth estimate as benign/likely-benign (green), pathogenic/likely-
477 pathogenic (magenta), conflicting interpretations (yellow), or as an asterisk below their growth estimate if they are
478 of uncertain significance (dark grey). The boxed panel depicts the stratification of normalized growth estimates,
479 based on thresholds derived from ClinVar clinical annotation, into functionally unimpaired (green), indeterminate
480 (gray), or functionally impaired (magenta) ranges.

481 **Experimental models of homozygous and compound heterozygous genotypes**

482 Loss of function mutations in *PSAT1* cause disease that ranges from severe (Neu-Laxova
483 syndrome 2, NLS2) to milder forms (PSAT deficiency, PSATD) of the disorder. These are
484 collectively referred to as NLS2/PSATD, or individually when clinical severity is discussed
485 specifically. Because NLS2/PSATD is an autosomal recessive disease, clinical manifestation
486 depends on the enzymatic function conferred by the combination of *PSAT1* alleles in the
487 patient's genome. The ability to generate stable diploid yeast cells by mating haploids allows us
488 to model diploid human *PSAT1* genotypes in yeast and quantitively assess the function of allele
489 pairs. As in our previous study [27], we sought to use our growth assay to functionally assess
490 pairwise combinations of protein-coding variants across all reported unique patient (and carrier
491 parent) genotypes currently available (**Supplemental Table S4**). Since our previous study [27],
492 six additional reports [20,52–56] have added descriptions of 20 new NLS2/PSATD patients and

twelve unique genotypes to the disease literature (**Supplemental Table S4**). To experimentally measure the functional impact of these allele combinations in our yeast assay, we constructed diploid strains harboring *yPSAT1* allele combinations encoding the same pair of PSAT1 amino acid sequences as the human genotypes. We then assayed the growth of these strains relative to homozygous wild type (*yPSAT1*) and null (*ser1*Δ0) on solid minimal medium lacking serine. Previously constructued *yPSAT1* diploid strains encoding homozygous A99V or S43R, and the compound heterzygote A99V / S179L [27], were also included for comparision. Although the growth of some strains modelling patient genotypes were close to those recpitulating some carrier genotypes, all modelled patient genotypes (**Supplemental Table S5**) displayed reduced growth relative to their respective carrier parents and the homozygous wildtype (**Figure 7**.**B**). Consistent with our previous study [27], we also observed good agreement between the degree of *PSAT1* functional impairment in our diploid assay and disease severity (**Figure 7**.**B**). Diploids corresponding to genotypes from NLS2 patients had normalized growth from 0% to 77%, and diploids corresponding to genotypes from PSATD patients had a higher normalized growth range of 70% to 88%.

