Deep Mutational Scanning in Disease-related Genes with Saturation Mutagenesis-Reinforced Functional Assays (SMuRF)

Kaiyue Ma[1,*], Shushu Huang[1,7], Kenneth K. Ng[1,7], Nicole J. Lake[1], Soumya Joseph[2], Jenny Xu[3], Angela Lek[1,4], Lin Ge[1,5], Keryn G. Woodman[1], Katherine E. Koczwara[1], Justin Cohen[1], Vincent Ho[1], Christine L. O'Connor[1], Melinda A. Brindley[6,8], Kevin P. Campbell[2,8] & Monkol Lek[1,8,9,*].

[1]Department of Genetics, Yale School of Medicine, New Haven, CT, USA
[2]Howard Hughes Medical Institute, Senator Paul D. Wellstone Muscular Dystrophy Specialized Research Center, Department of Molecular Physiology and Biophysics and Department of Neurology, Roy J. and Lucille A. Carver College of Medicine, The University of Iowa, Iowa City, IA, USA
[3]Yale University, New Haven, CT, USA
[4]Muscular Dystrophy Association, Chicago, IL, USA
[5]Department of Neurology, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China
[6]Department of Infectious Diseases, Department of Population Health, University of Georgia, Athens, GA, USA.
[7]Equal second authors
[8]Senior Authors
[9]Lead Contact
[*]Correspondence: kaiyue.ma@yale.edu (K.M.), monkol.lek@yale.edu (M.L.)

Summary
Interpretation of disease-causing genetic variants remains a challenge in human genetics. Current costs and complexity of deep mutational scanning methods hamper crowd-sourcing approaches toward genome-wide resolution of variants in disease-related genes. Our framework, Saturation Mutagenesis-Reinforced Functional assays (SMuRF), addresses these issues by offering simple and cost-effective saturation mutagenesis, as well as streamlining functional assays to enhance the interpretation of unresolved variants. Applying SMuRF to neuromuscular disease genes *FKRP* and *LARGE1*, we generated functional scores for all possible coding single nucleotide variants, which aid in resolving clinically reported variants of uncertain significance. SMuRF also demonstrates utility in predicting disease severity, resolving critical structural regions, and providing training datasets for the development of computational predictors. Our approach opens new directions for enabling variant-to-function insights for disease genes in a manner that is broadly useful for crowd-sourcing implementation across standard research laboratories.

Keywords
Deep mutational scanning, saturation mutagenesis, dystroglycanopathies, genetic diseases, cost-effective variant interpretation, high-throughput functional assays, diagnostics, variant effect prediction, muscular dystrophies, variants of uncertain significance

Introduction

Recent advancement in Next-generation sequencing (NGS) and population studies[1–4] together with establishment of large biobanks have improved disease-associated variant detection and interpretation[5–7]. However, for many genetic disease patients, it remains a challenge to pinpoint their specific disease-causing variant(s), hampering access to approved treatments, participation in clinical trials, options for disease management and family planning[8,9]. The emergence of gene therapies has further heightened the importance of finding diagnoses for genetic disease patients as an initial and crucial prerequisite towards clinical trial readiness. The variants discovered in patients that are difficult to interpret are classified as variants of uncertain significance (VUS), which is often a result of insufficient evidence to determine pathogenicity[10]. The rate of VUS observed in disease-associated genes is typically higher in individuals of non-European ancestry due to limited diversity in biomedical databases[11].

To provide an additional line of evidence for the interpretation of variants, deep mutational scanning (DMS) was proposed to unbiasedly generate functional scores for all possible variants[12]. DMS employs a pooled model cell population consisting of large amounts of variants, which are subsequently characterized with appropriate high-throughput functional assays[13]. The collective functional read-out from each variant in the gene of interest gives rise to the generation of an "Atlas of Variant Effects"[14]. However, despite its promise, the high cost and technical complexity of performing DMS discourage many researchers, especially those in modestly funded rare disease labs.

In saturation mutagenesis, well-established methods include the insertion of variant-carrying tiles[15], saturation genome editing (multiplex homology-directed repair)[16], reversibly-terminated inosine mutagenesis[17], and recently emerged saturation prime editing[18]. However, these methods often come with limitations like intensive labor requirements, high expenses, disparate variant representation, and limited spanning regions. Hence, there is a need for a simple and cost-effective saturation mutagenesis method that provides comprehensive and unbiased coverage. Programmed Allelic Series (PALS) was previously developed as a saturation mutagenesis method with relatively simple steps and unbiased coverage[19]. PALS had the potential to reduce costs by using minimized-length oligos instead of long DNA fragments to introduce variants. However, its reliance on special reagents has limited its widespread adoption in DMS studies. A previous optimization of PALS removed the need for special reagents but transformed the method from requiring only a single-tube reaction to necessitating independent reactions for each targeted codon, thus significantly increasing the labor involved[20].

Besides saturation mutagenesis, there are challenges in the development of robust and accurate functional assays. Typical functional assays used in DMS to enrich variants include growth assays and flow cytometry assays[21]. Growth assays are primarily suitable for genes significantly affecting cell viability or growth rates and may not apply to other genes. Flow cytometry assays convert gene function-related biological signals into quantifiable fluorescence. However, they are relatively costly due to cell sorter reliance, particularly when saturation mutagenesis efficiency is low thereby requiring an increased number of cells to be sorted. Second, precisely controlling protein expression levels to mimic physiological relevance poses another challenge, as overexpression of the targeted gene potentially mitigates variant pathogenic effects in certain

cases[22], thereby diminishing assay sensitivity for variant interpretation. Lastly, barcoding is a common practice in DMS studies to facilitate variant detection using short-read NGS. However, this process typically necessitates an additional round of NGS to assign the barcodes to the variants, incurring extra expenses[23].

Addressing these challenges is crucial for applying DMS to disease-related genes. Dystroglycanopathies are a set of rare autosomal recessive diseases with clinical heterogeneity ranging from brain malformation in Walker-Warburg syndrome (WWS) to milder muscular symptoms in Limb-Girdle Muscular Dystrophies (LGMDs)[24]. Pathogenic variants in *DAG1*, the gene that encodes alpha-dystroglycan (α-DG), and genes encoding enzymes involved in α-DG glycosylation disrupt the binding between α-DG and extracellular matrix ligands, which compromises the muscle cell integrity and leads to dystroglycanopathies[25]. The most severe dystroglycanopathy cases can lead to miscarriage and neonatal deaths, highlighting the critical need for a better understanding of the clinical significance of variants in genetic testing[26–28].

Saturation Mutagenesis-Reinforced Functional assays (SMuRF) was developed as a flexible DMS framework to address challenges that were identified in both saturation mutagenesis and functional assays. We built upon the core concept of PALS but achieved saturation mutagenesis without the need for specialized reagents, equipment, or labor-intensive steps, resulting in a method named Programmed Allelic Series with Common procedures (PALS-C). Since our method only requires minimal-length oligos, a regular plasmid template, restriction enzymes, PCR polymerase, and a Gibson assembly kit, PALS-C is ready to apply in most molecular biology laboratories. We demonstrated the versatility of SMuRF by applying it to two key enzyme genes, *FKRP* and *LARGE1*, involved in the glycosylation of α-DG. For this implementation of SMuRF, we employed lentiviral delivery to introduce variants into cells for the flow cytometric functional assay, achieving high variant coverage. In addition, we utilized the weak UbC promoter[29], and streamlined the workflow to control gene expression levels, ensuring precise variant assessment. Lastly, we implemented a block-by-block strategy throughout the saturation mutagenesis and functional assay, eliminating the need for barcoding and significantly reducing costs.

Results

Establishing the functional assay in SMuRF

SMuRF was developed with the aim of creating a DMS framework that is adaptable for various genes by employing appropriate assays. Initially, the functional assay was established to characterize all possible coding single nucleotide variants (SNVs) in genes related to dystroglycanopathies. Most variants of α-DG glycosylation enzymes, including FKRP and LARGE1, lack clinical reports and those reported remain poorly interpreted (Figure S1A), with many unique to the families[30]. FKRP adds the second ribitol-5-phosphate (Rbo5P) to the Rbo5P tandem[31] while LARGE1 is responsible for adding the repeated disaccharide units of matriglycan[32] (Figure S1B). New drugs[33], gene therapies[34,35] and cell therapies[36] are being actively developed for dystroglycanopathies related to FKRP or LARGE1, emphasizing the need for improved variant interpretation to facilitate patient enrollment in these gene-specific trials.

To develop a functional assay suitable for DMS, we initially identified α-DG hypoglycosylation as the molecular phenotype associated with dystroglycanopathies and the focal point for assessment. The IIH6C4 antibody, extensively used in α-DG-related research and clinical diagnoses, specifically binds to the matriglycan chain of glycosylated α-DG Core M3, enabling the quantification of α-DG glycosylation levels[37–50]. Fibroblasts from patients with dystroglycanopathies have been characterized using the IIH6C4 antibody in a fluorescence flow cytometry (FFC) assay[51]. The human haploid cell line, HAP1, has become a widely utilized platform in α-DG-related research, covering various areas such as dystroglycanopathy gene discovery[52], enzymatic functions[42], and α-DG binding properties and functions[53]. Additionally, previous studies have established the compatibility of HAP1 cells with FFC[54]. We modified the IIH6C4 FFC assay originally designed for fibroblasts to be suitable for HAP1 cells while improving its sensitivity to differentiate between variants (Figures S1C and S1D). Subsequently, this IIH6C4 assay was adapted for fluorescence-activated cell sorting (FACS) and applied to generate functional scores for all possible coding SNVs in *FKRP* and *LARGE1* (Figure 1A).

Streamlining the saturation mutagenesis and FACS assay in SMuRF

To engineer the cell line platforms for the saturation mutagenesis, the endogenous gene of interest (GOI), *FKRP* or *LARGE1*, was knocked out to make *GOI*-KO HAP1 lines, which lack the endogenous *GOI* function (Figures S1F and S1G). Since α-DG glycosylation level is the biological signal quantified by the functional assay, α-DG overexpression was achieved by Lenti-*DAG1* transduction, as a way to increase assay sensitivity. (Figures S1H-J). Subsequent experiments were performed using monoclonal *GOI*-KO Lenti-*DAG1* HAP1 lines. Since most known dystroglycanopathy cases are caused by missense variants[55], we performed saturation mutagenesis for all possible coding SNVs. We first constructed the plasmids carrying the wild-type (WT) GOIs. To better control the expression level (Figures S1K and S1L), we employed the weak promoter UbC for GOI expression[29], creating Lenti-UbC-*FKRP*-EF1α-*BSD* and Lenti-UbC-*LARGE1*-EF1α-*BSD* (Figure 1B).

Next, we performed PALS-C to introduce all possible coding SNVs using the WT plasmids as templates (Figure 1C). The variants were introduced initially as 64-nt reverse PCR primers. The short length of the oligos allowed for a low cost in synthesizing the oligo pool. We adopted a block-by-block strategy where we divided the GOI variants into multiple non-overlapping blocks (6 for *FKRP* and 10 for *LARGE1*). For each GOI, PALS-C initiates with one single pool of oligos and eventually generates an isolated lentiviral plasmid pool for each block. PALS-C steps are massively multiplexed: Step1 requires only a single-tube reaction; the following steps can be done in a single-tube reaction for each block (STAR Methods). All downstream experiments following PALS-C, up to the NGS, were conducted individually for each block. This block-by-block strategy allowed us to employ short-read NGS to examine variant enrichment while avoiding the requirement of an additional NGS to assign barcodes to variants spanning the entire CDS and the expenses associated with it[20,56,57].

Variant representation in the lentiviral plasmid pools generated by PALS-C was evaluated using a shallow NGS service, according to which, more than 99.6% of all possible SNVs were represented (Figures S2A-D). The plasmid pools were packaged into lentiviral particles, which were subsequently delivered into the platform cells through transduction. Once transduced cells

were sufficiently expanded following drug selection, they were sorted using FACS to isolate cells with either high or low glycosylation levels. Staining conditions and gating parameters in FACS were optimized with mini-libraries (Figure S2E). α-DG glycosylation level was quantified by IIH6C4-FITC signal (Figure 1D, and Figures S2F-G). The FACS events of each group achieved a minimum of ~1000 × coverage.

Genomic DNA from each group of each block was used to build the multiplexed sequencing library using a 3-round PCR strategy (Figure 1E). Raw NGS datasets were analyzed with our customized analytical pipeline "Gargamel-Azrael" to generate SMuRF scores for all variants (Figure 1F). Essentially, the SMuRF score is the normalized relative enrichment of a variant in the FACS groups (STAR Methods). High SMuRF scores indicate normal functionality of variants in α-DG glycosylation while low scores indicate deleterious effects.

SMuRF recapitulated and expanded the knowledge gained from variant databases

Three biological transduction replicates were performed for each gene to confirm the reproducibility of the workflow (Figure S2H and Table S1). After combining biological replicates using DiMSum[58], SMuRF scores were generated for all possible coding SNVs of *FKRP* (4455) and *LARGE1* (6804) (Figures S3A and S3B, Tables S2 and S3), and those with high confidence were used for downstream analysis (*FKRP*: 4325, 97.1%; LARGE1: 6473, 95.1%). SMuRF scores align with the anticipated patterns of different variant types (Figures 2A and 2B). *FKRP* synonymous variants have a median of -0.05, with a 95% confidence interval (CI) of -0.07~-0.02, while *LARGE1* synonymous variants have a median of -0.07 (95% CI: -0.09~-0.04). The nonsense variants consistently exhibit low SMuRF scores, with sparse outliers observed. *FKRP* nonsense variants have a median of -2.34 (95% CI: -2.43~-2.22), while *LARGE1* nonsense variants have a median of -2.51 (95% CI: -2.57~-2.45). Two noteworthy outliers among the nonsense variants are *FKRP* c.1477G>T (p.Gly493Ter) (SMuRF = -0.44) and *LARGE1* c.2257G>T (p.Glu753Ter) (SMuRF = 0.15). These two are the nonsense variants positioned closest possible to their respective canonical stop codons. The relatively high SMuRF scores of these two variants suggest that their impact on the enzymatic function is negligible in the context of the CDS constructs. Furthermore, since both variants are in the last exon of their respective transcripts, it is also unlikely for them to be substantially influenced by nonsense-mediated decay (NMD) *in vivo*[59].

Interestingly, the start-loss variants exhibit markedly low SMuRF scores, significantly lower than those observed in most nonsense variants (p-value = 2e-5, *FKRP*; 1.4e-5, *LARGE1*). *FKRP* start-loss variants have a median of -3.21 (95% CI: -3.52~-3.14), while *LARGE1* start-loss variants have a median of -3.19 (95% CI: -3.84~-2.99). While this observation does not necessarily indicate stronger negative effects of the start-loss variants, compared to that of the nonsense variants, it indicates that, at least in the context of the *FKRP* and *LARGE1* SMuRF CDS constructs, there is a lack of effective genetic compensation to counter the start-loss variants, such as functional downstream alternative start codons[60]. The homozygous start-loss variant *FKRP* c.1A>G (SMuRF = -3.29) has been reported to be associated with WWS, the most severe FKRP-related disorder. This variant has been documented in two cases, with one resulting in the unfortunate death of a child at the age of 6 days and the other leading to a terminated

pregnancy[28]. Therefore, the start-loss variants may warrant increased attention in genetic testing protocols.

When compared to allele frequency (AF) data in The Genome Aggregation Database (gnomAD, v4.0.0)[6,61], SMuRF scores aligned with the selection against pathogenic variants: low allele frequency variants (AF < 1.5e-05, representing allele count of 1 or 2 in gnomAD v4 genome sequencing data), exhibited a wide range of functional scores, while variants with higher frequency showed functional scores converging towards the WT score as common variants are depleted of pathogenic variants (Figures 2C, 2D, S3C and S3D).

The α-DG glycoepitope, as well as the enzymes involved in its glycosylation, are largely conserved within Metazoa[62]. Evolutionary conservation scores, such as PhyloP scores, indicate the degree of conservation of a variant derived from multiple sequence alignments across species[63]. When compared with PhyloP scores calculated from 100 vertebrates, SMuRF demonstrated the evolutionary tolerance of relatively harmless variants and the selection against damaging variants in both *FKRP* and *LARGE1* (Figures S3E and S3F), with a tendency for missense variants to be more disruptive at the more conserved sites (rho = -0.44, *FKRP*; -0.25, *LARGE1*).

SMuRF is sensitive and accurate in predicting variant pathogenicity and severity

ClinVar is a public archive of reports of human variants[64], where the variants were classified according to clinical supporting evidence into different categories including Benign (B), Benign/likely benign (B/LB), Likely benign (LB), Likely pathogenic (LP), Pathogenic/likely pathogenic (P/LP), Pathogenic (P), and Variants of Uncertain Significance (VUS). SMuRF scores correlate well with clinical classification in ClinVar (Figures 3A, 3B, S3G, and S3H). Furthermore, dystroglycanopathies associated with *FKRP* encompass a spectrum of diseases with varying severity, including severe cases like WWS and muscle-eye-brain disease (MEB), intermediate cases like congenital muscular dystrophies (CMD), and relatively mild cases like LGMDR9 (LGMD2I)[24,28]. To explore whether SMuRF scores reflect the disease severity, we employed a naive additive model where the functional scores of the variants on both alleles were summed to calculate the biallelic functional score[65,66]. We aggregated data from 8 well-curated cohorts and compared them with SMuRF scores (Table S4)[28,47,55,67–71]. The combined functional scores of the mild cases were significantly higher compared to those of the intermediate and severe cases, as expected (Figure 3C). Another strategy to evaluate the biallelic variant effect is using the higher allelic score to represent the compound heterozygous variant pair[56]. This model produces similar results and conclusions as the naive additive model (Figure S3I). Additionally, SMuRF scores showed a correlation with the reported disease onset age (Figure 3D), where high-function variants were associated with later onset (rho = 0.72; 0.70, male; 0.73, female). Cox proportional hazards analysis revealed a statistically significant association between the combined SMuRF scores and the onset (p-value < 0.003). The estimated hazard ratio was 0.28 (95% CI: 0.12 - 0.64), indicating that higher functional scores are correlated with a substantially lower risk of earlier disease onset. Specifically, the negative coefficient (-1.27) suggests that as the functional score decreases, the likelihood of early onset increases. The model's concordance index of 0.67 reflects a moderate level of predictive accuracy.

These analyses indicate the potential utility of the SMuRF scores for providing another line of evidence for clinical variant interpretation. According to the standards and guidelines developed jointly by the American College of Medical Genetics (ACMG) and the Association for Molecular Pathology (AMP), the SMuRF scores can potentially be applied as functional evidence under the PS3/BS3 criterion. This criterion pertains to "well-established" functional assays that demonstrate whether a variant has abnormal or normal gene/protein function, respectively[72,73]. Therefore, we employed the SMuRF scores to label the *FKRP* variants as SMuRF Benign, SMuRF Mild, SMuRF Intermediate and SMuRF severe, and *LARGE1* variants as SMuRF Benign and SMuRF Pathogenic (Tables 1, S2, S3 and STAR Methods). It is important to note that clinical variant interpretation involves the application of various criteria, and functional data such as SMuRF scores are just one piece of the contributing evidence, which is not intended to be used as a decisive classification in isolation. For *FKRP*, the SMuRF labeling was performed based on 3 boundary scores. The boundary score between SMuRF Benign and SMuRF Mild (-0.87) was determined according to the inflection point between the ClinVar P/LP and ClinVar B/LB densities. The boundary score between SMuRF Mild and SMuRF Intermediate (-1.62) was determined according to the peak value of the mild cases from the 8 well-curated cohorts. The boundary score between SMuRF Intermediate and SMuRF Severe (-2.24) was determined according to the peak value of the severe cases from the 8 well-curated cohorts. For *LARGE1*, the SMuRF labeling was performed based on 1 boundary score. The boundary score between SMuRF Benign and SMuRF Pathogenic (-1.32) was determined according to the inflection point between the ClinVar P/LP and ClinVar B/LB densities.

SMuRF scores can be employed to validate and improve computational predictors

In addition to directly assisting in variant classification, SMuRF scores can also be used to validate and improve computational predictors. Computational prediction is currently an active area of research with a wealth of methods recently developed, including CADD[74], metaSVM[75], REVEL[76], MVP[77], EVE[78], MutScore[79], PrimateAI-3D[80], ESM1b[81], MAVERICK[82], and AlphaMissense[83,84]. To compare SMuRF with these computational predictors, we assessed the receiver operating characteristic (ROC) curves for all methods using the P, P/LP, and LP variants in ClinVar as the positive reference set. For the negative reference set, we used missense variants observed in the gnomAD v4 genome sequencing database. Variants present in both ClinVar P, P/LP, LP and gnomAD v4 were excluded from the negative reference set. A higher Area Under Curve (AUC) value indicates better discriminatory ability in classifying pathogenic variants (Figures 4A, 4B and Table S5). SMuRF outperforms all computational methods for *LARGE1* (AUC = 0.96), followed by REVEL (AUC = 0.88). For *FKRP*, two predictors, REVEL (AUC = 0.88) and EVE (AUC = 0.86) exhibit comparable performance to SMuRF (AUC = 0.86). It is noted that REVEL was trained with Human Gene Mutation Database (HGMD) pathogenic variants, which have an overlap with ClinVar variants, potentially contributing to its higher AUC when evaluated on the ClinVar dataset. We checked the correlation between the predictors and SMuRF. Most predictors assigned higher scores to pathogenic variants, and hence are negatively correlated with SMuRF scores (except ESM1b). Among all the predictors examined, AlphaMissense has the strongest correlation with SMuRF (rho = -0.70, *FKRP*; -0.54, *LARGE1*) (Figures 4C-E, S4A and Table S6). REVEL, among the predictors, demonstrates the best performance according to the ROC curves and also exhibits a relatively good correlation with SMuRF (rho = -0.65, *FKRP*; -0.47, *LARGE1*) (Figure S4B).

To explore the potential benefit of combining experimental and predicted scores for more accurate variant interpretation, we selected three *LARGE1* VUSs from ClinVar with concordant designations by both SMuRF and AlphaMissense: c.2175C>G p.Ile725Met (concordant likely benign; SMuRF = 0.43; AlphaMissense = 0.0733), c.1060T>G p.Cys354Gly (concordant likely pathogenic; -2.03; 0.9962), c.2153G>C p.Arg718Pro (concordant likely pathogenic; -2.56; 0.9995). Anti-matriglycan Western blot and laminin overlay results confirmed the expected glycosylation activity of these 3 variants, along with the variants with decisive clinical classification (Figure 4F). This approach showcases the strategy for enhancing variant interpretation by corroborating DMS scores with prediction scores.

