



# A rapid solubility assay of protein domain misfolding for pathogenicity assessment of rare DNA sequence variants

Corey L. Anderson, PhD<sup>1,2</sup>, Tim C. Routes, BSc<sup>1,2</sup>, Lee L. Eckhardt, MD<sup>1,2</sup>, Brian P. Delisle, PhD<sup>3</sup>, Craig T. January, MD, PhD<sup>1,2</sup> and Timothy J. Kamp, MD, PhD<sup>1,2,4</sup>

**Purpose:** DNA sequencing technology has unmasked a vast number of uncharacterized single-nucleotide variants in disease-associated genes, and efficient methods are needed to determine pathogenicity and enable clinical care.

**Methods:** We report an *E. coli*-based solubility assay for assessing the effects of variants on protein domain stability for three disease-associated proteins.

**Results:** First, we examined variants in the Kv11.1 channel PAS domain (PASD) associated with inherited long QT syndrome type 2 and found that protein solubility correlated well with reported in vitro protein stabilities. A comprehensive solubility analysis of 56 Kv11.1 PASD variants revealed that disruption of membrane trafficking, the dominant loss-of-function disease mechanism, is largely determined by domain stability. We further validated this assay by using it to identify second-site suppressor PASD variants

that improve domain stability and Kv11.1 protein trafficking. Finally, we applied this assay to several cancer-linked P53 tumor suppressor DNA-binding domain and myopathy-linked Lamin A/C Ig-like domain variants, which also correlated well with reported protein stabilities and functional analyses.

**Conclusion:** This simple solubility assay can aid in determining the likelihood of pathogenicity for sequence variants due to protein misfolding in structured domains of disease-associated genes as well as provide insights into the structural basis of disease.

*Genetics in Medicine* (2020) 22:1642–1652; <https://doi.org/10.1038/s41436-020-0842-1>

**Keywords:** sequence variant; protein misfolding; Kv11.1; LMNA; P53

## INTRODUCTION

A pressing challenge in medicine is to understand the consequences of the enormous genetic variability in humans demonstrated by exome and genome sequencing projects and in clinical genetics.<sup>1</sup> The classification of low frequency DNA coding variants as disease causing is critical for the diagnosis, management, and prognostication of patients and families who harbor these variants. There is currently a dearth of functional data to support the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP)<sup>1,2</sup> classification of variants as benign, likely benign, likely pathogenic, pathogenic, or uncertain significance. With sequence variants in about 3000 genes underlying over 4000 Mendelian phenotypes as well as the growing number of cancer-linked somatic variants, the biochemical and physiological characterization for all variants is impractical.<sup>3</sup> Consequently, dozens of sequence and structure-based bioinformatics tools have been developed to predict their functional impact.<sup>1,4</sup> In many proteins, the most common disease mechanism introduced by sequence variants is domain instability, which can lead to protein misfolding and aggregation (Fig. S1).<sup>5,6</sup> However,

pathogenicity prediction tools such as FoldX,<sup>7</sup> a popular protein stability calculator, have limitations and are often unreliable<sup>8</sup> highlighting the continued need for experimental characterization.

Assessment of protein stability using bacterial expression systems has been a focus of intense effort given applications to develop commercial proteins and biopharmaceuticals. Native and overexpressed proteins in bacteria can misfold and form insoluble protein aggregates (inclusion bodies). For the *E. coli* protein HypF, variants predicted to be destabilizing have increased aggregation compared with wild type (WT).<sup>9</sup> Further, a high-throughput bacteria colony-based screen demonstrated a strong correlation between cellular aggregation and thermal stability for ten proteins from several different organisms.<sup>10</sup> These observations focused on engineering more stable proteins for production, and few studies have examined the ability of such assays to evaluate human sequence variants recombinantly expressed in *E. coli*.<sup>11,12</sup> Based on these studies and the simple methods available for screening protein solubility in bacteria, we set out to test on a larger scale and with several different proteins whether the presence of sequence variants causing protein misfolding

<sup>1</sup>Department of Medicine, University of Wisconsin–Madison, Madison, WI, USA; <sup>2</sup>Cellular and Molecular Arrhythmias Research Program, University of Wisconsin–Madison, Madison, WI, USA; <sup>3</sup>Department of Physiology, University of Kentucky–Lexington, Lexington, KY, USA; <sup>4</sup>Department of Cell and Regenerative Biology, University of Wisconsin–Madison, Madison, WI, USA. Correspondence: Corey L. Anderson ([clanders@medicine.wisc.edu](mailto:clanders@medicine.wisc.edu)) or Timothy J. Kamp ([tjk@medicine.wisc.edu](mailto:tjk@medicine.wisc.edu))

Submitted 4 March 2020; revised 11 May 2020; accepted: 12 May 2020

Published online: 1 June 2020

could be efficiently assessed in *E. coli* by simply measuring the amount of soluble protein by immunoblot without the need for additional denaturation steps or bulky tags. We focused on missense (single-nucleotide) variants since these are the most common variants and their detrimental effects are less predictable than the severe changes to protein structure and function such as insertions, frame-shifts, deletions, and premature truncations.

We chose three protein domains in human disease-associated genes with sequence variants reported to cause loss of function and have decreased thermal stabilities as benchmarks for testing the solubility assay. First, we tested variants in the Kv11.1 channel PASD responsible for a major component of cardiac action potential repolarization and associated with long QT syndrome type 2 (LQT2).<sup>13–16</sup> The most common mechanism underlying LQT2-associated loss of Kv11.1 channel function is impaired protein trafficking to the surface membrane.<sup>16</sup> Furthermore, trafficking for many of these variants can be improved with a variety of interventions including reduced culture temperature or culture with high affinity Kv11.1 channel blockers,<sup>16</sup> suggesting different degrees of protein stability. We took advantage of having trafficking data for over 60 PASD variants already reported to use for comparison with our solubility assay, toward providing a structural basis for different Kv11.1 trafficking phenotypes and helping to identify LQT2-PASD suppressor variants. Next, we tested our solubility assay on several oncogenic variants in the DNA-binding domain (DBD) of the tumor suppressor P53 protein. P53 accounts for half of all cancers with at least 1200 distinct variants identified.<sup>17–20</sup> Finally, we tested variants in the immunoglobulin D domain (IgD) of Lamin A/C, a major component of the nuclear envelope, associated with cardiomyopathy, muscular dystrophy, lipodystrophy, and premature aging.<sup>21,22</sup> Remarkably, we found solubility to be a good proxy for misfolding of structural domains in three different disease-associated proteins. This simple solubility assay should have broad utility toward rapid and efficient DNA sequence variant classification regarding pathogenicity via protein domain misfolding as well as gaining structural insights into disease and therapies.

