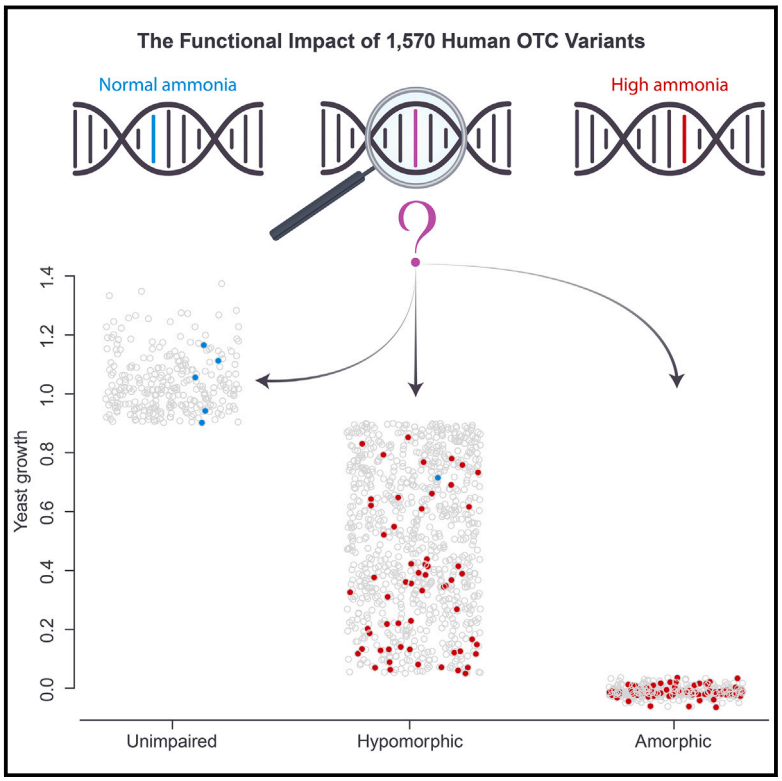
The functional impact of 1,570 individual amino acid substitutions in human OTC

Graphical abstract



Authors

Russell S. Lo, Gareth A. Cromie, Michelle Tang, ..., Nicholas Ah Mew, Andrea Gropman, Aimée M. Dudley

Correspondence

aimee.dudley@gmail.com

OTC deficiency is the most common urea cycle disorder. Here, we measure the functional impact of 1,570 single-nucleotide-variant-accessible amino acid substitutions in human OTC. Our assay distinguishes benign from pathogenic variants and neonatal from late-onset disease variants, supporting its use as PS3 evidence under the current ACMG guidelines.

Lo et al., 2023, The American Journal of Human Genetics *110*, 863–879
May 4, 2023
https://doi.org/10.1016/j.ajhg.2023.03.019

The functional impact of 1,570 individual amino acid substitutions in human OTC

Russell S. Lo,[1] Gareth A. Cromie,[1] Michelle Tang,[1] Kevin Teng,[1,7] Katherine Owens,[1,2,8] Amy Sirr,[1] J. Nathan Kutz,[2] Hiroki Morizono,[3,4] Ljubica Caldovic,[3,4] Nicholas Ah Mew,[3,4] Andrea Gropman,[3,4,5,6] and Aimée M. Dudley[1,*]

Summary

Deleterious mutations in the X-linked gene encoding ornithine transcarbamylase (*OTC*) cause the most common urea cycle disorder, OTC deficiency. This rare but highly actionable disease can present with severe neonatal onset in males or with later onset in either sex. Individuals with neonatal onset appear normal at birth but rapidly develop hyperammonemia, which can progress to cerebral edema, coma, and death, outcomes ameliorated by rapid diagnosis and treatment. Here, we develop a high-throughput functional assay for human OTC and individually measure the impact of 1,570 variants, 84% of all SNV-accessible missense mutations. Comparison to existing clinical significance calls, demonstrated that our assay distinguishes known benign from pathogenic variants and variants with neonatal onset from late-onset disease presentation. This functional stratification allowed us to identify score ranges corresponding to clinically relevant levels of impairment of OTC activity. Examining the results of our assay in the context of protein structure further allowed us to identify a 13 amino acid domain, the SMG loop, whose function appears to be required in human cells but not in yeast. Finally, inclusion of our data as PS3 evidence under the current ACMG guidelines, in a pilot reclassification of 34 variants with complete loss of activity, would change the classification of 22 from variants of unknown significance to clinically actionable likely pathogenic variants. These results illustrate how large-scale functional assays are especially powerful when applied to rare genetic diseases.

Introduction

Urea cycle disorders are inborn errors of metabolism caused by deficiencies in any of the eight proteins in the pathway responsible for the conversion of nitrogen, a waste product of protein metabolism, to urea. Accumulation of nitrogen, in the form of ammonia, to toxic levels in the blood and brain leads to symptoms that include vomiting, lethargy, behavioral abnormalities, cerebral edema, seizures, coma, and death. Ornithine transcarbamylase deficiency (OTC deficiency [MIM: 311250]) is the most common urea cycle disorder, accounting for approximately half of all urea cycle disorder cases.[1] OTC is a nuclear-encoded mitochondrial protein that converts ornithine and carbamoyl phosphate to citrulline in human liver cells. As an X-linked disorder, severe disease more often presents in males but can also occur in females harboring pathogenic alleles. Disease severity has both genetic and environmental components.[2] While severe loss-of-function mutations in *OTC* (MIM: 300461) are often associated with neonatal presentation, age of onset can vary, even among individuals harboring the same *OTC* allele.[3–5] Late-onset presentation is often associated with hyperammonemia-triggering events, such as high protein intake, prolonged fasting, surgery, exposure to organic chemicals, pregnancy, and administration of corticosteroids or valproate.[6] Severe outcomes can be prevented by rapid diagnosis and administration of nitrogen scavengers.[7] However, individuals with severe, neonatal-onset disease often require liver transplantation.[8]

Presymptomatic diagnosis of OTC deficiency is challenging. Because *de novo* mutations underlie 20%–30% of OTC deficiencies[9] and only a minority of variants are recurrent, family history often cannot be used to inform diagnosis. Additionally, the newborn screening assays for OTC deficiency have technical limitations. Low citrulline levels are a hallmark of OTC deficiency and a biochemical assay is used to quantitate citrulline using dried blood spot cards. However, citrulline is an unstable metabolite,[10] potentially generating false positive assay results. Low citrulline levels can also be non-specific, occurring, for example, when a newborn is premature or protein restricted.[11] Additionally, in severely affected males, hyperammonemia can occur rapidly after birth, often before newborn screening results are available. In females, the heterogeneity of liver cells due to X-inactivation confounds enzymatic assays from liver biopsies.

DNA sequencing provides an alternative approach for diagnosis that avoids many of these problems, potentially

[1]Pacific Northwest Research Institute, Seattle, WA, USA; [2]Department of Applied Mathematics, University of Washington, Seattle, WA, USA; [3]Center for Genetic Medicine Research, Children's National Research Institute, Children's National Hospital, Washington, DC, USA; [4]Department of Genomics and Precision Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, DC, USA; [5]Department of Neurology, Division of Neurogenetics and Neurodevelopmental Disabilities, Children's National Hospital, Washington, DC, USA; [6]Center for Neuroscience Research, Children's National Research Institute, Children's National Hospital, Washington, DC, USA
[7]Present address: University of Washington Medical Center, Seattle, WA, USA
[8]Present address: Fred Hutchinson Cancer Research Institute, Seattle, WA, USA
*Correspondence: aimee.dudley@gmail.com
https://doi.org/10.1016/j.ajhg.2023.03.019.

reducing the morbidity and mortality associated with delayed treatment or underdiagnosis. In fact, *OTC* was added to the American College of Medical Genetics and Genomics secondary findings gene list (ACMG SF v2.0),[12] which specifies a small set of highly actionable genes to be evaluated when an individual undergoes clinical exome and genome sequencing for any reason. For DNA sequencing to inform diagnosis and clinical management, any variants detected need to be clinically interpretable. However, only a small minority of variants observed in the human exome currently have benign or pathogenic interpretations. Instead, DNA sequencing often returns sequence variants of uncertain significance (VUSs), many of them novel, which cannot be used as a basis for diagnosis or treatment. Therefore, there is a critical need for methods that can aid clinical interpretation of genetic variation, particularly for highly actionable genes such as *OTC*.

High-throughput functional assays are one approach to addressing this problem at scale. Deep mutational scanning (DMS) or multiplexed assays of variant effect (MAVE) approaches have been applied to the comprehensive analysis of amino acid or nucleotide variants in synthetic peptides, protein domains, whole proteins, transcriptional promoters, RNA transcripts, splice sites, and DNA replication origins.[13–15] Because fundamental biological processes are generally conserved between humans and other species, it is often possible to leverage the experimental tractability of model organisms to perform low cost, high-throughput studies on massively parallel scales.

Many human protein-coding sequences can functionally replace (complement) their orthologs in the yeast, *Saccharomyces cerevisiae*.[16–24] This is especially true for genes that encode core metabolic processes such as nutrient biosynthesis and utilization.[23] The enzymes in the urea cycle responsible for arginine biosynthesis are highly conserved between humans and yeast (Figure 1A). At the protein level, human OTC is 34.5% identical and 54.5% similar to its yeast ortholog Arg3, and both function as homotrimers. Here, we present a complementation assay in which growth of yeast cells in the absence of arginine serves as a proxy for the function of human OTC. We use this assay at scale to measure relative growth scores for 1,570 SNV-accessible single amino acid substitutions across the length of the protein. Our results, based on multiple biological and experimental replicates of sequence-confirmed clones, are highly reproducible. A large proportion of these missense variants encode proteins exhibiting reduced protein function in our assay, including many that show complete loss of function. Variants within this extremely low range of activity include the majority of amino acid substitutions at catalytic residues, variants resulting from SNVs with pathogenic or likely pathogenic clinical classifications, and variants resulting from SNVs associated with severe disease (neonatal onset) in the literature. On this basis, our dataset satisfies the criteria of a well-validated functional assay, and we highlight examples in which its use as supporting evidence enables the reclassification of VUSs.

Material and methods

Plasmid and strain construction

All *S. cerevisiae* strains used in this study (Table S1) were derived from the isogenic lab strain FY4.[25] Unless noted, strains were grown in rich YPD medium (1% yeast extract, 2% peptone, and 2% glucose) or minimal (SD) medium (without amino acids, 2% glucose) with standard media conditions and methods for yeast genetic manipulation.[26]

To construct yeast strains harboring *OTC* variants, we first deleted the *ARG3* open reading frame from FY4 and replaced it with a selectable kanamycin resistance gene from pFA6a-kanMX6.[27] In this strain, expression of the kanamycin resistance gene was under control of the *S. cerevisiae ARG3* promoter and the *A. gossypii* TEF terminator.

We optimized the *OTC* coding sequence (GenBank: NM_000531.6) for expression in yeast (hereafter *yOTC*, GenBank: ON872185) by using a custom, in-house method designed to match the codon usage frequency of *S. cerevisiae*. Briefly, at positions where amino acids are conserved between the yeast and human proteins, the yeast (S288c reference) codon was used. In non-conserved locations, codons encoding the appropriate OTC amino acid were chosen to be as similar as possible to the usage frequency of the codons at corresponding positions in *ARG3*.