SMuRF highlighted the critical structural regions

The currently known disease-related mutations in *FKRP* and *LARGE1* are distributed throughout their entire sequences (Figure S5A), and only limited critical structural sites have been identified and associated with specific disease mechanisms. SMuRF can contribute to highlighting critical structural regions in the enzymes. The protein structures of both FKRP and LARGE1 have been previously studied. FKRP is known to have a stem domain (p.1-288) and a catalytic domain (p.289-495)[85]. SMuRF scores revealed that missense variants in the catalytic domain are generally more disruptive than those in the stem domain (p-values < 2.22e-16) (Figure 5A and 5B). Furthermore, it has been reported that a zinc finger loop within the catalytic domain (p.289-318) plays a crucial role in FKRP enzymatic function[85]. SMuRF analysis demonstrated that missense variants in the zinc finger loop (SMuRF median: -1.50; 95% CI: -2.00~-1.10) exhibit greater disruption, compared to variants in the remaining region of the catalytic domain (SMuRF median: -1.19; 95% CI: -1.29~-1.08), though not statistically significant (p-value = 0.17).

LARGE1 has two catalytic domains: a xylose transferase (XylT) domain (p.138-413) and a glucuronate transferase (GlcAT) domain (p.414-756)[86]. They are each responsible for adding one unit of the polysaccharide matriglycan chain, which consists of alternating xylose and glucuronate units. SMuRF revealed that the missense variants in both catalytic domains tend to be significantly more disruptive than the variants in the N-terminal domain (p-values < 2.22e-16) (Figure 5C and 5D). Interestingly, SMuRF also showed that the missense variants in the XylT domain (SMuRF median: -0.70; 95% CI: -0.79~-0.62) tend to be more disruptive than those in the GlcAT domain (SMuRF median: -0.42; 95% CI: -0.45~-0.38) (p-value < 2.22e-16). A previous IIH6C4 western blot experiment revealed a similar observation, demonstrating that mutations deactivating the GlcAT domain, but not the XylT domain, can generate a faint band indicative of glycosylated matriglycan[86]. Moreover, AlphaMissense also presents the same effects, where missense variants in the XylT domain tend to be more disruptive (Figure S4C and Supplemental Note).

To explore the variants' effects on different structural regions, we first mapped the mean SMuRF scores of SNV-generated single amino acid substitutions onto the 1D structures of the enzymes (Figures 5E and 5F). The heatmaps highlight the functional importance of the catalytical domains. We further mapped the same mean SMuRF scores onto the 3D structures of the enzymes (Figures 5G and 5H) (FKRP: PDB 6KAM; LARGE1: PDB 7UI7), thereby highlighting the critical regions that are susceptible to missense disruptions. SMuRF confirmed the functional

importance of p.Cys318 in FKRP (mean SMuRF = -2.15), which is required for Zn2+ binding in the zinc finger loop[87]. The p.Cys318Tyr variant (SMuRF = -2.23) has been reported to be associated with WWS[70]. SMuRF also highlighted the functional importance of p.Phe473 in FKRP (mean SMuRF = -1.99), which is located in a small hydrophobic pocket essential for CDP-ribitol substrate binding within the catalytic domain[87]. Three important amino acids in the FKRP stem domain were labeled on the 3D structure as well: p.Tyr88 (mean SMuRF = -2.69) and p.Ser221 (mean SMuRF = -1.00), which are both situated at the subunit-subunit interface involved in FKRP tetramerization *in vivo* (Figure S5B), and p.Leu276 (mean SMuRF = -0.61), which interacts with the catalytic domain[85]. p.Tyr88Phe is likely associated with disease[88], and has a low SMuRF score (-3.26). p.Ser221Arg was associated with CMD-MR (MR: mental retardation)[89]. All three p.Ser221Arg SNVs have low SMuRF scores (c.661A>C: -2.15; c.663C>A: -2.21; c.663C>G: -2.17). Moreover, c.663C>A was examined in the mini-library screen and presented low function (Figure S2E). p.Leu276Ile is a founder mutation in the European population[90], which is commonly associated with milder symptoms[91]. Interestingly, it has a relatively higher SMuRF score (-0.44) and performed more similarly to the benign variants rather than other pathogenic variants in the mini-library screen (Figure S2E). In addition, SMuRF highlighted the importance of p.Asp242 (mean SMuRF = -2.61) and p.Asp244 (mean SMuRF = -2.34) in LARGE1, which are crucial for XylT activity, as well as p.Asp563 (mean SMuRF = -1.18) and p.Asp565 (mean SMuRF = -1.57), which are required for GlcAT activity[86]. Variants affecting different enzyme domains may require distinct treatment approaches[92]. SMuRF can assist in selecting appropriate treatments for different variants by highlighting critical regions in different domains.

Lastly, we separated the amino acids into five groups: “negatively charged”, “positively charged”, “polar, uncharged”, “non-polar, aromatic”, and “non-polar, aliphatic”. Subsequently, the mean SMuRF scores of the SNV-induced substitutions were plotted for each residue according to their corresponding group (Figure 5I and 5J). Together, these figures provide an intuitive grasp of the patterns of constraint throughout the primary sequence and potentially aid in helping interpret variants in critical regions such as the catalytical domains and the signal anchor regions (*e.g.*, p.1-44 of FKRP)[93,94].

Validations confirmed SMuRF findings in the myogenic context

One caveat of SMuRF is that the HAP1 platform cell line, although widely used in α-DG-related studies, may not fully reflect the clinical relevance of dystroglycanopathies, which primarily affect neuromuscular tissues[95]. To address this issue, we generated myogenic platform cell lines by engineering MB135, a human control myoblast cell line[96]. Endogenous *FKRP* or *LARGE1* were knocked out respectively in the MB135 cell line. Monoclonal homozygous KO lines were established for both genes (Figure S6A). Despite being incompatible with the flow cytometric assay (Figure S6B), the KO MB135 myoblasts were effectively utilized for individual variant validation using an immunofluorescence (IF) assay that we developed. The *FKRP*-KO and *LARGE1*-KO MB135 myoblasts were rescued by different individual variants using lentivirus and differentiated into myotubes for IIH6C4 IF staining (Figures 6A, 6B and S6C). The results were consistent with the SMuRF scores, the mini-library screen (Figure S2E) and the ClinVar reports.

To further validate the SMuRF scores, we developed an orthogonal assay that is independent of the IIH6C4 antibody to examine α-DG glycosylation level. Proper glycosylation of α-DG is crucial for the viral entry of Lassa fever virus (LASV) and its utility as an assay to study dystroglycanopathies has been demonstrated in previous studies[97]. LASV glycoprotein complex (LASV-GPC) has been employed to generate recombinant vesicular stomatitis virus (rVSV-LASV-GPC) as a safer agent for investigating LASV entry[98,99]. rVSV-LASV-GPC was utilized in a previous gene-trap screen in HAP1 cells to identify crucial genes involved in α-DG glycosylation, where cells with dysfunctional α-DG glycosylation genes exhibited increased resistance to rVSV-LASV-GPC infection, resulting in their enrichment in the population[52]. We increased the sensitivity of this VSV assay and employed it to validate interesting findings from the flow-cytometry assay (Figure 6C). Instead of using rVSV-LASV-GPC, whose genome contains the LASV-GPC coding sequence, allowing the virus to re-enter cells, we utilized pseudotyped ppVSV-LASV-GPC. The ppVSV cannot re-enter cells as it is pseudotyped using a LASV-GPC helper plasmid and lacks a viral glycoprotein coding sequence in its genome (STAR Methods). The *GOI*-KO MB135 myoblasts were first transduced with Lenti-*GOI* variants and selected with blasticidin. Subsequently, half of the cells were infected by the ppVSV-LASV-GPC, while the other half was kept as the non-infected control.

We mixed 10 Lenti-*FKRP* variants with Lenti-WT-*FKRP* to rescue *FKRP*-KO MB135 (Figure 6D), and mixed 11 Lenti-*LARGE1* variants with Lenti-WT-*LARGE1* to rescue *LARGE1*-KO MB135 (Figure 6E). *FKRP* c.135C>T (p.Ala45=; SMuRF = 0.17; AF = 0.14) and *LARGE1* c.435C>T (p.Ala145=; SMuRF = -0.43; AF = 0.41) have the highest AFs in gnomAD v4, which were used as high-function references. *FKRP* c.663C>A (p.Ser221Arg; SMuRF = -2.21) and *LARGE1* c.992C>T (p.Ser331Phe; SMuRF = -2.58) are Pathogenic in ClinVar, which were used as low-function references. Nonsense variants with high SMuRF scores, as well as synonymous outliers with either low or high scores, were included to be validated (inclusion criteria: Supplemental Method). The ppVSV assay supports the notion that the 3'-most nonsense variants, *FKRP* c.1477G>T (p.Gly493Ter; SMuRF = -0.44) and *LARGE1* c.2257G>T (p.Glu753Ter; SMuRF = 0.15), can retain the enzymatic functions, while other nonsense variants (including those with 2nd highest SMuRF scores for both genes) did not exhibit high functionality in the ppVSV assay. With the successful application of the ppVSV assay in small-scale variant interpretation experiments, we have laid the foundation for optimizing the conditions to make the assay applicable for large-scale, all-possible variant interpretation experiments (Supplemental Note).

Discussion

In this study, we introduced SMuRF as a framework for DMS studies and demonstrated its utility in improving variant interpretation in genetic diseases. We developed an accessible saturation mutagenesis method and functional assays with the sensitivity and accuracy required by DMS studies. We generated high-confidence functional scores for over 97% and 95% of all possible coding SNVs of *FKRP* and *LARGE1*, respectively. The SMuRF scores enable severity prediction, which was validated with well-curated patient reports. SMuRF also resolved critical protein structural regions susceptible to missense disruptions, aiding in variant interpretation.

The SMuRF framework builds upon prior DMS approaches but distinguishes itself by prioritizing accessibility and simplicity. The PALS-C method of saturation mutagenesis can be performed in a pooled fashion using common reagents and an unmodified plasmid template. In contrast, other similar methods either require a large number of individual reactions[20], or require a specialized template (*e.g.*, uracil-containing single-stranded DNA)[19,56,100,101]. Our method has avoided these needs by optimizing the cloning steps. In addition, we employed a block-by-block approach of introducing variants and then performing sorting on each block. This approach avoids assigning barcodes for each variant and the expenses associated with it. Lastly, we have made detailed protocols and code available for designing and cloning of libraries and downstream data processing, which can be non-trivial barriers of entry for researchers.

Our study highlighted the need to control expression levels when studying enzymes. In developing our assay, we observed that over-expression of a *FKRP* mutant can compensate for its enzyme deficiency. We used the weaker UbC promoter, which is better correlated with endogenous expression and avoids this compensation. Overcoming these expression challenges in *FKRP* and *LARGE1*, SMuRF can be readily adapted for studying other enzymes involved in α-DG Core M3 glycosylation (Figure S1B). Additionally, variants in POMGNT1, an enzyme that mainly participates in the glycosylation of Core M1 and M2, can perturb the IIH6C4 signal[37] (Figure S1B), potentially by serving as an "enzymatic chaperone" for FKTN[62,102–104]. However, variants in GnT-Vb/IX, an enzyme that participates in Core M2 glycosylation, are unlikely to alter IIH6C4 signal[105]. In summary, SMuRF is readily applicable to at least 12 other dystroglycanopathy enzymes, and the PALS-C saturation mutagenesis method can be easily adapted for other genes.

Notably, the MaveDB database aggregates data from published DMS studies, but so far, only about 50 human genes have been recorded[106]. Charting the "Atlas of Variant Effects" across the entire human genome should be a crowdsourced effort, thus requiring the associated technologies and resources used in DMS to be accessible to as many laboratories as possible. We developed the SMuRF workflow to reduce DMS cost and complexity to encourage crowd-sourcing efforts required for coverage across thousands of disease genes.

Limitations of the study

In this study, we decided to explore all possible SNVs, therefore not all amino acid substitutions were represented. However, population sequencing data has demonstrated that multiple SNVs in the same codon (i.e. multi-nucleotide variants) are extremely rare[7], and the majority of VUS are SNVs[55], thus justifying that representing all possible SNVs is more practical and cost-effective when data is to be used to interpret variants observed in sequenced individuals. Additionally, it is worth noting that the SMuRF workflow can be extended to analyze small-sized variants beyond SNVs, including single amino acid substitutions, small insertions, or deletions. Future research covering these variants can provide further functional insights.

It is important to consider certain factors when utilizing scores from SMuRF or other DMS studies. The majority of dystroglycanopathies are recessive and patients may harbor compound heterozygous variants, which raises the need for further investigation on how to apply DMS functional scores in interpreting such cases. In this study, we explored a naive additive model

where the biallelic functional scores were calculated by the simple addition of the SMuRF scores of the variants on both alleles. This model demonstrated a promising correlation between SMuRF scores and disease severity (Figures 3C and 3D). The higher-allelic score model was also tested, which agreed with the additive model (Figure S3I). However, future experiments that can better model biallelic effects, such as preinstallation of one allele, are required to improve these analyses. Additionally, our results emphasized the significance of well-curated reports in predicting disease severity (Figure S3J). Our findings revealed a correlation between the disease onset age and SMuRF scores (Figures 3D and S3K). However, creatine kinase (CK) values did not show a significant correlation with SMuRF scores (Figure S3L). This observation is consistent with the knowledge in the field that CK levels can fluctuate with activity and decrease when muscle mass is lost over time[107]. Additionally, it is important to acknowledge the need for a larger patient dataset to enhance the predictive accuracy of SMuRF scores for patient phenotypes, including age of onset and disease severity. In our study, we evaluated 54 cases, with 6 cases classified as "Severe". The limited sample size restricts the robustness and reliability of our conclusions. Future cohort research expanding the dataset, will help improve the accuracy.

The use of SMuRF scores for the interpretation of genetic variants also poses challenges in the following aspects. Firstly, the IIH6C4 assay may only capture one of the multiple functions associated with FKRP and LARGE1. For example, FKRP is also known to participate in the glycosylation of fibronectin[108]. Second, the presence of paralogs may impact the clinical relevance of the SMuRF scores. For instance, *LARGE2*, a paralog of *LARGE1* resulting from a duplication event first observed in Chondrichthyes[62], may act as an effective modifier in *LARGE1*-related diseases[109–111]. Third, SMuRF analysis for nonsense variants has inherent limitations. In the context of the CDS constructs used in SMuRF, a nonsense mutation does not trigger exon-junction complex (EJC)-enhanced NMD. However, in the correct multi-exon genomic context, when a nonsense mutation is located upstream of the last EJC, it undergoes EJC-dependent NMD, further reducing any residual functions that the truncated proteins may possess[112].

Despite the limitations of this study, the SMuRF scores hold potential for being considered as functional evidence for the PS3/BS3 criterion following ACMG/AMP guidelines. Lastly, a careful curation of evidence from all available sources, including but not limited to clinical reports, DMS, computational predictors, will aid in the generation of a comprehensive and accurate variant interpretation guidebook.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
IIH6C4 antibody	Sigma-Aldrich Dr. Kevin Campbell	05-593 (Discontinued)
Rabbit anti-Mouse IgM FITC Secondary Antibody	Invitrogen	31557
Primary Anti-FKRP antibody	Abcam	ab220059
Primary Anti-Vinculin antibody	Sigma-Aldrich	V9131
Anti-laminin antibody	Sigma-Aldrich	L9393

Anti-rabbit HRP conjugated secondary antibody	Cell Signaling Technology	7074S
Anti-mouse HRP conjugated secondary antibody	Cell Signaling Technology	7076S
IRDye 800 CW Goat anti-mouse IgM (μ chain specific) Secondary Antibody	Li-COR Biosciences	926-32280
anti-rabbit IgG (H + L); donkey polyclonal	LI-COR Biosciences	926-32213
Bacterial and virus strains		
Endura Electrocompetent Cells	Lucigen	60242-1
rVSV-LASV-GPC viral particles	Dr. Melinda A. Brindley	
ppVSVΔG-VSV-G viral particles	Dr. Melinda A. Brindley	
ppVSV-LASV-GPC-Generation1	This paper	
ppVSV-LASV-GPC-Generation2	This paper	
ppVSV-LASV-GPC-Generation3	This paper	
Chemicals, peptides, and recombinant proteins		
SpCas9 2NLS Nuclease	Synthego	
SE Cell Line Nucleofector Solution	Lonza	PBC1- 00675
NEBuilder HiFi DNA Assembly Master Mix	NEB	E2621
T4 DNA Ligase	NEB	M0202
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	1725271
TransIT-LT1 Transfection Reagent	Mirus	MIR 2300
Lenti-X Concentrator	Takara	631232
Lenti-X GoStix Plus	Takara	631280
Blasticidin S HCl	Gibco	A1113903
Puromycin Dihydrochloride	Gibco	A1113803
Versene	Gibco	15040066
Viobility 405/452 Fixable Dye	Miltenyi Biotec	130-130-420
MACS BSA Stock Solution	Miltenyi Biotec	130-091-376
autoMACS Rinsing Solution	Miltenyi Biotec	130-091-222
Human BD Fc Block	BD Pharmingen	564220
Gelatin	Sigma-Aldrich	G9391
PFA	Sigma-Aldrich	158127
Bovine Serum Albumin	Sigma Aldrich	A9647
Antifade Mounting Medium with DAPI	Vector Laboratories	H1500
NH4Cl	Sigma-Aldrich	A9434
Q5 High-Fidelity DNA Polymerase	NEB	M0491SVIAL
Type2M enzyme DpnI	NEB	R0176S
Q5 Reaction Buffer	NEB	B9027SVIAL
Q5 High GC Enhancer	NEB	B9028AVIAL
dNTPs	NEB	N0447
BsmBI-v2	NEB	R0739S
BsaI-HFv2	NEB	R3733S
MspJI	NEB	R0661S
XbaI	NEB	R0145S
BamHI-HF	NEB	R3136S
Protease Inhibitor Cocktail	Sigma-Aldrich	P8340-1ML
RIPA lysis and extraction buffer	Pierce	PI89900
4x Laemlli	BioRad	1610747
PageRuler Plus Prestained Protein Ladder	ThermoFisher Scientific	26619

Clarity Max ECL substrate	BioRad	1705062
Clarity ECL substrate	BioRad	1705060
UDP-xylose	Sussex Research Laboratories	Custom order
UDP-glucuronate	Sigma-Aldrich	U6751
4-methylumbelliferone-glucuronate-xylose	Sussex Research Laboratories	Custom order
Laminin (Natural, mouse) Lam-111	Gibco	23017-015
Critical commercial assays		
Synthetic Single Guide RNA Kit	Synthego	
PrimeScript RT Reagent Kit	Takara	RR037
Phusion High-Fidelity DNA Polymerase	NEB	M0530
NEBuilder	NEB	E2621
NucleoSpin Gel and PCR Clean-Up Kit	Takara	740609
PureLink Midiprep Kit	Invitrogen	K210014
PureLink Genomic DNA Mini Kit	Invitrogen	K182002
SE Cell Line 4D-Nucleofector X Kit S	Lonza	V4XC-1032
BioRad DC Assay II Kit	BioRad	5000112
Experimental models: Cell lines		
HAP1	Horizon Discovery	C631
DAG1-KO HAP1	Horizon Discovery	HZGHC000120c016
HEK293T	ATCC	CRL-3216
MB135	Dr. Stephen J. Tapscott	
FKRP-KO HAP1 (c.181Adup)	This paper	
FKRP-KO MB135 (c.181Adup)	This paper	
LARGE1-KO HAP1 (c.121_214del)	This paper	
LARGE1-KO MB135 (c.121_214del)	This paper	
Oligonucleotides		
FKRP-KO and *LARGE1*-KO sgRNAs, genotyping primers	This paper (See Method S1)	
DAG1, *FKRP*, *LARGE1* CDS sequences, primers for CDS plasmid construction	This paper (See Method S2)	
RT-PCR, RT-qPCR, real-time PCR primers	This paper (See Method S3)	
Primers for PALS-C	This paper (See Method S4)	
R1 oligo lists	This paper	https://github.com/leklab/Balthazar/tree/main/PALS-C_oligos
Primers for plasmid pool QC	This paper (See Method S6)	
Primers for lentiviral sequence isolation, *FKRP* Sanger sequencing, *FKRP* Amplicon-EZ sequencing, *LARGE1* Sanger sequencing	This paper (See Method S7)	
Primers for PCR1, PCR2	This paper (See Method S8)	
Primers for NGS library construction for VSV related experiments	This paper (See Method S10)	
Recombinant DNA		
pLV-hTERT-IRES-hygro	Hayer et al.[104]	Addgene Plasmid #85140

lentiCRISPR v2	Sanjana et al.[113]	Addgene Plasmid #52961
lentiCas9-Blast	Sanjana et al.[113]	Addgene Plasmid #52962
pAAV-UbC-eGFP-F	Dr. Pantelis Tsoulfas	Addgene Plasmid #71545
lentiGuide-Puro	Sanjana et al.[113]	Addgene Plasmid #52963
Lenti-DAG1	This paper	Addgene Plasmid #205149
Lenti-UbC-*FKRP*-EF1a-BSD	This paper	Addgene Plasmid #205150
Lenti-UbC-*LARGE1*-EF1a-BSD	This paper	Addgene Plasmid #205151
psPAX2	Dr. Didier Trono	Addgene Plasmid #12260
pMD2.G	Dr. Didier Trono	Addgene Plasmid #12259
LASV-GPC	Dr. Melinda Brindley	
Software and algorithms		
R	R Core Team	https://www.r-project.org/contributors.html
Scripts for Oligo design, pre-SMuRF experiments	This paper	https://github.com/lekllab/Balthazar
Raw NGS data analytical pipeline	This paper	https://github.com/lekllab/Gargamel
Scripts for generating SMuRF scores and downstream analyses	This paper	https://github.com/lekllab/Azrael
ChatGPT 3.5	OpenAI	https://chat.openai.com/g/g-F00faAwkE-open-a-i-gpt-3-5
Gemini	Google	https://gemini.google.com/
Other		
0.45 μm PES filter	Thermo Scientific	165-0045
40 μm Cell Strainer	Falcon	352340
Thermanox Coverslips	Thermo Scientific	174969
SMuRF scores	This paper	https://www.mavedb.org/experiments/urn:mavedb:00001197-a
FFC and FACS datasets	This paper	FlowRepository (FR-FCM-Z6LL, FR-FCM-Z6LM, FR-FCM-Z6LN, FR-FCM-Z6LP, FR-FCM-Z6LQ, FR-FCM-Z6LR, FR-FCM-Z6LX, FR-FCM-Z77U, FR-FCM-Z794)

NGS raw data	This paper	Sequence Read Archive of the National Center for Biotechnology Information (accession: PRJNA993285).
Online protocol	This paper	https://www.protocols.io/view/using-saturation-mutagenesis-reinforced-functional-cwvvxe66/materials
4-15% mini-protean TGX stain free protein gel	BioRad	4568084
Nitrocellulose Membrane	BioRad	1704158
WGA resin	Vector Laboratories	AL-1023

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Monkol Lek (monkol.lek@yale.edu).