## MATERIALS AND METHODS

### Bioinformatics

Structural models were created in Pymol using Protein Data Bank (PDB) ID 5K7L for EAG1, PDB: 1BYW for the PASD,<sup>13</sup> PDB: 1IFR for the IgD,<sup>21</sup> and PDB: 1TSR for the DBD.<sup>17</sup> Three models (PDB 1BYW, 4HP9, 4HQA)<sup>13,23</sup> were used as inputs for FoldX PASD calculations<sup>7</sup> using the YASARA molecular graphics interface. PDB structures were first refined with the FoldX “RepairPDB” function and then predictions were made using default parameters. For the IgD and DBD, PDB 1IFR and 1TSR were used as inputs for FoldX calculations using the iRDP web server, respectively. Amino acid conservation scores were calculated with ConSurf using default parameters. Scores are ranked from 1 (variable) to 9

(conserved). Relative solvent accessibilities were calculated using ASAview. Variants were also evaluated using the stability prediction web servers: I-Mutant, Eris, CUPSAT, MuPro, PolyPhen, PoPMuSiC, and MutPred. Variants were binned using the PolyPhen descriptors “probably damaging” ( $\Delta\Delta G$  values  $\geq 2.0$ ), “possibly destabilizing” ( $0 < \Delta\Delta G < 2$ ) or benign ( $\Delta\Delta G \leq 0$ ). CUPSAT and I-Mutant gave results with opposite sign, so values were reversed to simplify comparisons. MutPred percentages  $\geq 0.75$  were assigned probably destabilizing, 0.5–0.75 possibly destabilizing, and  $\leq 0.5$  benign. References for all bioinformatics tools are in the supporting supplement.

### Expression constructs

All variants were made using the QuikChange II XL kit from Agilent (Santa Clara, CA) using primers designed with the Agilent Primer Design Program. Primers were obtained from Integrated DNA Technologies (Coralville, IA). Templates for mutagenesis were pcDNA3-Kv11.1 as previously published,<sup>16</sup> pEGFP-p53 (Addgene #12091),<sup>24</sup> and pcDNA3-GFP-LaminA-N195K (Addgene #32708),<sup>25</sup> which was mutated back to WT. Restriction digest analysis was used to test the integrity of all constructs and all variants were verified by sequencing at the University of Wisconsin (UW) Biotechnology Center. For *E. coli* expression constructs, polymerase chain reaction (PCR) was used to amplify the PASD (amino acids 2–135), IgD (amino acids 435–553), and DBD (amino acids amino acids 92–292) for ligation independent cloning (LIC) into a 6× His-tagged pET3 plasmid as previously described.<sup>26</sup>

### Protein expression and solubility assay

Single colonies of BL21 (DE3) cells transformed with each 6× His-tagged mutant construct were grown overnight (13–18 hours) at 37 °C in 2 ml of autoinduction media (0.5% glycerol, 0.5% glucose, 0.2%  $\alpha$ -lactose, 25 mM  $\text{Na}_2\text{HPO}_4$ , 25 mM  $\text{KH}_2\text{PO}_4$ , 50 mM  $\text{NH}_4\text{Cl}$ , 5 mM  $\text{Na}_2\text{SO}_4$ , and 2 mM  $\text{MgSO}_4$ ). An equal number of cells (1.5 ml max) were harvested, washed (50 mM Tris, 150 mM NaCl, pH 7.5) once, resuspended in 100  $\mu\text{L}$  of wash buffer and lysed for 10 minutes in lysis buffer (1× Cell Lytic B® [Sigma]) in wash buffer and 100  $\mu\text{M}$  PMSF. Then, 5  $\mu\text{L}$  of total cell lysate was diluted in 25  $\mu\text{L}$  of wash buffer and added to an equal volume of 2× sample buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 50 mM DTT and 0.005% Bromophenol Blue) before western blot. Soluble protein was collected from the supernatant after a 15,000g spin for 10 minutes and diluted in equal amounts of 2× sample buffer for western blot or serially diluted 1:2 for dot blot (1  $\mu\text{L}$ ). All samples were boiled for 1–2 minutes before 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (50× more soluble protein was loaded by volume than total protein) and transferred to nitrocellulose paper. All blots were first blocked for 10 minutes in blocking buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% Tween-20, and 10% dry milk) and detected with anti His-HRP antibodies (Santa Cruz Biotech for PASD and IgD or Rockland Immunochemicals Inc. for DBD). Densitometry was

performed using ImageJ (NIH) to quantify immunoblots. For dot blots, one representative row from each serially diluted dot blot was quantified ( $n \geq 3$ ) (see Fig. S3 for representative examples).

### Kv11.1 trafficking

HEK 293 cells were cultured at 37 °C and transfected with the Kv11.1 variants using Lipofectamine 2000 (Invitrogen; ThermoFisher). Immature and mature Kv11.1 protein bands were detected by immunoblot analysis of whole-cell lysates. Briefly, lysates were mixed with an equal amount of Laemmli sample buffer, separated by 7% SDS-PAGE, and detected with an antibody to the distal C-terminus as previously described.<sup>16</sup>

### Kv11.1 function

Stable cell lines were generated by transfecting HEK cells with mutant Kv11.1 pcDNA3 and selecting in G418 as previously described.<sup>16</sup> Single colonies of G418 resistant cells were then tested for Kv11.1 expression by immunoblot. Cell lines that gave a robust 155-kD band on immunoblot were used for electrophysiological analysis. Kv11.1 current was measured using the whole-cell patch clamp technique as previously described.<sup>16</sup> Voltage protocols are described in Fig. 3 and data analysis was done using pCLAMP 8.0 (Axon Instruments) and Origin (6.0 Microcal).

### Lamin A/C aggregation

HEK 293 cells of similar confluence were cultured at 37 °C and transfected with the GFP-tagged constructs described above using Lipofectamine 2000 (Invitrogen; ThermoFisher). Cells were imaged the following day at 20× or 40× magnification using the EVOS FL Imaging System (ThermoFisher). The percentage of cells with aggregates was obtained by counting at least 100 cells from several different fields of view averaged over 4–6 transfections.

### Statistics

All data are presented as mean  $\pm$  SD. One-way analysis of variance (ANOVA) was used for statistical analysis followed by the Tukey post hoc test.  $P < 0.05$  was considered statistically significant.