The mitochondrial leader sequence (amino acids 2–32) of the human OTC was omitted from *yOTC*, and *yOTC* was placed under the control of the native *ARG3* promoter. Hereafter, the stated variant amino acid positions refer to positions of native, full-length OTC. To avoid disrupting the regulation of the neighboring essential yeast gene (*TRL1*), which shares a 186 bp terminator sequence with *ARG3*, we retained the *TRL1* terminator and introduced a *TRP2* terminator sequence upstream of the drug marker to control the termination of the *yOTC* mRNA (Figure 1B).

A plasmid containing *yOTC* was constructed as follows. A DNA fragment containing the *ARG3* promoter (*pARG3*) and wild-type *yOTC* was synthesized (Genewiz) and cloned upstream of the *TRP2* terminator and *NatMX6* drug marker cassette in a pUC19-based vector with the NEB HiFi DNA Assembly Cloning Kit (New England Biolabs). This AB614_yOTC plasmid (GenBank: ON872185) was used as a template for variant library construction.

Construction of variant library

We designed the variant library to capture the amino acid substitutions resulting from all SNV-accessible missense mutations, excluding the start and stop codons, in the human *OTC* cDNA sequence (GenBank: NM_000531.6). These SNV-accessible amino acid substitutions were defined relative to the native human *OTC* cDNA sequence, not the *yOTC* codon-optimized sequence. The nine possible SNVs at each native codon can produce 4–7 unique amino acid substitutions. *yOTC* derivatives encoding the complete set of these unique amino acid substitutions (n = 1,879) were synthesized (Twist Biosciences) as a *pARG3-yOTC-tTRP2-NatMX* variant library in which each well of a 96-well plate contained an approximately equal pool of genotypes encoding each of the 4–7 amino acid substitutions at a given amino acid position (supplemental methods).

Separate PCR amplification reactions and subsequent integrative yeast transformations were then performed for each Twist well, i.e., a separate transformation for each amino acid position tested (n = 323). Approximately 500 ng of each PCR-amplified

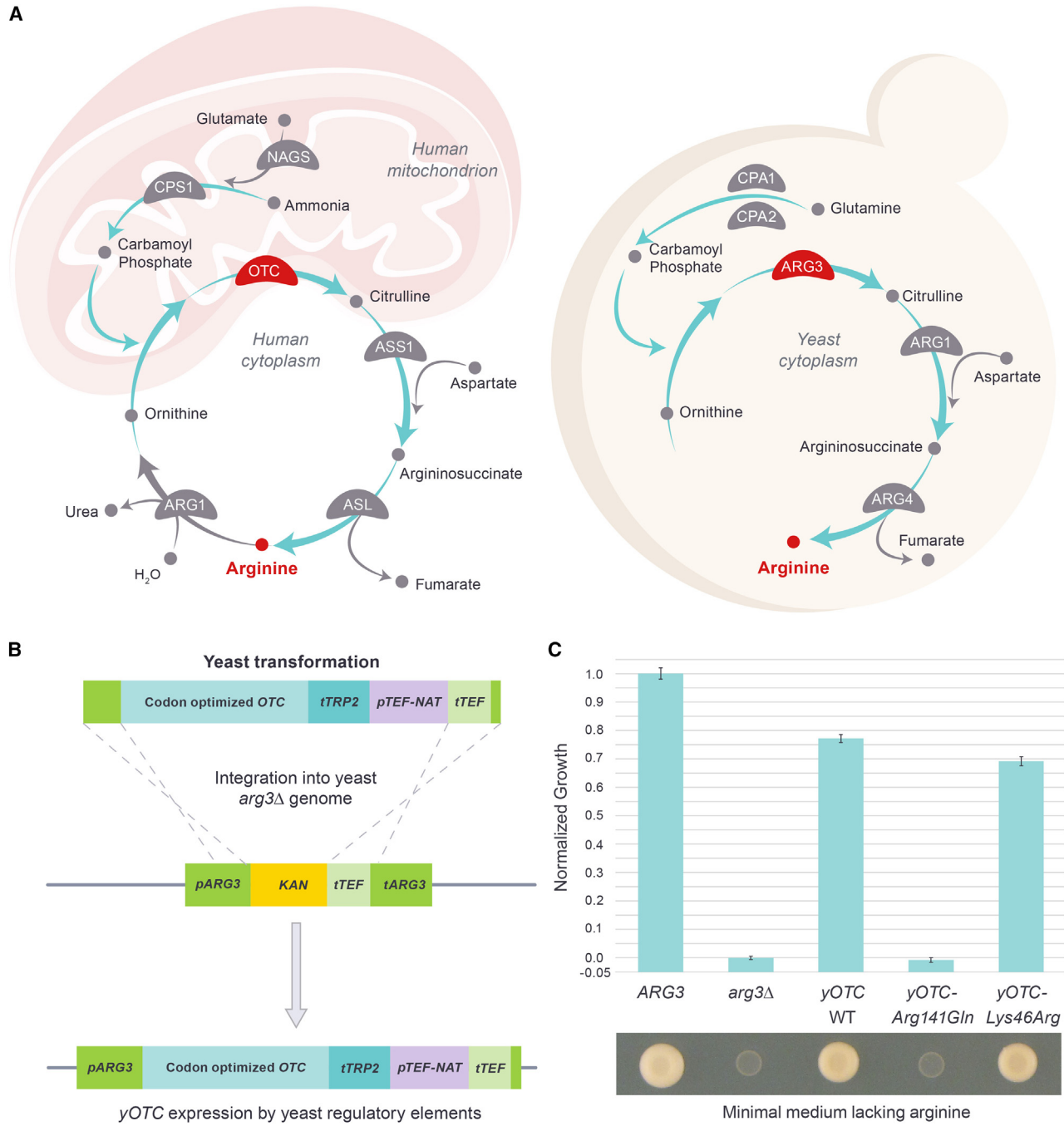


Figure 1. A yeast-based functional assay for human OTC amino acid substitutions

(A) Comparison of the urea cycle in humans (left) and the arginine biosynthesis pathway in *S. cerevisiae* (right).

(B) Yeast strain construction. The *arg3*Δ strain was transformed with PCR products containing either wild-type or variant alleles of the yeast codon-optimized human *OTC* coding sequence (*yOTC*), which is adjacent to a nourseothricin resistance drug marker (NAT). Homology-directed integration at the *ARG3* locus places *yOTC* under the control of the yeast *ARG3* promoter and *TRP2* terminator.

(C) Yeast growth in the absence of arginine as a quantitative assay for OTC function. In the absence of arginine, yeast cells harboring the native *OTC* ortholog (*ARG3*) grow robustly, but cells harboring a precise gene deletion (*arg3*Δ) do not. Yeast cells harboring wild-type *yOTC* or a benign variant (p.Lys46Arg) as the sole source of ornithine transcarbamylase activity grow at 77% and 69% of *ARG3*, respectively. Yeast cells harboring a pathogenic variant (p.Arg141Gln) are unable to grow. Growth on minimal medium is calculated as the product of the area and the intensity from plates imaged after 3 days. Growth estimates for each genotype ± standard deviations are displayed relative to *ARG3* (set to 1.0) and *arg3*Δ (set to 0). Six independent isolates of each strain (*ARG3*, *arg3*Δ*0*, wild-type *yOTC*, *yOTC-Arg141Gln*, and *yOTC-Lys46Arg*) were assayed in triplicate.

amino acid variant pool was transformed into a MATa haploid deletion strain (*arg3*Δ*0*) via standard methods. From each transformation, single colonies were isolated such that a total of 5,700 individual transformants were arrayed into 96-well plates containing rich medium. This number of colonies was chosen such that approximately three independent transformants of each variant

would be isolated and assayed. For downstream phenotype normalization, each library plate also contained replicates of the same control strains: two deletion (*arg3*Δ0) and four wild-type (*yOTC*). Stocks were maintained as individual strains in 96-well format.

Variant library sequence confirmation

In our strain construction pipeline, for each transformant in a given well of a 96-well plate, the target codon harboring the amino acid substitution is known, but the specific variant is not. To determine this, we used a custom MinION (Oxford Nanopore Technologies) long-read sequencing pipeline (described in detail in supplemental methods, script in Methods S1). Briefly, individual transformants were pooled in groups of 16, such that no target codon position was represented more than once in a single pool. Each pool was then sequenced using Oxford Nanopore Flongle flow cells (R9.4.1). At each target codon in each pool, the most frequent potential variant codon was identified (candidate variant) as well as the second most frequent variant (second variant). Because we know which target codon corresponds to which transformant, this allows us to associate each candidate variant with a single transformant.

The frequency and pattern of MinION sequencing errors is highly variable and depends on the context of surrounding bases. These errors can occur at high frequencies, potentially generating spurious matches to variant codons. However, the error patterns are also reproducible per given DNA sequence, allowing us to develop an error model for each variant, describing the frequency with which it is generated by sequencing errors. This frequency can be compared to the frequency observed for each candidate variant in the pooled sequencing, allowing true variants to be distinguished from sequencing noise.

We performed quality control on candidate variants by using the error model along with minimum variant frequency thresholds (supplemental methods). Candidates were also flagged for later removal if any missense or nonsense secondary mutations, or insertions or deletions (indels), were present in the *yOTC* sequence. The end result of this process was that each transformant was assigned either a high-confidence call for the variant present in that transformant or an "NA" call that resulted in the removal of that transformant from further consideration.

Validating the MinION variant-calling pipeline with Illumina sequencing

Long-read sequencing technologies, such as Oxford Nanopore, allow us to identify both the *yOTC* variant present in each transformant and also any secondary mutations elsewhere in the gene. To validate our Oxford Nanopore pipeline, we performed amplicon sequencing by using highly accurate, but short-read, Illumina sequencing, to identify the variant in each transformant (supplemental methods). These calls were then compared to the result from the Oxford Nanopore pipeline. To accommodate Illumina read length limits, the *yOTC* sequence was divided up into four ~300 bp overlapping segments. The specific segment to be amplified and sequenced corresponded to the known codon position of the variant in each transformant. Comparison of the MinION and Illumina results revealed excellent agreement (99.8%) for the MinION calls passing quality control (Figure S1). The accession numbers for the Illumina and MinION read sequences generated for this study are SRA: PRJEB61166 and SRA: PRJEB60279.

Measuring the growth of individual isolates

To prepare for phenotyping, each isolate was picked and grown in rich medium supplemented with the selectable drug and then pinned onto solid rich medium containing 3% glycerol as the carbon source (YPG), to select against cells that were slow growing as a result of loss of respiratory function. Cells were then pinned from YPG solid medium back into non-selective rich liquid medium (YPD) and grown to saturation overnight at 30°C. Next, cultures were pinned onto solid omniplates containing minimal (SD) medium (lacking arginine) with a Biomek i7 robot outfitted with a V&P Scientific 96 floating-pin replicator head. Pinning onto SD plates was done in triplicate, followed by growth at 30°C for 72 h. 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, 1/40 s exposure) were acquired as .jpg files. Growth phenotypes were extracted from the images via a custom pipeline, PyPL8 (https://github.com/lacyk3/PyPl8). Briefly, a pseudo patch "volume" for each pinned isolate was generated, consisting of the product ("IntDen") of the patch area (obtained via thresholding with Otsu's method[28] or circle detection[29]) and the mean patch pixel intensity (grayscale) (described in detail in supplemental methods).