Materials Availability

- Plasmids generated in this study have been deposited to Addgene: Lenti-*DAG1* (205149), Lenti-UbC-*FKRP*-EF1α-BSD (205150), and Lenti-UbC-*LARGE1*-EF1α-BSD (205151)

Data and Code Availability

- NGS raw data have been deposited at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and are publicly available as of the date of publication. Accession numbers are listed in the Key Resources Table.
- SMuRF scores have been deposited on MaveDB. The link is listed in the Key Resources Table.
- FFC and FACS datasets have been deposited on FlowRepository. IDs are listed in the Key Resources Table.
- All original code has been deposited on GitHub (https://github.com/leklab) and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Online protocol has been deposited at https://www.protocols.io/view/using-saturation-mutagenesis-reinforced-functional-cwvvxe66/materials.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Lines

- Wildtype HAP1 (C631) and *DAG1*-KO HAP1 (HZGHC000120c016) cells (male lacking Y chromosome) were ordered from Horizon Discovery. All HAP1 cells were cultured at 37ºC in Iscove's Modified Dulbecco's Medium (IMDM) (Gibco, 12440053) with 10% Fetal Bovine Serum (FBS, R&D Systems, S11150) and 1x Antibiotic-Antimycotic (Anti-anti, Gibco, 15240062). The medium was replaced every 2 days, unless otherwise stated. HAP1 cells tend to grow into multi-layers; hence, to keep the cells in optimal status, TrypLE Express Enzyme (Gibco, 12605010) was used to passage the cells to maintain the cells in healthy confluency (30-90%). HAP1 cells used in SMuRF were immortalized using lentivirus packaged with pLV-hTERT-IRES-hygro (Addgene, 85140), a gift from Dr. Tobias Meyer.
- HEK293T cells (female) were cultured at 37ºC in DMEM (Gibco, 11995065) with 10% FBS and 1x Anti-anti. The medium was replaced every 2 days, unless otherwise stated.
- MB135 cells (female) were cultured at 37ºC in Ham's F-10 Nutrient Mix (Gibco, 11550043) with 20% FBS, 1x Anti-anti, 51 ng/ml dexamethasone (Sigma- Aldrich, D2915) and 10 ng/mL basic fibroblast growth factor (EMD/Millipore, GF003AF-MG). The medium was replaced every 2 days, unless otherwise stated. MB135 cells were differentiated in Skeletal Muscle Differentiation Medium (PromoCell, C-23061) with 1x Anti-anti. The differentiation medium was replaced every 4 days, unless otherwise stated.

METHOD DETAILS

CRISPR RNP nucleofection

Synthetic Single Guide RNA (sgRNA) Kits and SpCas9 2NLS Nuclease were ordered from Synthego. RNP complexes were prepared in SE Cell Line Nucleofector Solution (Lonza, PBC1-00675) and delivered into cells with a Lonza 4D-Nucleofector. The program used for HAP1 was EN-138; the program used for MB135 was CA-137. Single clones were isolated from pooled nucleofected cells and genotyped by targeted Sanger sequencing. sgRNA sequences, RNP complex preparation conditions, and genotyping primers are provided in Method S1. *FKRP*-KO HAP1 carries a 1-bp insertion (c.181Adup); *FKRP*-KO MB135 is homozygous for the same mutation. *LARGE1*-KO HAP1 carries a 94-bp deletion (c.121_214del); *LARGE1*-KO MB135 is homozygous for the same mutation.

Plasmid construction

Lenti-*DAG1* plasmid used the backbone of lentiCRISPR v2, which was a gift from Dr. Feng Zhang (Addgene, 52961). *DAG1* coding exons were cloned from human genome DNA by PCR. Lenti-*FKRP* plasmids and Lenti-*LARGE1* plasmids used the backbone of lentiCas9-Blast, which was a gift from Dr. Feng Zhang (Addgene, 52962). *FKRP* coding exon was cloned from HAP1 genome DNA. *LARGE1* coding sequence was cloned from HEK293T cDNA. HEK293T carries a *LARGE1* mutation (c.1848G>A) on one allele, which was removed from the Lenti-*LARGE1* plasmids to make the pooled variant library. The removal of this mutation used the same strategy as the introduction of individual variants to the lentiviral plasmids for the mini-libraries: briefly, a short localized region was cut with restriction enzymes from the wildtype plasmid and 2 variant-carrying inserts, each covering 1 of 2 sides of this region were inserted. The UbC promoter was cloned from pAAV-UbC-eGFP-F, which was a gift from Dr. Pantelis Tsoulfas (Addgene, 71545). The EF-1α promoter was taken from lentiGuide-Puro, which was a gift from Dr. Feng Zhang (Addgene, 52963). BSD-WPRE was from lentiCas9-Blast. The lentiviral

plasmids used for the pooled library contain a UbC-driven gene-of-interest CDS and an EF-1α-driven *BSD*. *BSD* encodes Blasticidin S deaminase (BSD), which confers blasticidin resistance in transduced cells. Plasmid assemblies were achieved either with NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621) or T4 DNA Ligase (M0202). Cloning details of plasmid construction and the list of plasmids deposited to Addgene are provided in Method S2.

RT-PCR and RT-qPCR

RT-PCR and RT-qPCR were performed following manufacturers' manuals. PrimeScript RT Reagent Kit (Takara, RR037) was used for cDNA synthesis. Phusion High-Fidelity DNA Polymerase (NEB, M0530) was used for PCR reactions. SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, 1725271), Hard-Shell 96-Well PCR Plates (Bio-Rad, HSP9601), Plate Sealing Film (Bio-Rad, MSB1001) and Bio-Rad C1000 Touch Thermal Cycler were used for qPCR experiments. Primers are provided in Method S3.

Lentivirus packaging and transduction

Lentivirus was packaged by HEK293T cells. For a 10-cm dish (90% confluency), 1.5 mL Opti-MEM (Gibco, 31985062), 10 µg psPAX2 (Addgene, 12260), 2 µg pMD2.G (Addgene, 12259), 9 µg lentiviral plasmid, and 50 µL TransIT-LT1 Transfection Reagent (Mirus, MIR 2300) were mixed at room temperature for 15 mins and then added to the cells. 3.5 mL DMEM was added to the cells. 72 hrs later, the supernatant in the dish was filtered with 0.45 µm PES filter (Thermo Scientific, 165-0045), mixed with 5 mL Lenti-X Concentrator (Takara, 631232) and rocked at 4 °C overnight. The viral particles were then spun down (1800 ×g, 4 °C, 1hr) and resuspended in 200 µL DMEM. Lentivirus was titrated with Lenti-X GoStix Plus (Takara, 631280). For lentiviral transduction, the cells to be transduced were plated in wells of plates. One day after seeding, the medium was replaced and supplemented with polybrene (final conc. 8 µg/mL). Lentivirus was then added to the wells for a spinfection (800 ×g, 30 °C, 1hr). One day post-transduction, the medium was replaced, and drug selection was started if applicable. For constructs with BSD, Blasticidin S HCl (Gibco, A1113903, final conc. 5 µg/mL) was used for drug selection. For constructs with PuroR, Puromycin Dihydrochloride (Gibco, A1113803, final conc. 1 µg/mL) was used. Drug selection was performed for 10-14 days.

PALS-C cloning for saturation mutagenesis

Each variant of all possible CDS SNVs (Figure S2A) was included in a 64-bp ssDNA oligo. The oligos were synthesized (one pool per GOI) by Twist Bioscience. PALS-C is an 8-step cloning strategy to clone lentiviral plasmid pools from the oligos. An elaborate protocol can be found in Method S4. Briefly, the oligos were used as PCR reverse primers, which were annealed to the plasmid template and extended towards the 5' end of the gene of interest. The resulting products of each block were isolated using block-specific primers. Then the variant strands were extended towards the 3' end to get the full-length sequences, which were subsequently inserted into the plasmid backbone using NEBuilder (NEB, E2621). The purifications for PALS-C steps were done with NucleoSpin Gel and PCR Clean-Up kit (Takara, 740609). Final assembled products were delivered to Endura Electrocompetent Cells (Lucigen, 60242-1) via electrotransformation (Bio-Rad Gene Pulser II). Transformed bacteria were grown overnight and plasmid pools were extracted using the PureLink Midiprep Kit (Invitrogen, K210014). To check library complexity, colony forming units (CFUs) were calculated and a minimum 18 × coverage was achieved for the plasmid pool of each block of *FKRP* and *LARGE1*. Variants that created new type2S enzyme

recognition sites tended to be underrepresented in the pool. These variants are reported in Method S5.

Plasmid pools QC and saturation mutagenesis

Quality control (QC) was performed for the plasmid pools using the Amplicon-EZ service provided by GENEWIZ (Figure S2B and Method S6). 99.6% of the SNVs of both genes were represented in the plasmid pools (Figure S2C). Lentivirus of each block was packaged by HEK293T cells in one 10-cm dish. Small-scale pre-experiments were performed to determine the viral dosage for optimal separation. GoStix Value (GV) quantified by the Lenti-X GoStix App (Takara) was used to scale the titer of each block to be the same. GV is subject to viral-packaging batch effects; hence, lentiviral pools of all blocks were packaged at the same time using the reagents and helper plasmids of the same batch. Depending on the specific batch and packaging system, 1e3-1e5 GV×µL of lentivirus was used for each block. Considerations for controlling the Multiplicity of Infection (MOI) of the lentiviral pool were discussed in Note S1. For each block, 600k HAP1 cells or 200k MB135 cells were plated in a well of a 6-well plate for transduction. The cell number was counted with an Automated Cell Counter (Bio-Rad, TC20). The cell number for each block was expanded to more than 30M for FACS.

Mini-libraries for the FACS assay

Mini-libraries of variants were employed to examine and optimize the separation of the FACS assay. Lentiviral constructs were cloned and packaged individually for 8 *FKRP* variants and 3 *LARGE1* variants in addition to the wild-type constructs. These lentiviral particles were mixed to make a mix-9 *FKRP* mini-library and a mix-4 *LARGE1* mini-library. Conditions of transduction, staining, sorting and gDNA extraction were optimized using the mini-libraries. Relative enrichment of variants was defined as the ratio of the variants' representation in the high-glycosylation sample to their representation in the low-glycosylation group, which was quantified with either Sanger sequencing or Amplicon-EZ NGS (Method S7).

Fluorescence flow cytometry and FACS

Reagent volumes were determined based on sample size. Below, the staining for samples of one gene block is described as an example. The cells were washed twice with DPBS (Gibco, 14190144), digested with Versene (Gibco, 15040066), and counted. 30M cells were used for staining, which was performed in a 15 mL tube. The cells were spun down (700 ×g, 4 °C, 15 mins) and resuspended in 3mL DPBS supplemented with 30 µL Viobility 405/452 Fixable Dye (Miltenyi Biotec, 130-130-420). All the following steps were done in the dark. The sample was gently rocked at room temperature for 30 mins, and then 7 mL PEB buffer (1 volume of MACS BSA Stock Solution, Miltenyi Biotec, 130-091-376 ;19 volumes of autoMACS Rinsing Solution, Miltenyi Biotec, 130-091-222) was added to the tube. The cells were spun down (700 ×g, 4 °C, 15 mins) and resuspended in 3mL DPBS supplemented with 30 µL Human BD Fc Block (BD Pharmingen, 564220). The sample was gently rocked at room temperature for 30 mins, and then 7 mL DPBS was added. The cells were spun down (700 ×g, 4 °C, 10 mins) and resuspended in 3mL MAGIC buffer (5% FBS; 0.1% NaAz w/v; 10% 10× DPBS, Gibco, 14200166; water, Invitrogen, 10977015) supplemented with 15 µL IIH6C4 antibody (Sigma-Aldrich, 05-593, discontinued; or antibody made in Dr. Kevin Campbell's lab). The sample was gently rocked at 4 °C for 20 hrs. 7 mL MAGIC buffer was added before the cells were spun down (700 ×g, 4 °C, 10 mins) and resuspended in 3 mL MAGIC buffer supplemented with 60 µL Rabbit anti-Mouse

IgM FITC Secondary Antibody (Invitrogen, 31557). The sample was gently rocked at 4 °C for 20 hrs. 7 mL DPBS was added to the sample before the cells were spun down (700 ×g, 4 °C, 10 mins), resuspended with 4 mL DPBS and filtered with a 40 µm Cell Strainer (Falcon, 352340). Note: IIH6C4 antibody from Santa Cruz (sc-73586) is not effective for the experiments listed here.

Gating parameters

Fluorescence flow cytometry (FFC) experiments were performed with a BD LSR II Flow Cytometer; FACS experiments were performed with a BD FACSAria Flow Cytometer. Forward scatter (FSC) and side scatter (SSC) were used to exclude cell debris and multiplets. Singlets were isolated for downstream analysis. Pacific Blue (450/50 BP) or an equivalent channel was used to detect the Viobility 405/452 Fixable Dye and isolate the live cells for analysis. FITC (530/30 BP), GFP (510/20 BP) or an equivalent channel was used to detect the FITC secondary antibody signal. 20k events were recorded for each block to decide the gating parameters. For FACS, the top ~20% of the cells were isolated as the high-glycosylation group and the bottom ~40% of the cells were isolated as the low-glycosylation group. The .fcs files, the FlowJo .wsp files and the software interface reports of the sorter were made available on FlowRepository. A minimum ~1000 × coverage (*e.g.*, 750k cells harvested for a block with 750 variants) was achieved for both groups of each block.

NGS library construction

The cells were spun down (800 ×g, 4 °C, 10 mins), and gDNA was harvested from each sample with PureLink Genomic DNA Mini Kit (Invitrogen, K182002). A 3-step PCR library construction was performed to build the sequencing library. Step 1: lentiviral sequence isolation. A pair of primers specific to the lentiviral backbone was used to amplify the lentiviral CDS sequences of each sample. Step 2: block isolation. Each primer contained a 20-bp flanking sequence of the specific block and a partial Illumina adaptor sequence. Forward primers contained the barcodes to distinguish the high-glycosylation group and the low-glycosylation group. Step 3: adaptor addition. Step 2 products were multiplexed and the rest of the Illumina adaptor was added to the amplicons. An elaborate protocol is provided in Method S8. The NGS libraries were sequenced using Psomagen's HiSeq X service. ~400M reads were acquired per library.

Variant counting and scoring

Raw sequencing data was processed with the “Gargamel” pipeline. Variant counting was performed after the data cleaning process that essentially removed the reads with more than one variant. Enrichment of a variant (E_var) in a FACS group is calculated as a ratio of the count of the variant (c_var) to the total count (c_total) at the variant site:

$$E_var = c_var / c_total$$

Enrichment of the WT (E_WT) is calculated separately for each block. E_WT is calculated as a ratio of the number of the reads without variant (r_WT) to the number of the reads with one or no variant (r_clean).

$$E_WT = r_WT / r_clean$$

Relative enrichment (rE) is a ratio of the enrichment in the high-glycosylation group to the enrichment in the low-glycosylation group:

$$rE_var = E_var_high/E_var_low$$
$$rE_WT = E_WT_high/E_WT_low$$

The functional score of a variant (in one biological replicate) is calculated as the ratio of its relative enrichment to that of the WT sequence in the corresponding block:

$$Functional_score = rE_var/rE_WT$$

Count of variants and reads were generated from raw sequencing data using the analytical pipeline deposited in the GitHub repository Gargamel. SMuRF score is generated by combing 3 biological replicates (see below).

Fitness score combination using DiMSum

The DiMSum pipeline was employed to process the variant counts to generate a fitness score across all biological replicates for *FKRP* and *LARGE1*. For data preprocessing, the input files were prepared by block to account for block-specific WT variant counts. Variants with a count of 0 were replaced by the WT value. The variant counts for each replicate within each gene block were processed through DiMSum (https://github.com/lehner-lab/DiMSum/) from the STEAM stage. Default settings were used for the pipeline, aside from *--sequenceType noncoding* to calculate the scores on a variant-level instead of AA-level. The pipeline aggregated these counts to generate a fitness score for each variant. DiMSum incorporates both the frequency of each variant and the observed effects on fitness across all replicates and blocks, producing a merged fitness score for *FKRP* and *LARGE1*. To ensure accurate and reliable fitness estimates, DiMSum employs an advanced error modeling approach. The model addresses multiple sources of error to produce robust fitness scores. Firstly, Poisson-based errors are used to account for the high variance in fitness estimates that arise from low sequencing depth. Since count data often exhibit over-dispersion relative to a Poisson distribution, DiMSum integrates both multiplicative and additive error terms to model this variance effectively. Multiplicative errors, which scale with the variant sequencing counts, address variability arising from workflow-related inconsistencies. In contrast, additive errors are independent of sequencing counts and affect all variants uniformly, accounting for differential handling across replicates. These error components are summarized into a variable, sigma (σ), which is used to adjust fitness estimates on a variant-level. The primary output from the DiMSum pipeline is the 1) "DiMSum score" – fitness score from the direct output of the DiMSum pipeline, converted to the log2 scale, and 2) sigma – a numerical value representing the reliability of the variant count based on error modeling.

Fitness score normalization

To normalize the fitness scores, we used all synonymous variants as the reference, assuming these variants should have no impact and thus a fitness score of 1. We first transformed the fitness scores of the synonymous variants from their natural logarithm form to a linear scale using the exponential function. For each block, we calculated the mean fitness score of the synonymous variants, treating this mean as the scaling factor. Next, we normalized the fitness scores of all variants by transforming their scores back to the linear scale, dividing by the block's scaling factor, and then applying a log2 transformation to the normalized values to get the

"SMuRF score". This approach ensures that the scores are adjusted relative to the neutral impact of synonymous variants, allowing for consistent comparisons across different blocks.

Confidence score generation and classification

Following the DiMSum pipeline for variant score combination and error modeling, each identified variant was assigned a sigma (σ) value representing the estimated error. A higher sigma indicates a higher degree of error in the variant assessment. To classify variants into high and low confidence groups, we applied a threshold based on the 1.5 * interquartile range (IQR) rule, a commonly used outlier detection method. First, we calculated the first quartile (Q1) and third quartile (Q3) of the sigma distribution across all variants. The interquartile range (IQR) was then determined as IQR = Q3 - Q1. The upper threshold for high confidence variants was set at Q3 + 1.5*IQR. Any variant with a sigma value exceeding this threshold was classified as low confidence, while variants below the threshold were designated as high confidence. For all downstream analyses, low confidence variants were excluded to ensure the robustness of results. This approach provided an objective method to filter out potentially unreliable sites prior to further examination and interpretation of the data.

ClinVar classification

ClinVar classification of the variants was downloaded from the ClinVar webpage through its "Download" option as tabular files. The data were restructured using customized scripts to facilitate the relevant analyses. The version we downloaded was from April 20th, 2023. The ClinVar classification of the variants was included in Tables S2 and S3.

Comparison of SMuRF score to *in silico* predictors

To compare the performance of SMuRF with various computational predictors, we conducted a receiver operating characteristic (ROC) curve analysis. This analysis assesses the ability of each method to identify pathogenic variants, using the ClinVar annotations as the reference standard. For *FKRP* and *LARGE1*, we used a dataset containing annotated missense variants with functional scores from SMuRF and computational predictors including CADD, metaSVM, REVEL, MVP, MutScore, PrimateAI-3D, EVE, ESM1b, and MAVERICK. First, we subsetted the data to include only missense variants. We then further restricted the dataset to include only those variants scored by all computational predictors to ensure a fair comparison. The positive reference set comprised missense variants classified as pathogenic (P), likely pathogenic (LP), or pathogenic/likely pathogenic (P/LP) in ClinVar. The negative reference set consisted of missense variants that are observed in gnomAD v4 genomes and not classified as P, LP, or P/LP in ClinVar. To perform the ROC curve analysis, we employed the pROC package in R. First, we created a binary response variable indicating whether each variant is part of the positive reference set or the negative reference set. Using the roc function from the pROC package, we computed ROC curves for each tool. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The area under the curve (AUC) value was computed for each tool, providing a quantitative measure of their discriminatory ability.

Correlation of SMuRF scores with disease onset

To investigate the correlation between SMuRF scores and disease onset, a Cox proportional hazards model was employed. The dataset included SMuRF scores for biallelic variants and the

corresponding ages of disease onset. Entries with missing values for these variables were excluded. The combined functional score for each patient was calculated as logarithm base 2 of the sum of the exponentiated individual SMuRF scores for each of the biallelic variants.

The age of onset was used as a quantitative variable in months. An event indicator variable was defined to distinguish between disease onset at birth (0 months) and later onset (>0 months). This binary event indicator was coded as 1 for non-zero onset months and 0 for onset at birth.

The Cox proportional hazards model was applied using the *coxph* function from the *survival* package in R. The model evaluated the association between the combined functional score and the age of disease onset. In this analysis, the age of onset was treated as the time variable, the occurrence of the event (disease onset) was treated as a binary outcome variable, and the combined functional score (SMuRF score) was the predictor variable.

After fitting the model, the summary statistics were extracted and interpreted to assess the significance and strength of the association between the functional scores and disease onset. This included evaluating the hazard ratio and its confidence intervals, as well as the p-value for the predictor variable. These results provided insight into how the combined functional score influenced the likelihood of earlier disease onset, offering a quantitative measure of its predictive capacity.