## RESULTS

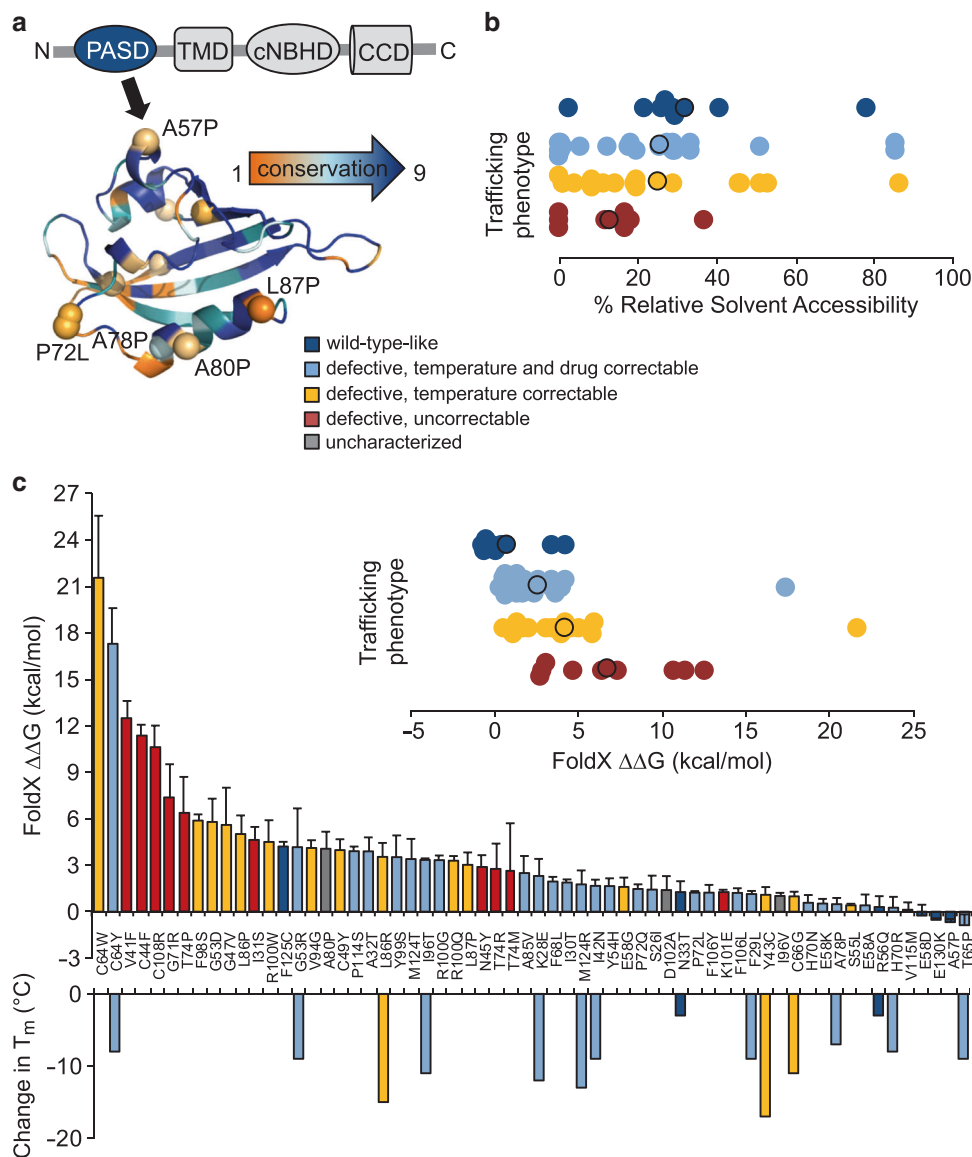
### Assessing Kv11.1 PASD misfolding

Kv11.1 is a large multidomain membrane protein containing a 110-amino acid N-terminal intracellular PASD harboring over 60 sequence variants putatively associated with LQT2 (Fig. 1a). We have previously used a membrane trafficking assay in HEK 293 cells to show that LQT2-associated variants exhibit different trafficking efficiencies that can be grouped as follows: (1) trafficking deficient and uncorrectable, (2) trafficking deficient but correctable by culturing cells at reduced temperature, (3) trafficking deficient but correctable with reduced temperature and Kv11.1 channel blockers, and (4) variants that traffic similar to WT. Because domain

misfolding leads to deficient trafficking, we first sought to see if computational models could predict the results of published trafficking assays for LQT2 variants in the PASD.<sup>16</sup> Since destabilizing variants are typically overrepresented at evolutionarily conserved residues,<sup>27</sup> in hydrophobic, buried regions in the core of the domain,<sup>5</sup> and are more chemically severe,<sup>28</sup> we used the ConSurf and ASAView bioinformatics tools to determine their conservation score (scored 1–9) and relative solvent accessibility (RSA), respectively. Using ConSurf, LQT2 residues averaged 7.1 (9 being the highest) compared with 6.1 for all residues (Fig. 1a and Table S1). Most of the variants at highly variable residues involve proline—a unique amino acid that forces structural rigidity and backbone strain on PASD folding—or result in other large physicochemical changes (e.g., S26I, I30T). Using ASAView, LQT2 residues averaged 26% solvent accessibility compared with 32% for all residues (Table S1) with the most severe variants (i.e., trafficking defective and uncorrectable) being buried (Fig. 1b).

Using the structure-based stability prediction tools FoldX, PolyPhen, MUpro, CUPSAT, Eris, and I-Mutant 2.0, we found that  $\geq 75\%$  of LQT2-PASD variants are “possibly to probably destabilizing” (Fig. S2 and Table S1) but with a great deal of scatter. Overall, FoldX is the best predictor of stability when compared with trafficking phenotype but is unreliable when individual variants are compared with reported changes in melting temperatures underscoring the need for biochemical characterization (Fig. 1c).

To assess the stability of these variants, we expressed the PASD with extra N-terminal aliphatic helix region (amino acids 2–135) and tested their solubility as a proxy. To benchmark this method, we tested 14 PASD missense variants reported in two different studies with thermal stabilities that correlate with their Kv11.1 trafficking properties (i.e., thermally unstable PASD variants are trafficking defective)<sup>14,15</sup> and three variants we previously categorized as likely benign (E58D, V115M, and F125C) (Fig. 2a).<sup>16</sup> Fig. 2b shows the immunoblot results for nine of the reported proteins where total protein levels were all similar to WT but the amount of soluble protein varied using western blot (Fig. 2b and Table S1). Overall, destabilized PASD variants with lower melting temperatures and trafficking defects are less soluble than WT in contrast to trafficking competent but dysfunctional R56Q and N33T as well as WT-like trafficking and functional E58D, V115M, and F125C (Fig. 2b–d).<sup>16</sup> Further, the level of solubility largely trended with stability (i.e., the more destabilized the variant, the less soluble it is) (Fig. 2c). To simplify the misfolding assay further, we also assessed solubility by dot blot, which had a good correlation to western blot analysis ( $R^2 = 0.82$ ) (Fig. 2b–d). We then used our solubility assay to perform a comprehensive analysis of all PASD variants previously characterized<sup>16</sup> and found that trafficking phenotype largely correlates with PASD solubility (Fig. 2e). This result demonstrates a simple and rapid way of assessing misfolding for PASD variants and could replace Kv11.1 trafficking as an assay for misfolding and pathogenicity.



**Fig. 1 Bioinformatics analysis of Kv11.1 PASD domain (PASD) variants.** (a) ConSurf calculated scores color-coded from orange (highly variable) to blue (highly conserved) for each residue. (b) ASAView calculated relative solvent accessibility percentages for variants grouped by trafficking phenotype. Circles outlined in black indicate the average. (c) FoldX calculated stabilities color-coded by trafficking phenotype for each variant. Bars represent mean  $\pm$  SD from three different structures (PDB: 1BYW, 4HQA, and 4HP9). Inset shows values grouped by trafficking phenotype. Circles outlined in black indicate the average. Reported changes in melting temperature ( $T_m$ ) are shown below for comparison.<sup>14,15</sup>