Data normalization and relative growth estimation

Raw phenotypic values were normalized, quality control filters were applied to each isolate, and a final relative growth estimate for each *yOTC* variant (across all isolates) was determined (described in detail in supplemental methods, processing script in Methods S2). Briefly, normalization steps to account for the effects of plate-to-plate variation, relative growth of neighboring patches, and plate edge effects were applied to the data with a custom script (Methods S2). Isolates with (non-synonymous) secondary mutations were removed from the dataset. Any *yOTC* variants that had discordant growth estimates between isolates were also removed. Finally, we used a linear model, fit with weighted least-squares regression, to estimate the relative growth of each *yOTC* variant, on a scale with growth of null controls set to 0 and growth of wild-type *yOTC* set to 1 (supplemental methods and Methods S2).

Curation of *OTC* sequence variant clinical significance

Criteria developed for the interpretation of clinical significance of sequence variants by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP)[30] were applied to the *OTC* benign variant (p.Leu166Phe) and to variants with growth rates below 5% of the wild-type that were annotated either as VUSs, likely pathogenic (LP) variants, or variants with conflicting interpretation (CI) in the ClinVar database[31] (May 20, 2022). Variant Curation Interface was used for variant classification.[32] The following moderate criteria were used for classification of *OTC* sequence variants: PM1, located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation; PM2, allele frequency in human population databases; PM5, missense change at an amino acid residue where a different pathogenic or likely pathogenic missense change has been seen before; PS4_Moderate, one male or female proband with disease-specific phenotype (PP4).[30] The following supporting criteria were used for classification of *OTC* sequence variants: PP3, multiple lines of computational evidence with REVEL score > 0.75 support a deleterious effect on the gene or gene product; PP4,

disease-specific phenotype—elevated plasma glutamine or elevated ammonia AND low/normal citrulline, elevated urine orotic acid, and enzyme activity <20% in patient-derived cells from a male proband. The PP4 and PS4 criteria were not applied to the same proband. Variant Curation Interface applies classification criteria to determine the clinical significance of each sequence variant according to the rules described in Richards et al.[30] Growth rates of each *OTC* sequence variant were applied as PS3_Supporting level of evidence for variant pathogenicity on the basis of the number of known benign and pathogenic variants used as controls.[33]

Permutation p values

For calculations of permutation p values, 1 million samples of the appropriate size were drawn, without replacement, from the appropriate subset of assay results. These samples were used to calculate a null distribution of medians, to which the test median was compared.

Conservation and structure analysis

Conservation scores of OTC (PDB: 1OTH) were downloaded from ConSurf-DB server.[34] Surface residues of 1OTH were identified by running findSurfaceResidues.py (https://pymolwiki.org/index.php/FindSurfaceResidues) with default cutoff exposed area of 2.5 Å in PyMOL version 1.8.2. We consider interior residues output from findSurfaceResidues.py as buried residues. Structural visualization was conducted in PyMOL 1.8.2. A Chi-squared test was performed in R to determine significant association of conserved residues (ConSurf conservation score < 0) in the amorphic range of median yeast growth (median estimate ≤ 0.05).

Results

Establishing an *in vivo* OTC functional assay in yeast

We set out to create a quantitative, yeast-based assay to measure the impact of amino acid substitutions on the activity of human OTC. Our assay is based on the ability of human OTC to functionally replace its yeast ortholog (Arg3) in *S. cerevisiae*. Human *OTC* contains ten exons and nine introns but, like most genes in the yeast genome, *ARG3* is intronless.[35] Therefore, we started with the *OTC* cDNA sequence (from GenBank: NM_000531.6), which encodes the 354 amino acid OTC precursor. This sequence includes a 32 amino acid mitochondrial leader sequence that is cleaved from the N terminus of OTC following mitochondrial import.[36] Because the Arg3 enzyme functions in the yeast cytosol,[37] we omitted amino acids 2–32 from the human *OTC* coding sequence. However, to simplify comparisons with the clinical literature and databases, we refer to variant amino acid positions within this protein by using their position in native, full-length OTC.

We next sought to match the gene expression of *OTC* to that of its yeast ortholog as follows. First, we generated a codon-harmonized version of *OTC* that encodes the human protein but more closely matches the codon utilization frequency of *ARG3* than does the original *OTC* cDNA sequence (material and methods). A single copy of this harmonized sequence, hereafter *yOTC*, was integrated into the (haploid) yeast genome at the *ARG3* locus, under the control of the native *ARG3* promoter (Figure 1B). This experimental design ensured stable maintenance of *yOTC* (wild-type and variant derivatives) at a uniform copy number in all cells and placed *yOTC* in the native regulatory environment of the yeast ortholog.

Because ornithine transcarbamylase activity is required for yeast to grow in the absence of exogenously supplied arginine, growth on minimal medium (lacking arginine) is a quantitative measure of enzyme function that is amenable to high-throughput analysis. To quantify the ability of *yOTC* to functionally replace *ARG3*, we compared growth of the parental strain (FY4, expressing yeast *ARG3*) to that of the same strain background lacking any ornithine transcarbamylase activity (*arg3Δ0::NATMX*) or harboring the human OTC expression construct (*yOTC*) at the *ARG3* locus as the sole source of enzyme activity. Each strain was grown under a non-selective condition (rich medium containing arginine), replica pinned to minimal medium lacking arginine, and grown at 30°C for 72 h. As expected, in the absence of arginine, FY4 grew robustly while the *arg3* deletion was unable to grow, displaying only the faint patch of cells deposited by the initial replica pinning (Figure 1C). In agreement with a previous plasmid-based study that expressed the full-length protein-coding sequence in yeast,[38] human ornithine transcarbamylase expressed from our integrative construct (*yOTC*) was able to complement an *arg3* deletion. In our assay, complementation was robust, conferring 77% growth relative to the FY4 strain expressing the yeast ortholog Arg3 (Figure 1C).

As an initial test of how well activity in yeast agreed with activity in humans, we constructed and individually assayed a small set of well-characterized variants with pathogenic or benign clinical significance calls in ClinVar.[31] We expected that amino acid changes matching those encoded by known pathogenic missense variants would impair OTC activity, resulting in poor growth, potentially down to the complete lack of growth conferred by the *arg3* deletion. In contrast, amino acid changes corresponding to those of benign missense variants should result in high growth, potentially up to the high level conferred by wild-type *yOTC*. The pathogenic missense SNV c.422G>A (p.Arg141Gln) (GenBank: NM_000531.6) encodes an amino acid substitution that abolishes enzymatic activity without altering protein abundance.[39] In our assay, the strain harboring this amino acid substitution behaved like the *arg3* deletion mutant and failed to grow on minimal medium (Figure 1C). The most common benign missense variant reported in gnomAD is c.137A>G (p.Lys46Arg) (GenBank: NM_000531.6). In our assay, this amino acid variant functionally complemented the *arg3* deletion to 90% of the growth of the strain harboring the wild-type human construct (*yOTC*) and to 69% of the original yeast strain harboring the yeast ortholog (*ARG3*) (Figure 1C). Thus, *yOTC* complements a deletion of *ARG3*, and the activities of amino acid substitutions corresponding to a small number of well-characterized pathogenic and benign missense variants are consistent with expectations.

Assaying SNV-accessible OTC amino acid substitutions at scale

With a validated functional assay in hand, we proceeded with the construction and analysis of a full-scale variant library. While it is possible to construct variant libraries that sample all 19 amino acid substitutions at a given residue, we focused on the subset of missense variants that are most likely to arise in the human population. Thus, our library design included all single amino acid substitutions (four to seven per residue) that are accessible by a single-nucleotide variant (SNV) in the reference human *OTC* sequence. While our experimental design measures the impact on protein function of the amino acid substitutions encoded by the SNVs, it does not directly test nucleotide variation itself. As such, the assay is not an informative readout for the effects of nucleotide-based phenomena such as synonymous mutations and mRNA splicing.

Yeast cells were transformed with a library of *yOTC* derivatives, each encoding a single amino acid substitution. Individual transformants were isolated and arrayed into 96-well plates, ensuring that each strain in each 96-well plate is an independently constructed biological isolate of the variant it contains (material and methods). The number of colonies picked (n = 5,700) was chosen such that on average, three biological isolates for each substitution would be collected and assayed. The identity of the altered codon in each transformant was then determined by sequencing the full length of *yOTC* with a custom long read (Oxford Nanopore) pipeline. The ability of this pipeline to accurately determine the *yOTC* variant present in a given strain was confirmed with Illumina sequencing, with 99.8% agreement between the altered codon identified by the Oxford Nanopore pipeline and that identified by Illumina sequencing (material and methods). The ability to sequence the entire protein-coding sequence with the long read pipeline also allowed us to identify transformants with secondary mutations in *yOTC* and remove from the final dataset any transformants with indels or non-synonymous or nonsense secondary mutations. After this filtering step, the remaining genotypes included 1,592 out of 1,879 (85%) of the total possible SNV-accessible amino acid substitutions in *OTC*.

We next measured the growth of each strain individually by robotically pinning cells arrayed in 96-well format to solid medium. In contrast to pooled competitive fitness-based approaches, phenotyping in this manner provides a stable estimate of activity for each *yOTC* variant. Specifically, individual measurements are independent of the composition of a particular competitive pool and are not subject to relative fitness changes as the mean pool fitness increases over time.[40] Individual strain phenotyping also facilitates the integration of the current dataset with additional strain phenotyping carried out in the future. Assaying future strains with a defined set of controls allows the additional data to be placed on the same quantitative scale and be compared directly to the earlier dataset.

Briefly, the growth assay was performed as follows. First, strains were grown to saturation in 96-well plates containing rich liquid medium (containing arginine). Then, cells were robotically replica pinned in triplicate to solid agar plates containing minimal medium (lacking arginine) and grown for 72 h. Next, growth was measured as the product of the area and average pixel intensity of each patch, with a custom script (material and methods). After normalization to account for plate-to-plate, edge, and neighbor patch effects on growth, we used a linear model to estimate the growth of each yeast *yOTC* genotype (using all transformants with that genotype). Finally, growth values were rescaled so that wild-type *yOTC* growth was set to a value of 100% and the *arg3* deletion strain was set to a value of 0% (Table S7) (material and methods). The results were highly reproducible. For genotypes with more than one isolate, the correlation between the genotype growth estimate and the individual isolate growth estimates had an r^2 value 0.982 (Figure S2). A small number of genotypes were identified that displayed inconsistent growth between independent isolates (Figure S2) (material and methods). These genotypes were removed from further analysis, leaving a final dataset of 1,570 amino acid substitutions in *OTC*, representing 84% of all SNV-accessible amino acid substitutions, and for the majority (77%) of these, multiple independent transformants were isolated and individually assayed (Table S7).

A high proportion of amino acid substitutions impair OTC activity

The amino acid substitutions tested conferred a wide range of growth values spanning the range between the wild-type (*yOTC*) and null (*arg3*) controls but, quite strikingly, 27% of amino acid substitutions exhibited relative growth values below 5% of wild-type activity (Figure 2). The approximately normal distribution of these low values, centered on the null control, is consistent with complete loss of OTC function in our assay combined with a small amount of measurement noise. We therefore defined substitutions falling in this range as amorphic. The remaining amino acid substitutions exhibited a relatively uniform distribution across the rest of the growth range. Within this group, we defined amino acid substitutions having growth >90% as functionally unimpaired and those falling in the range 5%–90% as hypomorphic. Unexpectedly, there was no frequency peak of amino acid substitutions centered around the wild-type level of growth in the unimpaired range and only 18% of amino acid substitutions had growth >90% in our assay (Figure 2).