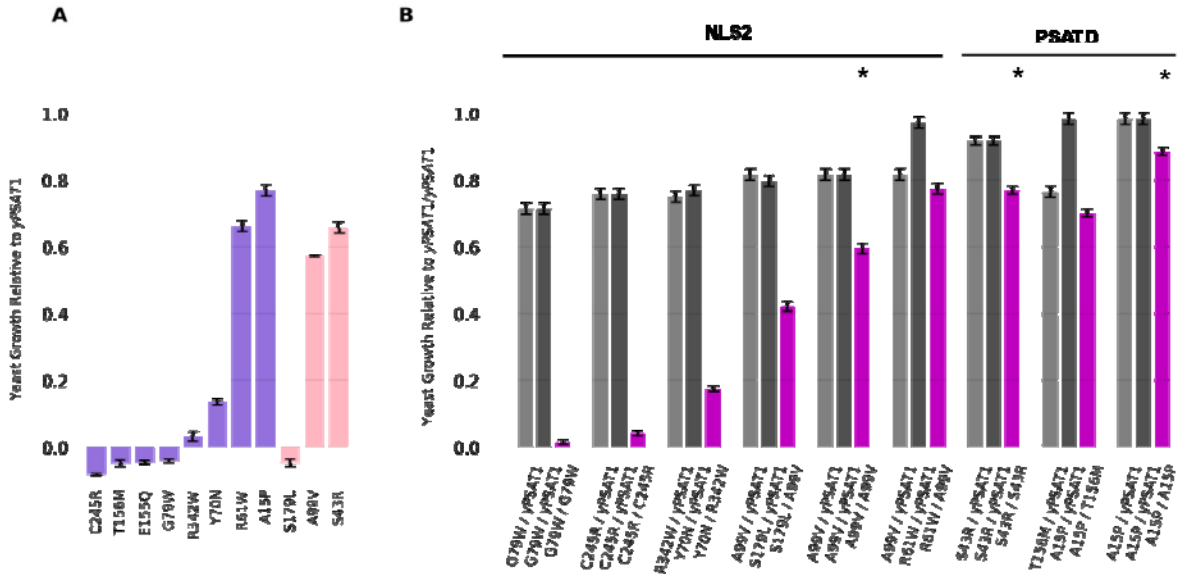


Figure 7: Modeling PSAT activity in patient genotypes. Both bar plots show the mean normalized growth estimate and standard errors. **A)** Haploid growth estimates for novel (purple) [20,52,54,56] and established [18,57] (pink) missense alleles in patient amino acid substitutions. **B)** Diploid yeast models of patient trios. Carrier parents are shown in light and dark grey, and the offspring in magenta. Recurrent carrier genotypes are repeated for ease of reference relative to the offspring. The clinical severity (NLS2 or PSATD) associated with patient genotypes is indicated. Asterisks denote offspring genotypes that have been reported in multiple unrelated families.

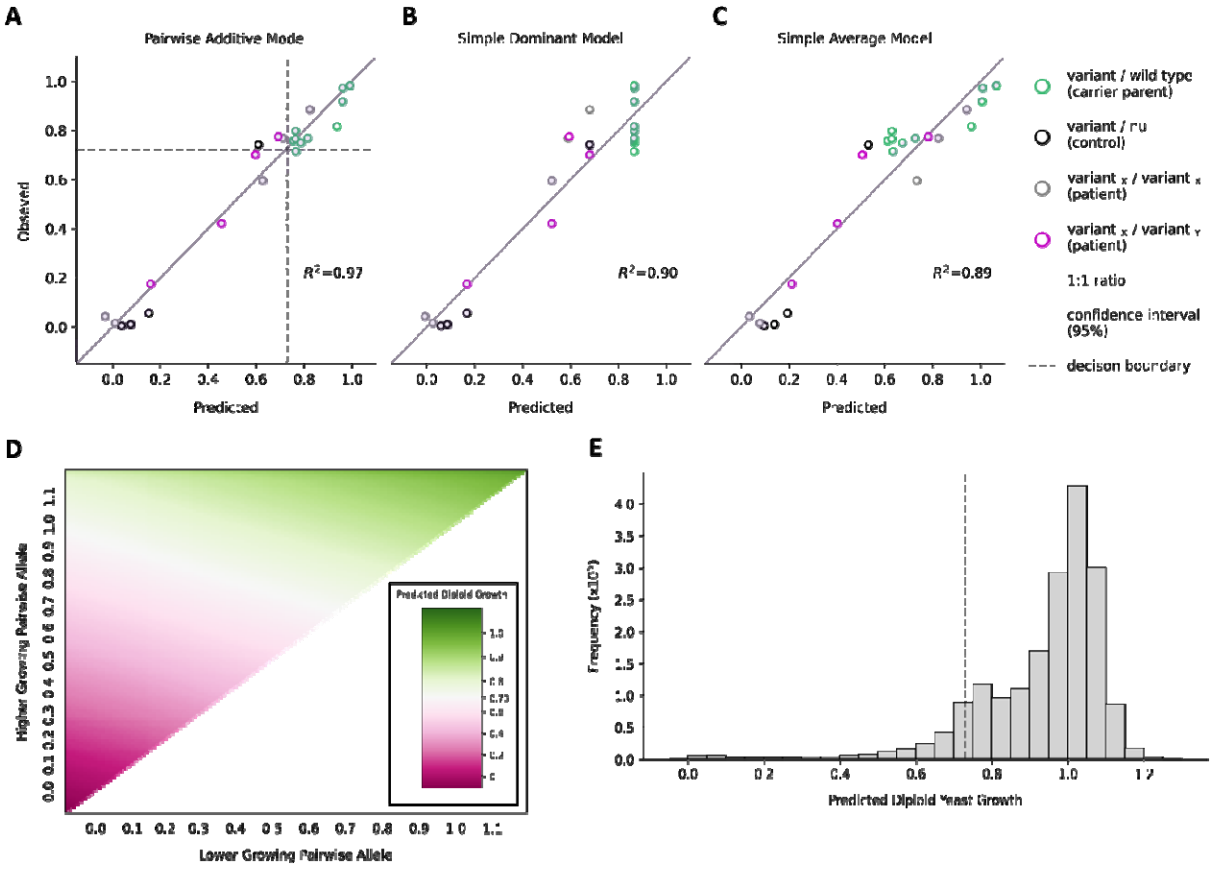
Our assay results are also consistent with clinical stratification within the group of NLS2 and PSATD patients. For PSATD, the A15P / A15P homozygote is the least impaired in our assay (77% growth) and is also the mildest form of PSATD reported to date. A recent report [56] described two unrelated teenagers that, other than congenital or childhood ichthyosis, were