Immunoblotting for LARGE1 enzymatic activity

Cell pellets expressing LARGE1 variants were resuspended in 50 mM tris pH 7.6, 150 mM NaCl, 1% Triton X-100 containing protease inhibitors (PMSF, benzamide, aprotinin, leupeptin and pepstatin A) and rotated at 4℃ for 1 hr. The cleared lysate was incubated with WGA resin (Vector laboratories) for 1 hr at 4℃ with rotation. The resin was washed two times with the same buffer with reduced triton X-100 (0.1%). The third wash was completed in matriglycan reaction buffer (20 mM PIPES pH 6.6, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$ and 2 mM $MnCl_2$). An equivalent volume as resin of a reaction master mix to a final concentration of 1 mM UDP-xylose (Sussex Research Laboratories), 1 mM UDP-glucuronate (Sigma-Aldrich), and 0.4 mM 4-methylumbelliferone-glucuronate-xylose (Sussex Research Laboratories) in reaction buffer containing protease inhibitors was added. The mixture was incubated at 37℃ for 72 hrs. The reaction was terminated with 50 mM EDTA. The cleared supernatant was resolved on a strong anion exchange column using a gradient of increasing anionic strength up to 500 mM NaCl over a background of 50 mM ammonium phosphate pH 6.0. The resin was combined with SDS loading buffer and resolved on a homemade 3-15% gradient SDS-PAGE at 60 V for 16 hrs at room temperature. Proteins were stained with Bio-Safe Coomassie G-250 Stain according to the manufacturer's instructions (Bio-Rad) and imaged on the Li-COR. The proteins were transferred to PVDF membrane (800 mA for 5 hrs at 4℃), blocked with 2% milk in 50 mM tris pH 7.6 and 75 mM NaCl (low salt-TBS) and incubated with anti-matriglycan primary antibody (IIH6C4) overnight at 4℃ with rocking. The membrane was washed three times with low salt-TBS prior to incubation with anti-mouse IgM secondary antibody for 30 minutes at room temperature with rocking. The membrane was washed three times and imaged on a Li-COR. Another PVDF membrane was incubated with 5% milk dissolved in LBB buffer (10 mM triethanolamine, 140 mM NaCl, 1 mM $CaCl_2$ and 1 mM $MgCl_2$). The membrane was washed three times with 3% BSA in LBB buffer and then incubated with 7.5 nM mouse laminin (Lam-111) dissolved in LBB

overnight at 4℃. The membrane was washed three times with LBB and then incubated with anti-laminin antibody dissolved in 5% milk-LBB overnight at 4℃. The membrane was washed three times with LBB and incubated with anti-rabbit IgG for 30 minutes at room temperature. The membrane was washed three times and imaged on a Li-COR.

Visualization of data on protein 3D structures

Protein 3D structures from Protein Data Bank (PDB) were visualized using UCSF ChimeraX v1.3[114]. The crystal structure of human FKRP (PDB:6KAM)[85] and the electron microscopy structure of LARGE1 (PDB:7UI7)[86] were used. Figures displaying domain locations or the log2 mean missense score per residue on 3D structures were generated using custom ChimeraX command files. Domain coordinates displayed are per Ortiz-Cordero *et al*. for FKRP[87], and Joseph *et al*. for LARGE1[86]. The log2 of the mean missense score per residue was calculated using a custom Python script.

Immunofluorescence

15 mm round Thermanox Coverslips (Thermo Scientific, 174969) were placed in the wells of 24-well plates. To coat the coverslips, 0.1% gelatin (Sigma-Aldrich, G9391) was added to the wells and immediately removed. After the coverslips were air-dried, 250k MB135 cells were resuspended in 0.5 mL growth medium and seeded into each well. One day after plating the cells, the medium was changed to the differentiation medium, and cells were differentiated for 3-7 days until mature myotubes were formed. The cells were washed with DPBS and fixed with 4% PFA (Sigma-Aldrich, 158127) for 10 mins at room temperature. The cells were blocked with 2% Bovine Serum Albumin (BSA, Sigma-Aldrich, A9647) at room temperature for 1 hr before undergoing incubation with the IIH6C4 antibody (1:200 in 2% BSA, Sigma-Aldrich, 05-593, discontinued) at 4 °C for 20 hrs. The cells were then washed with DPBS before undergoing incubation with the secondary antibody (1:100 in 2% BSA, Invitrogen, 31557) at room temperature for 2 hrs in the dark. Antifade Mounting Medium with DAPI (Vector Laboratories, H1500) was dropped onto Microscope Slides (Fisher Scientific, 22-037-246). The coverslips with cells were washed again with DPBS, put facedown onto the drops on the slides, and kept at room temperature for 30 mins in the dark. Pictures were taken with a Revolve ECHO microscope. Note: IIH6C4 antibody from Santa Cruz (sc-73586) is not effective for the experiments listed here.

Packaging and infection of rVSV / ppVSV

rVSV-LASV-GPC viral particles, ppVSVΔG-VSV-G viral particles, and the LASV-GPC plasmid were obtained from Dr. Melinda Brindley. To package ppVSV-LASV-GPC viral particles, HEK293T cells were transfected with the LASV-GPC plasmid and then transduced with ppVSVΔG-VSV-G viral particles. The resulting particles were referred to as ppVSV-LASV-GPC-Generation1. A new batch of LASV-GPC transfected HEK293T cells were subsequently transduced with ppVSV-LASV-GPC-Generation1 to produce ppVSV-LASV-GPC-Generation2, reducing residual VSV-G in the pseudotyped particles. Similarly, later generations can be packaged. The experiments in this study utilized ppVSV-LASV-GPC-Generation2 and 3. The 50% tissue culture infectious dose (TCID50) of the VSV was determined using the Spearman-Karber method[115,116]. Lentiviral transduction and blasticidin drug selection were performed in the same manner as those in the FACS assay. Afterward, cells were divided into two groups (~1M cells each): a no-infection group and an infection group. rVSV infection was

conducted at an approximate MOI of 0.5. NH4Cl (Sigma-Aldrich, A9434, final conc. 5mM) was added during the infection and subsequent recovery. After 60 hours of infection, the medium was replaced, and the cells were allowed to recover for 12 hours before harvesting. ppVSV infection was performed at an approximate MOI of 1~3, and the infected cells were recovered to ~1M prior to harvesting. Detailed ppVSV packaging protocol and considerations for application are provided in Method S10.

Validation using the ppVSV assay

10 Lenti-*FKRP* variants were mixed with Lenti-WT-*FKRP* to rescue *FKRP*-KO MB135. 11 Lenti-*LARGE1* variants were mixed with Lenti-WT-*LARGE1* to rescue *LARGE1*-KO MB135. The lentiviral plasmids were individually constructed and validated using the whole-plasmid sequencing service provided by Plasmidsaurus. The individual plasmids were mixed to package the *FKRP* mix11 and *LARGE* mix12 lentiviral pools. Lentiviral transduction, drug selection, and ppVSV infection were performed similarly as stated above. After harvesting the cells, the lentiviral GOI sequence was amplified with the PCR1 protocol in Method S8 and sequenced with nanopore sequencing provided by Plasmidsaurus. The functional score was quantified by the ratio of a variant's enrichment in the non-infected group to its enrichment in the ppVSV-infected group. The list of the variants included in the mini-libraries and other details are provided in Method S11.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis

Two-sided Wilcoxon tests were performed with the "ggsignif" R package. Spearman's rank correlation coefficients were calculated with the "cor.test" function in R. Multiple comparisons between groups were performed using analysis of variance (ANOVA) followed by Bonferroni post hoc test through GraphPad Prism 10.2.2.

ADDITIONAL RESOURCES

Detailed experimental protocols: https://www.protocols.io/view/using-saturation-mutagenesis-reinforced-functional-cwvvxe66/materials.

Acknowledgments

We thank Sander Pajusalu and Heather Best for their efforts which helped lead to this project. We thank Nicholas Johnson and the GRASP LGMD consortium for granting us access to their deidentified patient data, which was funded by ML Bio Solutions. We thank Stephen Tapscott for the gift of MB135 cells. This work is supported by a grant to M.L. from the Muscular Dystrophy Association (629095) for "improved clinical interpretation of rare variants in muscle

diseases". K.P.C. is an investigator of the Howard Hughes Medical Institute. This work was supported in part by a Paul D. Wellstone Muscular Dystrophy Specialized Research Center grant (1U54NS053672 to K.P.C.). M.A.B. was supported by the National Institute of Allergy And Infectious Diseases of the National Institutes of Health under Award Number R01AI139238 (MAB). S.H. is supported by a Muscular Dystrophy Association Development Grant (MDA 963708). J.C. is supported by The Chris Carrino Foundation for FSHD and NIH/5F32AR079892-02. We thank Yale Flow Cytometry for their assistance with FFC and FACS services. The Core is supported in part by an NCI Cancer Center Support Grant # NIH P30 CA016359.

Author contributions

K.M. and M.L. conceived and designed this study. M.L. supervised the experiments and analyses of this study. K.M. designed and performed the experiments for the establishment, application, and validation of SMuRF, with the help of S.H. and other authors. K.K.N., K.M., and M.L. developed the analytical pipeline. S.H. and K.M. performed the IF experiments. S.J. conducted the matriglycan-related validation experiments. K.M. and A.L. created the lentiviral plasmid pools. K.G.W. and K.E.K. participated in the early development of this study. V.H. and C.L.O'C. assisted in performing the experiments. S.J. and K.P.C. participated in the analysis of IIH6C4-related experiments. M.A.B. participated in designing and interpreting the VSV experiments. K.M., K.K.N., N.J.L. and M.L. performed computational analyses, with the help of J.X. and L.G.. K.M., K.K.N., N.J.L., J.X., J.C., and M.L. wrote this manuscript with the help of other authors.

Declaration of interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies

We used ChatGPT 3.5 and Gemini to improve the readability and language in this manuscript. The manuscript was first drafted by us and polished with the tools sentence-by-sentence where we deemed necessary. We then reviewed and finalized the text. We take full responsibility for the contents of this manuscript.

Supplemental information

Document S1. Figures S1–S7, Table S6 and Note S1

Table S1. Datasets collected for the generation of SMuRF scores.

Table S2. SMuRF scores and other scores used in this manuscript for *FKRP*.

Table S3. SMuRF scores and other scores used in this manuscript for *LARGE1*.

Table S4. Restructured data from 8 well-curated FKRP cohorts, related to Figure 3C and 3D.

Table S5. Variant inclusion details for the generation of the ROC curves, related to Figure 4A and 4B.

References

1. Fridman, H., Yntema, H.G., Mägi, R., Andreson, R., Metspalu, A., Mezzavila, M., Tyler-Smith, C., Xue, Y., Carmi, S., Levy-Lahad, E., et al. (2021). The landscape of autosomal-recessive pathogenic variants in European populations reveals phenotype-specific effects. Am. J. Hum. Genet. *108*, 608–619. 10.1016/j.ajhg.2021.03.004.

2. Balick, D.J., Jordan, D.M., Sunyaev, S., and Do, R. (2022). Overcoming constraints on the detection of recessive selection in human genes from population frequency data. Am. J. Hum. Genet. *109*, 33–49. 10.1016/j.ajhg.2021.12.001.

3. Barton, A.R., Hujoel, M.L.A., Mukamel, R.E., Sherman, M.A., and Loh, P.-R. (2022). A spectrum of recessiveness among Mendelian disease variants in UK Biobank. Am. J. Hum. Genet. *109*, 1298–1307. 10.1016/j.ajhg.2022.05.008.

4. Schmenger, T., Diwan, G.D., Singh, G., Apic, G., and Russell, R.B. (2022). Never-homozygous genetic variants in healthy populations are potential recessive disease candidates. NPJ Genom. Med. *7*, 54. 10.1038/s41525-022-00322-z.

5. Boycott, K.M., Hartley, T., Biesecker, L.G., Gibbs, R.A., Innes, A.M., Riess, O., Belmont, J., Dunwoodie, S.L., Jojic, N., Lassmann, T., et al. (2019). A diagnosis for all rare genetic diseases: the horizon and the next frontiers. Cell *177*, 32–37. 10.1016/j.cell.2019.02.040.

6. 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. 10.1038/s41586-020-2308-7.

7. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. 10.1038/nature19057.

8. Blöß, S., Klemann, C., Rother, A.-K., Mehmecke, S., Schumacher, U., Mücke, U., Mücke, M., Stieber, C., Klawonn, F., Kortum, X., et al. (2017). Diagnostic needs for rare diseases and shared prediagnostic phenomena: Results of a German-wide expert Delphi survey. PLoS ONE *12*, e0172532. 10.1371/journal.pone.0172532.

9. Libell, E.M., Bowdler, N.C., Stephan, C.M., Zimmerman, M.B., Gedlinske, A.M., and Mathews, K.D. (2021). The outcomes and experience of pregnancy in limb girdle muscular dystrophy type R9. Muscle Nerve *63*, 812–817. 10.1002/mus.27184.

10. 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. 10.1016/j.ajhg.2017.07.014.

11. Caswell-Jin, J.L., Gupta, T., Hall, E., Petrovchich, I.M., Mills, M.A., Kingham, K.E., Koff, R., Chun, N.M., Levonian, P., Lebensohn, A.P., et al. (2018). Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. Genet. Med. *20*, 234–239. 10.1038/gim.2017.96.

12. Livesey, B.J., and Marsh, J.A. (2020). Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. Mol. Syst. Biol. *16*, e9380. 10.15252/msb.20199380.

13. Wei, H., and Li, X. (2023). Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. Front. Genet. *14*, 1087267. 10.3389/fgene.2023.1087267.

14. Fowler, D.M., Adams, D.J., Gloyn, A.L., Hahn, W.C., Marks, D.S., Muffley, L.A., Neal, J.T., Roth, F.P., Rubin, A.F., Starita, L.M., et al. (2023). An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. *24*, 147. 10.1186/s13059-023-02986-x.

15. Majithia, A.R., Tsuda, B., Agostini, M., Gnanapradeepan, K., Rice, R., Peloso, G., Patel, K.A., Zhang, X., Broekema, M.F., Patterson, N., et al. (2016). Prospective functional classification of all possible missense variants in PPARG. Nat. Genet. *48*, 1570–1575. 10.1038/ng.3700.

16. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature *513*, 120–123. 10.1038/nature13695.

17. Haller, G., Alvarado, D., McCall, K., Mitra, R.D., Dobbs, M.B., and Gurnett, C.A. (2016). Massively parallel single-nucleotide mutagenesis using reversibly terminated inosine. Nat. Methods *13*, 923–924. 10.1038/nmeth.4015.

18. Erwood, S., Bily, T.M.I., Lequyer, J., Yan, J., Gulati, N., Brewer, R.A., Zhou, L., Pelletier, L., Ivakine, E.A., and Cohn, R.D. (2022). Saturation variant interpretation using CRISPR prime editing. Nat. Biotechnol. *40*, 885–895. 10.1038/s41587-021-01201-1.

19. Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S., and Shendure, J. (2015). Massively parallel single-amino-acid mutagenesis. Nat. Methods *12*, 203–206, 4 p following 206. 10.1038/nmeth.3223.

20. Jia, X., Burugula, B.B., Chen, V., Lemons, R.M., Jayakody, S., Maksutova, M., and Kitzman, J.O. (2021). Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. Am. J. Hum. Genet. *108*, 163–175. 10.1016/j.ajhg.2020.12.003.

21. Findlay, G.M. (2021). Linking genome variants to disease: scalable approaches to test the functional impact of human mutations. Hum. Mol. Genet. *30*, R187–R197. 10.1093/hmg/ddab219.

22. Tucker, J.D., Lu, P.J., Xiao, X., and Lu, Q.L. (2018). Overexpression of mutant FKRP restores functional glycosylation and improves dystrophic phenotype in FKRP mutant mice. Mol. Ther. Nucleic Acids *11*, 216–227. 10.1016/j.omtn.2018.02.008.

23. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol. *18*, 150. 10.1186/s13059-017-1272-5.

24. Meilleur, K.G., Zukosky, K., Medne, L., Fequiere, P., Powell-Hamilton, N., Winder, T.L., Alsaman, A., El-Hattab, A.W., Dastgir, J., Hu, Y., et al. (2014). Clinical, pathologic, and mutational spectrum of dystroglycanopathy caused by LARGE mutations. J. Neuropathol. Exp. Neurol. *73*, 425–441. 10.1097/NEN.0000000000000065.

25. Kanagawa, M. (2021). Dystroglycanopathy: from elucidation of molecular and pathological mechanisms to development of treatment methods. Int. J. Mol. Sci. *22*. 10.3390/ijms222313162.

26. Borisovna, K.O., Yurievna, K.A., Yurievich, T.K., Igorevna, K.O., Olegovich, K.D., Igorevna, D.A., Timofeevna, B.T., Vyacheslavovna, Z.N., Ivanovna, S.E., Alekseevich, S.P., et al. (2019). Compound heterozygous POMGNT1 mutations leading to muscular dystrophy-dystroglycanopathy type A3: a case report. BMC Pediatr. *19*, 98. 10.1186/s12887-019-1470-2.

27. Geis, T., Rödl, T., Topaloğlu, H., Balci-Hayta, B., Hinreiner, S., Müller-Felber, W., Schoser, B., Mehraein, Y., Hübner, A., Zirn, B., et al. (2019). Clinical long-time course, novel mutations and genotype-phenotype correlation in a cohort of 27 families with POMT1-related disorders. Orphanet J. Rare Dis. *14*, 179. 10.1186/s13023-019-1119-0.

28. Van Reeuwijk, J., Olderode-Berends, M.J.W., Van den Elzen, C., Brouwer, O.F., Roscioli, T., Van Pampus, M.G., Scheffer, H., Brunner, H.G., Van Bokhoven, H., and Hol, F.A. (2010). A homozygous FKRP start codon mutation is associated with Walker-Warburg syndrome, the severe end of the clinical spectrum. Clin. Genet. *78*, 275–281. 10.1111/j.1399-0004.2010.01384.x.

29. Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.-Z., and Lahn, B.T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS ONE *5*, e10611. 10.1371/journal.pone.0010611.

30. Johnson, K., Bertoli, M., Phillips, L., Töpf, A., Van den Bergh, P., Vissing, J., Witting, N., Nafissi, S., Jamal-Omidi, S., Łusakowska, A., et al. (2018). Detection of variants in dystroglycanopathy-associated genes through the application of targeted whole-exome sequencing analysis to a large cohort of patients with unexplained limb-girdle muscle weakness. Skelet. Muscle *8*, 23. 10.1186/s13395-018-0170-1.

31. Kanagawa, M., Kobayashi, K., Tajiri, M., Manya, H., Kuga, A., Yamaguchi, Y., Akasaka-Manya, K., Furukawa, J.-I., Mizuno, M., Kawakami, H., et al. (2016). Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. Cell Rep. *14*, 2209–2223. 10.1016/j.celrep.2016.02.017.

32. Goddeeris, M.M., Wu, B., Venzke, D., Yoshida-Moriguchi, T., Saito, F., Matsumura, K., Moore, S.A., and Campbell, K.P. (2013). LARGE glycans on dystroglycan function as a tunable matrix scaffold to prevent dystrophy. Nature *503*, 136–140. 10.1038/nature12605.

33. Wu, B., Drains, M., Shah, S.N., Lu, P.J., Leroy, V., Killilee, J., Rawls, R., Tucker, J.D., Blaeser, A., and Lu, Q.L. (2022). Ribitol dose-dependently enhances matriglycan expression and improves muscle function with prolonged life span in limb girdle muscular dystrophy 2I mouse model. PLoS ONE *17*, e0278482. 10.1371/journal.pone.0278482.

34. Vannoy, C.H., Leroy, V., and Lu, Q.L. (2018). Dose-Dependent Effects of FKRP Gene-Replacement Therapy on Functional Rescue and Longevity in Dystrophic Mice. Mol. Ther. Methods Clin. Dev. *11*, 106–120. 10.1016/j.omtm.2018.10.004.

35. Yonekawa, T., Rauckhorst, A.J., El-Hattab, S., Cuellar, M.A., Venzke, D., Anderson, M.E., Okuma, H., Pewa, A.D., Taylor, E.B., and Campbell, K.P. (2022). Large1 gene transfer in older myd mice with severe muscular dystrophy restores muscle function and greatly improves survival. Sci. Adv. *8*, eabn0379. 10.1126/sciadv.abn0379.

36. Dhoke, N.R., Kim, H., Selvaraj, S., Azzag, K., Zhou, H., Oliveira, N.A.J., Tungtur, S., Ortiz-Cordero, C., Kiley, J., Lu, Q.L., et al. (2021). A universal gene correction approach for FKRP-associated dystroglycanopathies to enable autologous cell therapy. Cell Rep. *36*, 109360. 10.1016/j.celrep.2021.109360.

37. Yoshida-Moriguchi, T., and Campbell, K.P. (2015). Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. Glycobiology *25*, 702–713. 10.1093/glycob/cwv021.

38. Panicucci, C., Fiorillo, C., Moro, F., Astrea, G., Brisca, G., Trucco, F., Pedemonte, M., Lanteri, P., Sciarretta, L., Minetti, C., et al. (2018). Mutations in GMPPB Presenting with Pseudometabolic Myopathy. JIMD Rep. *38*, 23–31. 10.1007/8904_2017_25.

39. Barone, R., Aiello, C., Race, V., Morava, E., Foulquier, F., Riemersma, M., Passarelli, C., Concolino, D., Carella, M., Santorelli, F., et al. (2012). DPM2-CDG: a muscular dystrophy-dystroglycanopathy syndrome with severe epilepsy. Ann. Neurol. *72*, 550–558. 10.1002/ana.23632.

40. Lefeber, D.J., de Brouwer, A.P.M., Morava, E., Riemersma, M., Schuurs-Hoeijmakers, J.H.M., Absmanner, B., Verrijp, K., van den Akker, W.M.R., Huijben, K., Steenbergen, G., et al. (2011). Autosomal recessive dilated cardiomyopathy due to DOLK mutations results from abnormal dystroglycan O-mannosylation. PLoS Genet. *7*, e1002427. 10.1371/journal.pgen.1002427.

41. Hu, H., Li, J., Gagen, C.S., Gray, N.W., Zhang, Z., Qi, Y., and Zhang, P. (2011). Conditional knockout of protein O-mannosyltransferase 2 reveals tissue-specific roles of O-mannosyl glycosylation in brain development. J. Comp. Neurol. *519*, 1320–1337. 10.1002/cne.22572.

42. Walimbe, A.S., Okuma, H., Joseph, S., Yang, T., Yonekawa, T., Hord, J.M., Venzke, D., Anderson, M.E., Torelli, S., Manzur, A., et al. (2020). POMK regulates dystroglycan function via LARGE1-mediated elongation of matriglycan. eLife *9*. 10.7554/eLife.61388.

43. Endo, Y., Dong, M., Noguchi, S., Ogawa, M., Hayashi, Y.K., Kuru, S., Sugiyama, K., Nagai, S., Ozasa, S., Nonaka, I., et al. (2015). Milder forms of muscular dystrophy associated with POMGNT2 mutations. Neurol. Genet. *1*, e33. 10.1212/NXG.0000000000000033.

