HHS Public Access

Author manuscript

J Inherit Metab Dis. Author manuscript; available in PMC 2021 July 01.

Published in final edited form as:

J Inherit Metab Dis. 2020 July ; 43(4): 758–769. doi:10.1002/jimd.12227.

A yeast-based complementation assay elucidates the functional impact of 200 missense variants in human *PSAT1*

Amy Sirr[1], **Russell S. Lo**[1], **Gareth A. Cromie**[1], **Adrian C. Scott**[1], **Julee Ashmead**[1], **Mirutse Heyesus**[1], **Aimée M. Dudley**[1,*]

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

Abstract

Background: Defects in serine biosynthesis caused by 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 prenatally, can ameliorate and in some cases even prevent symptoms, knowledge of pathogenic variants is medically actionable.

Methods: Here, we describe a functional assay that leverages the evolutionary conservation of an enzyme in the serine biosynthesis pathway, phosphoserine aminotransferase, and the ability of the human protein coding sequence (*PSAT1*) to functionally replace its yeast ortholog (*SER1*).

Results: Results from our quantitative, yeast-based assay agree well with clinical annotations and expectations based on the disease literature. Using this assay, we have measured the functional impact of the 199 *PSAT1* variants currently listed in ClinVar, gnomAD, and the literature.

Conclusions: We anticipate that the assay could be used to comprehensively assess the functional impact of all SNP-accessible amino acid substitution mutations in *PSAT1,* a resource that could aid variant interpretation and identify potential NLS carriers.

[*]**Corresponding Author:** Aimée M. Dudley, Ph.D., Pacific Northwest Research Institute, 720 Broadway, Seattle, WA 98122, aimee.dudley@gmail.com.

Author contributions

Amy Sirr, Russell Lo, Adrian Scott, Gareth Cromie, and Aimée Dudley conceived and designed the experiments.
Amy Sirr, Russell Lo, Julee Ashmead and Mirutse Heyesus performed the experiments.
Amy Sirr, Russell Lo, Adrian Scott and Gareth Cromie analyzed the data.
Gareth Cromie performed the bioinformatics.
Amy Sirr, Russell Lo, Adrian Scott, Gareth Cromie and Aimée Dudley wrote the paper.

Competing Interests

Aimée Dudley is a scientific advisor with a financial interest in Fenologica Biosciences, Inc., a company that develops instrumentation and analytical tools for the analysis of microbial colonies.
Adrian Scott has a financial interest in Fenologica Biosciences, Inc.
Gareth Cromie received funding from Twist Bioscience, Inc., which paid for his travel and accommodations while attending the Twist Bioscience User Group Meeting at PEGS Europe 2019.
Amy Sirr, Russell S. Lo, Julee Ashmead and Mirutse Heyesus declare that they have no competing interests.

Informed Consent and Animal Rights

This article does not contain any studies with human or animal subjects performed by the any of the authors.

Keywords

Yeast; Serine Deficiency; Neu-Laxova Syndrome; Inborn Errors of Metabolism; Variant of Uncertain Significance; Human Gene Complementation

INTRODUCTION

Inborn errors of metabolism (IEMs) are a class of genetic diseases for which knowledge of pathogenic variants often has high clinical utility.[1–4] Although dietary restriction or supplementation can positively impact patient health, timely recognition and targeted treatments are crucial. Medical intervention for IEMs is confounded by the fact that many have common symptoms but radically different treatments. Unfortunately, most polymorphisms detected in human disease genes are variants of uncertain significance (VUS)[5; 6] that cannot inform diagnosis or treatment.[7] Thus, our limited ability to relate genotype to phenotype is a major obstacle for precision medicine. As the use of whole exome and genome sequencing (WES/WGS) in health care increases, the number of VUS will skyrocket.[8] While computational methods for predicting variant effect could scale with the problem, their use in variant interpretation is limited[7] because they often give inaccurate or conflicting results.[9] As such, large-scale functional assays are the only methods for variant interpretation currently poised to match the pace of variant discovery.[8]

A key component of functional assay design is choice of organism and cell type. Because cell types can differ in the amount of a specific biochemical activity they require, not all human cell types can be used to assay the function of all human genes. However, orthologous genes often perform similar (even identical) functions in different organisms, and many human protein coding sequences are able to functionally replace (complement) their orthologs in heterologous systems, such as yeast.[10–15] Results of two large-scale studies suggest that it may be possible to establish functional assays for 40–50% of the 4,368 human genes with a yeast ortholog, and this success rate is predicted to be even higher (80–90%) for metabolic genes.[14; 16] As such, yeast-based assays are a rapid, cost effective means of measuring the functional impact of human variation in numerous genes known to cause IEMs, and these “surrogate genetic”[13] platforms could provide much needed information to aid the interpretation of rare variants.[17]

Serine deficiency disorders are highly actionable IEMs.[18–20] These rare autosomal recessive diseases are caused by loss of function mutations in any of the three genes in the phosphorylated serine biosynthesis pathway (*PHGDH* [MIM: 606879], *PSAT1* [MIM: 610936], and *PSPH* [MIM: 172480]). Serine is central to processes such as cellular proliferation, folate metabolism, phospholipid metabolism and the formation of the neuromodulators D-serine and glycine. As a result, serine limitation has profound effects on central nervous system development and function, and serine deficiency disorders are characterized by severe neurological symptoms including microcephaly, intractable seizures and intellectual disability.[18; 19; 21–23] Case studies have shown that when diagnosed early, maternal and newborn serine supplementation may reduce and, in some instances, prevent the onset of these devastating symptoms.[24–29] However, because accurate assessment of

serine levels generally requires measuring metabolite concentrations in cerebrospinal fluid and comparison to closely age-matched controls,[20; 24; 30] early diagnosis relies heavily on family history or genotype information. Thus, this is a disease for which knowledge of the functional significance of a gene variant can drastically alter clinical outcome.

Deleterious mutations in the human phosphoserine aminotransferase gene (*PSAT1)*, which encodes the second enzyme in the serine biosynthesis pathway, are linked to the serine deficiency disorder known as PSAT1 deficiency (PSATD [MIM: 610992]),[25] or in the most severe cases Neu-Laxova Syndrome 2 (NLS2 [MIM: 616038]).[18] The PSAT1 enzyme is a pyridoxal phosphate (PLP) dependent homodimer that is highly conserved from bacteria to humans, and the human cDNA sequence is able to functionally replace its orthologous gene, *SER1*, in *Saccharomyces cerevisiae*.[31] We have developed a yeast-based complementation assay for assessing the functional effects of genetic variation in *PSAT1* and calibrated it by introducing disease alleles into yeast individually and in combinations that have been observed in PSATD and NLS2 (hereafter PSATD/NLS2) patients and their carrier parents. We have also assayed missense alleles from gnomAD[6] and ClinVar[32]. Our yeast assay results agree well with the variant interpretations curated in ClinVar and suggest that 22% of missense alleles in the r2.0.2 build of the gnomAD database are loss of function mutations with potential clinical relevance.

METHODS

Plasmid and strain construction

S. cerevisiae strains used in this study (Table S1) are derived from the isogenic strains FY4 (MAT**a**) and FY5 (MATα).[33] Unless noted, strains were grown at 30° C in rich (YPD) medium (1% yeast extract, 2% peptone, 2% glucose) or minimal (SD) medium (without amino acids, 2% glucose) using standard media conditions and methods for genetic manipulation.[34]

A list of known *PSAT1* disease alleles (Table 1) was compiled from published reports, the Online Mendelian Inheritance in Man (OMIM) database and the Saudi Human Genome Program (SHGP)[35] and regulatory, splicing, and copy number variants were removed. The *PSAT1* sequence used for this study is based on the amino acid sequence for *PSAT1* isoform 1 (GenBank: NP_478059.1) and was optimized for expression in yeast (hereafter *yPSAT1*, GenBank: MN654100, MN654101) using a custom method designed to match the codon usage frequency of the yeast ortholog as follows. At amino acid positions conserved between the yeast and human sequences, the codon sequence for the yeast ortholog, *SER1*, in the S288c reference sequence was selected.[36] In non-conserved locations, codons for the appropriate PSAT1 amino acid were chosen to be as similar as possible to the usage frequency of the codons at corresponding *SER1* positions. Alleles containing a frameshift mutation utilized this same *yPSAT1* sequence up to the frameshift and the remaining open reading frame (ORF) of human *PSAT1* isoform 1 cDNA (GenBank: KJ898759.1) thereafter such that the sequence between the frameshift and first stop codon in the new reading frame encodes the frameshifted human protein sequence.