519 developmentally normal and displayed neuropathy onset at ages 16 and 17. For NLS2, the two
520 genotypes with the lowest growth in our assay (G79W / G79W and C245R / C245R) correspond
521 to patients displaying severe NLS2 [20]. Three of the remaining NLS2 genotypes modelled here
522 (Y70N / R342W, A99V / S179L, A99V / R61W) display higher growth in our assay and
523 correspond to patients described as having moderate NLS expression [18,20,52]. Among these,
524 Y70N / R342W was the most impaired in our assay (18% growth) and the patient had a
525 postnatal survival of 8 weeks [52], while the individual harboring R61W / A99V (77% growth,
526 least impaired) had the longest NLS2-associated postnatal survival described to date (4 months)
527 [20]. Patients with genotypes corresponding to the A99V homozygote (60% growth), displayed
528 variable postnatal survival ranging from 1 day to 9 weeks [18–20]. Together, these results
529 highlight the potential value of our model organism assay for variant interpretation in the
530 context of a diploid *PSAT1* genotype.

531 **Predicting biallelic effects from individual allele measurements**

532 Experimentally assessing the level of PSAT activity associated with all allele-pairs would require
533 the construction of 1.83M diploid yeast strains. As a more feasible alternative, we evaluated
534 whether a relationship exists between individual haploid estimates and their resulting diploid
535 growth in combination. For homozygotes, a linear relationship was seen between haploid and
536 diploid growth estimates (**Figure 7.A** and **Figure 7.B** and **Supplemental Figure S2**; R^2=0.93). To
537 extend this analysis to heterozygotes, we fit a model to our experimental diploid dataset that
538 includes patient and carrier genotypes (n = 23), predicting diploid growth as a linear
539 combination of the allele with the higher and the allele with the lower haploid growth values
540 (**Figure 8.A**). This model assumed a uniform degree of dominance of the higher growing allele
541 over the lower growing allele and explained 97% of variance in the growth of the diploids. To
542 compare how this model performed relative to alternatives with simpler assumptions, we
543 generated models that assumed either complete dominance of the higher-growing allele or an
544 equal impact of both alleles (**Figure 8.B** and **Figure 8.C**). Our full model performed significantly
545 better than either the alternative models (Anova, Df=1; p=<5.4x10^{-6} and p<2.3x10^{-6},
546 respectively). Leave one out cross validation of the pairwise model displayed an excellent ability
547 to predict diploid growth from haploid measurements (RMSE = 0.0698 and MAE=0.0566).