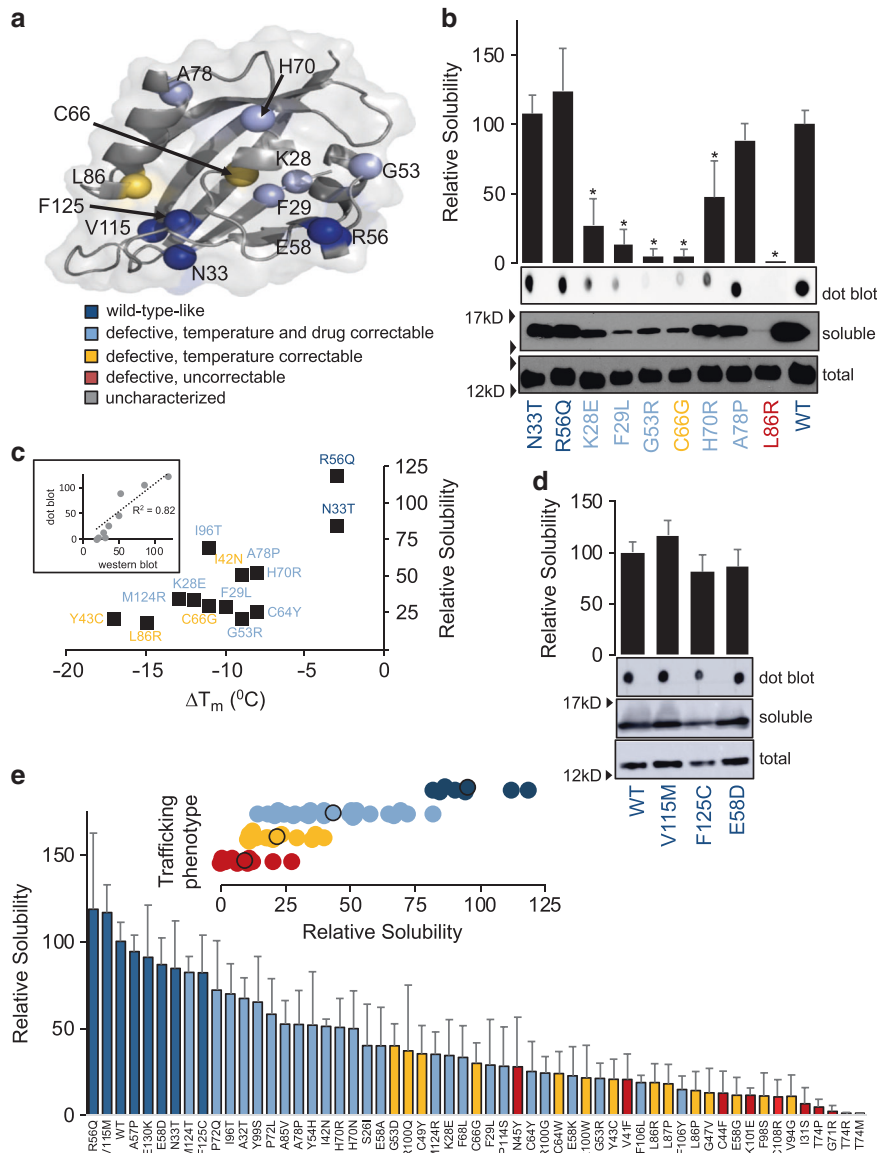
### Rational design of LQT2-PASD suppressor variants

Destabilization can result from loss of noncovalent interactions (Fig. S1). For example, the LQT2-linked variants V41F and C64Y/W introduce larger side chains at conserved and buried residues and likely cause PASD destabilization through conformational strain (Fig. 3a, b). To test this, we used FoldX and the well-established doublet on immunoblot<sup>16</sup> to assess PASD stability and Kv11.1 trafficking in HEK cells, respectively. Variants at V41 and C64 show larger  $\Delta\Delta G$  (more destabilizing) and a more severe trafficking phenotype as the hydrophobic side-chain volume increases, respectively (Fig. 3c, d and Table S2).

To further support conformational strain of V41F and C64W, we assessed the solubility of each variant and found

that as the hydrophobic side-chain volume increases, solubility tends to decrease (Fig. 3e). V41A likely decreases stability by causing a cavity in the core. However, C64A, C64M, and C64I traffic normally but show decreased solubility revealing some limitations to this assay. One possible explanation is that PASD misfolding may be compensated for through domain–domain interactions for some variants. Overall, these results show that this assay can largely be used to assess the stability of engineered variants, which in turn can help to identify second-site suppressor variants. It could also help determine misfolding for different missense variants at disease-linked residues.

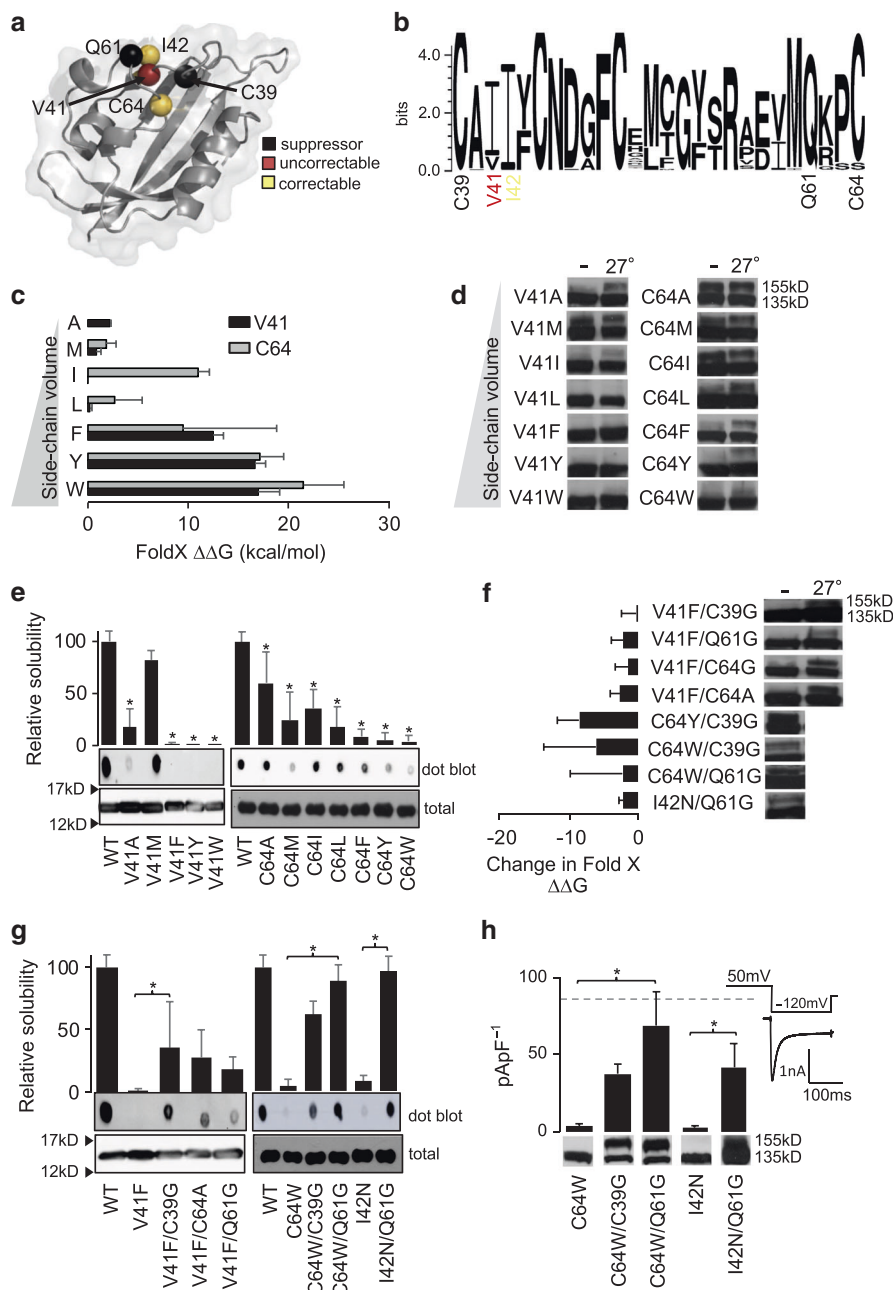
To mitigate conformational strain and improve PASD stability and consequently Kv11.1 trafficking, our strategy was