44. Stevens, E., Carss, K.J., Cirak, S., Foley, A.R., Torelli, S., Willer, T., Tambunan, D.E., Yau, S., Brodd, L., Sewry, C.A., et al. (2013). Mutations in B3GALNT2 cause congenital muscular dystrophy and hypoglycosylation of α-dystroglycan. Am. J. Hum. Genet. *92*, 354–365. 10.1016/j.ajhg.2013.01.016.

45. Willer, T., Lee, H., Lommel, M., Yoshida-Moriguchi, T., de Bernabe, D.B.V., Venzke, D., Cirak, S., Schachter, H., Vajsar, J., Voit, T., et al. (2012). ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. Nat. Genet. *44*, 575–580. 10.1038/ng.2252.

46. Ujihara, Y., Kanagawa, M., Mohri, S., Takatsu, S., Kobayashi, K., Toda, T., Naruse, K., and Katanosaka, Y. (2019). Elimination of fukutin reveals cellular and molecular pathomechanisms in muscular dystrophy-associated heart failure. Nat. Commun. *10*, 5754. 10.1038/s41467-019-13623-2.

47. Lee, A.J., Jones, K.A., Butterfield, R.J., Cox, M.O., Konersman, C.G., Grosmann, C., Abdenur, J.E., Boyer, M., Beson, B., Wang, C., et al. (2019). Clinical, genetic, and pathologic characterization of FKRP Mexican founder mutation c.1387A>G. Neurol. Genet. *5*, e315. 10.1212/NXG.0000000000000315.

48. Manya, H., Yamaguchi, Y., Kanagawa, M., Kobayashi, K., Tajiri, M., Akasaka-Manya, K., Kawakami, H., Mizuno, M., Wada, Y., Toda, T., et al. (2016). The Muscular Dystrophy Gene TMEM5 Encodes a Ribitol β1,4-Xylosyltransferase Required for the Functional Glycosylation of Dystroglycan. J. Biol. Chem. *291*, 24618–24627. 10.1074/jbc.M116.751917.

49. Willer, T., Inamori, K.-I., Venzke, D., Harvey, C., Morgensen, G., Hara, Y., Beltrán Valero de Bernabé, D., Yu, L., Wright, K.M., and Campbell, K.P. (2014). The glucuronyltransferase B4GAT1 is required for initiation of LARGE-mediated α-dystroglycan functional glycosylation. eLife *3*. 10.7554/eLife.03941.

50. Longman, C., Brockington, M., Torelli, S., Jimenez-Mallebrera, C., Kennedy, C., Khalil, N., Feng, L., Saran, R.K., Voit, T., Merlini, L., et al. (2003). Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of alpha-dystroglycan. Hum. Mol. Genet. *12*, 2853–2861. 10.1093/hmg/ddg307.

51. Stevens, E., Torelli, S., Feng, L., Phadke, R., Walter, M.C., Schneiderat, P., Eddaoudi, A., Sewry, C.A., and Muntoni, F. (2013). Flow cytometry for the analysis of α-dystroglycan glycosylation in fibroblasts from patients with dystroglycanopathies. PLoS ONE *8*, e68958. 10.1371/journal.pone.0068958.

52. Jae, L.T., Raaben, M., Riemersma, M., van Beusekom, E., Blomen, V.A., Velds, A., Kerkhoven, R.M., Carette, J.E., Topaloglu, H., Meinecke, P., et al. (2013). Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. Science *340*, 479–483. 10.1126/science.1233675.

53. Sheikh, M.O., Capicciotti, C.J., Liu, L., Praissman, J., Ding, D., Mead, D.G., Brindley, M.A., Willer, T., Campbell, K.P., Moremen, K.W., et al. (2022). Cell surface glycan engineering reveals that matriglycan alone can recapitulate dystroglycan binding and function. Nat. Commun. *13*, 3617. 10.1038/s41467-022-31205-7.

54. Beigl, T.B., Kjosås, I., Seljeseth, E., Glomnes, N., and Aksnes, H. (2020). Efficient and crucial quality control of HAP1 cell ploidy status. Biol. Open *9*. 10.1242/bio.057174.

55. Song, D., Dai, Y., Chen, X., Fu, X., Chang, X., Wang, N., Zhang, C., Yan, C., Zheng, H., Wu, L., et al. (2021). Genetic variations and clinical spectrum of dystroglycanopathy in a large cohort of Chinese patients. Clin. Genet. *99*, 384–395. 10.1111/cge.13886.

56. Weile, J., Sun, S., Cote, A.G., Knapp, J., Verby, M., Mellor, J.C., Wu, Y., Pons, C., Wong, C., van Lieshout, N., et al. (2017). A framework for exhaustively mapping functional missense variants. Mol. Syst. Biol. *13*, 957. 10.15252/msb.20177908.

57. Fowler, D.M., Stephany, J.J., and Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. *9*, 2267–2284. 10.1038/nprot.2014.153.

58. Faure, A.J., Schmiedel, J.M., Baeza-Centurion, P., and Lehner, B. (2020). DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. Genome Biol. *21*, 207. 10.1186/s13059-020-02091-3.

59. Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat. Genet. *48*, 1112–1118. 10.1038/ng.3664.

60. Benitez-Cantos, M.S., Yordanova, M.M., O'Connor, P.B.F., Zhdanov, A.V., Kovalchuk, S.I., Papkovsky, D.B., Andreev, D.E., and Baranov, P.V. (2020). Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context. Genome Res. *30*, 974–984. 10.1101/gr.257352.119.

61. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2022). A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. BioRxiv. 10.1101/2022.03.20.485034.

62. Bigotti, M.G., and Brancaccio, A. (2021). High degree of conservation of the enzymes synthesizing the laminin-binding glycoepitope of α-dystroglycan. Open Biol. *11*, 210104. 10.1098/rsob.210104.

63. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121. 10.1101/gr.097857.109.

64. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980-5. 10.1093/nar/gkt1113.

65. Sun, S., Weile, J., Verby, M., Wu, Y., Wang, Y., Cote, A.G., Fotiadou, I., Kitaygorodsky, J., Vidal, M., Rine, J., et al. (2020). A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. Genome Med. *12*, 13. 10.1186/s13073-020-0711-1.

66. Li, C., Wilborn, J., Pittman, S., Daw, J., Alonso-Pérez, J., Díaz-Manera, J., Weihl, C.C., and Haller, G. (2023). Comprehensive functional characterization of *SGCB* coding variants predicts pathogenicity in limb-girdle muscular dystrophy type R4/2E. The Journal of Clinical Investigation.

67. Awano, H., Saito, Y., Shimizu, M., Sekiguchi, K., Niijima, S., Matsuo, M., Maegaki, Y., Izumi, I., Kikuchi, C., Ishibashi, M., et al. (2021). FKRP mutations cause congenital muscular dystrophy 1C and limb-girdle muscular dystrophy 2I in Asian patients. J. Clin. Neurosci. *92*, 215–221. 10.1016/j.jocn.2021.08.014.

68. Unnikrishnan, G., Polavarapu, K., Bardhan, M., Nashi, S., Vengalil, S., Preethish-Kumar, V., Valasani, R.K., Huddar, A., Nishadham, V., Nandeesh, B.N., et al. (2023). Phenotype Genotype Characterization of FKRP-related Muscular Dystrophy among Indian Patients. J. Neuromuscul. Dis. *10*, 615–626. 10.3233/JND-221618.

69. Brown, S.C., Fernandez-Fuente, M., Muntoni, F., and Vissing, J. (2020). Phenotypic Spectrum of α-Dystroglycanopathies Associated With the c.919T>a Variant in the FKRP Gene in Humans and Mice. J. Neuropathol. Exp. Neurol. *79*, 1257–1264. 10.1093/jnen/nlaa120.

70. Beltran-Valero de Bernabé, D., Voit, T., Longman, C., Steinbrecher, A., Straub, V., Yuva, Y., Herrmann, R., Sperner, J., Korenke, C., Diesen, C., et al. (2004). Mutations in the FKRP gene can cause muscle-eye-brain disease and Walker-Warburg syndrome. J. Med. Genet. *41*, e61. 10.1136/jmg.2003.013870.

71. Louhichi, N., Triki, C., Quijano-Roy, S., Richard, P., Makri, S., Méziou, M., Estournet, B., Mrad, S., Romero, N.B., Ayadi, H., et al. (2004). New FKRP mutations causing congenital muscular dystrophy associated with mental retardation and central nervous system abnormalities. Identification of a founder mutation in Tunisian families. Neurogenetics *5*, 27–34. 10.1007/s10048-003-0165-9.

72. 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. 10.1186/s13073-019-0690-2.

73. 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. 10.1038/gim.2015.30.

74. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315. 10.1038/ng.2892.

75. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137. 10.1093/hmg/ddu733.

76. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am. J. Hum. Genet. *99*, 877–885. 10.1016/j.ajhg.2016.08.016.

77. Qi, H., Zhang, H., Zhao, Y., Chen, C., Long, J.J., Chung, W.K., Guan, Y., and Shen, Y. (2021). MVP predicts the pathogenicity of missense variants by deep learning. Nat. Commun. *12*, 510. 10.1038/s41467-020-20847-0.

78. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y., and Marks, D.S. (2021). Disease variant prediction with deep generative models of evolutionary data. Nature *599*, 91–95. 10.1038/s41586-021-04043-8.

79. Quinodoz, M., Peter, V.G., Cisarova, K., Royer-Bertrand, B., Stenson, P.D., Cooper, D.N., Unger, S., Superti-Furga, A., and Rivolta, C. (2022). Analysis of missense variants in the human genome reveals widespread gene-specific clustering and improves prediction of pathogenicity. Am. J. Hum. Genet. *109*, 457–470. 10.1016/j.ajhg.2022.01.006.

80. Gao, H., Hamp, T., Ede, J., Schraiber, J.G., McRae, J., Singer-Berk, M., Yang, Y., Dietrich, A., Fiziev, P., Kuderna, L., et al. (2023). The landscape of tolerated genetic variation in humans and primates. BioRxiv. 10.1101/2023.05.01.538953.

81. Brandes, N., Goldman, G., Wang, C.H., Ye, C.J., and Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. Nat. Genet. *55*, 1512–1522. 10.1038/s41588-023-01465-0.

82. Danzi, M.C., Dohrn, M.F., Fazal, S., Beijer, D., Rebelo, A.P., Cintra, V., and Züchner, S. (2023). Deep structured learning for variant prioritization in Mendelian diseases. Nat. Commun. *14*, 4167. 10.1038/s41467-023-39306-7.

83. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. 10.1126/science.adg7492.

84. Ljungdahl, A., Kohani, S., Page, N.F., Wells, E.S., Wigdor, E.M., Dong, S., and Sanders, S.J. (2023). AlphaMissense is better correlated with functional assays of missense impact than earlier prediction algorithms. BioRxiv. 10.1101/2023.10.24.562294.

85. Kuwabara, N., Imae, R., Manya, H., Tanaka, T., Mizuno, M., Tsumoto, H., Kanagawa, M., Kobayashi, K., Toda, T., Senda, T., et al. (2020). Crystal structures of fukutin-related protein (FKRP), a ribitol-phosphate transferase related to muscular dystrophy. Nat. Commun. *11*, 303. 10.1038/s41467-019-14220-z.

86. Joseph, S., Schnicker, N.J., Xu, Z., Yang, T., Hopkins, J., Watkins, M., Chakravarthy, S., Davulcu, O., Anderson, M.E., Venzke, D., et al. (2022). Structure and mechanism of LARGE1 matriglycan polymerase. BioRxiv. 10.1101/2022.05.12.491222.

87. Ortiz-Cordero, C., Magli, A., Dhoke, N.R., Kuebler, T., Selvaraj, S., Oliveira, N.A., Zhou, H., Sham, Y.Y., Bang, A.G., and Perlingeiro, R.C. (2021). NAD+ enhances ribitol and ribose rescue of α-dystroglycan functional glycosylation in human FKRP-mutant myotubes. eLife *10*. 10.7554/eLife.65443.

88. Liang, W.-C., Hayashi, Y.K., Ogawa, M., Wang, C.-H., Huang, W.-T., Nishino, I., and Jong, Y.-J. (2013). Limb-girdle muscular dystrophy type 2I is not rare in Taiwan. Neuromuscul. Disord. *23*, 675–681. 10.1016/j.nmd.2013.05.010.

89. Topaloglu, H., Brockington, M., Yuva, Y., Talim, B., Haliloglu, G., Blake, D., Torelli, S., Brown, S.C., and Muntoni, F. (2003). FKRP gene mutations cause congenital muscular dystrophy, mental retardation, and cerebellar cysts. Neurology *60*, 988–992. 10.1212/01.wnl.0000052996.14099.dc.

90. Poppe, M., Cree, L., Bourke, J., Eagle, M., Anderson, L.V.B., Birchall, D., Brockington, M., Buddles, M., Busby, M., Muntoni, F., et al. (2003). The phenotype of limb-girdle muscular dystrophy type 2I. Neurology *60*, 1246–1251. 10.1212/01.wnl.0000058902.88181.3d.

91. Krag, T.O., and Vissing, J. (2015). A New Mouse Model of Limb-Girdle Muscular Dystrophy Type 2I Homozygous for the Common L276I Mutation Mimicking the Mild Phenotype in Humans. J. Neuropathol. Exp. Neurol. *74*, 1137–1146. 10.1093/jnen/74.12.1137.

92. Cataldi, M.P., Lu, P., Blaeser, A., and Lu, Q.L. (2018). Ribitol restores functionally glycosylated α-dystroglycan and improves muscle function in dystrophic FKRP-mutant mice. Nat. Commun. *9*, 3448. 10.1038/s41467-018-05990-z.

93. Esapa, C.T., Benson, M.A., Schröder, J.E., Martin-Rendon, E., Brockington, M., Brown, S.C., Muntoni, F., Kröger, S., and Blake, D.J. (2002). Functional requirements for fukutin-related protein in the Golgi apparatus. Hum. Mol. Genet. *11*, 3319–3331. 10.1093/hmg/11.26.3319.

94. Esapa, C.T., McIlhinney, R.A.J., Waite, A.J., Benson, M.A., Mirzayan, J., Piko, H., Herczegfalvi, Á., Horvath, R., Karcagi, V., Walter, M.C., et al. (2023). Misfolding of fukutin-related protein (FKRP) variants in congenital and limb girdle muscular dystrophies. Front. Mol. Biosci. *10*, 1279700. 10.3389/fmolb.2023.1279700.

95. Taniguchi-Ikeda, M., Morioka, I., Iijima, K., and Toda, T. (2016). Mechanistic aspects of the formation of α-dystroglycan and therapeutic research for the treatment of α-dystroglycanopathy: A review. Mol. Aspects Med. *51*, 115–124. 10.1016/j.mam.2016.07.003.

96. Jagannathan, S., Shadle, S., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K., and Tapscott, S.J. (2016). Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Hum. Mol. Genet. 10.1093/hmg/ddw271.

97. Joseph, S., and Campbell, K.P. (2021). Lassa Fever Virus Binds Matriglycan-A Polymer of Alternating Xylose and Glucuronate-On α-Dystroglycan. Viruses *13*. 10.3390/v13091679.

98. Acciani, M., Alston, J.T., Zhao, G., Reynolds, H., Ali, A.M., Xu, B., and Brindley, M.A. (2017). Mutational Analysis of Lassa Virus Glycoprotein Highlights Regions Required for Alpha-Dystroglycan Utilization. J. Virol. *91*. 10.1128/JVI.00574-17.

99. Lay Mendoza, M.F., Acciani, M.D., Levit, C.N., Santa Maria, C., and Brindley, M.A. (2020). Monitoring Viral Entry in Real-Time Using a Luciferase Recombinant Vesicular Stomatitis Virus Producing SARS-CoV-2, EBOV, LASV, CHIKV, and VSV Glycoproteins. Viruses *12*. 10.3390/v12121457.

100. Wrenbeck, E.E., Klesmith, J.R., Stapleton, J.A., Adeniran, A., Tyo, K.E.J., and Whitehead, T.A. (2016). Plasmid-based one-pot saturation mutagenesis. Nat. Methods *13*, 928–930. 10.1038/nmeth.4029.

101. Firnberg, E., and Ostermeier, M. (2012). PFunkel: efficient, expansive, user-defined mutagenesis. PLoS ONE *7*, e52031. 10.1371/journal.pone.0052031.

102. Manya, H., Kuwabara, N., Kato, R., and Endo, T. (2020). FAM3B/PANDER-Like Carbohydrate-Binding Domain in a Glycosyltransferase, POMGNT1. Methods Mol. Biol. *2132*, 609–619. 10.1007/978-1-0716-0430-4_52.

103. Xiong, H., Kobayashi, K., Tachikawa, M., Manya, H., Takeda, S., Chiyonobu, T., Fujikake, N., Wang, F., Nishimoto, A., Morris, G.E., et al. (2006). Molecular interaction between fukutin and POMGnT1 in the glycosylation pathway of alpha-dystroglycan. Biochem. Biophys. Res. Commun. *350*, 935–941. 10.1016/j.bbrc.2006.09.129.

104. Kuwabara, N., Manya, H., Yamada, T., Tateno, H., Kanagawa, M., Kobayashi, K., Akasaka-Manya, K., Hirose, Y., Mizuno, M., Ikeguchi, M., et al. (2016). Carbohydrate-binding domain of the POMGnT1 stem region modulates O-mannosylation sites of α-dystroglycan. Proc Natl Acad Sci USA *113*, 9280–9285. 10.1073/pnas.1525545113.

105. Lee, J.K., Matthews, R.T., Lim, J.-M., Swanier, K., Wells, L., and Pierce, J.M. (2012). Developmental expression of the neuron-specific N-acetylglucosaminyltransferase Vb (GnT-Vb/IX) and identification of its in vivo glycan products in comparison with those of its paralog, GnT-V. J. Biol. Chem. *287*, 28526–28536. 10.1074/jbc.M112.367565.

106. Esposito, D., Weile, J., Shendure, J., Starita, L.M., Papenfuss, A.T., Roth, F.P., Fowler, D.M., and Rubin, A.F. (2019). MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. Genome Biol. *20*, 223. 10.1186/s13059-019-1845-6.

107. Baird, M.F., Graham, S.M., Baker, J.S., and Bickerstaff, G.F. (2012). Creatine-kinase- and exercise-related muscle damage implications for muscle performance and recovery. J. Nutr. Metab. *2012*, 960363. 10.1155/2012/960363.

108. Wood, A.J., Lin, C.H., Li, M., Nishtala, K., Alaei, S., Rossello, F., Sonntag, C., Hersey, L., Miles, L.B., Krisp, C., et al. (2021). FKRP-dependent glycosylation of fibronectin regulates muscle pathology in muscular dystrophy. Nat. Commun. *12*, 2951. 10.1038/s41467-021-23217-6.

109. Fujimura, K., Sawaki, H., Sakai, T., Hiruma, T., Nakanishi, N., Sato, T., Ohkura, T., and Narimatsu, H. (2005). LARGE2 facilitates the maturation of alpha-dystroglycan more effectively than LARGE. Biochem. Biophys. Res. Commun. *329*, 1162–1171. 10.1016/j.bbrc.2005.02.082.

110. Moore, C.J., Goh, H.T., and Hewitt, J.E. (2008). Genes required for functional glycosylation of dystroglycan are conserved in zebrafish. Genomics *92*, 159–167. 10.1016/j.ygeno.2008.05.008.

111. Esser, A.K., Miller, M.R., Huang, Q., Meier, M.M., Beltran-Valero de Bernabé, D., Stipp, C.S., Campbell, K.P., Lynch, C.F., Smith, B.J., Cohen, M.B., et al. (2013). Loss of LARGE2 disrupts functional glycosylation of α-dystroglycan in prostate cancer. J. Biol. Chem. *288*, 2132–2142. 10.1074/jbc.M112.432807.

112. Metze, S., Herzog, V.A., Ruepp, M.-D., and Mühlemann, O. (2013). Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. RNA *19*, 1432–1448. 10.1261/rna.038893.113.

113. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784. 10.1038/nmeth.3047.

114. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. 10.1002/pro.3943.

115. Tesla, B., Powers, J.S., Barnes, Y., Lakhani, S., Acciani, M.D., and Brindley, M.A. (2022). Temperate conditions limit zika virus genome replication. J. Virol. *96*, e0016522. 10.1128/jvi.00165-22.

116. Ramakrishnan, M.A. (2016). Determination of 50% endpoint titer using a simple formula. World J. Virol. *5*, 85–86. 10.5501/wjv.v5.i2.85.

Figure:

Figure 1: Streamlining the saturation mutagenesis and FACS assay in SMuRF.

(A) A universal workflow of SMuRF. SMuRF accompanies saturation mutagenesis with functional assays. Here, saturation mutagenesis is achieved by delivering variant lentiviral particles to the engineered HAP1 platform where the endogenous gene of interest (GOI) was knocked out and stable *DAG1* overexpression was established through lentiviral integration. A fluorescence-activated cell sorting (FACS) assay was employed to separate the high-function population and the low-function population.

(B) Lenti-*GOI* constructs used for the saturation mutagenesis. The *GOI* CDS expression is driven by a weak promoter UbC.

(C) PALS-C is simple and accessible to most molecular biological laboratories. To accommodate the requirements of downstream short-read NGS, the *GOI* variants were separated into multiple blocks (6 blocks for *FKRP* and 10 blocks for *LARGE1*). PALS-C generates block-specific lentiviral plasmid pools from 1 oligo pool per GOI. The steps are massively multiplexed: Step 1 requires only a single-tube reaction; the following steps can be done in a single-tube reaction for each block. Step 8 (not shown) uses electrotransformation to deliver the assembled plasmid pools to bacteria for amplification.

(D) A representative example shows the gating strategy; 20k flow cytometry events of FKRP block1 were recorded and reanalyzed with FlowJo.

(E) A 3-round PCR strategy to build the NGS library. Samples from the high and low glycosylation groups were barcoded differently in PCR2. PCR2 products of all samples were multiplexed for a single PCR3 reaction.

(F) A universal pipeline to generate SMuRF scores from raw NGS data. Steps colored yellow indicate employment of customized scripts. Cleaning is a critical step where the reads carrying co-occurred variants are filtered out.

Figure 2: SMuRF recapitulated and expanded the knowledge gained from variant databases.

(A and B) High confidence SMuRF scores align with variant types (A, *FKRP*; B, *LARGE1*). The mean of synonymous variants of each block was used to normalize the scores across blocks. The box boundaries represent the 25th/75th percentiles, with a horizontal line indicating the median and a vertical line marking an additional 1.5 times interquartile range (IQR) above and below the box boundaries. p-values were calculated using the two-sided Wilcoxon test. Counts of variants were labeled below the boxes.