The fact that the human protein-coding sequence did not fully complement loss of the yeast ortholog (77% of *ARG3* growth, Figure 1C) suggests that our assay has the potential to detect hypermorphic variants exhibiting increased growth up to at least the level of the *ARG3* strain. Interestingly, among the 16 amino acid substitutions with the highest growth (top 1%), ten were substitutions at positions 143, 324–326, or 349–350 (Figure S3).

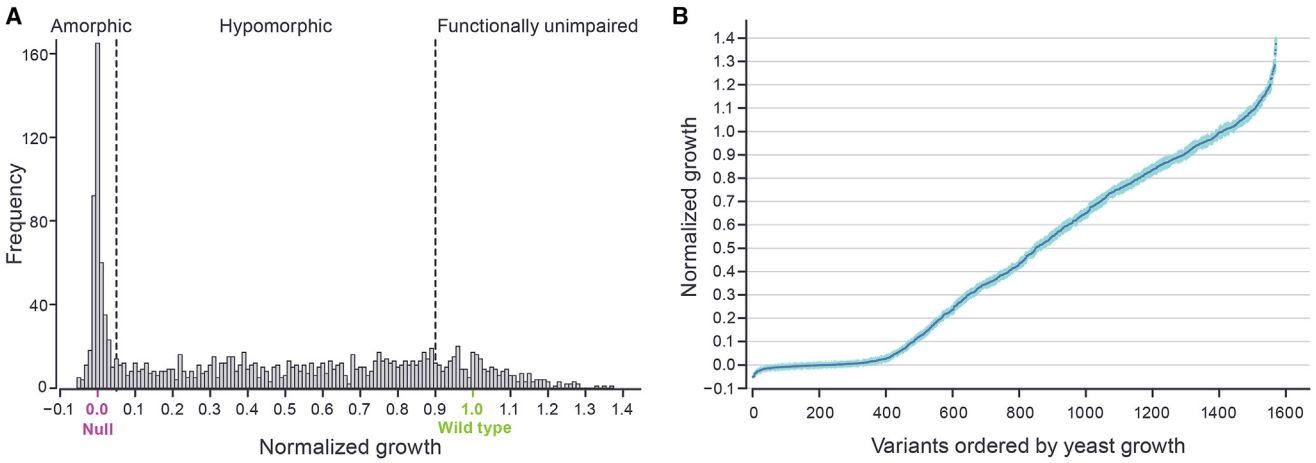


Figure 2. Quantitative growth measurements of 1,570 OTC variants

(A) Frequency of *yOTC* variants exhibiting varying levels of normalized growth. Variant growth is reported relative to that of the *arg3* deletion (normalized growth value = 0, marked in purple) and the wild-type *yOTC* (normalized growth value = 1, marked in green). The 0.05 and 0.9 lines used to demarcate the amorphic, hypomorphic, and unimpaired growth ranges are indicated with dashed gray lines.

(B) Relative growth of each ordered *yOTC* variant plotted with ± SE of the growth estimates.

Assay results agree with expectations based on protein function, evolutionary conservation, and variant representation in the human population

We next examined the functional impact of amino acid substitutions in our assay with respect to the evolutionary conservation of residues and the protein structure of OTC. Substitutions causing loss of function were spread relatively evenly across the linear sequence of the protein and most positions exhibited high levels of intolerance (Figures 3A and S3). Because OTC is an evolutionarily ancient enzyme, we expected conserved residues to be highly sensitive to mutation and variable residues to be more tolerant to mutation. Consistent with this, we observed a highly significant ($p = 3.1 \times 10^{-8}$) but weak (Spearman's $\rho = 0.3$) correlation between the ConSurf[34] relative conservation scores and our median growth estimates for each residue (Figures 3A and S4); with more conserved sites showed lower median growth. In particular, we observed a strong enrichment ($p = 7 \times 10^{-4}$, Chi-squared test) of the most highly intolerant substitution class (median values in the amorphic range) at conserved residues (negative ConSurf scores) (Figure S4).

Because buried residues are important for proper protein folding, we expected them to be both highly conserved and highly sensitive to amino acid substitution in OTC. When we mapped the conservation scores and the median yeast growth onto the protein structure, the majority of both the highly conserved and mutation-sensitive residues were indeed concentrated in the internal parts of the protein (Figures 3B, 3C, and S4). In particular, high levels of conservation and sensitivity to substitution were observed around the substrate-binding and catalytic "pocket" of OTC (Figures 3B and 3C). In contrast, the least conserved and most substitution-tolerant residues were found mostly on the protein surface (Figures 3B, 3C, and S4).

While OTC shows a general intolerance to amino acid substitutions, we expected that altering the amino acids at active sites of the enzyme would have a particularly high likelihood of impairing OTC function and therefore inhibiting growth in our assay. OTC has several known active sites involved in interaction with the substrates carbamoyl phosphate and ornithine for enzymatic conversion to citrulline.[41,42] Interactions with carbamoyl phosphate involve the SxRT (Ser90, Arg92, Thr93) and HPxQ (His168, Pro169, Gln171) motifs as well as residues His117, Arg141, and Arg330. Residues Asp263, Leu163, and Asn199 interact with ornithine while the HCLP (His302, Cys303, Leu304, and Pro305) motif participates in interactions with both substrates. The SMG motif (Ser267, Met268, and Gly269) is involved in the capping of the active site pocket. Most of these active site residues were very sensitive to the effect of amino acid substitutions (Figure 4A). The median growth of substitutions at all active site residues was only 13.2%, which is significantly, and very substantially, lower than the global median of 41.7% ($p < 1 \times 10^{-3}$, one-sided permutation test). In particular, 77% of substitutions at catalytic residues were amorphic (<5% growth), significantly greater than the 27% of all substitutions that fall in this range ($p = 1.0 \times 10^{-8}$, one-sided exact binomial test). Thus, as expected, active sites were intolerant to amino acid substitutions in our assay.

One notable exception to the pattern of sensitivity at conserved functional residues is the SMG motif, which is tolerant to amino acid substitutions in our assay (Figure 4A). This motif is part of a loop from amino acids 264–276 that swings to close the active site upon binding of carbamoyl phosphate and ornithine (Figure 4B). The full length of this mobile loop is tolerant to amino acid substitutions (Figure S5), suggesting that its function,

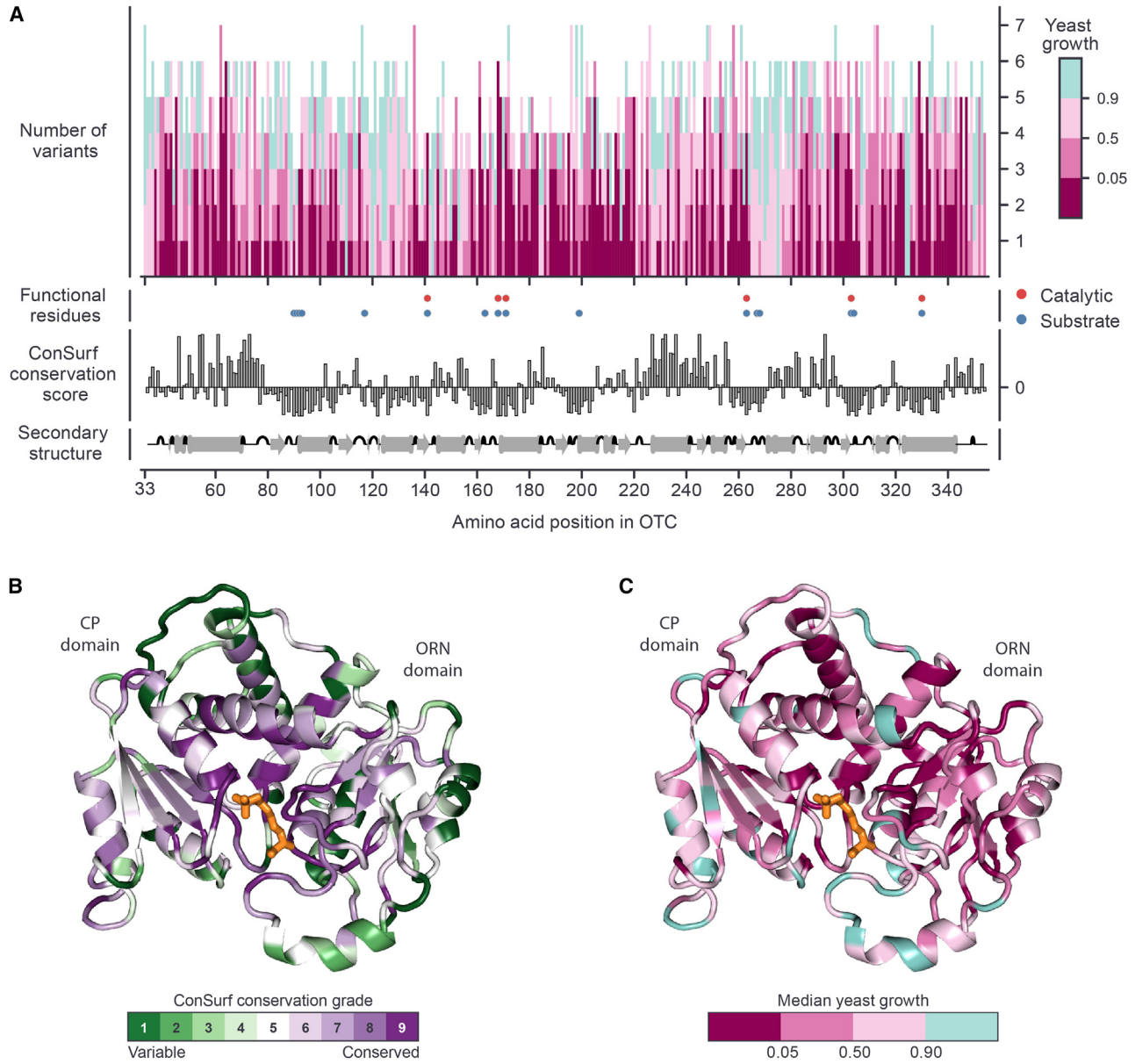


Figure 3. Missense variant effect map across the length of human OTC compared to structural features and conservation
(A) The top plot depicts the binned growth of the 2–7 substitutions introduced at each amino acid position in our library, ordered from highest (top) to lowest (bottom) growth. Below are the positions of OTC functional residues (substrate binding and catalysis), ConSurf[34] evolutionary conservation scores (PDB: 1OTH),[41] and OTC secondary structure. More negative ConSurf scores indicate more conserved positions. Helices and beta-sheets are depicted as gray cylinders and arrows, respectively, and turns are shown as upward half-coils.
(B) Protein structure of OTC (PDB: 1OTH) colored by binned ConSurf conservation score.
(C) Protein structure of OTC (PDB: 1OTH) colored by binned median yeast functional assay score.

while important for the activity of OTC in human mitochondria, may not be required for OTC activity in the context of the yeast cytoplasm.

Finally, we evaluated the agreement of our results with variant representation in the human population. Because of the severity of this X-linked disease, we expect that *OTC* missense SNVs that result in deleterious amino acid substitutions will be underrepresented. Consistent with this, substitutions corresponding to human missense SNVs present in gnomAD[43] (May 20, 2022) have significantly higher median levels of growth than do all substitutions (71.1% versus 40.4%; $p < 10^{-3}$, one-sided permutation test). In particular, substitutions corresponding to gnomAD SNVs are very strongly and significantly ($p = 2.5 \times 10^{-8}$, one-sided exact binomial test) depleted among substitutions showing <5% growth in our assay, with 1.5% of gnomAD substitutions falling below this cutoff versus 27% of all substitutions (Figure S6A). Interestingly, for growth values above 5% in our assay, the distributions of gnomAD substitutions and all substitutions appear similar (Figure S6B).