548



549

550 **Figure 8**: Experimental and predicted biallelic growth values as clinical classifiers. Observed versus fitted values
551 shown for three linear regression models (pairwise additive **(A)**, simple dominant **(B)**, and simple average **(C)**)
552 predicting diploid growth (allele pairs) as a function of haploid growth (single alleles). Each circle represents a
553 unique biallelic combination. Patient, carrier, and control strains are labeled as indicated. A diagonal line of perfect
554 correspondence (1:1 ratio) and the coefficient of determination (R^2) for each model are included for ease of
555 comparison. Decision boundaries for observed (72%) versus predicted (73%) diploid growth as a binary classifier
556 for identifying patient genotypes modeled in our diploid assay are shown as horizontal and vertical dotted grey
557 lines in **(A),** respectively. Confidence intervals (95%) for each model are shaded in purple. **D)** Heatmap of the
558 predicted diploid growth value for all possible pairwise combinations of haploid allele measurements. **E)** Frequency
559 of predicted diploid growth values (in 5% bins) for all ~1.8M combinations of alleles. The decision boundary of 73%
560 is shown as a vertical dotted line.

561 We next compared the results of applying logistic regression to experimental and predicted
562 diploid growth values to produce binary classifiers distinguishing (low growing) genotypes from
563 patients diagnosed with NLS2/PSATD (n=9) versus (high growing) carrier parent genotypes in
564 our trio models (n=10). The threshold growth values at the decision boundary (p=0.5) were
565 similar for both approaches: 72% for experimental and 73% for predicted diploid growth.
566 Although the decision boundary using predicted values performed better and misclassified only
567 one patient genotype (A15P / A15P) versus a total of 4 for the model using experimental values
568 (**Figure 8.A**). It is likely the haploid estimates used to make predictions on diploid growth were
569 less noisy as they were generated from a larger number of measurements compared to our set
570 of experimental diploids. These results suggest that diploid growth values predicted from
571 haploid measurements perform well in clinical classification of human *PSAT1* genotypes.

572 On this basis, we computed the predicted diploid growth from all unique pairwise combinations
573 (~1.8M) of haploid estimates from 1,914 *yPSAT1* missense alleles, wild type (*yPSAT1*), and null
574 (*ser1*Δ0) using the pairwise-additive model (**Figure 8.A**). We then classified whether the
575 resulting values fell above or below the disease classification boundary of 73% growth. A large
576 majority (89%) of genotypes fell above the decision boundary threshold, suggesting these
577 genotypes would not lead to NLS2/PSATD (**Figure 8.D** and **Figure 8.E**). However, over 200K
578 *yPSAT1* biallelic combinations displayed predicted diploid growth below the classification
579 threshold, consistent with these genotypes resulting in disease (**Figure 8.E**).

580 **Discussion**

581 Here we presented a comprehensive functional analysis of SNVs in human PSAT. While this data
582 shows strong agreement with the structure-function features of the enzyme, clinical
583 interpretation relies both on knowledge of functional impact of a variant and the relationship to
584 clinical data. Specifically, we leveraged the small number of variants with existing clinical
585 interpretation to provide evidence for the likely clinical impact of other variants in similar
586 ranges of the assay. As additional clinical information become available through identification
587 of new patients and additional research into the natural history of the disease, the ability to
588 make clinical inferences using our dataset will improve. In the three years since our previous

589 study [27], additional patients were described in the disease literature. Not only was there good
590 agreement between the results of our assay and disease severity, but this additional clinical
591 information also allowed us to extend the range of our assay associated with clinical outcomes.