To construct yeast strains harboring disease alleles in the context of *yPSAT1*, the *SER1* open reading frame was first deleted from the FY4 MAT**a** lab strain using a selectable drug marker[37] (Figure 1A and Table S1). Next, a series of plasmids containing the *yPSAT1* wild type (GenBank: MN654100) or genetic variants (Table S2), including known disease alleles, was constructed as follows. Synthetic *yPSAT1* constructs (IDT, Genewiz) (GenBank: MN654100 positions 1379–2809) were assembled (NEB HiFi Assembly) into the 4717 bp gel-purified vector backbone of *Bcl*I digested pAS-SER1-S.[37] All assembled PSAT1 plasmids include homology to sequence upstream of the *SER1* start codon, the *yPSAT1* variant, sequence encompassing the *SER1* transcriptional terminator, the kanMX4 selectable drug marker[38] and homology to the region downstream of the *SER1* terminator to direct integration into the genome. This 3.4 kb fragment was liberated from the vector backbone by *Sna*BI-*Sap*I double digestion and integrated into haploid *ser1* deletion (*ser1*Δ0) strains at the yeast *SER1* locus by homologous recombination. Two isolates that contained the allele of interest but lacked secondary mutations in either *yPSAT1* or the yeast regulatory sequences (Sanger sequencing, Genewiz) were selected for further analysis (Table S1). Haploid *yPSAT1* strains of the opposite mating type (used in the construction of diploid strains) were obtained by mating the MAT**a** strain containing the desired allele to the otherwise isogenic MATα lab strain (FY5) followed by standard tetrad analysis and Sanger sequencing (Genewiz) to confirm the inclusion of the correct *yPSAT1* allele (Table S1). *yPSAT1* compound heterozygous diploids were generated by mating MAT**a** and MATα strains harboring the appropriate *yPSAT1* alleles and micromanipulating zygotes.[34] At least two independent isolates of each *yPSAT1* genotype were confirmed to be diploids by PCR of the mating type locus (Table S1).

The *yPSAT1* variant library (Table S3), which included all missense gnomAD r2.0.2 *PSAT1* variants,[6] was synthesized as follows. Plasmid pAS59 (GenBank: MN654101) was constructed using the HiFi Assembly kit (NEB) to assemble a *Bam*HI, *Pst*I cut pUC19 backbone with a 3.1 kb *yPSAT1* construct. The *yPSAT1* construct includes 77 bp of *SER1* 5' UTR homology, the wild type *yPSAT1* ORF, 463 bp of the *SER1* terminator, the natMX drug marker,[39] and 168 bp of sequence downstream of the *SER1* terminator. The 3.8 kb fragment liberated by a *Bsp*QI-*Bsp*HI double digest of pAS59 was gel purified and used as the *yPSAT1* wild type template for library construction (Twist Bioscience). Each variant synthesized (Table S3) consists of the 3.1 kb *yPSAT1* construct (described above) with an additional 19 bp of *SER1* homology on both the 5' and 3' ends, which were added during variant library construction. This additional sequence (Table S4) was used to prime a 15-cycle PCR amplification that was specific for the variant constructs over the original (wild type) template. Approximately 500 ng of each PCR-amplified variant was transformed into a *ser1*Δ0 strain using standard methods. Four individual yeast transformants per variant were picked into 1 mL deep 96-well plates containing rich media supplemented with 100 μg/ml nourseothricin and grown for 2 days at 30° C. Sequence confirmation (by amplicon sequencing) is described below.

Four alleles (N180K, S285Y, H19D and T132R) were constructed by site-directed mutagenesis as follows. Primers incorporating the desired change (Table S4) were phosphorylated with T4 polynucleotide kinase (NEB) for 1 hour at 37° C under

manufacturer's conditions followed by a 20 minute 65° C heat inactivation. The phosphorylated oligos were used to PCR amplify the pAS59 template with the iProof polymerase (BioRad). Following *Dpn*I digestion, the PCR-amplified plasmid was transformed into *E. coli* (Table S2). The *yPSAT1* variants were then liberated from the plasmid backbone with *Not*I-HF (NEB). Ends were trimmed using Mung Bean Nuclease (NEB) prior to transformation into yeast. Sanger sequencing (Genewiz) was used to identify at least two isolates (Table S1) that contained the allele of interest but lacked secondary mutations in *yPSAT1* and the yeast regulatory sequences.

Variant library sequence confirmation

To confirm that the *yPSAT1* variants integrated into the yeast genome harbored the expected alleles but lacked any secondary mutations, we performed amplicon sequencing using the KiloSeq method.[40] Briefly, the portion of the expression construct containing the *yPSAT1* variant ORF, the *SER1* promoter, and the *SER1* terminator was PCR-amplified from yeast genomic DNA as two separate, overlapping fragments using 3' primers (Table S4) that also contained well-specific barcode sequences. Amplicons from each plate of our variant library were pooled and subjected to a Nextera tagmentation reaction (Illumina) as described[40] to generate a set of fragments with Nextera adapters on the ends of each random break. Selective PCR amplification on this pool of fragments was performed to generate a library of amplicons that each contained a well-specific barcode and an i7 index in the 3' end and a Nextera adapter and a plate-specific i5 index on the 5' ("ladder") end. Paired-end sequencing was performed on an Illumina NextSeq 500. In each pair of reads, a 25 bp read sequenced the inline well-specific barcode, and a 125 bp read sequenced a fragment of the variant ORF. The i5 and i7 indexes were used to demultiplex at the plate-level and the inline barcode was used to demultiplex reads at the well-level using fastq-multx (v 1.03) from ea-utils [41]. Reads were then aligned to the *yPSAT1* sequence using BWA (v0.7.9)[42], allowing four mismatches and using quality trimming with a threshold of Phred = 20. SAMtools (v0.1.17)[43] was used to generate a pileup file for each strain using the -C 50 and -q 20 parameters.

Because of PCR-based template switching[44] during plate-level barcoding, contaminating reference base calls, which are common to the great majority of strains on each plate, were observed in each strain, but they were not as frequent as the true variants. Therefore, the pileup file was parsed so that at each position in the reference genome, if the reference base was seen at >80% frequency, it was reported as such; otherwise, the most frequent non-reference base (along with its frequency) was reported. This identified the non-reference variant(s) in each strain, which were compared to the expected sequence, including the expected non-reference bases. Strains lacking the expected variant base(s) or harboring additional mutations were removed from further analysis.

Growth Assays and Model Fitting

The function of *yPSAT1* variants were assessed as follows. For each allele, 2–4 independently constructed, sequence verified strains were grown to saturation in rich (YPD) liquid medium in 96-well plates and pinned onto multiple (n= 4–6) agar plates containing synthetic minimal medium (SD, 2% glucose), which lacks serine.[34] Each plate of haploid

strains contained replicates of the same control strains: the unmodified FY4 yeast strain (n=2), null controls (n=2), wild type yPSAT1 (n=4), variant A99V (n=2), and variant V149M (n=2). Each plate of diploid strains contained replicates of the same diploid control strains: the unmodified FY4/ FY5 laboratory yeast strain (n=2) and wild type yPSAT1/ yPSAT1 (n=3). After 3 days of growth at 30° C, the plates were photographed with a mounted Canon PowerShot SX10 IS compact digital camera under consistent lighting, camera to subject distance, and zoom. Images (ISO200, f4.5, 1/40sec exposure) were acquired as jpg files. The area (in pixels) of each replica-pinned patch (Tables S5 and S6) was extracted with ImageJ[45] using a custom script (Supplemental File S1). Variant strain patches exhibiting minimal growth were visible as faint patches, but not reliably detected using our ImageJ quantitation pipeline. In these cases, images were imported into Photoshop and the visually detectable cell patches were measured using the Quick Selection Tool. Linear modelling was performed in R using a custom script (Supplemental File S2).