(C and D) SMuRF revealed functional constraints based on variants reported in gnomAD v4.0.0 genome sequencing data (C, *FKRP*; D, *LARGE1*): Low allele frequency variants had diverse functional scores, while high allele frequency variants converged towards wild-type (WT) due to selection pressures (Gray box: AF < 1.5e-05). Dashed lines represent WT functional score. Dots were jittered with geom_jitter (width = 0.05, height = 0.05).

Figure 3: SMuRF improved the scope of clinical interpretation of genetic variants.

(A and B) SMuRF scores correlate well with clinical classification in ClinVar (**A,** *FKRP*; **B,** *LARGE1*). (B/LB: Benign, Benign/Likely benign or Likely benign in ClinVar; VUS: Uncertain significance in ClinVar; P/LP: Pathogenic, Pathogenic/Likely pathogenic or Likely pathogenic in ClinVar.) Counts of variants were labeled below the violins.

(C) Real patient data from eight well-curated cohorts demonstrated that SMuRF scores have the potential to predict disease severity. The additive SMuRF scores of the variant pairs associated with mild cases were significantly higher than those of intermediate and severe cases. Counts of cases were labeled below the violins. p-values were calculated using the Wilcoxon test. FS1, the SMuRF functional score of the variant on Allele1; FS2, the SMuRF functional score of the variant on Allele2.

(D) The SMuRF scores are correlated with the disease onset age. Dashed trendlines represent linear regression. Spearman's rank correlation rho: 0.72 (all data), 0.70 (male), 0.73 (female).

Figure 4: SMuRF scores can be employed to validate and improve computational predictors.

(A and B) Receiver operating characteristic (ROC) curves of SMuRF and computational predictors (A, *FKRP*; B, *LARGE1*). AUC: Area Under Curve. Higher AUC indicates better performance in classifying pathogenic variants.

(C) The correlation coefficient was calculated between SMuRF scores and scores generated by computational predictors. Figure depicts absolute correlation coefficient.

(D and E) Among all the predictors examined, AlphaMissense has the strongest correlation with SMuRF (rho = -0.70, *FKRP*; -0.54, *LARGE1*). Density was calculated with contour_var = "count" in R. White dashed lines represent linear regression.

(F) IIH6C4 blots indicate matriglycan synthesis activity of selected LARGE1 enzyme variants. The laminin overlay provides a different probe for matriglycan, with an arrow indicating the expected band size. The protein loading was controlled by Coomassie gel. Experiments were conducted with the myogenic cell line, MB135. Control: WT MB135. KO Rescue: endogenous *LARGE1* was knocked out and the cells were resecured with individual lentiviral transduction. P.Ala145= (SMuRF = -0.43) has the highest AF (0.41) in gnomAD v4, which was used as a high-function reference. P.Ser331Phe (SMuRF = -2.58) is Pathogenic in ClinVar, which was used as a low-function reference.

Figure 5: SMuRF highlights the critical structural regions.

(A) SMuRF scores showed higher functional disruption by missense variants in the catalytic domain of *FKRP* compared to the stem domain. The zinc finger loop (Znf) within the catalytic domain exhibited greater disruption by missense variants. Box plots depict the 25th/75th percentiles (box boundaries), median (horizontal line), and an additional 1.5 times IQR (vertical line) above and below the box boundaries. p-values were calculated using the two-sided Wilcoxon test. Counts of variants were labeled below the violins. Dashed lines represent WT functional score.

(B) The SMuRF scores of synonymous *FKRP* variants in different domains.

(C) Missense variants in the catalytic domains of *LARGE1* showed higher disruption compared to the N-terminal domain. Missense variants in the XylT domain were more disruptive than those in the GlcAT domain.

(D) The SMuRF scores of synonymous *LARGE1* variants in different domains.

(E and F) Mean SMuRF scores were utilized to map SNV-generated single amino acid substitutions onto the 1D structures of the enzymes (E, *FKRP*; F, *LARGE1*). The mean SMuRF score per amino acid residue was calculated and visualized using a color scale, where red indicates positions sensitive to substitutions and green is tolerated.

(G and H) Mean SMuRF scores were utilized to map SNV-generated single amino acid substitutions onto the 3D structures of the enzymes (G, *FKRP*; H, *LARGE1*). The crystal structure of human FKRP (PDB:6KAM, codon: 45-495) and the electron microscopy structure of LARGE1 (PDB:7UI7, codon: 34-756) were used. Same color scale is employed as E and F.

(I and J) Heatmap representation of the mean SMuRF scores for each codon (I, *FKRP*; J, *LARGE1*). Amino acids were categorized into five groups: nonpolar, aliphatic (G, A, V, L, M, I); polar, uncharged (S, T, C, P, N, Q); positively charged (K, R, H); negatively charged (D, E); and nonpolar, aromatic (F, Y, W). Each cell in the heatmap corresponds to a codon position (x-axis) and an amino acid group (y-axis). The black dots indicate the wild-type amino acid group for each residue. Grey squares denote the scenario where the amino acid change is not possible with a single SNV within the codon, and a red cross marks positions where variants were filtered out due to low confidence.

Figure 6: Validations confirmed SMuRF findings in the myogenic context.

(A) Validation of individual *FKRP* variants using an IIH6C4 IF assay. The myoblasts underwent transduction and drug selection, followed by differentiation into myotubes, which were subsequently used for IF. ".r" denotes lentiviral transduction of an individual variant. Blue: DAPI. Green: IIH6C4, the glycosylation level of α-DG. Red: SMuRF scores; White: ClinVar clinical significance. The brightness and contrast of the photos were adjusted in Adobe Photoshop with the same settings.

(B) Immunofluorescence intensity was quantified using integrated density (IntDen) of IIH6C4 relative to DAPI in differentiated myotubes by ImageJ. Analyses were conducted on 8 representative images. *p□<□0.05, ****p□<□0.0001 (compared with FKRP KO group). ##p□<□0.01, ###p□<□0.001, ####p□<□0.0001 (compared with WT.r group). Multiple comparisons between groups were performed using analysis of variance (ANOVA) followed by Bonferroni post hoc test through GraphPad Prism 10.2.2. Experiments were conducted independently three times.

(C) An orthogonal assay based on α-DG-dependent viral entry. Vesicular stomatitis virus (VSV) with Lassa fever virus glycoprotein complex (LASV-GPC) can infect cells in a glycosylated-α-

DG-dependent manner. Variant enrichment before/after VSV infection can be used to quantify their performances regarding α-DG glycosylation.

(D and E) The ppVSV assay can be employed to validate the findings from the flow cytometry assay (D, *FKRP*; E, *LARGE1*). 10 Lenti-*FKRP* variants were mixed with Lenti-WT-*FKRP* to rescue *FKRP*-KO MB135. 11 Lenti-*LARGE1* variants were mixed with Lenti-WT-*LARGE1* to rescue *LARGE1*-KO MB135. The functional score was quantified by the ratio of a variant's enrichment in the non-infected group to its enrichment in the ppVSV-infected group. A higher functional score indicates better performance in α-DG glycosylation. *FKRP* c.135C>T (p.Ala45=) and *LARGE1* c.435C>T (p.Ala145=) have the highest AFs in gnomAD v4. *FKRP* c.663C>A (p.Ser221Arg) and *LARGE1* c.992C>T (p.Ser331Phe) are Pathogenic in ClinVar. Biological replicate N = 3: Lentiviral transduction and ppVSV infection were both performed independently. Figures display mean values with SEM. Additional discussions for the ppVSV results are included in Supplemental Note. Multiple comparisons were performed using ANOVA and Dunnett's test.

Table 1. A potential SMuRF labeling.

FKRP:		
Clinvar clinical significance	**Label**	**Variant counts**
P/LP	SMuRF Benign	3
	SMuRF Mild	12
	SMuRF Intermediate	21
	SMuRF Severe	26
VUS	SMuRF Benign	285
	SMuRF Mild	68
	SMuRF Intermediate	39
	SMuRF Severe	31
B/LB	SMuRF Benign	275
	SMuRF Mild	13
	SMuRF Intermediate	0

	SMuRF Severe	1
Unclassified	SMuRF Benign	2328
	SMuRF Mild	436
	SMuRF Intermediate	360
	SMuRF Severe	427
LARGE1:		
Clinvar clinical significance	**SMuRF label**	**Variant counts**
P/LP	SMuRF Benign	0
	SMuRF Pathogenic	8
VUS	SMuRF Benign	224
	SMuRF Pathogenic	24
B/LB	SMuRF Benign	156
	SMuRF Pathogenic	0
Unclassified	SMuRF Benign	4747
	SMuRF Pathogenic	1314

SMuRF labeling is based on SMuRF functional scores, which provides an additional line of evidence for clinical variant interpretation. These labels are not meant to be used as a decisive classification in isolation. ClinVar clinical significance: P/LP: Pathogenic, Pathogenic/Likely pathogenic, and Likely pathogenic; B/LB: Benign, Benign/Likely benign, and Likely Benign; VUS: Uncertain significance, Conflicting interpretations of pathogenicity, and no interpretation for the single variant.

Figure S1: The background and preparation of the implementation of SMuRF

(A) The majority of possible coding SNVs of α-DG glycosylation enzyme-coding genes lack clinical interpretation (*e.g.*, no ClinVar record, April 20, 2023). Around half of the reported SNVs of these genes in ClinVar were classified as variants of uncertain significance.

(B) Functions of most α-DG glycosylation enzymes can be evaluated by the IIH6C4 antibody (blue texts). Bold arrows link enzymes to the modifications or glycan additions they catalyze; for instance, POMK catalyzes mannose phosphorylation. POMT1/2 participate the glycosylation of α-DG Core M1, M2 and M3. POMGNT1 mutations in patients manifest the hallmark of reduced IIH6C4 signal. GnT-Vb/IX mutations are not known to alter IIH6C4 signals. Hence, POMGNT1 but not GnT-Vb/IX can be evaluated by IIH6C4. Dol-P-Man, dolichol phosphate mannose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; Rbo5P, ribitol-5-phosphate; Xyl, xylose; GlcA, glucuronic acid; Gal, galactose.

(C) IIH6C4 fluorescence flow cytometry is compatible with the HAP1 platform. 1-hr room temperature staining was performed for both primary and secondary antibodies. The *DAG1*-KO HAP1 carries a 1-bp deletion (c.205Gdel). *FKRP*-KO HAP1 was established from a single clone of CRISPR RNP nucleofected cells and carries a 1-bp insertion (c.181Adup). FITC-A: IIH6C4; Pacific Blue-A: Viobility 405/452 Fixable Dye.

(D) 1-hr staining is not sensitive enough for a high-throughput FACS-based variant characterization, as the separation between positive signal and null signal is unsatisfactory.

(E) Optimized 4 °C overnight staining increased the assay sensitivity. Both staining protocols used the same samples: Lenti-*DAG1 FKRP*-KO HAP1 rescued by a mixture of Lenti-EF1a-*FKRP* and Lenti-EF1a-*FKRP*(c.826C>A).

(F and G) *GOI*-KO HAP1 monoclonal cell lines were isolated from pooled CRISPR RNP nucleofected cells. *FKRP*-KO HAP1 carries a 1-bp frameshifting insertion (c.181Adup). *LARGE1*-KO HAP1 carries a 94-bp frameshifting deletion (c.121_214del). The mutations were validated with Sanger sequencing.

(H) α-DG overexpression was achieved with Lenti-*DAG1*.

(I) RT-PCR experiment for *FKRP*-KO Lenti-*DAG1* monoclonal lines (FLmc#1 and FLmc#2). The same amount of RNA molecules was used for each reaction. Bands in lane 2-5 were the PCR products of the *DAG1* cDNA primers (225 bp). Bands in lanes 1&6 were 200-bp ladder bands. FLmc#1 was chosen as the cell line platform for downstream experiments.

(J) RT-qPCR experiment for *LARGE1*-KO Lenti-*DAG1* monoclonal lines. *HPRT1* primers were used as the housekeeping control for the ddCt quantification. The expression level was normalized to the *LARGE1*-KO sample. qPCR triplicates were set for each condition. LLmc#1 was chosen as the cell line platform for downstream experiments.

(K and L) These results were from experiments in the early development stage. The platform cell line is the *FKRP*-KO cell line without Lenti-*DAG1* transduction. Initially, we constructed Lenti-EF1a-*FKRP*-P2A-BSD. We later abandoned this design as the over-expression failed to generate the expected separation and the P2A fusion impeded our ability to study nonsense variants in an unbiased manner. Interestingly, the cells treated with the lentivirus carrying the p.Trp26Ter nonsense mutation unexpectedly survived the blasticidin drug selection, suggesting the existence of a genetic mechanism that could rescue this nonsense mutation, at least in the context of using this construct. (K) A mini-library comprising 7 *FKRP* variants and the WT sequence was employed in a FACS experiment. The over-expression masked the deleterious effects of the pathogenic variants, leading to a lack of significant enrichment differences of those variants between the high-glycosylation and low-glycosylation populations. (The enrichment was quantified by Sanger sequencing peaks; the enrichment of the WT was calculated based on the enrichments of other variants.) (L) Western blot was performed for individually transduced samples as well as the Mix8 sample. The lentiviral expression was excessively high compared to the endogenous *FKRP* expression in WT HAP1 (not detectable by WB). A smaller band (between 35 kDa and 55 kDa) was detected in the p.Trp26Ter sample, supporting the hypothesis of the existence of a downstream alternative start codon rescuing this mutation (likely, Met144, resulting in a 39.0 kDa product). The WB protocol is described in Method S9.

Figure S2: The implementation of SMuRF

(A) We designed and synthesized oligos for all possible CDS SNVs of *FKRP* and *LARGE1*. Stop-loss variants were intentionally excluded from the design due to their incompatibility with the lentivirus-based saturation mutagenesis strategy.

(B) Plasmid pool QC was performed using the Amplicon-EZ service. QC sequencing libraries were constructed using PCR amplification strategies for *FKRP* and *LARGE1*. BLK, block.

(C) 99.6% of all possible SNVs were represented in the plasmid pools *FKRP* and *LARGE1*. Color groups indicate the variant representation in the pools. If the variants are evenly represented, we expect a 0.13% (*FKRP* BLK1-5) or a 0.15% (*FKRP* BLK6; *LARGE1* blocks) representation for all variants.

(D) Per-variant frequency distribution along the CDS. x-axis represents the nucleotide sites.

(E) Mini-libraries (Method S7) were used to optimize the FACS assay to achieve expected separation of variants with known clinical significance; *FKRP* (top), *LARGE1* (bottom). Both Amplicon-EZ NGS and Sanger sequencing were used to quantify the variant enrichment for *FKRP*. Sanger sequencing was used to quantify the variant enrichment for *LARGE1*. Relative enrichment was defined as a ratio of a variant's representation in the high-glycosylation group to that in the low-glycosylation group. Higher relative enrichment indicates higher variant function. The *LARGE1* sequence was cloned from HEK293T cDNA.

HEK293T carries a heterozygous *LARGE1* mutation (c.1848G>A, p.Met616Ile). This variant was removed when building the *LARGE1* SMuRF lentiviral pools but was kept in the experiment depicted here. Sanger sequencing confirmed the frequency of this mutation in these four lentiviral constructs were: WT 35.1%, S331F 96.3%, D336= 2.0%, Q368*=2.0%.

(F and G) The high-glycosylation group (top ~20%) and low-glycosylation group (bottom ~40%) were gated based on the IIH6C4+FITC signal. Examples show the FACS gating parameters for all blocks of *FKRP* (F) and *LARGE1* (G). 20k events were recorded. Cell debris and multiplets were pre-excluded before the sorting.

(H) The functional scores from the biological replicates exhibited good pair-wise correlation. The Spearman's rank correlation rho is reported.

Figure S3: SMuRF scores were employed to characterize *FKRP* and *LARGE1* variants

(A and B) SMuRF scores align with variant types (A, *FKRP*; B, *LARGE1*): synonymous variants resemble wildtype, nonsense variants consistently have low scores, and start-loss variants exhibit markedly lower scores than nonsense variants. Noteworthy outliers include high-scoring nonsense variants at the end of coding sequences. The box boundaries represent the 25th/75th percentiles, with a horizontal line indicating the median and a vertical line marking an additional 1.5 times interquartile range (IQR) above and below the box boundaries. p-values were calculated using the two-sided Wilcoxon test. Counts of variants were labeled below the boxes.

(C and D) SMuRF revealed functional constraints based on variants reported in gnomAD v4.0.0 (both genome and exome sequencing data; C, *FKRP*; D, *LARGE1*): Low allele frequency variants had diverse functional scores, while high allele frequency variants converged towards wildtype due to selection pressures (Gray box: AF<1.5e-05). Dashed lines represent WT functional score. Dots were jittered with geom_jitter(width = 0.05, height = 0.05).

(E and F) SMuRF scores were compared with the phyloP score calculated from multiple alignments of 100 vertebrate species (E, *FKRP*; F, *LARGE1*). The higher the phyloP score is, the more conserved a genomic site is. The missense variants show clustering around the WT SMuRF score (0) at less conserved sites, while they tend to be more disruptive at more conserved sites. Density was calculated with contour_var = "count" in R.

(G and H) SMuRF scores correlate well with clinical classification in ClinVar (G, *FKRP*; H, *LARGE1*). B: Benign; B/LB: Benign/Likely benign; LB: Likely Benign; P: Pathogenic; P/LP: Pathogenic/Likely pathogenic; LP: Likely pathogenic; VUS: Uncertain significance. Counts of variants were labeled above the boxes. Dashed lines represent WT functional score.

(I) The higher SMuRF score of the two scores of a variant pair is used to represent the variant pair. Scores associated with mild cases were significantly higher than those of intermediate and severe cases. Counts of cases were labeled below the boxes. p-values were calculated using the Wilcoxon test. FS1, the SMuRF functional score of the variant on Allele1; FS2, the SMuRF functional score of the variant on Allele2.

(J) The SMuRF scores of variants classified as Pathogenic (P), Pathogenic/Likely Pathogenic (P/LP), and Likely Pathogenic (LP) in ClinVar did not show a significant pattern when grouped according to the reported conditions in ClinVar. (Left, *FKRP*; Right, *LARGE1*). WWS: Walker-Warburg syndrome. MDDG: muscular dystrophy dystroglycanopathies (left to right: type A1, type A5, type B5). LGMD: limb-girdle muscular dystrophy (left to right: not specified; type R9/2I). Other: not specified (left to right: Abnormality of the musculature, Cardiovascular phenotype). RP: retinitis pigmentosa. Counts of cases were labeled above the boxes.

(K) Variability in onset time among patients with homozygous p.Leu276Ile mutation.Data from The Genetic Resolution and Assessments Solving Phenotypes in LGMD consortium (GRASP LGMD) were integrated with the 8 well-curated cohorts. The GRASP LGMD cohort mainly includes patients with homozygous p.Leu276Ile mutation and patients with compound heterozygous mutations (with p.Leu276Ile on one allele). Patients with homozygous p.Leu276Ile mutation exhibited a diverse range of onset ages, suggesting that the severity of dystroglycanopathy in these individuals might be influenced by other factors, like genetic modifiers, in addition to the *FKRP* mutation.

(L) Creatine kinase (CK) values did not show a significant correlation with SMuRF scores. Spearman's rank correlation rho: -0.04 (all data), -0.09 (male), 0.03 (female). Light grey rectangle indicates the normal range in healthy people (24.00 - 204.00 IU/L).

Figure S4: Comparing SMuRF scores with the predictor scores.

(A) Top-ranked computational predictors correlate well with each other. Legend indicates the absolute Spearman’s rho.

(B) Among the predictors examined, those based on evolutionary conservation exhibit high Spearman's coefficients with SMuRF, such as EVE (*FKRP*: -0.67; *LARGE1*: -0.51) and PrimateAI-3D (*FKRP*: -0.69; *LARGE1*: -0.49). REVEL demonstrates the best performance according to the ROC curves and exhibits high Spearman's coefficient with SMuRF (*FKRP*: -0.65; *LARGE1*: -0.47). White dashed lines represent linear regression. Density was calculated with contour_var = "count" in R.

(C) AlphaMissense also exhibits functional differences between LARGE1 domains. Box plots depict the 25th/75th percentiles (box boundaries), median (horizontal line), and an additional 1.5 times IQR (vertical line) above and below the box boundaries. p-values were calculated using the Wilcoxon test. Counts of variants were labeled below the boxes.

Figure S5: SMuRF highlights the critical structural regions.

(A) The currently known disease-related SNVs in *FKRP* and *LARGE1* are dispersed throughout their entire sequences. Pathogenic, Pathogenic/Likely pathogenic, and Likely pathogenic variants in ClinVar (April 20, 2023) were labeled at their respective sites on the cDNA sequences.

(B) The SMuRF scores of SNV-generated single amino acid substitutions were projected onto the FKRP tetramer (PDB 6KAM). p.Tyr88 (mean SMuRF=-2.69) and p.Ser221 (mean SMuRF=-1.00) are both situated at the subunit-subunit interface involved in FKRP tetramerization.

(C) The SMuRF scores of SNV-generated single amino acid substitutions were projected to the LARGE1 dimer (PDB 7UI7).

(D) The SMuRF scores of SNV-generated single amino acid substitutions were projected to the FKTN-FKRP-RXYLT1 complex predicted with Alphafold2 LocalColabFold (https://github.com/YoshitakaMo/localcolabfold). Blue: FKTN; orange: RXYLT1.

Figure S6: Additional validations and analyses for SMuRF findings.

(A) Homozygous *GOI*-KO MB135 monoclonal cell lines were isolated from pooled CRISPR RNP nucleofected cells. *FKRP*-KO MB135 carries a 1-bp frameshifting insertion (c.181Adup). *LARGE1*-KO MB135 carries a 94-bp frameshifting deletion (c.121_214del). The mutations were validated with Sanger sequencing. These mutations are the same as the mutations in the *GOI*-KO HAP1 lines.

(B) MB135 is not compatible with the flow cytometric assay due to high background noise. Monoclonal Lenti-*DAG1 FKRP*-KO MB135 myoblasts were used in this IIH6C4 flow cytometric experiment. High background noise was observed in both the control sample where no Lenti-*FKRP* rescue was performed (UT), and the control sample where the Lenti-*FKRP*-rescued cells were stained with secondary antibody but without the IIH6C4 primary antibody (WT.r 2nd-only). No significant differences were observed among different variants. (Monoclonal *FKRP*-KO MB135 myoblasts were also tested and showed the same pattern).