592 Given that the therapeutic potential for this medically actionable disorder is highest before
593 patients become symptomatic, clinical sequencing can be a strong diagnostic tool if variant
594 interpretation is also informative. Sequencing based diagnostics are becoming more widely
595 applied in newborns, prenatally [58], and as a part of programs like Undiagnosed Disease
596 Program [59,60] or Deciphering Developmental Disorders [61]. The most recent AMCG
597 guidelines [62] recommend exome or genome sequencing as a first or second tier option for
598 patients displaying early neonatal (<1 year) congenital abnormalities, or for patients exhibiting
599 developmental delay or intellectual disability before the age of 18. Furthermore, large scale
600 genome sequencing efforts like the All of US research program and UK biobank [63] are
601 aggregating a depth of population level genomic and longitudinal health data. Given these
602 developments, studies such as ours that provide information both about individual alleles and
603 the functional impact of allele pairs will play an increasingly important role in realizing the
604 promise of precision medicine.

605 **References**

606 1. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational
607 constraint spectrum quantified from variation in 141,456 humans. Nature. Nature Research;
608 2020;581:434–43.

609 2. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant
610 interpretation using population databases: Lessons from gnomAD. Hum Mutat [Internet]. 2021;
611 Available from: https://onlinelibrary.wiley.com/doi/10.1002/humu.24309

612 3. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving
613 access to variant interpretations and supporting evidence. Nucleic Acids Res. Oxford University
614 Press; 2018;46:D1062–7.

615 4. Illsinger S, Das AM. Impact of selected inborn errors of metabolism on prenatal and neonatal
616 development. IUBMB Life. 2010;62:403–13.

617 5. Ferreira CR, Rahman S, Keller M, Zschocke J, Abdenur J, Ali H, et al. An international
618 classification of inherited metabolic disorders (ICIMD). J Inherit Metab Dis. John Wiley and Sons
619 Inc; 2021;44:164–77.

620 6. Sanderson S, Green A, Preece MA, Burton H. The incidence of inherited metabolic disorders
621 in the West Midlands, UK. Arch Dis Child. 2006;91:896–9.

622 7. Dionisi-Vici C, Rizzo C, Burlina AB, Caruso U, Sabetta G, Uziel G, et al. Inborn errors of
623 metabolism in the Italian pediatric population: a national retrospective survey. J Pediatr.
624 2002;140:321–7.

625 8. Gambello MJ, Li H. Current strategies for the treatment of inborn errors of metabolism.
626 Journal of Genetics and Genomics. Institute of Genetics and Developmental Biology; 2018. p.
627 61–70.

628 9. Morris AAM, Kožich V, Santra S, Andria G, Ben-Omran TIM, Chakrapani AB, et al. Guidelines
629 for the diagnosis and management of cystathionine beta-synthase deficiency. J Inherit Metab
630 Dis. Springer Netherlands; 2017. p. 49–74.

631 10. Stockler S, Plecko B, Gospe SM, Coulter-Mackie M, Connolly M, van Karnebeek C, et al.
632 Pyridoxine dependent epilepsy and antiquitin deficiency. Mol Genet Metab. 2011;104:48–60.

633 11. El-Hattab AW. Serine biosynthesis and transport defects. Mol Genet Metab. Academic Press
634 Inc.; 2016. p. 153–9.

635 12. de Koning TJ, Klomp LWJ, Oppen ACCV, Beemer PFA, Dorland L, Berg IETV den, et al.
636 Prenatal and early postnatal treatment in 3-phosphoglycerate-dehydrogenase deficiency.
637 Lancet. Elsevier B.V.; 2004;364:2221–2.

638 13. Jaeken J, Detheux M, van Maldergem L, Foulon M, Carchon H, van Schaftingen E, et al. 3-
639 Phosphoglycerate dehydrogenase deficiency: an inborn error of serine biosynthesis. Archives
640 ofDisease in Childhood. 1996.

641 14. Jaeken J, Detheux M, Fryns JP, Collet JF, Alliet P, van Schaftingen E. Phosphoserine
642 phosphatase deficiency in a patient with Williams syndrome. J Med Genet. BMJ Publishing
643 Group; 1997;34:594–6.

644 15. de Koning TJ, Duran M, van Maldergem L, Pineda M, Dorland L, Gooskens R, et al.
645 Congenital microcephaly and seizures due to 3-phosphoglycerate dehydrogenase de¢ciency:
646 Outcome of treatment with amino acids. J. Inherit. Metab. Dis. 2002.