Thus, although the proportion of all amino acid substitutions in OTC that are deleterious is high, our results

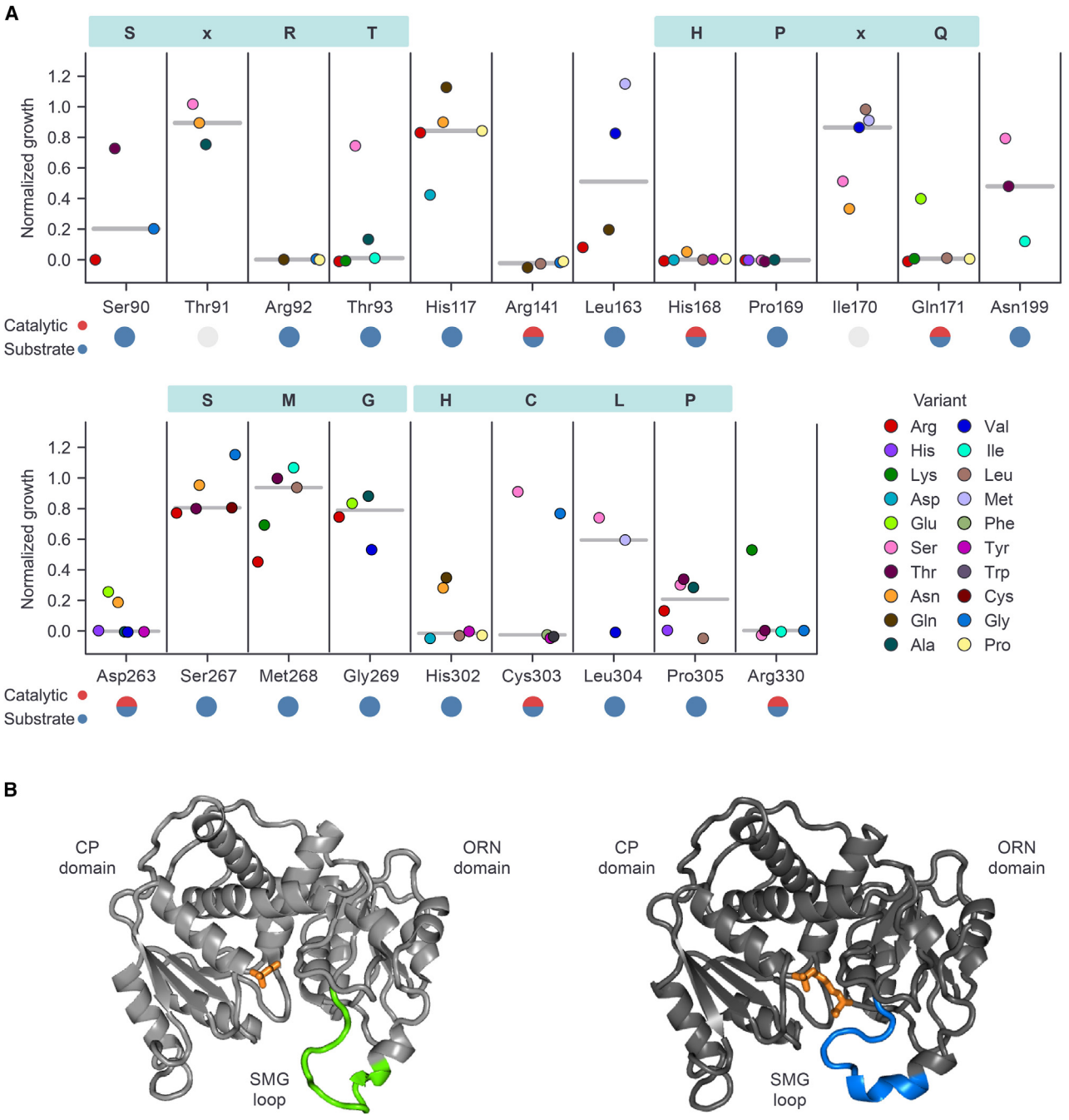


Figure 4. Effect of amino acid substitutions at functionally important residues

(A) The distribution of variant growth scores for residues in OTC motifs involved in substrate binding and catalysis. Each circle is colored according to the amino acid that is substituted as shown (variant). The median growth score for each position is shown as a horizontal gray bar.

(B) Protein structures of OTC highlighting the SMG loop at residues 264–276 (green and blue, left and right, respectively). The left structure shows OTC binding CP in orange (PDB: 1FVO). The structure on the right shows OTC binding PALO, an analog of CP-ornithine, with PALO in orange (PDB: 1OTH).

agree with expectations based on the known features of the protein structure and with variant representation in the human population.

Assay results closely agree with clinical annotations

To evaluate how our functional assay compared with human phenotypes, we assessed the growth scores of amino acid substitutions corresponding to those caused by missense SNVs with definitive clinical significance calls in the ClinVar database (Figure 5). In our assay, we expect that amino acid substitutions corresponding to benign missense variants will display high growth, while those corresponding to pathogenic missense variants will display reduced growth.

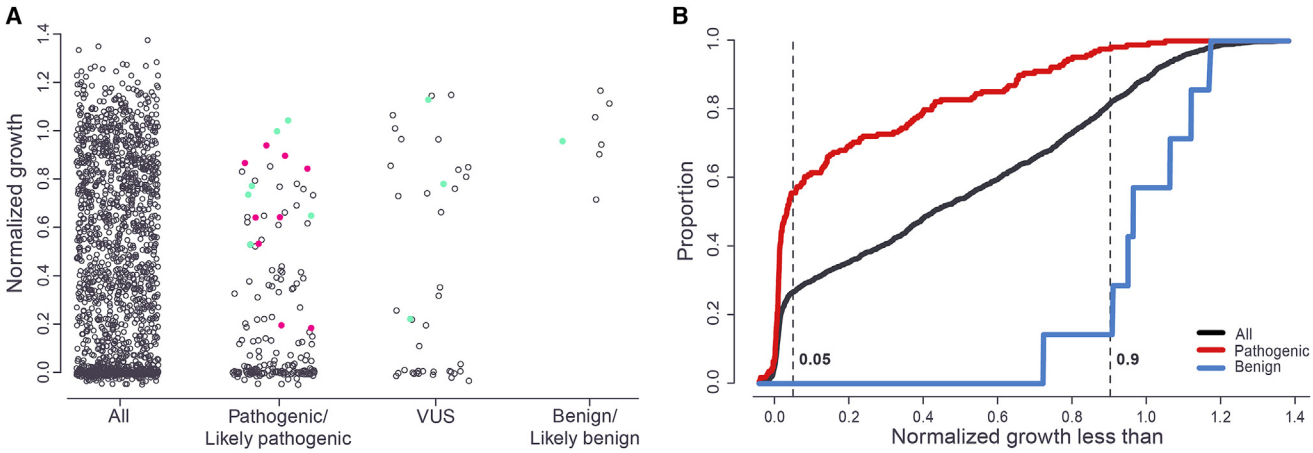


Figure 5. Yeast assay results for amino acid substitutions corresponding to variants with clinical significance calls
(A) Strip charts of normalized growth for all amino acid substitutions and those corresponding to ClinVar classification groups. Colored dots correspond to amino acid substitutions in the SMG loop (green) or that correspond to pathogenic/likely pathogenic missense SNVs predicted to impair splicing (pink).
(B) Comparison of the proportion of all amino acid substitutions and those with either ClinVar pathogenic/likely pathogenic or benign/likely benign annotations falling under a given level of normalized growth. Dotted lines indicate boundaries of amorphic, hypomorphic, and functionally unimpaired assay ranges.

ClinVar (May 20, 2022) contains twelve missense variants classified as benign or likely benign. Two of these encode substitutions (p.Met21Val and p.Arg23Gln) in the mitochondrial leader portion of the protein, which is not included in our assay. Seven of the remaining ten amino acid substitutions were present in our library. Six of these (p.Lys46Arg, p.Gly50Ala, p.The150Ile, p.Thr150Asn, p.His255Arg, and p.Gln270Arg) exhibited growth levels in the unimpaired range of our assay (>90%) and one (p.Leu166Phe, growth = 71%) in the hypomorphic range. No benign substitutions displayed growth in the amorphic range of our assay (<5%) (Figure 5). In contrast, for all substitutions, the proportions in the amorphic, hypomorphic, and unimpaired growth ranges were 27%, 55%, and 18%, respectively. This difference in distributions, where the benign variants showed less functional impairment, was significant ($p = 2.8 \times 10^{-4}$, exact multinomial test).

We next analyzed the set of amino acid substitutions in our library corresponding to SNVs with ClinVar pathogenic (n = 165 of 208) and pathogenic/likely pathogenic or likely pathogenic (n = 30 of 39) classifications. The majority of substitutions found in these groups (57%) displayed amorphic growth (<5%) (Figure 5), greatly enriched versus the 27% of all substitutions that fall in this growth range. The pathogenic-associated substitutions were also strongly depleted in the unimpaired (>90% growth) range of our assay; only 2% of pathogenic substitutions fell in this range versus 18% of all substitutions. The difference between the distribution of the pathogenic-associated substitutions versus all substitutions, where pathogenic-associated amino acid substitutions showed greater functional impairment, was highly significant ($p = 2.4 \times 10^{-6}$, exact multinomial test).

A recent study[44] characterized the functional impact of 71 pathogenic variants in *OTC* by using a mammalian cell-based spectrophotometric assay, where all of the tested variants demonstrated strong reductions in measured OTC activity (<50% of wild-type). 44 of these were missense variants encoding amino acid substitutions that were also tested in our assay. Among these, 23 (52%) had growth rates falling in our amorphic range, and all but one of the remaining variants fell in the hypomorphic range (<90% growth). Therefore, both assays confirm that these pathogenic amino acid substitutions lead to impaired OTC activity.

The expected correlation between pathogenicity and low growth in our assay is based on the assumption that the pathogenicity of a missense SNV is an effect of the amino acid substitution that it encodes. However, in addition to causing amino acid substitutions, missense SNVs in the first and last codons of an exon can also affect mRNA splicing. Because the intronless construct used in our assay is not subject to splicing effects, pathogenicity and growth in our assay may not be correlated for this small class of variants. For example, a missense variant pathogenic through its effect on splicing might encode an amino acid substitution with little effect on enzyme activity, leading to high growth in our assay.

To identify amino acid substitutions encoded by pathogenic missense SNVs that might impact splicing, we first identified all 76 possible missense SNVs in the last three bases of exons 2–9 and the first base of exons 3–10 of *OTC*. The junction between exons 1–2 was not assessed because exon 1 encodes the mitochondrial leader sequence, which is not present in our expression construct. *In silico* predictions for the effect of these variants on *OTC* mRNA splicing were then generated with MaxEntScan,[45] VarSeak (www.varSEAK.bio), and SpliceAI,[46] allowing us to classify variants as having a high, intermediate, or low probability of impairing splicing (Table S8). Among the SNVs with a high

probability of impairing splicing, we determined that ten also had ClinVar pathogenic or likely pathogenic classifications, and we identified nine amino acid substitutions in our assay corresponding to one of these SNVs. Four of these substitutions (p.Arg129His, p.Arg129Leu, p.Gln180His, and p.Lys289Asn) displayed high growth value in our assay (>84% growth), suggesting that the corresponding SNVs are pathogenic primarily due to their effect on splicing. The remaining five substitutions (p.Arg129Pro, p.Lys221Asn, p.Glu239Gly, p.Glu239Val, and p.Met335Ile) displayed impaired growth (<65%) in our assay (Table S7). Therefore, the corresponding SNVs both encode amino acid substitutions that have a deleterious effect on OTC function and are also likely to impair splicing, making their mode of pathogenicity ambiguous. Thus, for pathogenic missense SNVs with the potential to impact splicing, our data could be used to help disentangle the molecular mechanisms of pathogenicity.