To normalize for a phenomenon commonly seen in agar plate-based growth assays (increased growth of colonies on the outer rows and columns of each plate),[46] we assayed a set of five 96-spot control plates containing only wild type *yPSAT1* haploids and used this data to model the effect via a multiplicative model with separate coefficients estimated for edge-row and edge-column. These estimates were then used to normalize the growth values observed on the haploid test plates (Supplemental File S2). Because strains on the diploid test plates were pinned less densely than strains on the haploid plates, adjustment for edge effects did not improve the model and was therefore omitted. The effect of each variant on colony area was then assessed, using the growth values (edge-normalized in the case of the haploid strains). Evidence of unequal error variances was observed, and therefore a generalized least squares linear model was fit using the GLS command from the NMLE package. The model had the form Area~0+Variant+Plate, where “Plate” allows for plate-specific batch effects on growth. Separate models were fitted for the haploid and diploid datasets (Supplemental File S2).

For each dataset, the effect of each variant on growth (colony area) was compared to that of the wild type *yPSAT1* strain (haploid or homozygous diploid) by Wald-test to identify strains with growth significantly different to wild type (Supplemental File S2). This was done using linearHypothesis from the CAR package, based on an asymptotic Chi-squared statistic. P-values were multiple-hypothesis-corrected using the Holm method via the p.adjust command. For the haploid dataset, the growth of each variant was also compared to the *ser1*Δ (hereafter, “null control”).

For the haploid strains, the effect of each variant on growth was normalized so that the wild type *yPSAT1* control had a value of 1 and the null control had a value of zero. This range was used to rescale the colony areas of the remaining variants, giving a normalized ratio to the growth of wild type *yPSAT1* as follows. First, the area coefficient of the null control was subtracted from the area coefficient of each variant to give U_{TN} and from the *yPSAT1* wild type control to give U_{CN}. The standard error of each variant, SE_T, and the standard error of the null control, SE_N, were then combined to give the standard error of the null-adjusted area of each variant, SE_{TN}, calculated as $SE_{TN} = (SE_T^2 + SE_N^2)^{0.5}$. Similarly, the standard error of the null-adjusted *yPSAT1* control was estimated as $SE_{CN} = (SE_C^2 + SE_N^2)^{0.5}$. The

relative standard error of each null-adjusted variant, RSE_{TN}, was then estimated by dividing SE_{TN} by U_{TN} and the relative standard error of null-adjusted PSAT1, RSE_{CN}, was estimated by dividing SE_{CN}, by U_{CN}. For each variant, the final normalized ratio R_{CN} was then estimated as U_{TN}/U_{CN} x $(1+ RSE_{CN}^2)$. We express these values as ratios "Normalized Growth Estimate" (Table S7) as well as in percentage form, which for brevity we refer to as percentage of wild type *yPSAT1* activity. The standard error of the ratio R_{CN} was estimated as U_{TN}/U_{CN} x $(RSE_{TN}^2 + RSE_{CN}^2)^{0.5}$ as described previously.[47]

For the diploid strains, the effect of each variant on growth (the variant coefficient from the linear regression) was normalized so that the *yPSAT1* homozygous control had a value of 1 and the null estimate from the haploid experiment (2378 pixels, an approximation of the area of cell deposition by pinning) was set to value of zero. This range was used to rescale the colony areas of the remaining variants giving a normalized ratio to *yPSAT1* as follows. First, 2378 was subtracted from the area coefficient of each variant to give U_T and from the *yPSAT1* homozygous control to give U_C. The relative standard error RSE_T of each variant was then estimated by dividing the standard error of each strain, SE_T, by U_T and the relative standard error RSE_C of PSAT1 was estimated by dividing the standard error of PSAT1, SE_C, by U_C. For each variant, the normalized ratio R_C was then calculated as U_T/U_C x $(1+ RSE_C^2)$. We express these values as ratios "Normalized Growth Estimate" (Table S8) as well as in percentage form, which for brevity we refer to as percentage of wild type *yPSAT1* activity. The standard error of the ratio R_C was estimated as U_T/U_C x $(RSE_T^2 + RSE_C^2)^{0.5}$ as described previously.[47]

Accession numbers

The accession numbers for the wild type yPSAT1 plasmids reported in this paper are Genbank: MN654100 and MN654101.

RESULTS

Establishing an *in vivo PSAT1* functional assay

Previous work has shown that the human *PSAT1* cDNA sequence expressed from a high copy plasmid under the control of a strong promoter can functionally replace (complement) loss of its yeast ortholog (*SER1*), conferring growth on minimal medium lacking serine.[31] To reduce the experimental variability associated with plasmid expression, we chose to integrate a single copy of the *PSAT1* protein coding sequence into the genome of *S. cerevisiae*, which can be stably maintained as haploid or diploid cells. A yeast codon optimized version (*yPSAT1*) of human *PSAT1* isoform 1 (GenBank: NP_478059.1) was integrated into a haploid *ser1* null mutant (*ser1*Δ) at the *SER1* locus in a configuration that placed the human phosphoserine aminotransferase protein coding sequence under the *SER1* transcriptional promoter and 3' UTR (Figure 1). Because phosphoserine aminotransferase activity is required for yeast to grow in the absence of exogenously supplied serine,[48] colony growth on minimal medium (lacking serine) is a quantitative readout of enzyme function that is amenable to high throughput analysis. To test the extent to which our construct could complement loss of the yeast enzyme, we compared the growth of two independent isolates expressing *yPSAT1* to the growth of the wild type lab strain (harboring the yeast *SER1*

gene) and two independent *ser1*Δ isolates on defined minimal media in the presence or absence of exogenously supplied serine (Figures 2A and S1). While the *ser1*Δ strains were unable to grow on medium lacking serine, strains with the integrated copy of *yPSAT1* are able to functionally complement a *ser1* deletion at 89% the growth of the strain harboring the yeast ortholog.

To further assess the extent to which the yeast system could serve as a readout for human gene function, we assayed haploid strains containing *yPSAT1* alleles that corresponded to alleles with definitive clinical significance calls in the ClinVar database.[32] Because our assay is specific to the protein coding sequence, we limited our analysis to missense, frameshift, and stop gain mutations. This included a panel of strains containing all pathogenic missense mutations (A99V, D100A, S179L), the sole likely benign missense mutation in the database (R306C), and three pathogenic alleles that introduce frameshifts near the beginning, middle, and end of the protein (G36fs, D145fs, R342fs). The results (Figure 2B) demonstrated that all pathogenic alleles tested exhibited significantly reduced growth (relative to wild type *yPSAT1*) in the absence of serine, with S179L and all three frameshift mutants exhibiting growth that was statistically equivalent to that of the *ser1*Δ mutant. We also found that the likely benign allele (R306C) conferred levels of growth in the absence of serine that were not statistically different from growth conferred by the wild type *yPSAT1*. Thus, results from our quantitative yeast growth assay agree well with the variant interpretations curated in ClinVar.