647 16. Hart CE, Race V, Achouri Y, Wiame E, Sharrard M, Olpin SE, et al. Phosphoserine
648 aminotransferase deficiency: A novel disorder of the serine biosynthesis pathway. Am J Hum
649 Genet. University of Chicago Press; 2007;80:931–7.

650 17. de Koning TJ. Amino acid synthesis deficiencies. J Inherit Metab Dis. Springer Netherlands;
651 2017. p. 609–20.

652 18. Acuna-Hidalgo R, Schanze D, Kariminejad A, Nordgren A, Kariminejad MH, Conner P, et al.
653 Neu-laxova syndrome is a heterogeneous metabolic disorder caused by defects in enzymes of
654 the l-serine biosynthesis pathway. Am J Hum Genet. Cell Press; 2014;95:285–93.

655 19. El-Hattab AW, Shaheen R, Hertecant J, Galadari HI, Albaqawi BS, Nabil A, et al. On the
656 phenotypic spectrum of serine biosynthesis defects. J Inherit Metab Dis. Springer Netherlands;
657 2016;39:373–81.

658 20. Abdelfattah F, Kariminejad A, Kahlert AK, Morrison PJ, Gumus E, Mathews KD, et al.
659 Expanding the genotypic and phenotypic spectrum of severe serine biosynthesis disorders. Hum
660 Mutat. John Wiley and Sons Inc.; 2020;41:1615–28.

661 21. Moat S, Carling R, Nix A, Henderson M, Briddon A, Prunty H, et al. Multicentre age-related
662 reference intervals for cerebrospinal fluid serine concentrations: Implications for the diagnosis
663 and follow-up of serine biosynthesis disorders. Mol Genet Metab. 2010;101:149–52.

664 22. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J, et al. An extended set of yeast-
665 based functional assays accurately identifies human disease mutations. Genome Res.
666 2016;26:670–80.

667 23. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant
668 Interpretation: Functional Assays to the Rescue. Am J Hum Genet. Cell Press; 2017. p. 315–25.

669 24. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for
670 the interpretation of sequence variants: A joint consensus recommendation of the American
671 College of Medical Genetics and Genomics and the Association for Molecular Pathology.
672 Genetics in Medicine. Nature Publishing Group; 2015;17:405–24.

673 25. Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic
674 variants. Nat Protoc. Nature Publishing Group; 2016. p. 1782–7.

675 26. Weile J, Roth FP. Multiplexed assays of variant effects contribute to a growing genotype–
676 phenotype atlas. Hum Genet. Springer Verlag; 2018. p. 665–78.

677 27. Sirr A, Lo RS, Cromie GA, Scott AC, Ashmead J, Heyesus M, et al. A yeast-based
678 complementation assay elucidates the functional impact of 200 missense variants in human
679 PSAT1. J Inherit Metab Dis. John Wiley and Sons Inc.; 2020;43:758–69.

680 28. Winston F, Dollard C, Ricupero-Hovasse SL. Construction of a set of convenient
681 saccharomyces cerevisiae strains that are isogenic to S288C. Yeast. Chichester, UK: John Wiley
682 & Sons, Ltd; 1995;11:53–5.

683 29. Rose M, Winston F, and Hieter P. Methods in Yeast Genetics: A Laboratory Course Manual.
684 Cold Spring Harbor Laboratory Press; 1990.

685 30. Farrell CM, O'Leary NA, Harte RA, Loveland JE, Wilming LG, Wallin C, et al. Current status
686 and new features of the Consensus Coding Sequence database. Nucleic Acids Res.
687 2014;42:D865–72.

688 31. Illuxley C, Green ED, Dunbam I. Rapid assessment of S. cerevisiae mating type by PCR.
689 Trends in Genetics. 1990;6:236.