After accounting for pathogenic variants with likely effects on splicing, the only two remaining pathogenic/likely pathogenic amino acid substitutions with growth >90% in our assay (p.The264Asn and p.Met268Thr) are located in the SMG loop (Figure 5A). This is consistent with our earlier observation that highly conserved residues in this region are tolerant to amino acid substitutions (Figures 3, 4, and S5) and support the possibility that the function provided by this region of OTC in human mitochondria is not essential when OTC is localized in the yeast cytoplasm. Based on the region predicted to undergo a conformational change upon binding of ornithine (Figure 4B), we flagged resides located in the portion of the SMG loop from residues 264–276 as a region in which the yeast assay results might not be informative about human protein function.

Defining informative ranges of the assay to support clinical variant classification

The strong correlation between the clinical significance calls in ClinVar and the quantitative scale of our assay supports the validity of the yeast assay for assessing human protein function. By comparing the distribution of existing benign and pathogenic variants, it is possible to identify assay ranges that are consistent with pathogenic and benign clinical presentation. Although the number of amino acid substitutions associated with benign annotations is small, the degree of separation between the benign and pathogenic growth distributions observed in our assay suggests that a relatively clear classification boundary between the two clinical classes may exist. After removing from consideration amino acid changes associated with variants likely to influence splicing or that are present in the SMG loop, all remaining pathogenic-associated amino acid substitutions (154/154) had growth <86% in our assay. In contrast, all but one benign substitution (6/7) had growth >90% in our assay, and the outlier (p.Leu166Phe) had 71% growth (Figure 5). This suggests a cutoff between the two classes could be placed either just below 71%, in which case seven pathogenic variants would be misclassified as benign, or in the range 86%–90%, where one benign variant would be misclassified as pathogenic. Using a cutoff of 90% would have the advantage of both minimizing misclassification of existing clinically annotated variants and minimizing the potential to underdiagnose pathogenic variants for this severe and clinically actionable disease. In addition, the fact that one of a limited number of benign variants behaves as an outlier by our assay (p.Leu166Phe), suggests that the clinical classification of the corresponding SNV, c.498G>T (p.Leu166Phe) (GenBank: NM_000531.6), should be examined more closely. In fact, under the current ACMG guidelines and independent of our data, this variant would be reclassified as a VUS (Table S9).

On this basis, we chose a cutoff of 90% growth, below which we categorize substitutions as deleterious, impairing OTC function to a level sufficient for disease presentation, and above which we categorize substitutions as non-deleterious (Table S7). As such, the deleterious class corresponds to the combined amorphic and hypomorphic classes previously defined. Using this threshold, 1,294 (82%) of variants assayed exhibit functional impairment consistent with pathogenicity. Functional information that is consistent with the behavior of known pathogenic and benign alleles can be used as supporting information for clinical classification of the corresponding human SNVs. Our assay measuring the effect of amino acid substitutions on OTC activity meets this requirement, allowing it to be considered a well-validated functional assay for use as a PS3 supporting level of evidence according to the ACMG guidelines. In addition, we anticipate that as more clinical data becomes available over time, and in particular a larger set of benign variants are identified, the boundary between the benign and pathogenic distributions of values in our assay may become better resolved.

Variant assay values agree closely with *OTC* clinical stratification

In males, the presentation of OTC deficiency is classified into two groups defined by age of onset: neonatal and late-onset (more than 6 weeks of age). Amino acid substitutions corresponding to SNVs with neonatal onset are expected to more severely impair OTC activity than those corresponding to the late-onset alleles. In females, it has been suggested that disease presentation results from skewed patterns of X-inactivation affecting expression of a wild-type and a severe pathogenic SNV.[47–49] To examine how well our results agree with clinical stratification of OTC deficiency, we reviewed the disease literature and identified missense SNVs observed in instances of male neonatal, male late-onset, and female disease presentation. We then identified any amino acid substitutions tested in our assay that correspond to these SNVs, which comprised 69 neonatal male, 57 late-onset male, and 88 female substitutions (Table S7). No substitutions were present in more than one class.

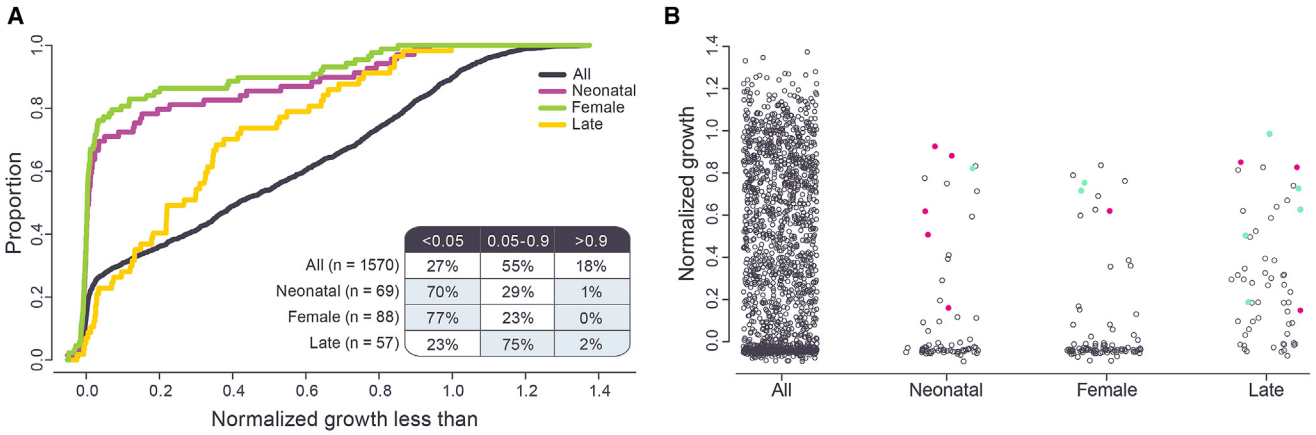


Figure 6. Clinical stratification of yeast assay results

(A) Comparison of the proportion of all amino acid substitutions falling under a given level of normalized growth versus those associated with neonatal, female, or late presentation. Table insert gives the proportion of each class falling in each growth range. Blue shading indicates enrichment and depletion relative to the whole population.

(B) Strip charts of normalized growth for all amino acid substitutions and those corresponding to disease presentation groups. Colored dots corresponding to amino acid substitutions in the SMG loop (green) or to missense SNVs predicted to impair splicing (pink) are shown.

Consistent with expectations, amino acid substitutions observed in the neonatal and female classes were associated with severe loss of activity in our assay. The two distributions were very similar (Figure 6A), showing strong enrichment in the amorphic range (<5% growth), depletion in the hypomorphic range (5%–90% growth), and strong depletion in the functionally unimpaired range (>90% growth). In particular, 70% of the neonatal class and 77% of the female class behaved as amorphs versus 27% of all substitutions while 1% of neonatal and 0% of female substitutions displayed unimpaired growth versus 18% of all variants. The difference between the growth distribution of all substitutions versus the neonatal and female substitutions considered as a single combined class was highly significant ($p = 2.2 \times 10^{-6}$, exact multinomial test); the neonatal- and female-associated amino acid substitutions showed greater functional impairment. The strong association between the two most severe disease classes and the amorphic range of our assay (<5% growth) suggests that amorphic missense variants should be under strong purifying selection in the human population. This is consistent with our observation that amorphic variants are strongly depleted in gnomAD (Figure S6A).

Consistent with their pathogenicity, substitutions in the late-onset class were also strongly depleted in the unimpaired range of our assay (Figure 6A), relative to all substitutions (2% versus 18%). However, while the neonatal and female categories showed strongest enrichment in the amorphic range, late-onset substitutions were most enriched in the hypomorphic growth range (5%–90% activity); 75% of late-onset substitutions fell in this range compared to 55% of all substitutions. The difference between the late-onset distribution and that of all substitutions is significant ($p = 3.1 \times 10^{-4}$, exact multinomial test), as is the difference between the late-onset distribution and that of the combined neonatal and female class ($p = 1.1 \times 10^{-11}$, Fisher's exact test). In fact, the late-onset substitutions are most strongly concentrated in the lower half of the hypomorphic range (5%–50% growth), as 51% fell in this range versus 29% of neonatal/female substitutions and 27% of all substitutions (Table S7). Therefore, late-onset presentation is associated with amino acid substitutions that have impaired activity in our assay but generally less impaired than those associated with neonatal and female presentation. Unlike amorphs, which appear to be under strong purifying selection (Figure S6A), variants associated with the hypomorphic range of our assay (5%–90% growth) show little sign of depletion in gnomAD (Figure S6B) and therefore little sign of strong purifying selection. This difference may be explained by the association between amorphs and severe disease presentation (both female and neonatal male), while hypomorphs are more strongly associated with a less severe disease phenotype, i.e., late-onset male-specific disease presentation (Figure 6).

Similar to the results for the ClinVar pathogenic class, several of the amino acid substitutions displaying the highest growth in each of the three disease stratification groups correspond to SNVs with a high probability of affecting splicing (Figure 6B). This is consistent with the pathogenicity of these variants resulting from splicing defects rather than the effect of the amino acid substitution on OTC activity. Several other high growing substitutions in the three disease groups occur in the SMG loop (Figure 6). This provides further evidence that the SMG loop, while important for OTC function in human mitochondria, does not appear to be needed for OTC function in the yeast cytoplasm.

Together, these results demonstrate a very close agreement between growth in our assay and disease

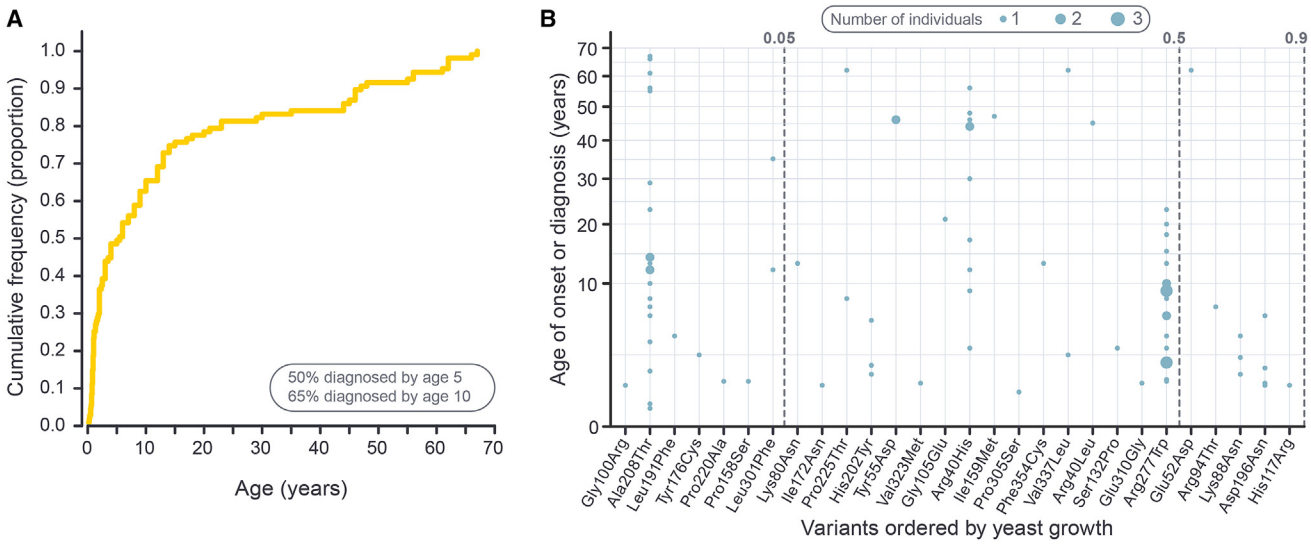


Figure 7. Age of onset in late-onset males

(A) Cumulative frequency versus age of onset/diagnosis for symptomatic, late-onset males.