**Fig. 2 Misfolding analysis of Kv11.1 PASD variants.** (a) Structure of the Kv11.1 PASD (PDB: 1BYW) with variants color-coded based on trafficking phenotype. (b) Representative immunoblots for recombinant PASD variants with reported melting temperatures expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of wild type [WT]) determined by dot blot ( $n \geq 3$ ). (c) Comparison of solubility determined by western blot and reported melting temperatures.<sup>14,15</sup> Inset compares relative solubilities of dot blot to western blot. (d) Representative immunoblots for recombinant PASD variants reported to be benign expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of WT) determined by western blot ( $n \geq 3$ ). (e) Bars  $\pm$  SD represent relative solubility (% of WT) determined by western blot for each variant ( $n \geq 3$ ). Inset shows relative solubility values grouped by trafficking phenotype. Circles outlined in black indicate the average. \* $P < 0.05$ , a significant reduction in solubility of variants compared with WT.

to mutate nearby residues to help relieve clashes created by the larger side chains. We chose the highly conserved residues C39 and C64, which would clash with larger side chains at C64 and V41 as well as Q61, which is in a nearby flexible loop (Fig. 3a, b). Supporting this rationale, FoldX predicted that mutating these second-site variants to a smaller more flexible glycine or alanine should improve stability (Fig. 3f, Table S2). Indeed, Fig. 3g shows that these second-site variants improve the solubility for V41F and C64Y/W, respectively. Furthermore, immunoblot analysis of transiently transfected HEK cells showed that C39G, C64A, and C64G corrected the trafficking of V41F at 27 °C while C64W and nearby I42N

were correctable at 37 °C consistent with the solubility assay (Fig. 3f). To directly measure surface expression, current densities of stably transfected HEK cells showed that LQT2-C64W ( $4 \pm 2$  pA/pF) can be significantly improved at 37 °C with C39G ( $38 \pm 6$  pA/pF) and to levels not statistically different than WT ( $98 \pm 19$  pA/pF) with Q61G ( $69 \pm 22$  pA/pF). Likewise, LQT2-I42N ( $3 \pm 1$  pA/pF) can also be significantly improved with Q61G ( $42 \pm 15$  pA/pF) (Fig. 3h). These results show that our solubility assay can be used to design second-site suppressor variants to help understand the structural basis of disease as illustrated in (Fig. S4). In addition, many genes contain common single-nucleotide



**Fig. 3 Rational design of PASD domain (PASD) suppressor variants.** (a) Structure of the Kv11.1 PASD (PDB: 1BYW) showing the location and (b) conservation of each variant characterized. (c) FoldX calculated stabilities for increasingly larger hydrophobic substitutions and V41 and C64 variants. Bars  $\pm$  SD represent  $\Delta\Delta G$  predictions from three different structures (PDB: 1BYW, 4HQA, 4HP9).<sup>23</sup> (d) Representative western blot of each full-length Kv11.1 variant expressed in HEK cells under normal conditions (-) or at reduced temperature (27 °C). A lack of a 155-kD bands indicates defective trafficking. (e) Representative immunoblots of recombinant PASD C64 and V41 hydrophobic substitutions expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of wild type [WT]) determined by dot blot ( $n \geq 3$ ). (f) Changes in FoldX stabilities are shown for second-site variants predicted to improve variant stability (smaller  $\Delta\Delta G$ ). Representative western blots of each full-length Kv11.1 variant expressed in HEK cells are also shown under normal conditions (-) or at reduced temperature (27 °C). (g) Representative immunoblots of recombinant PASD suppressor variants expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of WT) determined by dot blot ( $n \geq 3$ ). (h) Current densities and western blots for full-length Kv11.1 variants expressed in HEK cells. Inset shows voltage-clamp protocol with representative current trace. Bars  $\pm$  SD represent current density levels ( $n \geq 4$  cells). Dashed line indicates WT current density level previously reported,<sup>16</sup> which was performed at the same time and with the same intracellular and extracellular solutions as these variants. \* $P < 0.05$ , a significant reduction in solubility of variants compared with WT (e), increase in solubility of the double variant compared with the single variant (g) or increase in current density of the double variant compared with the single variant (h).

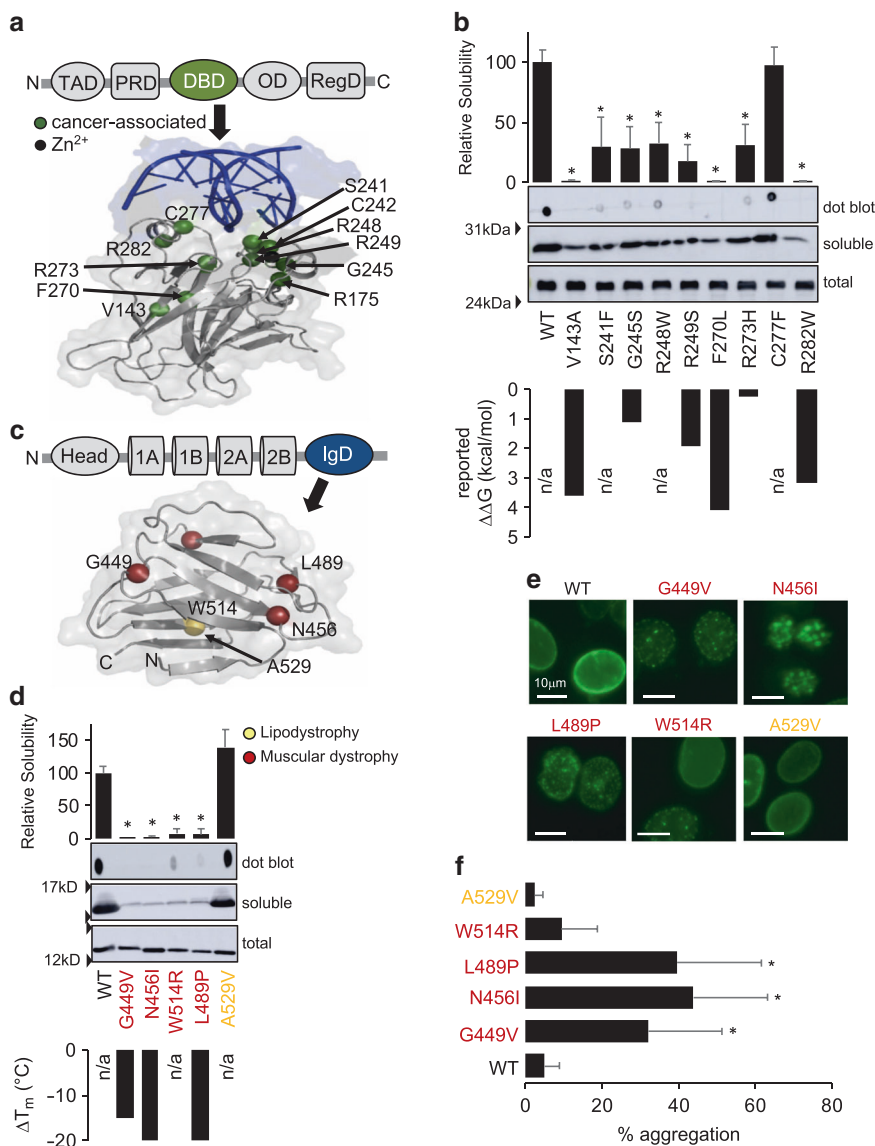
polymorphisms (SNPs) in the same protein domain as rare sequence variants, and so this assay can study the effects of rare sequence variants with different genetic backgrounds that could impact protein folding and pathogenicity.