Yeast models of patient trios

To evaluate the correlation between our yeast assay and human phenotypes, we next examined *PSAT1* alleles from the disease literature. As an autosomal recessive disease, PSATD/NLS2 manifests only in individuals harboring either homozygous or compound heterozygous loss of function mutations.[25] Because it is a rare disease, published information about disease-causing *PSAT1* alleles (Table 1) included only ten individuals with seven unique *PSAT1* genotypes and one case of suspected NLS2[35] (Fowzan S. Alkuraya, personal communication). While the activity of many of these alleles had been characterized, they had not been directly compared using the same assay. We constructed diploid panels of yeast strains harboring the *yPSAT1* allele combinations observed in each unique trio of carrier parents and their affected offspring (Tables 1 and S1). We then assayed their growth in the absence of serine relative to a strain homozygous for the wild type *yPSAT1*. Despite the fact that disease severity in humans can be influenced by supplementation therapy, genetic background effects, or other environmental factors, our results using isogenic diploid yeast strains grown under controlled environmental conditions (Figure 3 and Table S8) ranked the genotypes in an order that agreed well with disease severity reported in the literature. Yeast strains with complete loss of function growth phenotypes corresponded to the *PSAT1* genotypes of the most severe cases and the suspected case of PSATD/NLS2 in which the affected individuals were stillborn.[18; 35] Genotypes that produced growth values of 51%−69% of wild type *yPSAT1* activity in the yeast assay corresponded to the *PSAT1* genotypes of affected individuals that, without serine and glycine supplementation, were either stillborn or who died before the age of 7 months.

[18; 25] Genotypes that produced growth values of 79%−82% of wild type *yPSAT1* activity in the yeast assay corresponded to affected individuals who survived past 7 months of age.[24; 49]

Assaying variants of uncertain significance

With an assay in hand, we sought to test the functional impact of individual *PSAT1* missense variants of uncertain significance that are known to exist in the human population. We constructed a variant library of *yPSAT1* in haploid strains that included all (n=189) missense variants from the ~140,000 human genome or exome sequences present in the gnomAD database (version r2.0.2)[6] as well as VUS alleles from ClinVar[32] (Table S3). We also included an additional allele (N302S) that was observed in a clinical WES panel, but was not associated with a serine deficiency phenotype (Fowzan S. Alkuraya, personal communication). Our results (Figure 4 and Table S7) demonstrate that human *PSAT1* alleles of uncertain significance span a wide range of activity in our assay, from values equivalent to those of the null to values greater than wild type. This includes 59 alleles that show levels of activity that are significantly below that of wild type *yPSAT1* ($p < 0.05$, Table S7).

Functional information can be used to inform human variant interpretation.[7] However, the precise degree to which PSAT protein function needs to be impaired to negatively impact human health is not known. Because microbial growth rates are a continuous variable that can be sensitively measured across a wide dynamic range, growth-based assays in yeast may be able to detect modest decreases in activity that are not sufficient to cause human disease. Therefore, to benchmark the results of our assay, we considered alleles that had definitive calls in ClinVar and descriptions in the literature, which for *PSAT1* is a relatively small set. We then ranked all alleles based on their normalized growth values and used the rank positions of alleles with known effects to classify the others (Figure 4).

We began with the set of 59 alleles with significant growth impairment in our assay relative to wild type and examined their ranks relative to known pathogenic alleles. A99V exhibits the highest activity (58% of wild type activity in our assay) of any variant that has a pathogenic call in ClinVar and has been observed in the homozygous state in PSATD/NLS2 patients. Based on their ranking, 22 alleles have growth value estimates lower than that of A99V (Figure 4, dark magenta box and Table S7). This group comprises our most severe and highest confidence loss of function alleles and includes all frameshift mutations assayed, all homozygous alleles (S179L, A99V and S43R) that have been observed in PSATD/NLS2 patients, and the homozygous allele (G78A) in the suspected case of PSATD/NLS2. As such, alleles with functional scores in this range warrant closer examination as possible loss of function mutations especially as homozygotes or compound heterozygotes with other loss of function alleles in this range.

Next, we considered alleles with pathogenic calls in ClinVar that have growth estimates in our assay higher than that of A99V and which have not been observed as homozygotes. The only allele that meets these criteria is D100A, which exhibits 85% of wild type *yPSAT1* activity in our assay. D100A has been observed as a compound heterozygote with a frameshift allele in two PSATD/NLS2 patients and heterologously expressed protein harboring the D100A variant had reduced enzymatic activity *in vitro*.[25] Using this allele as the upper boundary for potential clinical significance partitioned an additional 20 alleles

(Figure 4, light magenta box and Table S7) into the loss of function range. Alleles with functional scores in this range warrant closer examination as possible loss of function mutations especially as compound heterozygotes with more severe loss of function mutations.

Since PSATD/NLS2 is an autosomal recessive disease, alleles with wild type levels of activity (equal to or greater than wild type *PSAT1*) are considered benign. We therefore annotated all alleles with normalized growth values of at least 1.0 in our assay as phenotypically neutral. By this metric, 53% (n= 100) of the alleles in gnomAD, three of the ClinVar alleles with ambiguous calls (VUS or conflicting) and the N302S fell in this neutral range (Figure 4). Only two *PSAT1* alleles (I123V and A234S) appear as homozygotes in gnomAD, which does not contain sequences from individuals with severe developmental syndromes such as PSATD/NLS2. As such, these polymorphisms are expected to be benign, and our results (101% and 109%, respectively, of wild type *yPSAT1* activity) are consistent with this expectation. While some alleles exhibit significantly greater than wild type activity in our assay, since PSATD/NLS2 is only scored in one direction (reduced function), this difference is presumed to lack clinical significance.

Finally, we considered any alleles ranked between D100A and wild type (n=47) as being of uncertain significance. This range includes four alleles (H19D, A15V, K274T and V290I) that are annotated as VUS in ClinVar. Of these only A15V, which has 97% of wild type *yPSAT1* activity in our assay, is reported in the disease literature (Table 1).[49] A15V was reported as compound heterozygous with a frameshift mutation with null activity in our assay in a single patient and has not, to our knowledge, been assayed independently. As such, we believe that its rank within the uncertain range of our assay is consistent with what is currently known about the allele. In the absence of additional patient data, the clinical significance of variants with functional impact scores in this range of our assay remains uncertain.

DISCUSSION

Here, we have developed an assay for measuring the functional impact of genetic variation in the human *PSAT1* protein coding sequence that is amenable to large-scale testing. Calibration of functional assays using alleles with known impact is crucial for assay validation.[50–52] Although our ability to accurately define a pathogenicity threshold or to perform extensive comparisons between the yeast assay and human disease phenotypes is limited by the fact that PSATD/NLS2 is an extremely rare disease, our assay agrees well with the reported disease severity of PSATD/NLS2 homozygous and compound heterozygous genotypes. We used the assay to test the functional impact of *PSAT1* missense variants of uncertain significance that are currently published. Our results suggest that 53% (n=100) of the alleles in the r2.0.2 build of the gnomAD database are likely benign and at least 22% (n=42) warrant closer consideration as possible loss of function mutations that could cause disease.