(B) Age of diagnosis/onset among late-onset males with missense SNVs and yeast growth scores for the matching amino acid substitutions. Substitutions corresponding to missense SNVs predicted to impair splicing or that are present in the SMG loop were omitted from this plot.

stratification. All three classes of disease-associated amino acid substitutions are depleted in the upper range of our assay, but those associated with neonatal and female presentation show complete, or near complete, loss of OTC function, while those associated with the late-onset class demonstrate less severe loss of OTC function.

Heterogeneity in age of diagnosis in the late-onset class

The correlation between age of disease onset and activity in our assay is strong. Amino acid changes corresponding to neonatal variants (<6 weeks old onset) display significantly lower growth than those corresponding to late-onset variants (>6 weeks old onset) (Figure 6). However, within the late-onset class, the age of diagnosis varies greatly. Therefore, we extended our analysis to investigate whether there was also a correlation between age of (late) onset and growth in our assay. To this end, we further resolved the exact age of onset, or diagnosis, in the late-onset class of males by performing an exhaustive literature search for these parameters (Table S10). Age of diagnosis varied from 6 weeks of age up to the 7th decade of life, with 50% occurring by age 5 and 65% by age 10 (Figure 7). However, 25% of diagnoses occurred in individuals older than 18, further helping to explain the lack of strong selection against variants falling in the hypomorphic range of our assay (5%–90% growth, Figure S6B) that are associated with late-onset male-specific presentation (Figure 6).

After removing late-onset substitutions located in the SMG loop or corresponding to SNVs with a high probability of affecting splicing, we observed no significant relationship between growth in our assay and age of (late) onset (Kendall's tau = −0.14; p = 0.061) (Figure 7B). Therefore, while late-onset variants are associated with moderate loss of function in OTC, the actual age of onset within the late class is, at most, only weakly determined by the relative severity of the OTC amino acid substitution. In addition, our results also highlighted a number of instances where people with the same amino acid substitution (p.Ala208Thr, p.Pro225Thr, p.Arg40His, p.Val337Leu, and p.Arg277Trp) displayed a wide range of age of onset (Figure 7B). This suggests that age of onset within the late class is highly variable, even for a single *OTC* allele. Both of these results are consistent with the very strong role that environment is known to play in instances of late-onset disease presentation, with specific environmental triggers frequently identified for instances of hyperammonemia in older individuals (Table S10). The unpredictable occurrence of these triggers together with additional factors, such as the individual's diagnostic odyssey, are likely to contribute to the variability of age of disease presentation.

Reclassifying VUSs with functional assay data

Finally, having identified functional thresholds in our data that agree well with clinical annotations and disease stratification, we set out to examine the impact that inclusion of our data might have on *OTC* variant reclassification. While formal reclassification of *OTC* variants in the ClinVar database is performed through a process[50] overseen by the Urea Cycle Disorders Variant Curation Expert Panel,[51] we performed a pilot reclassification of variants by using the current ACMG guidelines.[30]

We previously identified growth <90% in our assay as identifying the level of functional impairment necessary to cause disease. However, for our pilot reclassification, we used the highly conservative 5% threshold associated with complete loss of OTC function in our assay. This very stringent cutoff captures amino acid substitutions

Table 1. Reclassification of VUSs and likely pathogenic and conflicting variants according to current ACMG/AMP guidelines

	ClinVar	ACMG alone	ACMG with yeast data
VUSs	16	32	10
Likely pathogenic	17	2	24
Conflicting	1	0	0

corresponding to 57% of ClinVar pathogenic and likely pathogenic variants, 70% of neonatal variants, and 81% of those in affected females. In total, 27% of all amino acid substitutions that we tested had growth <5%, including 320 corresponding to missense variants not yet documented in ClinVar or gnomAD.

Accepting growth <5% in our assay as strong functional evidence for impaired OTC activity, we examined the effect this evidence would have on variant classification by using the current ACMG guidelines.[30] Of the 39 amino acid substitutions in our library that correspond to variants annotated as VUSs in ClinVar (n = 39 of 58), 16 had growth values below the 5% cutoff (Figure 5A). Additionally, 17 likely pathogenic variants and one variant with conflicting interpretations corresponded to amino acid substitutions with growth values below 5%. We reanalyzed these 34 variants according to current ACMG standards and guidelines by using the Variant Curation Interface[32] (material and methods). To control for potential differences in interpretation, variant reclassification was performed with and without the inclusion of the yeast functional data as PS3 functional evidence (Tables 1 and S9) by the same disease expert. Surprisingly, in the absence of our data, reclassification with the current ACMG guidelines resulted in a substantial reduction in the number of variants with definitive clinical significance calls (32 VUS and two likely pathogenic). However, inclusion of the yeast assay data reclassified 22 of these VUSs to likely pathogenic and did not change the classification of the two likely pathogenic alleles. For the ten variants that remained VUSs, the limitation was not the strength of the functional information, but rather limitations due to other criteria (Table S9), such as the number of documented individuals harboring the same rare allele. Taken together, these results provide an example of the importance of including functional evidence in variant reclassification, particularly as the rigorous ACMG standards are applied more widely.

Discussion

Eliminating the VUS category of clinical diagnosis by 2030 has been deemed one of the ten "highest-priority elements envisioned for the cutting-edge of human genomics."[52] Because computational prediction algorithms often give inaccurate or conflicting results,[53] large-scale functional assays are the only means of variant interpretation currently poised to match the pace of variant discovery. Here, we describe the development and application at scale of such an assay, a yeast-based quantitative growth assay for measuring the activity of human OTC. Because we assay single copies of gene variants integrated into a haploid yeast strain, our functional assay not only assesses the activity of individual OTC variants but also recapitulates the genotypes of males with single amino acid changes in the enzyme encoded by this X-linked gene. In total, we measured the effect of 1,570 single amino acid substitutions on OTC activity.

The validity of the assay is supported by the close agreement with existing information about OTC, both at the level of clinically characterized variants and protein structure. Amino acid substitutions corresponding to known pathogenic variants and substitutions affecting active sites in the protein both demonstrated impaired activity in our assay. In addition, the degree of impairment shows good agreement with clinical stratification of known disease variants. Amino acid substitutions corresponding to the most severe clinical variants (neonatal presentation in males) are highly enriched in the amorphic range of the assay. Additionally, our results support the long-standing hypothesis that highly deleterious variants in females can lead to disease presentation. Variants corresponding to amorphic substitutions, and therefore neonatal/female disease presentation, are very strongly depleted in gnomAD, suggesting they are under strong purifying selection.

Amino acid substitutions corresponding to male late-onset OTC deficiency variants also display significant impairment in our assay but on average have activity levels that are significantly higher than substitutions associated with severe variants. Interestingly, the lack of significant correlation between these more moderately impaired variants and finer scaled analysis of age of onset is consistent with observations that there is a strong environmental component to late-onset presentation. In these individuals, clear triggering events are often identified. These events include dietary changes, surgeries, infections, pregnancy, or exposure to valproate (an inhibitor of N-acetylglutamate synthase, which is upstream of OTC in the urea cycle) or glucocorticoids (which inhibit protein synthesis and promote catabolism). As these events are stochastic, age of late onset for moderate variants is also largely stochastic. This pattern is consistent with moderately impaired OTC variants being sufficient for baseline urea cycle activity but not sufficient under stress conditions. Thus, variants with moderately impaired growth values (5%–90%) are also actionable, as individuals with these variants may be able to avoid potential hyperammonemia-triggering factors, and physicians can more rapidly

diagnose metabolic crises should they occur. Variants corresponding to less impaired substitutions show little sign of depletion in gnomAD, suggesting they are not under strong purifying selection. This is consistent with the association between less severe, but still pathogenic, levels of functional impairment and late-onset male-specific disease presentation that is dependent on a triggering environmental event.

Although we saw excellent agreement between clinical information and activity in our yeast assay across most of the length of the protein, we identified one relatively small region in which known pathogenic alleles exhibited relatively high growth values. This region of the protein corresponded to a 13 amino acid domain that is thought (based on the crystal structure) to undergo a conformational change upon binding ornithine. Our data suggest that this region of the protein may have a role in human cells that is not required in yeast, and therefore in this region, our assay may not be informative for human protein function. In total, 68 of our 1,570 amino acid substitutions fell in this region. We have flagged these results in the dataset indicating that they should be interpreted with extreme care.

The results of our functional assay provide additional evidence for assignment of clinical significance to *OTC* missense variants classified as VUSs according to the ACMG criteria.[54] Current classifications of *OTC* variants in ClinVar were determined by submitters who have applied their own criteria for classification. However, the effort to standardize the clinical annotation process will undoubtedly come with some challenges, particularly for rare diseases where certain types of supporting evidence are unlikely to be available. As our preliminary reclassification of a small number of variants illustrated, one consequence of applying a set of uniform and relatively stringent criteria for variant reclassification is that some variants that previously had definitive calls could be reclassified as VUSs. Our reclassification exercise illustrates the impact that the inclusion of functional assay results can have when ACMG criteria are consistently applied, changing the classification of 22 variants that would otherwise be VUSs to the clinically actionable likely pathogenic class.[52] These results emphasize how large-scale functional assays are particularly powerful in the context of rare genetic diseases.

Data and code availability

The PyPL8 code used for automated extraction of yeast growth measurements is available at https://github.com/lacyk3/PyPl8 and is free for non-commercial use. Additional scripts used for determination of the *yOTC* variant present in each transformant and for phenotype data processing and normalization are included as Methods S1 and S2, respectively. The yeast growth scores for each variant and the standard error estimates are available at mavedb.org under accession number MAVEDB: urn:mavedb:00000112-a-1 and as Table S7 of the manuscript. This dataset and the methods used to generate it are free for non-commercial use. The dataset and methods described herein may be the subject of one or more US patent or patent applications. Raw sequence reads used to identify the variant and any potential secondary mutations are available at the Sequence Read Archive (SRA) under accession numbers SRA: PRJEB61166 and SRA: PRJEB60279. This data is free for both commercial and non-commercial use.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.03.019.

Acknowledgments

We thank Richard McLaughlin Jr., Douglas Fowler, Michael Xie, Nicholas Hurlburt, and Jesse Bloom for helpful discussions. We thank Julee Ashmead for help with manuscript and figure preparation. This work was funded in part by an NIH/NIGMS award (R01 GM134274 to A.M.D.). M.T. was supported in part by an NIH/NHGRI training grant (T32 HG00035). K.O. and J.N.K. acknowledge support from the National Science Foundation AI Institute in Dynamic Systems (grant number 2112085). The authors would like to thank Joachim Strub and the varSEAK Team, as well as the groups that provided their variant data.