### Assessing P53 DBD misfolding

P53 contains a DBD, which is a hotspot for missense variants including the six highest frequency cancer-associated P53 variants (Fig. 4a). Using the same ASAView, ConSurf, and FoldX analyses as above, we focused on 12 missense variants (V143A, R175H, S241F, C242S, G245S, R248Q, R248W,

R249S, F270L, R273H, C277F, R282W); 9 with reported  $\Delta\Delta G$  values (Table S3). We found that several variants are not buried (S241F, R248Q/W, R273H, C277F) or highly conserved (V143A and F270L) and that the correlation between measured and FoldX  $\Delta\Delta G$  was poor ( $R^2 = 0.12$ ) further underscoring the need for biochemical characterization to determine deleteriousness (Fig. S5 and Table S3).

We used our solubility assay to assess misfolding for 10 of the 12 missense variants analyzed above and found that while total protein expression levels are all similar to WT, the amount of soluble protein varied. Destabilized P53 variants



**Fig. 4 Properties of P53 DBD and Lamin A/C IgD variants.** (a) Structure of the P53 DBD (PDB: 1TSR) with cancer-associated variants in green. (b) Representative immunoblots of recombinant DBD variants expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of wild type [WT]) determined by dot blot ( $n \geq 3$ ). Reported  $\Delta\Delta G$  values are shown below for comparison.<sup>18–20</sup> (c) Structure of the Lamin A/C IgD (PDB: 1IFR) with variants color-coded based on disease. (d) Representative immunoblots of recombinant IgD variants expressed in *E. coli*. Bars  $\pm$  SD represent relative solubility (% of WT) determined by dot blot. (e) 40 $\times$  images of HEK nuclei after GFP-tagged full-length Lamin A variant overexpression. (f) Bars  $\pm$  SD represent the percentage of cells showing aggregation for each variant ( $n \geq 4$  transfections). Disease classifications from Universal Mutation Database (UMD). \* $P < 0.05$ , a significant reduction in solubility of variants compared with WT (b,d) or significant increase in aggregation (f). 1A–2B coiled-coil “rod” domains, DBD DNA-binding domain, IgD Immunoglobulin-like domain, n/a not available, OD oligomerization domain, PRD proline rich domain, RegD C-term regulatory domain, TAD transactivation domain.

with larger  $\Delta\Delta G$  values were less soluble (Fig. 4b and Table S3). All variants have reduced solubility except for C277F, which is not predicted to be destabilizing (i.e., not buried and FoldX  $\Delta\Delta G = -0.18$ ).

### Assessing Lamin A/C IgD misfolding

Lamin A/C contains an IgD, which is a hotspot for disease-associated missense variants (Fig. 4c). Using the same ASAView, ConSurf, and FoldX analyses as above, we focused on four well-characterized variants (G449V, L489P, N456I, W514R)<sup>22</sup> with decreased thermal stabilities reported and A529V; a homozygous variant not predicted to be destabilizing (i.e., not buried, highly variable residue, and FoldX  $\Delta\Delta G = -1.01$ ). All four variants are buried and predicted to be destabilizing by FoldX but only two are highly conserved (G449V and N456I) (Table S4). In contrast to the PASD and DBD, deleteriousness of these IgD variants was better predicted using these tools.

We used our solubility assay to assess misfolding of the IgD missense variants analyzed above and found that while total protein expression levels are all similar to WT, destabilized variants with reduced melting temperatures were less soluble than WT in contrast to A529V, which was similar to WT (Fig. 4d and Table S4).

Since some Lamin A variants aggregate upon ectopic expression,<sup>29</sup> we transiently transfected HEK cells with each GFP-tagged Lamin A constructs and compared mutant aggregation to WT. Representative images are shown in Fig. 4e. Except for A529V ( $2 \pm 2\%$ ), all variants showed an increased percentage of cells with nuclear aggregation to varying degrees (W514R [ $9 \pm 10\%$ ], G449V [ $32 \pm 17\%$ ], N456I [ $44 \pm 16\%$ ], and L489P [ $40 \pm 22\%$ ] compared with WT [ $5 \pm 2\%$ ]) (Fig. 4f and Table S4). These results again validate our solubility assay and suggest that IgD misfolding can cause Lamin A/C aggregation in the nucleus.

## DISCUSSION

In this study, we developed a simple *E. coli*-based solubility assay to assess the damaging effects of sequence variants in disease-associated genes on protein domain stability. By analyzing over 50 LQT2-associated Kv11.1 PASD variants with reported thermal stabilities<sup>14,15</sup> and functional consequences,<sup>16</sup> we showed that this assay largely predicts whether a variant will be destabilizing and thus cause deficient Kv11.1 trafficking. We also extend these findings to several disease-associated LMNA IgD and P53 DBD variants where, again, the solubility assay largely correlates with reported thermal stability studies (i.e., destabilizing variants are less soluble). Combined, these results suggest that our assay may have widespread value as a new, simpler approach to assist with the complex interpretive process of deciding the clinical relevance of rare sequence variants.<sup>29–34</sup>

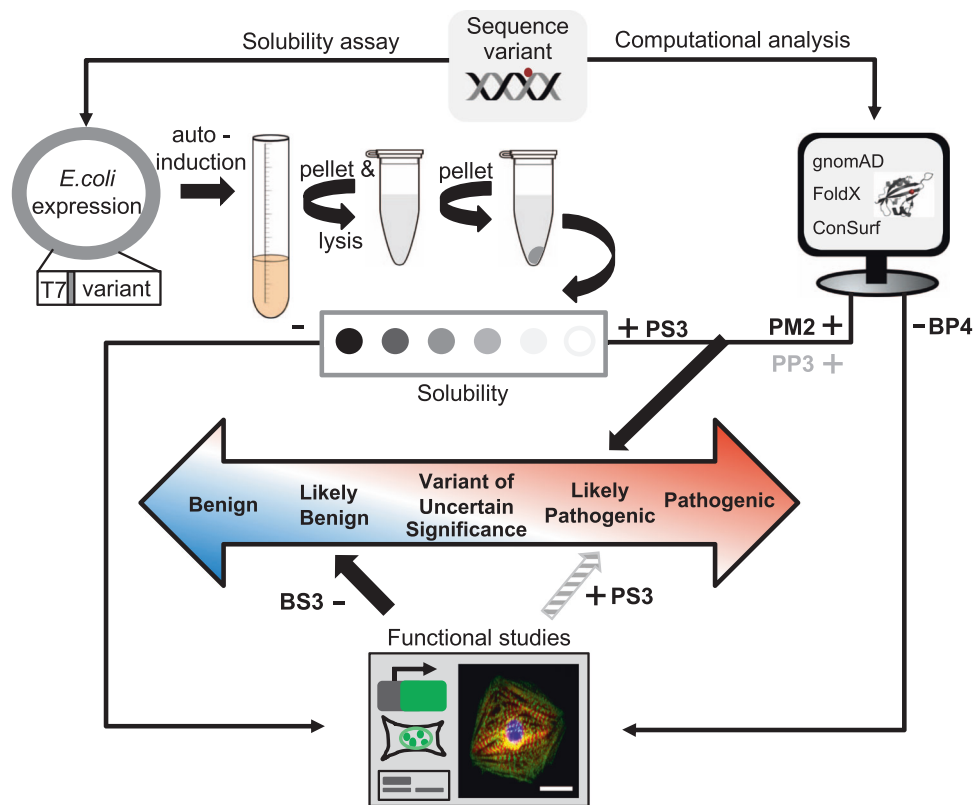
The broader application of this solubility assay across a range of Mendelian disorders will require further validation studies focused on missense variants in highly structured domains that have been studied with other established