(C) Validation of individual *LARGE1* variants using an IIH6C4 IF assay. ".r" denotes lentiviral transduction of an individual variant (Same lentivirus as used in Figure S2E). Green: IIH6C4, α-DG the glycosylation level. SMuRF scores are indicated in red. Immunofluorescence intensity was quantified using integrated density (IntDen) of IIH6C4 relative to DAPI by ImageJ. Analyses were conducted on 8 representative images from 2 of 3 independent experiments. ****$p < 0.0001$ (compared with LARGE1-KO group).

##p□<□0.01, ####p□<□0.0001 (compared with WT.r group). Multiple comparisons between groups were performed using analysis of variance (ANOVA) followed by Bonferroni post hoc test through GraphPad Prism 10.2.2.

Figure S7: Considerations for the implementation and further development of SMuRF.

(A) A small-scale FACS experiment utilizing an FKRP mini-library comprising WT, Trp26Ter, and Leu276Ile. y-axis is high/low enrichment normalized to Trp26Ter. Higher number indicates higher function. 1-fold transduction conferred lentiviral *FKRP* expression comparable to the endogenous *FKRP* expression in WT cells. Higher MOI appeared to reduce the functional difference between Trp26Ter and Leu276Ile.

(B-D) The VSV-based viral entry assays did not exhibit sufficient sensitivity for large-scale evaluation of variants. Dashed line indicates no enrichment difference between the VSV-infected group and the non-infected group. Black solid line represent linear regression for all variants. Other solid lines represent linear regression for different variant types. B, rVSV-LASV-GPC viral entry assay for *FKRP* (The trend lines of missense variants and all variants overlay with the trend line of synonymous variants). ppVSV-LASV-GPC viral entry assay for *FKRP* (C) and *LARGE1* (D; The trend line of all variants overlays with the trend line of missense variants). The trend lines indicated a tendency where start-loss variants were the most enriched in the infected group, suggesting a greater disruptive effect on α-DG glycosylation. Nonsense variants showed the next highest enrichment, followed by missense variants, while synonymous variants demonstrated the lowest level of enrichment. However, overall, the VSV assays lack sufficient sensitivity for accurate interpretation of variants.

(E) Two synonymous variants (red) in *LARGE1*, c.639T>C (p.Asn213=) (SMuRF=0.45) and c.642A>G (p.Lys214=) (SMuRF=0.64), may potentially have gain-of-function (GOF) effects by removing a poly(A) motif from the coding sequence. However, validation using the orthogonal ppVSV assay did not support the hypothesis of these 2 synonymous variants having GOF effects (Figure 6E).

(F) Higher synVep scores indicate stronger effects, either loss-of-function effects (LOF) or gain-of-function (GOF) effects. Light green: potential GOF variants. Red: Potential LOF variants.

(G) Comparing SMuRF scores with GOF scores generated by LoGoFunc, an ensemble machine learning-based predictor. LoGoFunc classifies the variants into 3 categories: GOF, LOF, or Neutral. LoGoFunc_GOF score is a probability for GOF.

(H) Nonsense variants in the GlcAT region of *LARGE1* generally exhibit higher SMuRF scores than those in the XylT region. Box plots depict the 25th/75th percentiles (box boundaries), median (horizontal line), and an additional 1.5 times IQR (vertical line) above and below the box boundaries. p-values were calculated using the two-sided Wilcoxon test. Counts of variants were labeled below the boxes. Dashed lines represent WT functional score.

Table S6: Spearman's rank correlation rho of SMuRF and the computational predictors.

Gene	AlphaMissense	PrimateAI-3D	EVE	REVEL	MutScore
FKRP	-0.6978703	-0.6903724	-0.6659107	-0.6507694	-0.6342445
LARGE1	-0.5407195	-0.4947909	-0.5132787	-0.4741103	-0.4700939

Gene	ESM1b	CADD	MAVERICK	metaSVM	MVP
FKRP	0.5655441	-0.5813337	-0.5043554	-0.33925	-0.3255476
LARGE1	0.4767206	-0.4000735	-0.2382434	-0.4111915	-0.2704415

Note S1: Considerations for implementation and future development of SMuRF.

The utilization of lentiviral delivery primarily stems from technical and economic considerations. Delivery of variants through this exogenous method is artificial, as it removes the genes' cis-regulatory elements and intron/exon structures. On-site saturation genome editing is preferred for maintaining genomic context. However, lentiviral systems offer benefits by increasing the useful event for flow cytometric assays. Currently, saturation genome editing is limited by its efficiency, which is heavily influenced by nature of specific genomic regions. Achieving comprehensive unbiased high coverage for saturation mutagenesis in certain genes is challenging. Particularly in flow cytometric assays, low efficiency results in longer sorting times and poses significant economic challenges. Lentiviral delivery presents a trade-off where genomic context is sacrificed for more comprehensive coverage.

The impact of the Multiplicity of Infection (MOI) of the lentiviral pool on the outcome of the functional characterization is profound (Figure S7A). Several factors might need to be considered here: (1) Higher MOI may render higher expression of the gene of interest (GOI), thus compensating for the pathogenic effects of some pathogenic variants. (2) Different levels of MOI result in different co-existing status of different variants in the same single cell. Whether co-existing variants can generate modifying effects through protein dimerization, and whether different co-existing combinations generate different modifying

effects (positive or negative), are unclear and require further research. In the SMuRF workflow, for the purpose of better separating variants with different levels of function, the MOI was controlled to ensure that the lentiviral GOI RNA level is comparable to the endogenous level in WT cells. The titer was determined in pre-experiments. Future improvement to the workflow entails identifying and employing the endogenous promoter core element as a replacement for the UbC promoter[1].

We utilized rVSV-LASV-GPC or ppVSV-LASV-GPC to infect *GOI*-KO MB135 myoblasts that had been rescued by all-possible Lenti-*GOI* variant pools. The rVSV genome incorporates the LASV-GPC coding sequence, allowing it to re-enter cells. In contrast, the ppVSV lacks this capability as it is pseudotyped using a LASV-GPC plasmid (STAR Methods). Likely due to the re-entering feature of rVSV, the rVSV screen did not yield meaningful results (Figure S7B). The ppVSV screen for both FKRP and LARGE1 showed a tendency where start-loss variants were the most enriched in the infected group, suggesting a higher disruptive effect on α-DG glycosylation. Nonsense variants were the next most enriched, followed by missense variants, while synonymous variants were the least enriched (Figures S7C and S7D). This tendency aligns with what we observed in the SMuRF FACS assay. However, the ppVSV assay, in general, lacked the sensitivity to distinguish differences among the variants in the whole-CDS scale. Therefore, we employed the ppVSV assay for the mini-library rescued myoblasts (Figures 6D and 6E). In the *FKRP* mini-library, we also included *FKRP* c.1174T>G (p.Phe392Val; SMuRF=-1.13; SMuRF_mild, Confidence=HIGH), which is a previously unknown variant and has been observed in a patient with compound mutations (Leu276Ile on the other allele). ppVSV assay supports the notion that this variant is mildly pathogenic.

Synonymous variants are typically considered to have limited effects on gene function[2]. In our study, we adopted the assumption that synonymous variants have no effects when normalizing our scores. However, it is possible that some synonymous variants can affect RNA motifs, structure, or splicing, leading to potential changes in gene function[3–5]. While the effects of synonymous variants on splicing cannot be evaluated within the SMuRF framework, their impact on certain RNA motifs can be captured by SMuRF. We selected two synonymous variants in *LARGE1*, c.639T>C (SMuRF=0.45) and c.642A>G (SMuRF score=0.64), for evaluation, which might have gain-of-function (GOF) effects by removing a poly(A) motif from the coding sequence (Figure S7E). The poly(A) signal is associated with transcription termination, and its removal may increase the level of functional transcripts[6]. However, validation using the orthogonal ppVSV assay did not support the hypothesis of these 2 synonymous variants having GOF effects (Figure 6E). Nevertheless, it's essential to note that even if this GOF effect is genuine, it may not fully apply to the endogenous *LARGE1* in its correct genomic context, despite the fact that this poly(A) motif is not separated by exon junctions *in vivo*.

synVep (synonymous Variant effect predictor) is a machine learning-based predictor for evaluating the effects of synonymous SNVs[5]. Higher synVep scores (range: 0-1) indicate stronger potential effects. We compared the synVep scores with the SMuRF scores for *FKRP* and *LARGE1* (Figure S7F). For *FKRP*, c.783A>T has an ultra-low SMuRF score and is also predicted to have strong effects by synVep (SMuRF=-3.42, synVep=0.99). For *LARGE1*, c.477C>A has an ultra-low SMuRF score and is also predicted to have strong effects by synVep (SMuRF=-2.47, synVep=0.99). *FKRP* c.783A>T, together with c.783A>C (SMuRF=-2.62, confidence=LOW, synVep=0.92) and c.783A>G (SMuRF=0.02, synVep=0.83), was included in our ppVSV test, as well as *LARGE1* c.690T>A (SMuRF=-1.28, synVep=0.95), c.693T>G (SMuRF=-1.33, synVep=0.98) and c.699C>G (SMuRF=-1.67, synVep=0.99). However, the results did not support any of them having LOF effects (Figures 6D and 6E). Further research, such as endogenous gene editing, is required to fully evaluate the effects of these synonymous variants.

LoGoFunc is an ensemble machine-learning predictor for pathogenic GOF, pathogenic LOF, and neutral genetic variants[7]. While it did not predict the existence of any GOF variants in *FKRP*, it did predict the existence of several GOF variants in *LARGE1*, including c.26G>C (p.Arg9Pro) (Figure S7G). We examined c.26G>C (SMuRF=0.18, LoGoFunc_GOF=0.36) with the ppVSV assay, but the result did not support the notion that this variant possesses GOF effects (Figure 6E).

Consistent with what was shown in Figure 5C, nonsense variants also generally exhibit higher SMuRF scores in the GlcAT region of *LARGE1* than those in the XylT region (p-value=0.00034, Figure S7H).

Since the SMuRF system employs the CDS of the gene, the nonsense mutations are expected to produce truncated protein without significantly affected by EJC-enhanced NMD. It may be assumed that the SMuRF score difference between the GlcAT region and the XylT region is partly technical. That is, in the absence of a functional GlcAT domain, a functional XylT domain can still add one single xylose unit to α-DG, which may be sufficient to be detected by IIH6C4, albeit generating a very weak signal. In this scenario, the IIH6C4 signal may not truthfully represent the physiological function, as it is uncertain how much function a single xylose possesses in comparison to the complete matriglycan chain. Nevertheless, AlphaMissense also presents a similar result, where missense variants generally exhibit lower AlphaMissense scores (higher functionality) in the GlcAT region of *LARGE1* than those in the XylT region (p-value<2.22e-16, Figure S4C). This suggests that the difference in these two domains is to some extent biologically relevant, that is the XylT domain may possess a more important functional role than the GlcAT domain.

It has been previously reported that FKTN, FKRP, and RXYLT1 (TMEM5) may form a complex[8]. The SMuRF scores of SNV-generated single amino acid substitutions were projected to the FKTN-FKRP-RXYLT1 complex predicted with Alphafold2[9,10] (Figure S5D). One interesting residue identified by SMuRF is p.Trp255 (mean SMuRF=-0.79), which is predicted to interact with FKTN p.Leu263. *FKTN* c.787C>G (p.Leu263Val) and *FKTN* c.788T>G (p.Leu263Arg) are variants of uncertain significance in ClinVar (Oct. 11th, 2023).

Method S1

FKRP-KO sgRNA:
GTTCGAGGCATTTGACAACG

LARGE1-KO sgRNAs (used as a mixture):
GATGGGATGGGGCTCGGCCC
GGTGCGCATGCGCGAGGTGG
TCCAGCGGTGACAGAGACAC

RNP preparation:
18 µL supplemented SE Solution + 6 µL 30 µM sgRNA + 1 µL 20 µM Cas9
Room temperature 10 mins
Add to 150 k cells (spun down 100 ×g, 10 mins; resuspended in 5 µL supplemented SE Solution)
SE Cell Line 4D-Nucleofector X Kit S was used (Lonza, V4XC-1032)

Pick single clones:
Allow the nucleofected cells to recover in the growth medium; plate the nucleofected cells sparsely and allow them to form monoclonal clusters; pick monoclonal cells under microscope.

Genotyping primers:
FKRP-GT-F: CATCACCCTCAACCTTCTGGTC
FKRP-GT-R: CATCAGGTACTAGGGCCACAAACTC
LARGE1-GT-F: GGCAATCGGGACTTTGGACA
LARGE1-GT-R: GCCTCGCCATGTAGTAAGGG

Method S2

Q5 High-Fidelity DNA Polymerase (NEB, M0491SVIAL) was used for the PCR reactions according to the manufacturers' manuals.

The *DAG1* CDS sequence in this project is:
ENST00000308775.7 but carrying a common missense variant c.41C>G (p.Ser14Trp)
The Allele Frequency of this variant is 0.9762 in gnomAD v3.1.2 and 0.9955 in gnomAD v4.0.0

The *FKRP* CDS sequence in this project is as in:
ENST00000318584.10

The *LARGE1* CDS sequence in this project is as in:
ENST00000354992.7

Primers for CDS plasmid construction:

DAG1-Exon2-F: GGTTCTAGAGCGCTGCCACCATGAGGATGTCTGTGGGCCTC
DAG1-Exon2-R: CCGCTGATACCTTGATGATATCTCCACTGGAGGCAAT
DAG1-Exon3-F: TATCATCAAGGTATCAGCGGCAGGGAAGGAG
DAG1-Exon3-R: AAGTTTGTTGCGCCGGATCCAGGTGGGACATAGGGAGGAGG
FKRP-clone-F (matching part): ATGCGGCTCACCCGCT
FKRP-clone-R (matching part): GCCGCTTCCCGTCAGA
LARGE1-clone-F (matching part): ATGCTGGGAATCTGCAGGG
LARGE1-clone-R (matching part): CTAGCTGTTGTTCTCGGCTGTGAG
UbC-clone-F (matching part): GGCCTCCGCGCCGG
UbC-clone-R (matching part): ACCAAGTGACGATCACAGCGATCC
EF1α-clone-F (matching part): GGCTCCGGTGCCCGTCA
EF1α-clone-R (matching part): GGTGGCCGTACGTCACGACA
BSD-WPRE-clone-F (matching part): ATGGCCAAGCCTTTGTCTCAAGAAG
BSD-WPRE-clone-R (matching part): GCGGGGAGGCGGCC
*The overhangs were designed according to different assembly schemes.

Two cloning strategies to introduce individual variant to the WT plasmid:
1. Cut and Swap: A short, localized region was cut with restriction enzymes from the wild-type plasmid to serve as the backbone. Two variant-carrying inserts were amplified using PCR. The variant of interest was introduced in the R primer for the 5' insert and the F primer for the 3' insert. The backbone and those two inserts were assembled with NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621)
2. Site-directed mutagenesis: The whole plasmid was amplified using a pair of primers carrying the variant of interest. The WT plasmid was removed from the reaction using Type2M enzyme DpnI (NEB, R0176S). F primers were designed as: (5'->3') 7-bp flanking sequence A + 1-bp variant + 7-bp flanking sequence B + 15-bp matching sequence; R primers were designed as (5'->3') 7-bp flanking sequence B + 1-bp variant + 7-bp flanking sequence A + 15-bp matching sequence.

Oxford Nanopore long-read sequencing services were provided by Plasmidsaurus to validate the plasmid sequences.

Plasmids deposited to Addgene:
Lenti-*DAG1*: 205149
Lenti-UbC-*FKRP*-EF1α-*BSD*: 205150
Lenti-UbC-*LARGE1*-EF1α-*BSD*: 205151

Method S3

RT-PCR, RT-qPCR and real-time PCR primers:
DAG1-cDNA-F: GATCTGCTACCGCAAGAAGC
DAG1-cDNA-R: ATGGTGTCCTGGTTCAGAGG
FKRP-rtPCR-F: TGGGCATCTACTTGGAGGAC
FKRP-rtPCR-R: TTGCTTTCGCTGTACTGCAC
LARGE1-cDNA-F: GAGCAGTGCTACAGAGACGTGT
LARGE1-cDNA-R: TGCCGTCATACTCCAGGAAGGT
LARGE1-gDNA-F: ATCAGCCTGGCCCTCTACC
LARGE1-gDNA-R: TGTAGGGAGTGCTGATGTGC

Method S4

Primers used for PALS-C:
PALS-C-universal-F1: GAACAGGCGAGGAAAAGTAG
PALS-C-universal-F2: GATCGTCACTTGGTACCGGTTCTAGA
PALS-C-universal-R2-FKRP: TGGCACTTTTCGGGGGATCCTC
PALS-C-universal-R2-LARGE1: TGGCACTTTTCGGGGGATCCCT

All possible SNVs were generated and visualized using the following scripts:
All possible SNVs package

R1 oligos were generated using the following Python scripts:
PALS_C_oligos_FKRP.py
PALS_C_oligos_LARGE1.py

R1 oligo structure (5'->3'):

8-nt block-specific adaptor
Type2S enzyme recognition site
5-nt block-specific insulator (the 3' 4-nt sequence does not appear in this block)
19-nt *GOI* downstream sequence
Variant
25-nt *GOI* upstream sequence

R1 oligo lists:
FKRP_oligo_to_order_out.fa
LARGE1_oligo_to_order_out.fa

Plasmid templates:
Lenti-UbC-FKRP-EF1α-BSD
Lenti-UbC-LARGE1-EF1α-BSD

Reagents:
Q5 Reaction Buffer (NEB, B9027SVIAL)
Q5 High GC Enhancer (NEB, B9028AVIAL)
10 mM dNTPs (NEB, N0447)
Q5 High-Fidelity DNA Polymerase (NEB, M0491SVIAL)
BsmBI-v2 (NEB, R0739S)
BsaI-HFv2 (NEB, R3733S)
DpnI (NEB, R0176S)
MspJI (NEB, R0661S)
XbaI (NEB, R0145S)
BamHI-HF (NEB, R3136S)
NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621)
Endura Electrocompetent Cells (Lucigen, 60242)

PALS-C Step1:

$$coverage = 10\text{^}6$$

$$oligo\ library\ input\ weight = (gene\ cds\ length) * coverage * (block\ number) * \frac{relative\ molecular\ mass\ of\ one\ oligo}{Avogadro\ constant}$$

$$plasmid\ template\ input\ weight = [(oligo\ library\ input\ weight) * (plasmid\ length) * 2]/(oligo\ length)$$

Q5 Rxn buffer	10 μL
Q5 Enh buffer	10 μL
10 mM dNTPs	1 μL
Oligo library	
Plasmid template	
Q5 polymerase*	2 μL
Water	To 50 μL*

1	98 °C	hot start
2	98 °C	4 min
3	Annealing temperature*	20 s
4	72 °C	Elongation time*
5	12 °C	forever

*Q5 polymerase is likely to be the limiting factor of Step1, volume of which requires optimization.

*Multiple 50-μL reactions are needed if product yield is insufficient for subsequent steps.
*Annealing temperature and elongation time are decided based on the product that is the most difficult to amplify. One universal condition was feasible for FKRP (70 °C; 40 s) and LARGE1 (68 °C ; 90 s); however, if a GOI requires drastically different conditions for different blocks, reactions with different conditions can be set accordingly.

PCR purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

PALS-C Step2:

Starting from Step2, the reactions are done independently for different blocks.

Q5 Rxn buffer	10 μL
Q5 Enh buffer	10 μL
10 mM dNTPs	1 μL
Purified Step1 product	*
10 μM Universal F1	2.5 μL
10 μM Block specific R1	2.5 μL
Q5 polymerase	0.5 μL
Water	To 50 μL

1	98 °C	hot start
2	98 °C	3 min
3	98 °C	8 s
4	Annealing temperature	20 s
5	72 °C	Elongation time
6	Repeat 3-5 for 34 more cycles	
7	72 °C	5 min
8	12 °C	Forever

*The input should be decided based on the position of the block. The more distant a block is from the 5' side, the more input is required. Evenly distributed input for all 6 blocks of FKRP generated enough yield for subsequent steps, while the 3' side blocks of LARGE1 required extra input. Step1-2 should be repeated if product yield is insufficient for subsequent steps.

PCR purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

PALS-C Step3:

Type2S enzyme*	2 μL
Purified Step2 product	1.2 μg
Reaction buffer	5 μL
Water	To 50 μL

Reaction temperature (Lid: 60 °C)	50 min
12 °C	Forever

*BsmBI was used for FKRP and BsaI was used for LARGE1. The type2S enzymes were picked to avoid the presence of their recognition sites within the CDS.

Electrophoresis in 1% agarose gel was performed and the bands within the correct size range were cut from the gel; gel purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

PALS-C Step4:

$$purified\ step3\ product\ input\ weight + plasmid\ template\ input\ weight = 1000\ ng$$

$$purified\ step3\ product\ input\ weight : plasmid\ template\ input\ weight$$
$$\approx\ step3\ product\ length : plasmid\ template\ length$$

$$The\ longest\ step3\ product\ length\ of\ a\ block\ can\ be\ used\ for\ the\ calculation.$$

Q5 Rxn buffer	10 µL
Q5 Enh buffer	10 µL
10 mM dNTPs	1 µL
Purified Step3 product	
Plasmid template	
Q5 polymerase*	2 µL
Water	To 50 µL

1	98 °C	hot start
2	98 °C	5 min
3*	72 °C	5 s
4*	66 °C	20 s
5	72 °C	Elongation time*
6	12 °C	forever

*Q5 polymerase is likely to be the limiting factor of Step4, volume of which requires optimization.
*Optional steps: the purpose is to enhance the annealing of all strands.
*The elongation time should be sufficient for the shortest strand to be elongated to the R2 primer site.

PCR purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

PALS-C Step5:

DpnI	0.5 µL
MspJI	0.5 µL
CutSmart	5 µL
Enzyme activator	1 µL
Purified Step4 product	500 ng
Water	To 50 µL

37 °C (Lid: 60 °C)	1 hr
12 °C	Forever

Column purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609). (Follow the PCR purification protocol.)

PALS-C Step6:

Q5 Rxn buffer	10 µL
Q5 Enh buffer	10 µL
10 mM dNTPs	1.5 µL
Purified Step5 product	100 ng
10 µM F2	2.5 µL
10 µM R2	2.5 µL
Q5 polymerase	1 µL
Water	To 50 µL

1	98 °C	hot start
2	98 °C	5 min
3	98 °C	6 s
4	Annealing temperature	20 s
5	72 °C	Elongation time
6	Repeat 3-5 for 34 more cycles	
7	72 °C	5 min
8	12 °C	Forever

Electrophoresis in 1% agarose gel was performed and the bands with the correct size were cut from the gel; gel purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

Important: Use 20 µL water or less to dissolve after gel purification; or use a vacuum concentrator to evaporate excess water.