Declaration of interests

A.M.D. is a scientific advisor with a financial interest in Fenologica Biosciences, Inc. H.M. is a founding member of Cogthera LLC. A.M.D., R.S.L., G.A.C., A.S., and L.C. are listed as inventors on US patents or patent applications that may be relevant to this work.

Received: November 15, 2022
Accepted: March 30, 2023
Published: May 4, 2023

References

1. Summar, M.L., Koelker, S., Freedenberg, D., Le Mons, C., Haberle, J., Lee, H.S., Kirmse, B.; European Registry and Network for Intoxication Type Metabolic Diseases; and Members of the Urea Cycle Disorders Consortium UCDC (2013). The incidence of urea cycle disorders. Mol. Genet. Metab. *110*, 179–180.
2. Caldovic, L., Abdikarim, I., Narain, S., Tuchman, M., and Morizono, H. (2015). Genotype-Phenotype Correlations in Ornithine Transcarbamylase Deficiency: A Mutation Update. J Genet Genomics *42*, 181–194.
3. Ausems, M.G., Bakker, E., Berger, R., Duran, M., van Diggelen, O.P., Keulemans, J.L., de Valk, H.W., Kneppers, A.L., Dorland, L., Eskes, P.F., et al. (1997). Asymptomatic and late-onset ornithine transcarbamylase deficiency caused by a A208T mutation: clinical, biochemical and DNA analyses in a four-generation family. Am. J. Med. Genet. *68*, 236–239.
4. van Diggelen, O.P., Zaremba, J., He, W., Keulemans, J.L., Boer, A.M., Reuser, A.J., Ausems, M.G., Smeitink, J.A., Kowalczyk, J., Pronicka, E., et al. (1996). Asymptomatic and late-onset ornithine transcarbamylase (OTC) deficiency in males of a five-generation family, caused by an A208T mutation. Clin. Genet. *50*, 310–316.
5. Gascon-Bayarri, J., Campdelacreu, J., Estela, J., and Reñé, R. (2015). Severe hyperammonemia in late-onset ornithine transcarbamylase deficiency triggered by steroid administration. Case Rep. Neurol. Med. *2015*, 453752.

6. Adam, M.P. (1993). GeneReviews. In GeneReviews((R)), M.P. Adam, D.B. Everman, G.M. Mirzaa, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K.W. Gripp, and A. Amemiya, eds. (NML).
7. Batshaw, M.L., MacArthur, R.B., and Tuchman, M. (2001). Alternative pathway therapy for urea cycle disorders: twenty years later. J. Pediatr. *138*, S46–S54. discussion S54-5.
8. Gerstein, M.T., Markus, A.R., Gianattasio, K.Z., Le Mons, C., Bartos, J., Stevens, D.M., and Mew, N.A. (2020). Choosing between medical management and liver transplant in urea cycle disorders: A conceptual framework for parental treatment decision-making in rare disease. J. Inherit. Metab. Dis. *43*, 438–458.
9. Rüegger, C.M., Lindner, M., Ballhausen, D., Baumgartner, M.R., Beblo, S., Das, A., Gautschi, M., Glahn, E.M., Grünert, S.C., Hennermann, J., et al. (2014). Cross-sectional observational study of 208 patients with non-classical urea cycle disorders. J. Inherit. Metab. Dis. *37*, 21–30.
10. Golbahar, J., Altayab, D.D., and Carreon, E. (2014). Short-term stability of amino acids and acylcarnitines in the dried blood spots used to screen newborns for metabolic disorders. J. Med. Screen *21*, 5–9.
11. Vasquez-Loarte, T., Thompson, J.D., and Merritt, J.L., 2nd. (2020). Considering Proximal Urea Cycle Disorders in Expanded Newborn Screening. Int. J. Neonatal Screen. *6*, 77.
12. Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., Eng, C., Evans, J.P., Herman, G.E., Hufnagel, S.B., Klein, T.E., Korf, B.R., 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. Genet. Med. *19*, 249–255.
13. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.
14. Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. *101*, 315–325.
15. Weile, J., and Roth, F.P. (2018). Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. Hum. Genet. *137*, 665–678.
16. Kruger, W.D., and Cox, D.R. (1995). A yeast assay for functional detection of mutations in the human cystathionine beta-synthase gene. Hum. Mol. Genet. *4*, 1155–1161.
17. 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. J. Biol. Chem. *284*, 28926–28934.
18. 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.
19. Mayfield, J.A., Davies, M.W., Dimster-Denk, D., Pleskac, N., McCarthy, S., Boydston, E.A., Fink, L., Lin, X.X., Narain, A.S., Meighan, M., and Rine, J. (2012). Surrogate genetics and metabolic profiling for characterization of human disease alleles. Genetics *190*, 1309–1323.
20. Greene, A.L., Snipe, J.R., Gordenin, D.A., and Resnick, M.A. (1999). Functional analysis of human FEN1 in Saccharomyces cerevisiae and its role in genome stability. Hum. Mol. Genet. *8*, 2263–2273.
21. Kato, S., Han, S.Y., 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. Proc. Natl. Acad. Sci. USA *100*, 8424–8429.
22. Sun, S., Yang, F., Tan, G., Costanzo, M., Oughtred, R., Hirschman, J., Theesfeld, C.L., Bansal, P., Sahni, N., Yi, S., et al. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res. *26*, 670–680.
23. Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte, E.M. (2015). Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science *348*, 921–925.
24. Yang, F., Sun, S., Tan, G., Costanzo, M., Hill, D.E., Vidal, M., Andrews, B.J., Boone, C., and Roth, F.P. (2017). Identifying pathogenicity of human variants via paralog-based yeast complementation. PLoS Genet. *13*, e1006779.
25. Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53–55.
26. Rose, M., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics: A Laboratory Course Manual (Cold Spring Harbor Laboratory Press).
27. Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943–951.
28. Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man Cybern. *9*, 62–66.
29. Duda, R.O., and Hart, P.E. (1972). Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM *15*, 11–15.
30. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., 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. Genet. Med. *17*, 405–424.
31. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., 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 Res. *46*, D1062–D1067.
32. Preston, C.G., Wright, M.W., Madhavrao, R., Harrison, S.M., Goldstein, J.L., Luo, X., Wand, H., Wulf, B., Cheung, G., Mandell, M.E., et al. (2022). ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. Genome Med. *14*, 6.
33. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. *12*, 3.
34. Ben Chorin, A., Masrati, G., Kessel, A., Narunsky, A., Sprinzak, J., Lahav, S., Ashkenazy, H., and Ben-Tal, N. (2020). ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. Protein Sci. *29*, 258–267.
35. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston,

M., et al. (1996). Life with 6000 genes. Science *274*, 546–563-567.
36. Horwich, A.L., Kalousek, F., Fenton, W.A., Pollock, R.A., and Rosenberg, L.E. (1986). Targeting of pre-ornithine transcarbamylase to mitochondria: definition of critical regions and residues in the leader peptide. Cell *44*, 451–459.
37. Urrestarazu, L.A., Vissers, S., and Wiame, J.M. (1977). Change in location of ornithine carbamoyltransferase and carbamoylphosphate synthetase among yeasts in relation to the arginase/ornithine carbamoyltransferase regulatory complex and the energy status of the cells. Eur. J. Biochem. *79*, 473–481.
38. Cheng, M.Y., Pollock, R.A., Hendrick, J.P., and Horwich, A.L. (1987). Import and processing of human ornithine transcarbamoylase precursor by mitochondria from Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *84*, 4063–4067.
39. Augustin, L., Mavinakere, M., Morizono, H., and Tuchman, M. (2000). Expression of wild-type and mutant human ornithine transcarbamylase genes in Chinese hamster ovary cells and lack of dominant negative effect of R141Q and R40H mutants. Pediatr. Res. *48*, 842–846.
40. Nguyen Ba, A.N., Cvijović, I., Rojas Echenique, J.I., Lawrence, K.R., Rego-Costa, A., Liu, X., Levy, S.F., and Desai, M.M. (2019). High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. Nature *575*, 494–499.
41. Shi, D., Morizono, H., Ha, Y., Aoyagi, M., Tuchman, M., and Allewell, N.M. (1998). 1.85-A resolution crystal structure of human ornithine transcarbamoylase complexed with N-phosphonacetyl-L-ornithine. Catalytic mechanism and correlation with inherited deficiency. J. Biol. Chem. *273*, 34247–34254.
42. Shi, D., Morizono, H., Aoyagi, M., Tuchman, M., and Allewell, N.M. (2000). Crystal structure of human ornithine transcarbamylase complexed with carbamoyl phosphate and L-norvaline at 1.9 A resolution. Proteins *39*, 271–277.
43. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
44. Scharre, S., Posset, R., Garbade, S.F., Gleich, F., Seidl, M.J., Druck, A.C., Okun, J.G., Gropman, A.L., Nagamani, S.C.S., Hoffmann, G.F., et al. (2022). Predicting the disease severity in male individuals with ornithine transcarbamylase deficiency. Ann. Clin. Transl. Neurol. *9*, 1715–1726.
45. Shamsani, J., Kazakoff, S.H., Armean, I.M., McLaren, W., Parsons, M.T., Thompson, B.A., O'Mara, T.A., Hunt, S.E., Waddell, N., and Spurdle, A.B. (2019). A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. Bioinformatics *35*, 2315–2317.
46. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535–548.e24.
47. Gobin-Limballe, S., Ottolenghi, C., Reyal, F., Arnoux, J.B., Magen, M., Simon, M., Brassier, A., Jabot-Hanin, F., Lonlay, P.D., Pontoizeau, C., et al. (2021). OTC deficiency in females: Phenotype-genotype correlation based on a 130-family cohort. J. Inherit. Metab. Dis. *44*, 1235–1247.
48. Yorifuji, T., Muroi, J., Uematsu, A., Tanaka, K., Kiwaki, K., Endo, F., Matsuda, I., Nagasaka, H., and Furusho, K. (1998). X-inactivation pattern in the liver of a manifesting female with ornithine transcarbamylase (OTC) deficiency. Clin. Genet. *54*, 349–353.
49. McCullough, B.A., Yudkoff, M., Batshaw, M.L., Wilson, J.M., Raper, S.E., and Tuchman, M. (2000). Genotype spectrum of ornithine transcarbamylase deficiency: correlation with the clinical and biochemical phenotype. Am. J. Med. Genet. *93*, 313–319.
50. Rivera-Muñoz, E.A., Milko, L.V., Harrison, S.M., Azzariti, D.R., Kurtz, C.L., Lee, K., Mester, J.L., Weaver, M.A., Currey, E., Craigen, W., et al. (2018). ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. Hum. Mutat. *39*, 1614–1622.
51. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen — The Clinical Genome Resource. N. Engl. J. Med. *372*, 2235–2242.
52. Green, E.D., Gunter, C., Biesecker, L.G., Di Francesco, V., Easter, C.L., Feingold, E.A., Felsenfeld, A.L., Kaufman, D.J., Ostrander, E.A., Pavan, W.J., et al. (2020). Strategic vision for improving human health at The Forefront of Genomics. Nature *586*, 683–692.
53. Tang, H., and Thomas, P.D. (2016). Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. Genetics *203*, 635–647.
54. Tavtigian, S.V., Greenblatt, M.S., Harrison, S.M., Nussbaum, R.L., Prabhu, S.A., Boucher, K.M., Biesecker, L.G.; and ClinGen Sequence Variant Interpretation Working Group ClinGen SVI (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet. Med. *20*, 1054–1060.