Experiments in human patients and mammalian models are crucial for understanding the function of human genes and the diseases that arise when they are perturbed. Understanding

the molecular mechanisms that underlie human disease requires human genetics and studies in mammalian systems. However, model organisms, such as yeast, play an important role in the ecosystem of human disease research, because although they have limits, they also have advantages. The experimental tractability of *S. cerevisiae* makes it possible to construct and assay the functional impact of individual alleles (in haploids) and allele combinations (in diploids) at unrivaled speed, cost, and throughput. Once a complementation assay, such as the one we describe here, has been established and validated, new alleles can be constructed and tested in under two weeks. Synthetic biology approaches permit assays of variants without access to patient samples and even prior to the detection of the alleles or allele combinations in the human population. The low cost and high throughput of yeast-based assays also facilitates variant testing at scale. For example, with a protein the size of PSAT1, it is feasible to construct and assay all SNP-accessible amino acid substitutions across the length of the protein. Thus, the rapid turn-around time of yeast-based assays and the ability to generate comprehensive tables of variant impact could play a major role in increasing the speed of variant interpretation, which is currently the major obstacle to using rapid whole genome and exome sequencing for genetic diagnosis.[53–57]

Variant interpretation, e.g. significance calls in ClinVar, is generally reported in the context of individual alleles. However, for rare variants and rare diseases, information about the compound heterozygous states is often needed, and the ability to mate haploid yeast strains permits these genotypes to be constructed for trivial amounts of time, cost, and effort. Functional data specific to a patient genotype may provide valuable insights for stratifying disease or guiding treatment. Because maternal serine levels can have profound effects on the developmental outcome of a fetus with an NLS genotype,[25] understanding the functional impact of the maternal genotype could help explain variation in developmental outcomes. For example, our assay predicts that a carrier mother's PSAT1 activity might vary depending on the severity of the mutant allele for which she is heterozygous. Also, while heterozygous carriers do not develop PSATD or NLS2, these individuals may have increased risk for other diseases that are underdiagnosed or late onset. Notably, a recent GWAS study identified two enzymes in the serine biosynthesis pathway, *PHGDH* and *PSPH*, as candidate genes involved in adult onset macular telangiectasia type 2[59] and low serine levels were found in a separate study of 125 patients with the disease.[58]

While our study focused on *PSAT1*, using colony growth as a proxy for enzyme function is generalizable to many human genes that complement deletions of their corresponding yeast orthologs. Although not every human gene, biological process, or disease can be modeled in yeast, those that can, including metabolic enzymes, are often important and of high clinical utility. As more functional assays are developed, establishing lookup tables cataloging the results of these assays[8] can provide important supporting evidence for variant interpretation. Thus, surrogate genetic assays offer a form of variant interpretation capable of "pre-screening" the functional effects of large numbers of alleles[59] which may provide clinicians with additional information in cases with limited clinical evidence. As such, these assays could play an important role in increasing the speed of diagnosis for genetic disorders that can reduce morbidity and mortality, guide palliative care decisions, and substantially lower hospitalization costs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Fowzan S. Alkuraya, Richard McLaughlin, Michael Xie, Ayman El-Hattab, Amal Alhashem, Patrick May, Jay Shendure, Isaiah Chuang, J. Lawrence Merritt II, Candace Myers, and Anna Scott for helpful discussions. This work was funded in part by a National Institutes of Health grant (R01 GM134274) to A.M.D.

Funding

This work was funded in part by a National Institutes of Health grant (R01 GM134274) to AMD.

References

1. Campistol J, and Plecko B (2015). Treatable newborn and infant seizures due to inborn errors of metabolism. Epileptic Disord 17, 229–242. [PubMed: 26234933]
2. El-Hattab AW (2015). Inborn errors of metabolism. Clin Perinatol 42, 413–439, x. [PubMed: 26042912]
3. Lanpher B, Brunetti-Pierri N, and Lee B (2006). Inborn errors of metabolism: the flux from Mendelian to complex diseases. Nat Rev Genet 7, 449–460. [PubMed: 16708072]
4. Vernon HJ (2015). Inborn Errors of Metabolism: Advances in Diagnosis and Therapy. JAMA Pediatr 169, 778–782. [PubMed: 26075348]
5. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, and Maglott DR (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research 42, D980–985.
6. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291. [PubMed: 27535533]
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. (2015). 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. Genetics in medicine : official journal of the American College of Medical Genetics 17, 405–424.
8. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, Shendure J, and Fowler DM (2017). Variant Interpretation: Functional Assays to the Rescue. American journal of human genetics 101, 315–325. [PubMed: 28886340]
9. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, Balakishnan B, Liang R, Zhang Y, Lyon S, et al. (2015). Comparison of predicted and actual consequences of missense mutations. Proceedings of the National Academy of Sciences of the United States of America 112, E5189–5198.
10. Greene AL, Snipe JR, Gordenin DA, and Resnick MA (1999). Functional analysis of human FEN1 in Saccharomyces cerevisiae and its role in genome stability. Human molecular genetics 8, 2263–2273. [PubMed: 10545607]
11. Hamza A, Tammpere E, Kofoed M, Keong C, Chiang J, Giaever G, Nislow C, and Hieter P (2015). Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. Genetics 201, 1263–1274. [PubMed: 26354769]
12. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, and Ishioka C (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proceedings of the National Academy of Sciences of the United States of America 100, 8424–8429. [PubMed: 12826609]
13. Mayfield JA, Davies MW, Dimster-Denk D, Pleskac N, McCarthy S, Boydston EA, Fink L, Lin XX, Narain AS, Meighan M, et al. (2012). Surrogate genetics and metabolic profiling for characterization of human disease alleles. Genetics 190, 1309–1323. [PubMed: 22267502]

14. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J, Theesfeld CL, Bansal P, Sahni N, Yi S, et al. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome research 26, 670–680. [PubMed: 26975778]

15. Trevisson E, Burlina A, Doimo M, Pertegato V, Casarin A, Cesaro L, Navas P, Basso G, Sartori G, and Salviati L (2009). Functional complementation in yeast allows molecular characterization of missense argininosuccinate lyase mutations. The Journal of biological chemistry 284, 28926–28934.

16. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, and Marcotte EM (2015). Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science 348, 921–925. [PubMed: 25999509]

17. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genetics in medicine : official journal of the American College of Medical Genetics 19, 249–255. [PubMed: 27854360]

18. Acuna-Hidalgo R, Schanze D, Kariminejad A, Nordgren A, Kariminejad MH, Conner P, Grigelioniene G, Nilsson D, Nordenskjold M, Wedell A, et al. (2014). Neu-Laxova syndrome is a heterogeneous metabolic disorder caused by defects in enzymes of the L-serine biosynthesis pathway. American journal of human genetics 95, 285–293. [PubMed: 25152457]

19. El-Hattab AW, Shaheen R, Hertecant J, Galadari HI, Albaqawi BS, Nabil A, and Alkuraya FS (2016). On the phenotypic spectrum of serine biosynthesis defects. J Inherit Metab Dis 39, 373–381. [PubMed: 26960553]

20. de Koning TJ, and Klomp LW (2004). Serine-deficiency syndromes. Curr Opin Neurol 17, 197–204. [PubMed: 15021249]

21. Tabatabaie L, Klomp LW, Berger R, and de Koning TJ (2010). L-serine synthesis in the central nervous system: a review on serine deficiency disorders. Mol Genet Metab 99, 256–262. [PubMed: 19963421]

22. Ducker GS, and Rabinowitz JD (2017). One-Carbon Metabolism in Health and Disease. Cell Metab 25, 27–42. [PubMed: 27641100]

23. Kalhan SC, and Hanson RW (2012). Resurgence of serine: an often neglected but indispensable amino Acid. J Biol Chem 287, 19786–19791.

24. Brassier A, Valayannopoulos V, Bahi-Buisson N, Wiame E, Hubert L, Boddaert N, Kaminska A, Habarou F, Desguerre I, Van Schaftingen E, et al. (2016). Two new cases of serine deficiency disorders treated with l-serine. European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society 20, 53–60. [PubMed: 26610677]

25. Hart CE, Race V, Achouri Y, Wiame E, Sharrard M, Olpin SE, Watkinson J, Bonham JR, Jaeken J, Matthijs G, et al. (2007). Phosphoserine aminotransferase deficiency: a novel disorder of the serine biosynthesis pathway. American journal of human genetics 80, 931–937. [PubMed: 17436247]

26. de Koning TJ, Klomp LW, van Oppen AC, Beemer FA, Dorland L, van den Berg I, and Berger R (2004). Prenatal and early postnatal treatment in 3-phosphoglycerate-dehydrogenase deficiency. Lancet 364, 2221–2222. [PubMed: 15610810]

27. De Koning TJ, Duran M, Van Maldergem L, Pineda M, Dorland L, Gooskens R, Jaeken J, and Poll-The BT (2002). Congenital microcephaly and seizures due to 3-phosphoglycerate dehydrogenase deficiency: outcome of treatment with amino acids. J Inherit Metab Dis 25, 119–125. [PubMed: 12118526]