functional tests. Assuming the assay is validated for a particular domain, a strategy for incorporating this solubility assay into a workflow for determining the pathogenicity of sequence variants in disease-associated genes is proposed based on ACMG standards of evidence for pathogenicity (Fig. 5). Because of the simplicity and efficiency of the solubility assay, it can rapidly contribute key evidence for the classification of many missense variants that otherwise would typically require more cumbersome functional assessment. In addition, the solubility assay could provide more precision to certain pathogenicity criteria such as when a novel missense change occurs at an amino acid residue where a different missense change was previously documented to be pathogenic (PM5). Although we anticipate that many missense variants in a wide range of disease-associated genes will be amenable to this scheme, there will be a significant fraction of variants outside of highly structured protein domains or in domains in which the solubility assay is not possible.

Our solubility assay might also be useful for applications other than misfolding/pathogenicity assessment for individual domain variants by combining a second variant. Such second-site variants could be used to generate suppressor variants to understand the structural basis of disease,<sup>35</sup> to assess the stability of engineered proteins important in biotechnology,<sup>10</sup> and to study the important effect of background variants that may act as “genetic modifiers.” For example, disease-associated variants can occur together with more common sequence variants in the same domain, and the second-site variant can act as a suppressor by inhibiting misfolding and thus blocking the disease manifestation. The solubility assay could efficiently screen multiple combinations of a disease-associated variant with a range of relevant second-site variants for suppressor function for improved prediction of pathogenicity in a given genetic background. Additionally, correcting misfolding and aggregation is a promising therapeutic goal by directly targeting domains with small stabilizing molecules or indirectly by modulating the cell’s proteostasis network.<sup>36,37</sup> By simply assessing solubility, our assay should help determine which variants are potential targets for these types of correction strategies.

In addition to describing a new method, we note several interesting biochemical observations. Kv11.1 misfolding, like many protein conformational diseases, leads to ER-associated degradation (Fig. S4).<sup>37</sup> All trafficking defective PASD variants studied here have decreased solubility compared with WT making this assay a good predictor of defective Kv11.1 trafficking and could potentially be applied to other trafficking-related diseases. It also supports our previous model of Kv11.1 misfolding that proposed the level of defective trafficking correlates with domain stability,<sup>16</sup> which we show here can be improved with stabilizing second-site variants (Fig. S4). Our findings also support domain misfolding and aggregation as the likely mechanism underlying many Lamin A IgD and P53 DBD variants. Interestingly, destabilized variants within the same IgD can have variable effects on Lamin A aggregation in HEK cells with





**Fig. 5 Proposed strategy for characterizing pathogenicity of variants.** Missense variants of interest are initially evaluated using computational analysis and the solubility/misfolding assay followed by further functional analysis as needed. Computational analysis is performed to determine the minor allele frequency of the variant in the general population (e.g., gnomAD), to predict if the variant is destabilizing (e.g., FoldX), and to determine if the variant is evolutionarily conserved (e.g., ConSurf). The rapid solubility assay is used for functional assessment. In the case of a sequence variant that is absent or extremely rare in the population (PM2) and is also insoluble (PS3), then per American College of Medical Genetics and Genomics (ACMG) guidelines<sup>1</sup> this variant would be categorized as likely pathogenic and hence clinically actionable. Computational models could provide further evidence of pathogenicity (PP3). In cases where variants have wild type (WT)-like solubilities, further functional studies (e.g., heterologous expression, patient-specific induced pluripotent stem cells [iPSCs]) are needed to exclude other mechanisms of pathogenesis than domain misfolding. If a variant is soluble, functional studies are negative (BS3) and computational models are negative (BP4), then it will be classified as likely benign.

W514R being more aggregation prone than the other destabilizing variants. Finally, we provide new stability insights into three previously uncharacterized P53 variants. We observed that C277F is soluble (stable) in contrast to S241F, R248W, and most other DBD variants characterized.<sup>18–20</sup> Thus, C277F likely would not benefit from targeted therapeutics designed to stabilize the DBD.

### Limitations

There are several limitations to this *E. coli*-based assay. Protein expression and solubility in bacteria can vary between different proteins and optimization may be needed (e.g., domain length, tag type and placement, growth and lysis conditions). Since protein expression levels can change between variants potentially confounding reduced solubility results, a second validation experiment should be performed to either test total expression or compare soluble and insoluble fractions<sup>12</sup> to rule out false positives. Further, this assay might lead to false negatives where the protein is soluble but loss of function is through some other mechanism (e.g., altered gating<sup>16</sup> or higher turnover<sup>38</sup>); however, when the

assay is incorporated as part of a multistage variant classification strategy (Fig. 5), such variants can be identified by other functional assays. Rare false positives were seen as with a few C64 variants that had decreased solubility, but normal Kv11.1 trafficking. This points to an important limitation of this assay in studying multidomain proteins in which domain–domain interactions can compensate for local domain misfolding or blunt quality control mechanisms.<sup>36,37</sup> Finally, this assay addresses variants only in small protein domains that are easily expressed in *E. coli* and not variants in other regions of disease-associated proteins, and this constraint limits the number of clinically relevant variants that can be studied. However, protein misfolding and aggregation are major disease mechanisms,<sup>5,6</sup> and so this assay should still be useful for characterizing a great number of targets and variants.

### Conclusion

In summary, we demonstrate a simple solubility assay for quickly assessing protein misfolding, which should be applicable to most soluble protein domains that can be

expressed in bacteria. Small culture volume and benchtop centrifugation also make this solubility assay amenable to higher throughput multiwell formats. Further, high-throughput mutagenesis methods<sup>32</sup> and *E. coli* protein aggregation protocols<sup>39</sup> are available that could be adapted to study domain sequence variants. Finally, this method in conjunction with *in silico* analysis will aid in determining whether putative disease-associated sequence variants are actionable,<sup>40</sup> possibly with strategies to correct misfolding or aggregation.<sup>36</sup>

## SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-020-0842-1>) contains supplementary material, which is available to authorized users.

## ACKNOWLEDGEMENTS

We sincerely thank Caleb Hintz, Ryan Childs, Catherine Kuzmicki, and Madilyn Anderson for technical assistance and Robert Stroud (University of California–San Francisco) for providing the modified pET-based LIC plasmid. This study was supported by National Institutes of Health (NIH) R01 HL07887 (T.J.K.), NIH R01 HL060723 (C.T.J.), American Heart Association (AHA) Competitive Catalyst Renewal Grant 17CCRG33700289 (B.P.D.), AHA Midwest Affiliate Predoctoral Fellowship (C.L.A.), Ruth L Kirschstein F32 HL128091 National Research Service Award (NRSA) postdoctoral fellowship (C.L.A.), and NIH R01 HL139738–01 (L.L.E.).