690 32. Lo RS, Cromie GA, Tang M, Teng K, Owens K, Sirr A, et al. The functional impact of 1,570
691 SNP-accessible missense variants in human OTC. Available from:
692 https://doi.org/10.1101/2022.10.26.513893

693 33. Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, et al. A series of PDB-
694 related databanks for everyday needs. Nucleic Acids Res. 2015;43:D364-8.

695 34. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of
696 hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577–637.

697 35. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
698 protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

699 36. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold
700 Protein Structure Database: massively expanding the structural coverage of protein-sequence
701 space with high-accuracy models. Nucleic Acids Res. 2022;50:D439–44.

702 37. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available
703 Python tools for computational molecular biology and bioinformatics. Bioinformatics.
704 2009;25:1422–3.

705 38. Goldenberg O, Erez E, Nimrod G, Ben-Tal N. The ConSurf-DB: pre-calculated evolutionary
706 conservation profiles of protein structures. Nucleic Acids Res. 2009;37:D323-7.

707 39. ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, et al. ConSurf-DB: An
708 accessible repository for the evolutionary conservation patterns of the majority of PDB
709 proteins. Protein Sci. 2020;29:258–67.

710 40. Murtas G, Marcone GL, Sacchi S, Pollegioni L. L-serine synthesis via the phosphorylated
711 pathway in humans. Cellular and Molecular Life Sciences. Springer Science and Business Media
712 Deutschland GmbH; 2020. p. 5131–48.

713 41. Koper K, Han SW, Pastor DC, Yoshikuni Y, Maeda HA. Evolutionary origin and functional
714 diversification of aminotransferases. Journal of Biological Chemistry. American Society for
715 Biochemistry and Molecular Biology Inc.; 2022.

716 42. John RA. Pyridoxal phosphate-dependent enzymes. Biochim Biophys Acta. 1995;1248:81–
717 96.

718 43. Eliot AC, Kirsch JF. Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary
719 considerations. Annu Rev Biochem. 2004. p. 383–415.

720 44. Liang J, Han Q, Tan Y, Ding H, Li J. Current advances on structure-function relationships of
721 pyridoxal 5'-phosphate-dependent enzymes. Front Mol Biosci. Frontiers Media S.A.; 2019.

722 45. Youn BAEK J, Youn JUN D, Taub D, Ho KIM Y. Characterization of human phosphoserine
723 aminotransferase involved in the phosphorylated pathway of L-serine biosynthesis. Biochem. J.
724 2003.

725 46. Singh RK, Kumar D, Gourinath S. Phosphoserine Aminotransferase has Conserved Active Site
726 from Microbes to Higher Eukaryotes with Minor Deviations. Protein Pept Lett. Bentham Science
727 Publishers Ltd.; 2021;28:996–1008.

728 47. Hester G, Stark W, Moser M, È rg Kallen Zora Markovic Â-Housley J, Jansonius JN. Crystal
729 Structure of Phosphoserine Aminotransferase from Escherichia coli at 2.3 A Ê Resolution:
730 Comparison of the Unligated Enzyme and a Complex with a a a-Methyl-L-Glutamate. J. Mol.
731 Biol. 1999.

732 48. Battula P, Dubnovitsky AP, Papageorgiou AC. Structural basis of l-phosphoserine binding to
733 Bacillus alcalophilus phosphoserine aminotransferase. Acta Crystallogr D Biol Crystallogr.
734 2013;69:804–11.

735 49. Sekula B, Ruszkowski M, Dauter Z. Structural analysis of phosphoserine aminotransferase
736 (isoform 1) from arabidopsis Thaliana– the enzyme involved in the phosphorylated pathway of
737 serine biosynthesis. Front Plant Sci. Frontiers Media S.A.; 2018;9.

738 50. Singh RK, Tomar P, Dharavath S, Kumar S, Gourinath S. N-terminal residues are crucial for
739 quaternary structure and active site conformation for the phosphoserine aminotransferase
740 from enteric human parasite E. histolytica. Int J Biol Macromol. Elsevier B.V.; 2019;132:1012–
741 23.