28. de Koning TJ, Duran M, Dorland L, Gooskens R, Van Schaftingen E, Jaeken J, Blau N, Berger R, and Poll-The BT (1998). Beneficial effects of L-serine and glycine in the management of seizures in 3-phosphoglycerate dehydrogenase deficiency. Ann Neurol 44, 261–265. [PubMed: 9708551]

29. Jaeken J, Detheux M, Van Maldergem L, Foulon M, Carchon H, and Van Schaftingen E (1996). 3-Phosphoglycerate dehydrogenase deficiency: an inborn error of serine biosynthesis. Arch Dis Child 74, 542–545. [PubMed: 8758134]

30. Moat S, Carling R, Nix A, Henderson M, Briddon A, Prunty H, Talbot R, Powell A, Wright K, Fuchs S, et al. (2010). Multicentre age-related reference intervals for cerebrospinal fluid serine

concentrations: implications for the diagnosis and follow-up of serine biosynthesis disorders. Mol Genet Metab 101, 149–152. [PubMed: 20692860]

31. Baek JY, Jun DY, Taub D, and Kim YH (2003). Characterization of human phosphoserine aminotransferase involved in the phosphorylated pathway of L-serine biosynthesis. Biochem J 373, 191–200. [PubMed: 12633500]
32. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic acids research 46, D1062–D1067. [PubMed: 29165669]
33. Winston F, Dollard C, and Ricupero-Hovasse SL (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast 11, 53–55. [PubMed: 7762301]
34. Rose MD, Winston FM, and Hieter P (1990). Methods in yeast genetics : a laboratory course manual.(Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).
35. Monies D, Abouelhoda M, AlSayed M, Alhassnan Z, Alotaibi M, Kayyali H, Al-Owain M, Shah A, Rahbeeni Z, Al-Muhaizea MA, et al. (2017). The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. Human genetics 136, 921–939. [PubMed: 28600779]
36. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. (1996). Life with 6000 genes. Science 274, 546, 563–547.
37. Sirr A, Scott AC, Cromie GA, Ludlow CL, Ahyong V, Morgan TS, Gilbert T, and Dudley AM (2018). Natural Variation in SER1 and ENA6 Underlie Condition-Specific Growth Defects in Saccharomyces cerevisiae. G3 8, 239–251. [PubMed: 29138237]
38. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285, 901–906. [PubMed: 10436161]
39. Goldstein AL, and McCusker JH (1999). Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast 15, 1541–1553. [PubMed: 10514571]
40. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, Wu Y, Pons C, Wong C, van Lieshout N, et al. (2017). A framework for exhaustively mapping functional missense variants. Molecular systems biology 13, 957. [PubMed: 29269382]
41. Aronesty E (2011). Command-line tools for processing biological sequencing data.
42. Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]
43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]
44. Kebschull JM, and Zador AM (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic acids research 43, e143.
45. Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671–675. [PubMed: 22930834]
46. Wagih O, Usaj M, Baryshnikova A, VanderSluis B, Kuzmin E, Costanzo M, Myers CL, Andrews BJ, Boone CM, and Parts L (2013). SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. Nucleic acids research 41, W591–596.
47. Holmes DT, and Buhr KA (2007). Error propagation in calculated ratios. Clinical biochemistry 40, 728–734. [PubMed: 17434158]
48. Melcher K, Rose M, Kunzler M, Braus GH, and Entian KD (1995). Molecular analysis of the yeast SER1 gene encoding 3-phosphoserine aminotransferase: regulation by general control and serine repression. Current genetics 27, 501–508. [PubMed: 7553933]
49. Glinton KE, Benke PJ, Lines MA, Geraghty MT, Chakraborty P, Al-Dirbashi OY, Jiang Y, Kennedy AD, Grotewiel MS, Sutton VR, et al. (2018). Disturbed phospholipid metabolism in serine biosynthesis defects revealed by metabolomic profiling. Molecular genetics and metabolism 123, 309–316. [PubMed: 29269105]
50. Mahmood K, Jung CH, Philip G, Georgeson P, Chung J, Pope BJ, and Park DJ (2017). Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. Human genomics 11, 10. [PubMed: 28511696]

51. Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, and Hauke J (2018). Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. BMC medical genomics 11, 35. [PubMed: 29580235]

52. Shendure J, and Fields S (2016). Massively Parallel Genetics. Genetics 203, 617–619. [PubMed: 27270695]

53. Berg JS, Agrawal PB, Bailey DB Jr., Beggs AH, Brenner SE, Brower AM, Cakici JA, Ceyhan-Birsoy O, Chan K, Chen F, et al. (2017). Newborn Sequencing in Genomic Medicine and Public Health. Pediatrics 139.

54. Miller NA, Farrow EG, Gibson M, Willig LK, Twist G, Yoo B, Marrs T, Corder S, Krivohlavek L, Walter A, et al. (2015). A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med 7, 100. [PubMed: 26419432]

55. Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, Yeung A, Peters H, Mordaunt D, Cowie S, et al. (2016). A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genet Med 18, 1090–1096. [PubMed: 26938784]

56. Willig LK, Petrikin JE, Smith LD, Saunders CJ, Thiffault I, Miller NA, Soden SE, Cakici JA, Herd SM, Twist G, et al. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. Lancet Respir Med 3, 377–387. [PubMed: 25937001]

57. Monies D, Abouelhoda M, Assoum M, Moghrabi N, Rafiullah R, Almontashiri N, Alowain M, Alzaidan H, Alsayed M, Subhani S, et al. (2019). Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. American journal of human genetics 104, 1182–1201. [PubMed: 31130284]

58. Gantner ML, Eade K, Wallace M, Handzlik MK, Fallon R, Trombley J, Bonelli R, Giles S, Harkins-Perry S, Heeren TFC, et al. (2019). Serine and Lipid Metabolism in Macular Disease and Peripheral Neuropathy. The New England journal of medicine 381, 1422–1433. [PubMed: 31509666]

59. Dunham MJ, and Fowler DM (2013). Contemporary, yeast-based approaches to understanding human genetic variation. Current opinion in genetics & development 23, 658–664. [PubMed: 24252429]

Synopsis:

Our yeast-based assay for human *PSAT1* function agrees well with clinical annotations and makes functional predictions for nearly 200 missense mutations.