## DISCLOSURE

T.J.K. serves as a consultant for Fujifilm Cellular Dynamics, which is a stem cell technology company. The other authors declare no conflicts of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–423.
- MacArthur D, Manolio T, Dimmock D, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508:469–476.
- Chong JX, Buckingham KJ, Jhangiani SN, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genetics*. 2015;97:199–215.
- Peterson TA, Doughty E, Kann MG. Towards precision medicine: advances in computational approaches for the analysis of human variants. *J Mol Biol*. 2013;425:4047–4063.
- Yue P, Li Z, Moulton J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*. 2005;353:459–473.
- Baets G, Doorn L, Rousseau F, Schymkowitz J. Increased aggregation is more frequently associated to human disease-associated mutations than to neutral polymorphisms. *PLoS Comput Biol*. 2015;11:e1004374.
- Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33 (Suppl 2):W382–W388.
- Khan S, Vihinen M. Performance of protein stability predictors. *Hum Mutat*. 2010;31:675–684.
- Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F. Investigating the effects of mutations on protein aggregation in the cell. *J Biol Chem*. 2005;280:10607–10613.
- Asial I, Cheng Y, Engman H, et al. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat Commun*. 2013;4:2901.
- Mayer S, Rüdiger S, Ang H, Joerger AC, Fersht AR. Correlation of levels of folded recombinant p53 in *Escherichia coli* with thermodynamic stability *in vitro*. *J Mol Biol*. 2007;372:268–276.
- Lee Y, Stiers KM, Kain BN, Beamer LJ. Compromised catalysis and potential folding defects in *in vitro* studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency. *J Biol Chem*. 2014;289:32010–32019.
- Cabral JH, Lee A, Cohen SL, Chait BT, Li M, Mackinnon R. Crystal structure and functional analysis of the HERG potassium channel N terminus. *Cell*. 1998;95:649–655.
- Harley CA, Jesus CS, Carvalho R, Brito RM, Morais-Cabral JH. Changes in channel trafficking and protein stability caused by LQT2 mutations in the PAS domain of the HERG channel. *PLoS One*. 2012;7:e32654.
- Ke Y, Ng CA, Hunter MJ, Mann SA, Heide J, Hill AP, et al. Trafficking defects in PAS domain mutantK<sub>v</sub>11.1 channels: roles of reduced domain stability and altered domain–domain interactions. *Biochem J*. 2013;454:69–77.
- Anderson CL, Kuzmicki CE, Childs RR, Hintz CJ, Delisle BP, January CT. Large-scale mutational analysis of Kv11.1 reveals molecular insights into type 2 long QT syndrome. *Nat Commun*. 2014;5:5535.
- Cho Y, Gorina S, Jeffrey P, Pavletich N. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*. 1994;265:346–355.
- Bullock AN, Henckel J, Fersht AR. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene*. 2000;19:1245–1256.
- Ang H, Joerger AC, Mayer S, Fersht AR. Effects of common cancer mutations on stability and DNA binding of full-length p53 compared with isolated core domains. *J Biol Chem*. 2006;281:21934–21941.
- Joerger A, Ang H, Fersht A. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc National Acad Sci U S A*. 2006;103:15056–15061.
- Dhe-Paganon S, Werner ED, Chi Y-I, Shoelson SE. Structure of the globular tail of nuclear lamin. *J Biol Chem*. 2002;277:17381–17384.
- Dialynas G, Shrestha OK, Ponce JM, et al. Myopathic lamin mutations cause reductive stress and activate the Nrf2/Keap-1 pathway. *PLoS Genet*. 2015;11:e1005231.
- Adaixo R, Harley CA, Castro-Rodrigues AF, Morais-Cabral JH. Structural properties of PAS domains from the KCNH potassium channels. *PLoS One*. 2013;8:e59265.
- Boyd SD, Tsai KY, Jacks T. An intact HDM2 RING-finger domain is required for nuclear exclusion of p53. *Nat Cell Biol*. 2000;2:563–568.
- Gilchrist S, Gilbert N, Perry P, Östlund C, Worman HJ, Bickmore WA. Altered protein dynamics of disease-associated lamin A mutants. *BMC Cell Biol*. 2004;5:46.
- Savage DF, Anderson CL, Robles-Colmenares Y, Newby ZE, Stroud RM. Cell-free complements *in vivo* expression of the *E. coli* membrane proteome. *Protein Sci*. 2007;16:966–976.
- Vitkup D, Sander C, Church GM. The amino-acid mutational spectrum of human genetic disease. *Genome Biol*. 2003;4:R72.
- Jackson HA, Accili EA. Evolutionary analyses of KCNQ1 and HERG voltage-gated potassium channel sequences reveal location-specific susceptibility and augmented chemical severities of arrhythmogenic mutations. *BMC Evol Biol*. 2008;8:188.
- Hübner S, Eam JE, Wagstaff KM, Jans DA. Quantitative analysis of localization and nuclear aggregate formation induced by GFP-lamin A mutant proteins in living HeLa cells. *J Cell Biochem*. 2006;98:810–826.
- Pricolo MR, Herrero-Galán E, Mazzaccara C, et al. Protein thermodynamic destabilization in the assessment of pathogenicity of a variant of uncertain significance in cardiac myosin binding protein C. *J Cardiovasc Transl Res*. 2020 Feb 7; <https://doi.org/10.1007/s12265-020-09959-6> [Epub ahead of print].
- Woods NT, Baskin R, Golubeva V, et al. Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *Npj Genom Medicine*. 2016;1:16001.
- Glazer AM, Kroncke BM, Matreyek KA, Yang T, Wada Y, Shields T, et al. Deep mutational scan of an SCN5A voltage sensor. *Circ Genom and Precis Med*. 2020;13:e002786.

33. Wiltshire T, Ducey M, Foo TK, et al. Functional characterization of 84 PALB2 variants of uncertain significance. *Genet Med*. 2020;22:622–632.
34. Ma N, Zhang J, Itzhaki I, et al. Determining the pathogenicity of a genomic variant of uncertain significance using CRISPR/Cas9 and human-induced pluripotent stem cells. *Circulation*. 2018;138:2666–2681.
35. Otsuka K, Kato S, Kakudo Y, Mashiko S, Shibata H, Ishioka C. The screening of the second-site suppressor mutations of the common p53 mutants. *Int J Cancer*. 2007;121:559–566.
36. Tao Y-X, Conn MP. Pharmacoperones as novel therapeutics for diverse protein conformational diseases. *Physiol Rev*. 2018;98:697–725.
37. Foo B, Williamson B, Young JC, Lukacs G, Shrier A. hERG quality control and the long QT syndrome. *J Physiol*. 2016;594:2469–2481.
38. Kanner SA, Jain A, Colecraft HM. Development of a high-throughput flow cytometry assay to monitor defective trafficking and rescue of long QT2 mutant hERG channels. *Front Physiol*. 2018;9:397
39. Cornvik T, Dahlroth SL, Magnusdottir A, Herman MD, Knaust R, Ekberg M, et al. Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*. *Nature Methods*. 2005;2:507–509.
40. Hunter J, Irving SA, Biesecker LG, et al. A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet Med*. 2016;18:1258–1268.