742 51. Denesyuk AI, Denessiouk KA, Korpela T, Johnson MS. Functional attributes of the phosphate
743 group binding cup of pyridoxal phosphate-dependent enzymes. J Mol Biol. Academic Press;
744 2002;316:155–72.

745 52. Ni C, Cheng RH, Zhang J, Liang JY, Wei RQ, Li M, et al. Novel and recurrent PHGDH and
746 PSAT1 mutations in Chinese patients with Neu-Laxova syndrome. European Journal of
747 Dermatology. John Libbey Eurotext; 2019;29:641–6.

748 53. Shapira Zaltsberg G, McMillan HJ, Miller E. Phosphoserine aminotransferase deficiency:
749 imaging findings in a child with congenital microcephaly. Journal of Maternal-Fetal and
750 Neonatal Medicine. Taylor and Francis Ltd; 2020;33:1033–5.

751 54. Debs S, Ferreira CR, Groden C, Kim HJ, King KA, King MC, et al. Adult diagnosis of congenital
752 serine biosynthesis defect: A treatable cause of progressive neuropathy. Am J Med Genet A.
753 John Wiley and Sons Inc; 2021;185:2102–7.

754 55. Serrano Olave A, Padín López A, Martín Cruz M, Monís Rodríguez S, Narbona Arias I, Jiménez
755 López JS. Prenatal Diagnosis of Neu–Laxova Syndrome. Diagnostics. MDPI AG; 2022;12:1535.

756 56. Shen Y, Peng Y, Huang P, Zheng Y, Li S, Jiang K, et al. Juvenile-onset PSAT1-related
757 neuropathy: A milder phenotype of serine deficiency disorder. Front Genet. Frontiers Media SA;
758 2022;13.

759 57. Brassier A, Valayannopoulos V, Bahi-Buisson N, Wiame E, Hubert L, Boddaert N, et al. Two
760 new cases of serine deficiency disorders treated with l-serine. European Journal of Paediatric
761 Neurology. W.B. Saunders Ltd; 2016;20:53–60.

762 58. Fu F, Li R, Yu Q, Wang D, Deng Q, Li L, et al. Application of exome sequencing for prenatal
763 diagnosis of fetal structural anomalies: clinical experience and lessons learned from a cohort of
764 1618 fetuses. Genome Med. BioMed Central Ltd; 2022;14.

765 59. Gahl WA, Wise AL, Ashley EA. The Undiagnosed Diseases Network of the National Institutes
766 of Health: A National Extension. JAMA. 2015;314:1797–8.

767 60. Macnamara EF, D'Souza P, Tifft CJ. The undiagnosed diseases program: Approach to
768 diagnosis. Ferreira C, editor. Transl Sci Rare Dis [Internet]. 2019;4:179–88. Available from:
769 https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/TRD-190045

770 61. Firth H v, Wright CF, DDD Study. The Deciphering Developmental Disorders (DDD) study.
771 Dev Med Child Neurol. 2011;53:702–3.

772 62. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome
773 and genome sequencing for pediatric patients with congenital anomalies or intellectual
774 disability: an evidence-based clinical guideline of the American College of Medical Genetics and
775 Genomics (ACMG). Genetics in Medicine. Springer Nature; 2021;23:2029–37.

776 63. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource
777 with deep phenotyping and genomic data. Nature. 2018;562:203–9.

778

779 **Declarations**

780 No human or animal subjects were used in this work.

781 **Competing interests**

782 A.M.D. is a scientific advisor to and has a financial interest in Fenologica Biosciences, Inc., a
783 company that develops instrumentation and analytical tools for the analysis of microbial
784 colonies.

785 **Funding**

786 This work was funded in part by NIH/ NIGMS award R01 GM134274 to A.M.D.

787 **Acknowledgements**

788 We thank Douglas Fowler, Maitreya Dunham, Fowzan Alkuraya, and Ayman El Hattab for
789 helpful discussions.

790