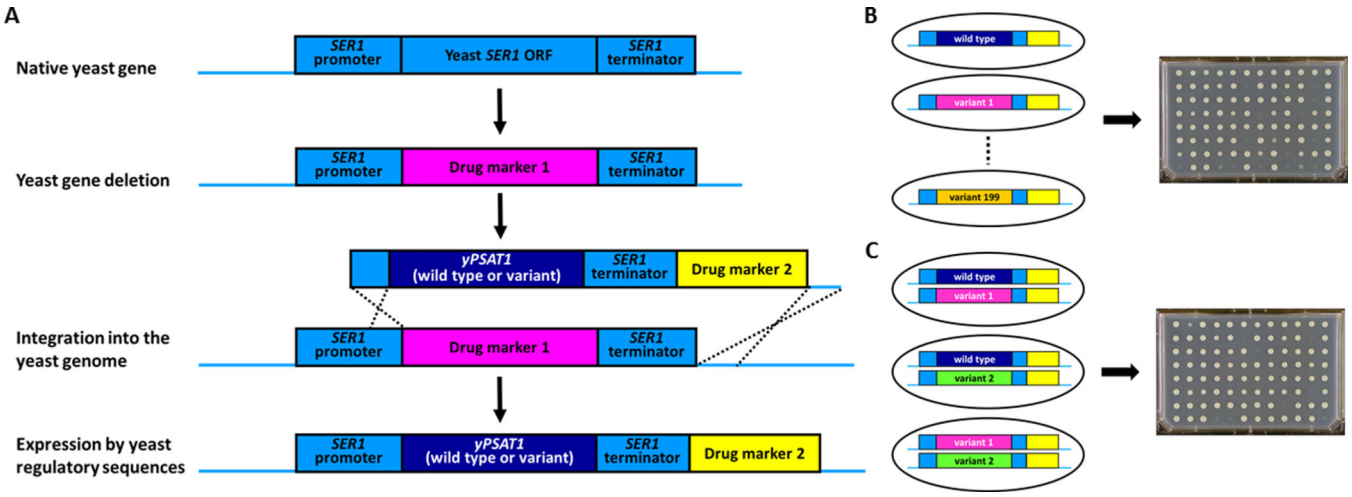

Figure 1. Experimental design

(A) Starting with a haploid FY4 laboratory yeast strain, the native yeast *SER1* open reading frame is deleted and replaced with a selectable drug marker. This strain serves as the null control strain. This null mutant is then transformed with PCR products containing either wild type or variant alleles of the yeast-optimized human *PSAT1* protein coding sequence (*yPSAT1*) and additional sequences that target its integration (by homologous recombination) into the yeast genome in a configuration that places its expression under the control of the orthologous yeast gene's regulatory elements (*SER1* promoter and *SER1* terminator). Haploid strains harboring the correct allele of *yPSAT1* and lacking secondary mutations are identified by the correct pattern of drug resistance and confirmed by PCR and amplicon sequencing of the entire reporter construct (Materials and Methods).

(b)Haploid yeast strains harboring each variant are assayed by growth in the absence of serine.

(C) Diploid yeast strains harboring compound heterozygotes are constructed by mating haploid strain that harbor the appropriate alleles and assayed by growth in the absence of serine.

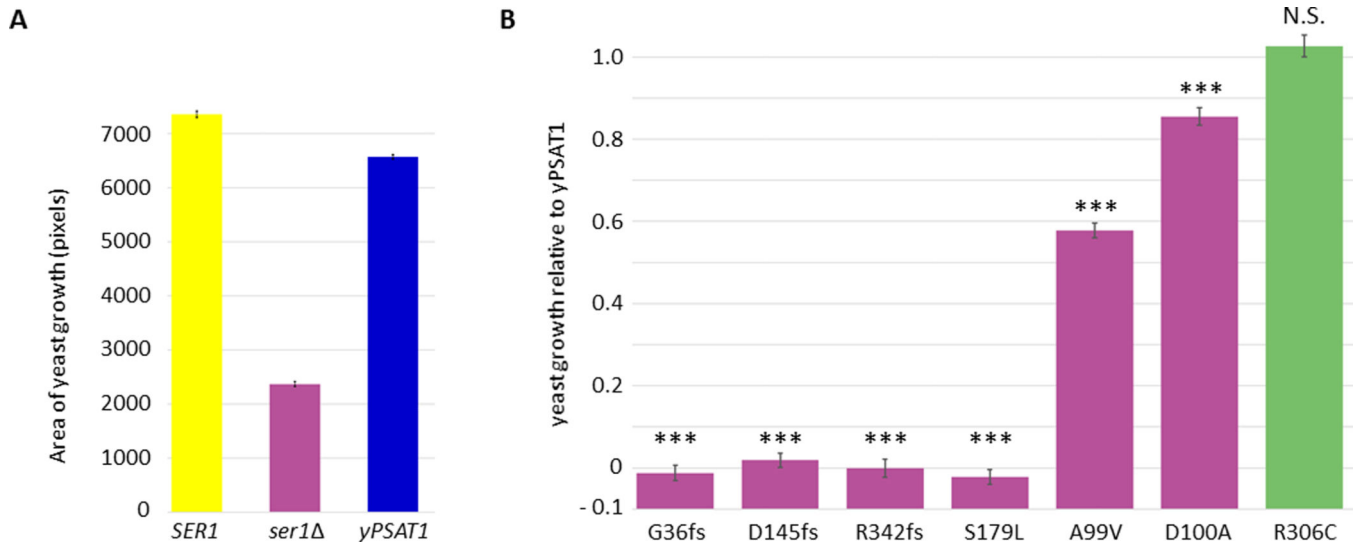

Figure 2. Establishing a yeast model for human *PSAT1* function.

(A) *yPSAT1* can functionally replace the yeast *SER1* as assayed by quantitative growth of yeast cells replica pinned onto minimal medium lacking serine. Values are the area of growth estimates for each genotype +/− their standard errors derived from fitting a linear model to these data (Materials and Methods). Assays were performed in replicate (2–4 independently constructed clones with ~50 technical replicates of each).

(B) Yeast assays of *PSAT1* alleles agree with clinical significance calls in ClinVar. Values shown are estimates of growth relative to *yPSAT1* +/− the standard error of the estimates and with *ser1*Δ set to 0 (Materials and Methods). Assays were performed in replicate (2–4 independently constructed clones with 4–6 technical replicates of each). Statistical significance between the variant and wild type *yPSAT1* was assessed by Wald-test (Materials and Methods). Results are indicated as $p < 0.001$ (***) or not significant (N.S.).

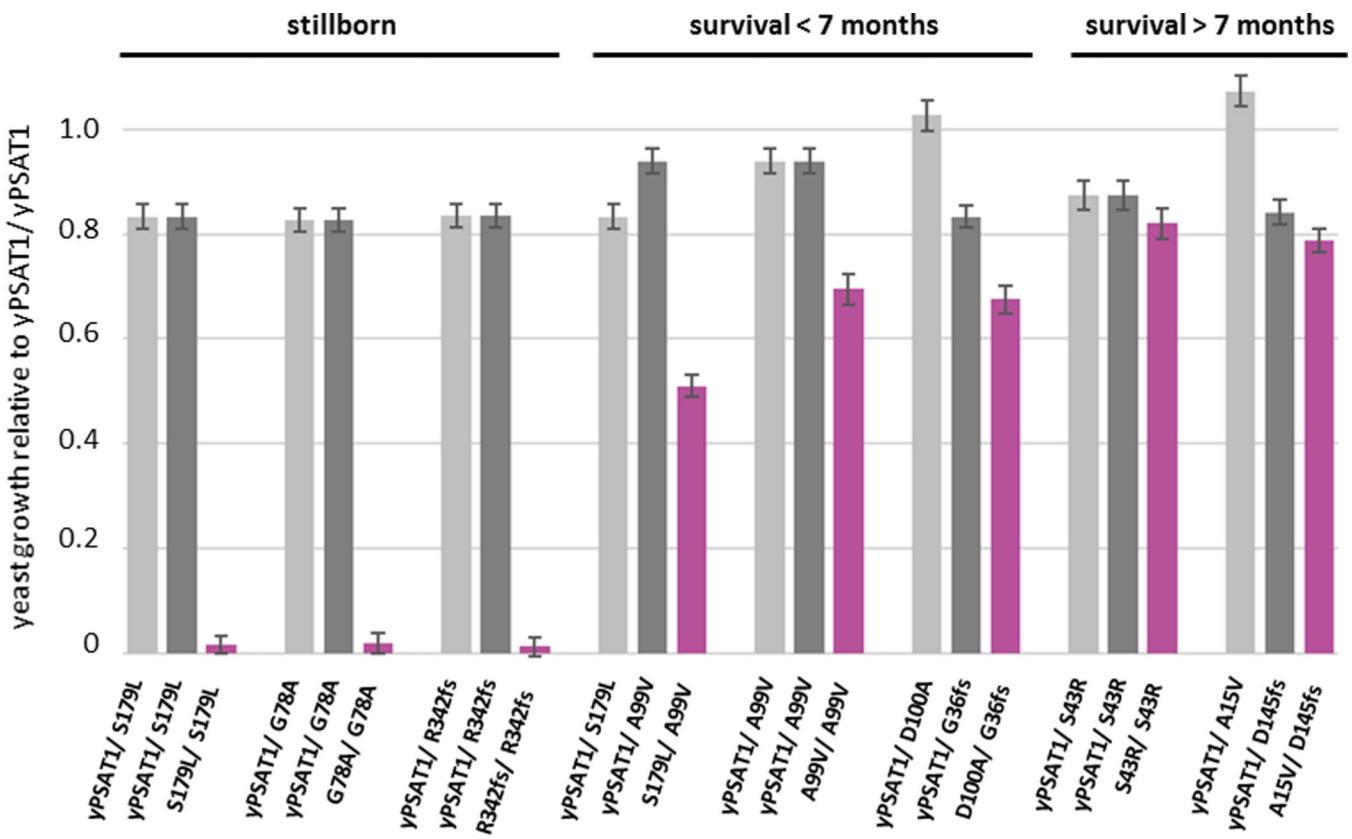

Figure 3. Yeast models of human *PSAT1* genotypes for all published disease trios.

Values shown are estimates of growth relative to a homozygous wild type (*yPSAT1*/*yPSAT1*) diploid +/− the standard error of the estimates and with *ser1*Δ set to 0 (Materials and Methods). Results are organized by trios of the genotypes of the parents (light and dark gray) and the affected individuals (magenta), where *yPSAT1* refers to the wild type allele. Assays were performed in replicate (2–3 independently constructed strains with 10 technical replicates of each). Genotypes that appear multiple times in the plot, i.e. the same parental genotype in a trio or in two different trio, are the same values plotted multiple times for ease of comparison.

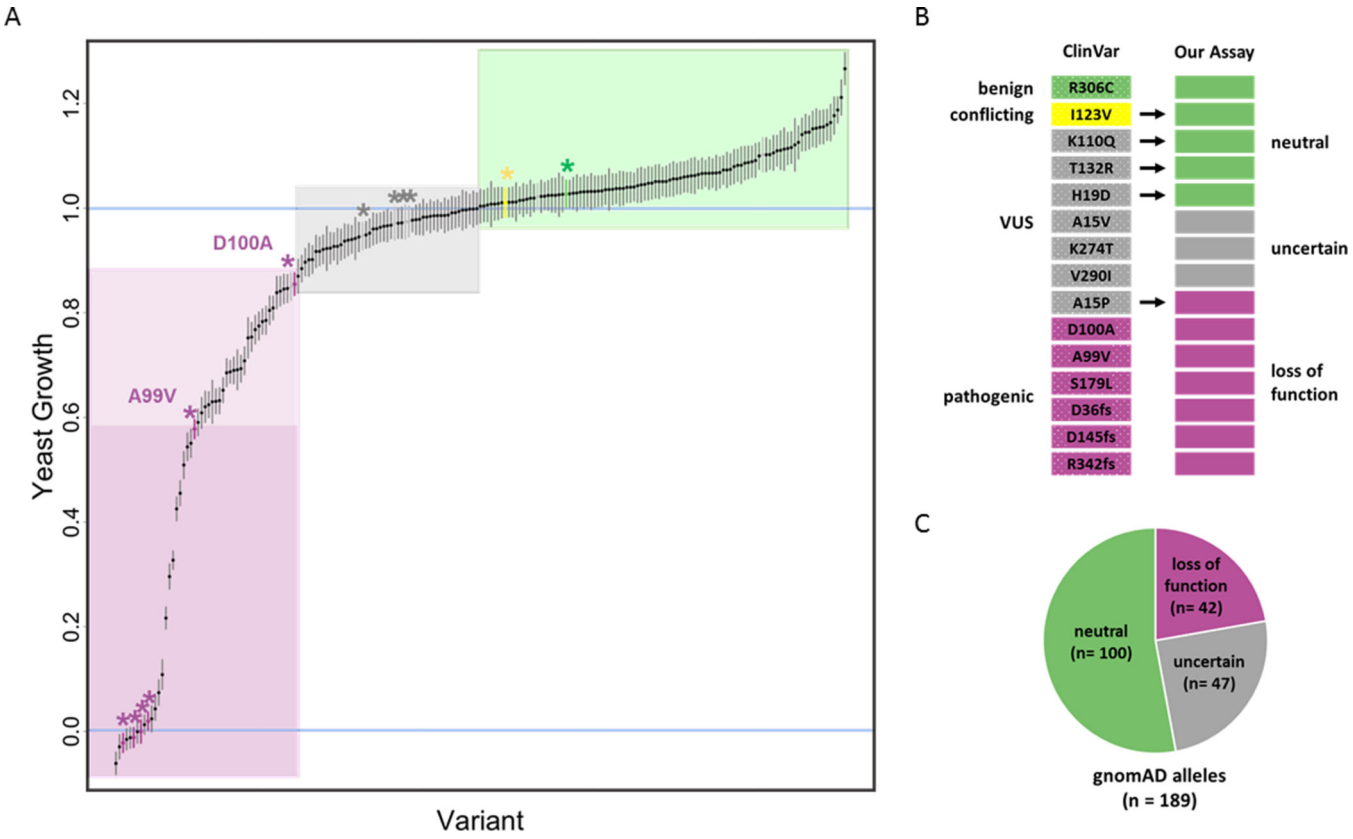

Figure 4. *PSAT1* alleles exhibit a range of activity.

(A) Relative growth of each yPSAT1 allele ordered (along the x-axis) and plotted +/− standard error of the estimate (Materials and Methods). Alleles from ClinVar are indicated by an asterisks and color coded (magenta = pathogenic, green = likely benign, yellow = conflicting, gray = VUS). Alleles with values greater or equal to 1.0 (wild type value) were flagged as “neutral” (green box). Alleles with values less than or equal to that of D100A were flagged as “loss of function” (magenta box). Values less than A99V were highlighted in dark magenta. This region includes the pathogenic variants S179L, G36fs, D145fs and Arg342fs (unlabeled, magenta asterisks), which are not statistically different from the null mutant. The gray box (“uncertain”) indicates alleles with values less than wild type, but greater than D100A.

(B) Comparison of ClinVar annotations and the results of our assay. Each bar represents one of the 15 *PSAT1* alleles in ClinVar. Color coding is: magenta = pathogenic/loss of function, green = likely benign/ neutral, yellow = conflicting, gray= VUS/uncertain.

(C) Annotation of gnomAD alleles by our assay. Pie chart representing the number of *PSAT1* alleles from gnomAD in each of our functional classification categories.

Table 1.

Published PSAT1 alleles from confirmed or suspected cases of PSATD/NLS2.

Patient	Parent Allele 1	Parent Allele 2
Hart_Patient_1[25]	c.299 A>C (p.Asp100Ala)	c.delG107 (p.Gly36Ala_fs*5)
Hart_Patient_2[25]	c.299 A>C (p.Asp100Ala)	c.delG107 (p.Gly36Alafs_*5)
Acuna-Hidalgo_Patient_1[18]	c.del1023_1027delinsAGACCT (p.Arg342Asp_fs*6)	c.del1023_1027delinsAGACCT (p.Arg342Asp_fs*6)
Acuna-Hidalgo_Patient_2[18]	c.296C>T (p.Ala99Val)	c.296C>T (p.Ala99Val)
Acuna-Hidalgo_Patient_3[18]	c.536C>T (p.Ser179leu)	c.536C>T (p.Ser179leu)
Acuna-Hidalgo_Patient_4[18]	c.296C>T (p.Ala99Val)	c.296C>T (p.Ala99Val)
Acuna-Hidalgo_Patient_5[18]	c.296C>T (p.Ala99Val)	c.296C>T (p.Ala99Val)
Acuna-Hidalgo_Patient_6[18]	c.296C>T (p.Ala99Val)	c.536C>T (p.Ser179leu)
Brassier_Patient_2[24]	c.129T>G (p. Ser43Arg)	c.129T>G (p. Ser43Arg)
Monies_16W-0250/16N-0116[35]	c.233G>C (p.Gly78Ala)	c.233G>C (p.Gly78Ala)
Glinton_Patient_4[49]	c.432delA (p.Asp145Met_fs*49)	c.44C>T (p.Ala15Val)