PALS-C Step7:

Prepare the backbone:

XbaI	1.5 µL
BamHI-HF	1.5 µL
Plasmid template	3 µg
CutSmart	5 µL
Water	To 50 µL

37 °C (Lid: 60 °C)	40 min
12 °C	Forever

Electrophoresis in 1% agarose gel was performed and the band with the correct size was cut from the gel; gel purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609).

For Gibson assembly:

NEBuilderMaster Mix	20 µL
Backbone	210 ng
Purified Step6 product	140 ng
Water	To 30 µL

50 °C (Lid: 60 °C)	60 min
12 °C	Forever

PALS-C Step8:

For each block:

Electrocompetent cells	40 µL
Assembly reaction	4 µL
Water	160 µL

Split the mixed sample into two pre-chilled 0.1 cm cuvettes (Bio-Rad, 1652089).
Use Bio-Rad Gene Pulser II: (ATTENTION: avoid bubbles) 25 µF; 200 Ohms; 1800 volts.
Add the sample to 900 µL recovery media per cuvette: 250 rpm, 1hr, 37°C.
Combine transformed bacteria from 2 cuvettes.
Add 1/500 volume of the bacteria to 200 µL LB broth and plate it on an ampicillin LB agar plate for quick estimation of complexity. (We achieved 19.1 × to 115.3 x coverage for all blocks)
$coverage = 500 * (colony\ number)/(variant\ number\ of\ a\ block)$
Seed the bacteria in 150 mL LB broth supplemented with ampicillin (100 µg/ml) for overnight culture.

Method S5

Scripts used to list underrepresented variants due to the limitation of PALS-C:
Cloning alert package

Group1 variants introduce a cut before the variants which will remove the variant from the top strand;
Group2 variants introduce a cut after the variants which might affect the elongation of the top strand.

FKRP:

site	codon	WT_nt	Variant_nt	classification	alert_group
69	23	T	C	Synonymous	group_2
628	210	C	G	Missense	group_2
6	2	G	T	Synonymous	group_2
1034	345	G	T	Missense	group_2
1151	384	G	T	Missense	group_2
145	49	G	C	Missense	group_2

303	101	G	C	Synonymous	group_2
465	155	G	C	Synonymous	group_2
573	191	C	A	Nonsense	group_1
781	261	G	A	Missense	group_1
252	84	C	G	Missense	group_1
909	303	C	G	Missense	group_1
1290	430	C	G	Missense	group_1
439	147	G	A	Missense	group_1
591	197	T	C	Synonymous	group_1
1187	396	A	C	Missense	group_1
357	119	C	G	Synonymous	group_1
879	293	C	G	Synonymous	group_1
975	325	C	G	Synonymous	group_1

LARGE1:

site	codon	WT_nt	Variant_nt	classification	alert_group
119	40	T	G	Missense	group_2
2176	726	T	G	Missense	group_2
982	328	C	G	Missense	group_2
167	56	C	T	Missense	group_2
1556	519	G	T	Missense	group_2
1713	571	G	T	Synonymous	group_2
1432	478	G	T	Missense	group_2
666	222	G	C	Synonymous	group_2
298	100	A	G	Missense	group_1
559	187	C	G	Missense	group_1
1082	361	C	G	Nonsense	group_1
1859	620	G	A	Missense	group_1
838	280	A	G	Missense	group_1
139	47	T	A	Missense	group_1
269	90	C	A	Missense	group_1
1468	490	G	A	Missense	group_1
1528	510	G	A	Missense	group_1
1963	655	G	A	Missense	group_1
191	64	G	C	Missense	group_1
227	76	A	C	Missense	group_1
338	113	A	C	Missense	group_1
815	272	A	C	Missense	group_1
1347	449	G	C	Missense	group_1
2261	754	A	C	Missense	group_1
26	9	G	C	Missense	group_1
331	111	A	C	Missense	group_1
411	137	A	C	Synonymous	group_1
1117	373	G	C	Missense	group_1
1797	599	A	C	Synonymous	group_1
1994	665	G	C	Missense	group_1

Method S6

Primers used for plasmid pool QC:
FKRP-QC-block1-F: GCCAGAACACAGGACCGGTTCTAGA
FKRP-QC-block1-R: GCCAGGGGCGGGTAGGGGAGCGTGT
FKRP-QC-block2-F: GCCCAGCCCGTGGTGGTGGCAGCCG
FKRP-QC-block2-R: CAGGCTGACGTTCAGGGCCAGGCAC
FKRP-QC-block3-F: CGGTTGCCACGGCCAACCCTGCCAG
FKRP-QC-block3-R: GCTCAGCCTTCCAGCGCGCGTGGGC
FKRP-QC-block4-F: GGCGCGCCAGCCCCCGCTGGCCACG
FKRP-QC-block4-R: AGCCAGTAGCGCACGCCCGCAGCCT
FKRP-QC-block5-F: GCCCGCTATGTGGTGGGCGTGCTGG
FKRP-QC-block5-R: GCCATTGCGGGGGTAGAAGGGCCAC
FKRP-QC-block6-F: AAAGCAACCACTTGCACGTGGACCT
FKRP-QC-block6-R: GAGAGAAGTTTGTTGCGCCGGATCC
LARGE1-NGS-blk1-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTCACTTGGTACCGGTTCTAGA

LARGE1-NGS-blk2-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTCGCGAGGTGGAGGAGGAGAA
LARGE1-NGS-blk3-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTCGGATACAATGCCAGCCGGG
LARGE1-NGS-blk4-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTGTCTGATGAAGCTTGTCCTG
LARGE1-NGS-blk5-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTACAGGGGTGATCCTGTTACT
LARGE1-NGS-blk6-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTCGTGTCTGATCTAAAGGTCA
LARGE1-NGS-blk7-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTGAGAGCGCTTCACTGTCCAC
LARGE1-NGS-blk8-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTATGAGCCGCCACAACGTGGG
LARGE1-NGS-blk9-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTACTGCGCTACCGGCTGTCCT
LARGE1-NGS-blk10-QC-F:
CCCTACACGACGCTCTTCCGATCTCCTACGTTAGGCTTTGGCTGGAACAAA
LARGE1-NGS-blk1-R:
GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCCTGCGGAGGGCGCGG
LARGE1-NGS-blk2-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTGACCAGGGTGACGACAT
LARGE1-NGS-blk3-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTGGCAGGAAGAGTCTTGGT
LARGE1-NGS-blk4-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTTCCGCAGCTTATCCAGA
LARGE1-NGS-blk5-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGGGGGAGTTCCAGTGAA
LARGE1-NGS-blk6-R:
GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGAAGTACAGGTGGGTGCG
LARGE1-NGS-blk7-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTGTACACGATGTGGTAG
LARGE1-NGS-blk8-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCCGCTTTTGACTTGGGGA
LARGE1-NGS-blk9-R: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCTCCATGATATGAGCCAC
LARGE1-NGS-blk10-R:
GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACTTTTCGGGGGATCCCTA
NGS-PCR3-F: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
NGS-PCR3-R:
CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCTCTT

For *FKRP* plasmid pool QC:
For the plasmid pool of each block:

Q5 Rxn buffer	10 µL
Q5 Enh buffer	10 µL
10 mM dNTPs	1 µL
Plasmid	~300 ng
10 µM F primer	2.5 µL
10 µM R primer	2.5 µL
Q5 polymerase	0.5 µL
Water	To 50 µL

1	98 °C	hot start
2	98 °C	3 min
3	98 °C	6 s
4	70 °C	15 s
5	72 °C	5 s
6	Repeat 3-5 for 32 more cycles	
7	72 °C	5 min
8	12 °C	Forever

Electrophoresis and gel purification (Takara, 740609) were performed.

The purified products were sent for Amplicon-EZ sequencing.

For *LARGE1* plasmid pool QC:
For the plasmid pool of each block:

Q5 Rxn buffer	10 µL
Q5 Enh buffer	10 µL
10 mM dNTPs	1 µL
Plasmid	~50 ng
10 µM F primer	2.5 µL
10 µM R primer	2.5 µL
Q5 polymerase	0.5 µL
Water	To 50 µL

1	98 °C	hot start
2	98 °C	3 min
3	98 °C	6 s
4	Annealing temperature	15 s
5	72 °C	7 s
6	Repeat 3-5 for 33 more cycles	
7	72 °C	5 min
8	12 °C	Forever

Annealing temperature	Blocks
61 °C	1, 4, 6, 7
64 °C	5, 9, 10
66 °C	2, 3, 8

Electrophoresis and gel purification (Takara, 740609) were performed.
The purified products were mixed for the following reaction:

Q5 Rxn buffer	10 µL
Q5 Enh buffer	10 µL
10 mM dNTPs	1 µL
Mixed purified products	100 ng
10 µM NGS-PCR3-F	2.5 µL
10 µM NGS-PCR3-R	2.5 µL
Q5 polymerase	1 µL
Water	To 50 µL

1	98 °C	hot start
2	98 °C	3 min
3	98 °C	6 s
4	72 °C	15 s
5	72 °C	8 s
6	Repeat 3-5 for 19 more cycles	
7	72 °C	5 min
8	12 °C	Forever

PCR purification (Takara, 740609) was performed and the sample was sent for Amplicon-EZ sequencing.

Scripts for the plasmid pool QC analytical pipelines:
FKRP plasmid QC package
LARGE1 plasmid QC package

Method S7

Variants in the mini-libraries/used for individual validation:

Variant	Clinical significance (ClinVar)
NM_024301.5(FKRP):c.77G>A (p.Trp26Ter)	Pathogenic/Likely pathogenic
NM_024301.5(FKRP):c.266C>T (p.Pro89Leu)	Pathogenic/Likely pathogenic
NM_024301.5(FKRP):c.427C>A (p.Arg143Ser)	Benign/Likely benign
NM_024301.5(FKRP):c.520A>T (p.Ser174Cys)	Benign/Likely benign
NM_024301.5(FKRP):c.663C>A (p.Ser221Arg)	Pathogenic
NM_024301.5(FKRP):c.822C>G (p.Ile274Met)	Conflicting interpretations of pathogenicity
NM_024301.5(FKRP):c.826C>A (p.Leu276Ile)	Pathogenic/Likely pathogenic
NM_024301.5(FKRP):c.1364C>A (p.Ala455Asp)	Pathogenic
NM_133642.5(LARGE1):c.992C>T (p.Ser331Phe)	Pathogenic

NM_133642.5(LARGE1):c.1008T>C (p.Asp336=)	Benign
NM_133642.5(LARGE1):c.1102C>T (p.Gln368Ter)	Pathogenic

Primers used for lentiviral sequence isolation:
PALS-C-universal-F2: GATCGTCACTTGGTACCGGTTCTAGA
PALS-C-universal-R2-FKRP: TGGCACTTTTCGGGGGATCCTC
PALS-C-universal-R2-LARGE1: TGGCACTTTTCGGGGGATCCCT

Primers used for *FKRP* Sanger sequencing:
FKRP-QC-block2-F: GCCCAGCCCGTGGTGGTGGCAGCCG
FKRP-QC-block3-F: CGGTTGCCACGGCCAACCCTGCCAG
FKRP-QC-block5-F: GCCCGCTATGTGGTGGGCGTGCTGG
FKRP-GT-R: CATCAGGTACTAGGGCCACAAACTC

Primers used for *FKRP* Amplicon-EZ sequencing:
Amplicon1:
PALS-C-universal-F2: GATCGTCACTTGGTACCGGTTCTAGA
FKRP-GT-R: CATCAGGTACTAGGGCCACAAACTC
Amplicon2:
FKRP-QC-block2-F: GCCCAGCCCGTGGTGGTGGCAGCCG
FKRP-QC-block3-R: GCTCAGCCTTCCAGCGCGCGTGGGC
Amplicon3:
FKRP-QC-block3-F: CGGTTGCCACGGCCAACCCTGCCAG
FKRP-QC-block3-R: GCTCAGCCTTCCAGCGCGCGTGGGC
Amplicon4:
FKRP-QC-block4-F: GGCGCGCCAGCCCCCGCTGGCCACG
FKRP-QC-block4-R: AGCCAGTAGCGCACGCCCGCAGCCT
Amplicon5:
FKRP-QC-block6-F: AAAGCAACCACTTGCACGTGGACCT
FKRP-CDS-backbone-junction-R: GTGGCACTTTTCGGGGGATCCTCAGCCGCTTCCCGTCAGACTC

Primers used for *LARGE1* Sanger sequencing:
LARGE1-block4-seq-F: TTATGGTCTGATGAAGCTTGTCCTG

Method S8

PCR1:

PCR1-F	GATCGTCACTTGGTACCGGTTCTAGA
PCR1-R (FKRP)	TGGCACTTTTCGGGGGATCCTC
PCR1-R (LARGE1)	TGGCACTTTTCGGGGGATCCCT

Reagent	Catalog	Volume or weight
Q5 Reaction Buffer	NEB, B9027SVIAL	10 µL
Q5 High GC Enhancer	NEB, B9028AVIAL	10 µL
10 mM dNTPs	NEB, N0447	1 µL
Q5 High-Fidelity DNA Polymerase	NEB, M0491SVIAL	1 µL
10µM PCR1-F		2.5 µL
10µM PCR1-R		2.5 µL
gDNA		0.3-1 µg

Nuclease-Free Water		To 50 µL

Step1	98 °C	Hot start
Step2	98 °C	3 mins
Step3	98 °C	8 s
Step4	68 °C	20 s
Step5	72 °C	45 s
Step 3-5, 35 cycles		
Step6	72 °C	5 mins
Step7	12 °C	Infinite

Electrophoresis in 1% agarose gel. Product band size is 1534 bp (FKRP) or 2317 bp (LARGE1). Gel purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609). Elute with 25 µL Nuclease-Free Water.

PCR2:

FKRP-blk1-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCACTTGGTACCGGTTCTAGA
FKRP-blk2-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCGCCCGTGGTGGTGGCAGCCG
FKRP-blk3-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCGCCACGGCCAACCCTGCCAG
FKRP-blk4-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCGCCAGCCCCCGCTGGCCACG
FKRP-blk5-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCTATGTGGTGGGCGTGCTGG
FKRP-blk6-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCAACCACTTGCACGTGGACCT
FKRP-blk1-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCACTTGGTACCGGTTCTAGA
FKRP-blk2-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGGCCCGTGGTGGTGGCAGCCG
FKRP-blk3-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGGCCACGGCCAACCCTGCCAG
FKRP-blk4-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGGCCAGCCCCCGCTGGCCACG
FKRP-blk5-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCTATGTGGTGGGCGTGCTGG
FKRP-blk6-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGAACCACTTGCACGTGGACCT
FKRP-blk1-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCGGGTAGGGGAGCGTGT
FKRP-blk2-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGACGTTCAGGGCCAGGCAC
FKRP-blk3-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTTCCAGCGCGCGTGGGC
FKRP-blk4-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAGCGCACGCCCGCAGCCT
FKRP-blk5-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCGGGGGTAGAAGGGCCAC
FKRP-blk6-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACTTTTCGGGGGATCCTCA
LARGE1-blk1-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCACTTGGTACCGGTTCTAGA

LARGE1-blk2-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCGCGAGGTGGAGGAG GAGAA
LARGE1-blk3-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCGGATACAATGCCAGC CGGG
LARGE1-blk4-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCGTCTGATGAAGCTTGT CCTG
LARGE1-blk5-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCACAGGGGTGATCCTGT TACT
LARGE1-blk6-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCCGTGTCTGATCTAAAG GTCA
LARGE1-blk7-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCGAGAGCGCTTCACTGT CCAC
LARGE1-blk8-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCATGAGCCGCCACAACG TGGG
LARGE1-blk9-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCACTGCGCTACCGGCTG TCCT
LARGE1-blk10-high-F	CCCTACACGACGCTCTTCCGATCTTACACGATCTAGGCTTTGGCTGGAA CAAA
LARGE1-blk1-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCACTTGGTACCGGTTCT AGA
LARGE1-blk2-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCGCGAGGTGGAGGAGG AGAA
LARGE1-blk3-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCGGATACAATGCCAGCC GGG
LARGE1-blk4-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGGTCTGATGAAGCTTGTC CTG
LARGE1-blk5-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGACAGGGGTGATCCTGTT ACT
LARGE1-blk6-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGCGTGTCTGATCTAAAGG TCA
LARGE1-blk7-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGGAGAGCGCTTCACTGTC CAC
LARGE1-blk8-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGATGAGCCGCCACAACGT GGG
LARGE1-blk9-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGACTGCGCTACCGGCTGT CCT
LARGE1-blk10-low-F	CCCTACACGACGCTCTTCCGATCTAAGTAGAGTAGGCTTTGGCTGGAAC AAA
LARGE1-blk1-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCCTGCGGAGGGC GCGG
LARGE1-blk2-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTGACCAGGGTGACGA CAT
LARGE1-blk3-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTGGCAGGAAGAGTCTT GGT
LARGE1-blk4-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTTCCGCAGCTTATCC AGA
LARGE1-blk5-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGGGGGAGTTCCAGT GAA
LARGE1-blk6-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGAAGTACAGGTGGGT GCG
LARGE1-blk7-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTGTACACGATGTGG TAG
LARGE1-blk8-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCCGCTTTTGACTTGG GGA
LARGE1-blk9-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCTCCATGATATGAGC

	CAC
LARGE1-blk10-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACTTTTCGGGGGATCC CTA

Reagent	Catalog	Volume or weight
Q5 Reaction Buffer	NEB, B9027SVIAL	10 µL
Q5 High GC Enhancer	NEB, B9028AVIAL	10 µL
10 mM dNTPs	NEB, N0447	1 µL
Q5 High-Fidelity DNA Polymerase	NEB, M0491SVIAL	1 µL
10µM PCR2-F		2.5 µL
10µM PCR2-R		2.5 µL
Purified PCR1 product		0.2-0.5 µg
Nuclease-Free Water		To 50 µL

Step1	98 °C	Hot start
Step2	98 °C	3 mins
Step3	98 °C	6 s
Step4	Annealing temperature	15 s
Step5	72 °C	7 s
Step 3-5, 25 cycles		
Step6	72 °C	5 mins
Step7	12 °C	Infinite

Annealing temperature:

61 °C	LARGE1-blk1, LARGE1-blk4, LARGE1-blk6, LARGE1-blk7
64 °C	FKRP-blk1, LARGE1-blk5, LARGE1-blk9, LARGE1-blk10
66 °C	LARGE1-blk2, LARGE1-blk3, LARGE1-blk8
68 °C	FKRP-blk6
72 °C	FKRP-blk2, FKRP-blk3, FKRP-blk4, FKRP-blk5

PCR purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609). Elute with 40 µL Nuclease-Free Water.

<u>PCR3:</u>

PCR3-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PCR3-R	CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCTCTT

Mix purified PCR2 products (200 ng each). Dilute the mixed sample to 11 ng/µL.

Reagent	Catalog	Volume or weight
Q5 Reaction Buffer	NEB, B9027SVIAL	10 µL
Q5 High GC Enhancer	NEB, B9028AVIAL	10 µL
10 mM dNTPs	NEB, N0447	1 µL
Q5 High-Fidelity DNA Polymerase	NEB, M0491SVIAL	1 µL
10µM PCR3-F		2.5 µL
10µM PCR3-R		2.5 µL
Mixed sample		23 µL

Step1	98 °C	Hot start
Step2	98 °C	3 mins
Step3	98 °C	6 s
Step4	72 °C	15 s

Step5	72 °C	8 s
Step 3-5, 25 cycles		
Step6	72 °C	5 mins
Step7	12 °C	Infinite

Set 3*50 µL reactions.

PCR purification with NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609). Elute with 50 µL Nuclease-Free Water for each column. The purified PCR3 product was used for the next-generation sequencing. Amplicon-EZ service (GENEWIZ) was used to check the sample quality and coverage. The sample was then sequenced using the HiSeq X service (Psomagen; size selection also performed there).

Method S9

Protein extraction and quantitation:

Ice-cold protein extraction lysis buffer made by adding 10 µL of protease inhibitor cocktail (Sigma Aldrich, P8340-1ML) to every 1 mL of RIPA lysis and extraction buffer (Pierce, PI89900) was used to treat the cells. After 15 minutes of rocking incubation on ice, the samples were centrifuged for 30 minutes at 16,000 × g at 4 °C in a refrigerated centrifuge (Eppendorf, 5404). The supernatant was transferred to a clean 1.5 mL microcentrifuge tube for the following steps.

Protein was quantitated using the BioRad DC Assay II kit (BioRad, 5000112). Using the BSA reagent from the kit, seven standards are prepared from 0–1.5 mg/mL in the protein extraction lysis buffer from above to create a standard curve. Measurements were done according to the instruction manual, and absorbances at 750nm were read on a plate reader (BioRad, #1681135). Concentration is calculated relative to the standard curve.

SDS-PAGE and western blotting:

Appropriate amount of protein extraction lysis buffer from above was added into each protein sample (~10 µg) to balance the samples to achieve an equal volume. 4x Laemlli (BioRad, 1610747) was supplemented with 10% b-mercaptoethanol. 1 µL of this supplemented buffer was then added to every 3 µL of the balanced sample. Samples were mixed by a brief vortex and then heated at ~95°C on a heat

block (BioRad, 1660571) for 5 minutes. Samples were cooled down on ice and then centrifuged briefly at max speed to ensure equal loading (Eppendorf, 5404).

3.5 µL of PageRuler Plus prestained protein ladder (ThermoFisher Scientific, 26619) and all the samples were then loaded onto a 4-15% mini-protean TGX stain free protein gel (*e.g.*, BioRad, 4568084) that had been placed into a Mini-PROTEAN tetra cell electrophoresis system (BioRad, 1658000). SDS-PAGE was performed at 120V for ~60 minutes using a PowerPac HC power supply (BioRad, 1645052). Upon completion of the SDS-PAGE, the gel was removed and transferred onto a nitrocellulose membrane (BioRad, 1704158) by the TransBlot Turbo System (BioRad, 1704150EDU) using the 7-minute transfer program (2.5 A constant, up to 25 V variable).

Vinculin was used as a loading control. Hence, the membrane was cut at the correct position for separate FKRP and vinculin WBs. Membrane blocking buffer was made of 5% Omniblock non-fat dry milk (American Bio, AB10109-00100) in 1× TBS-Tween 20 (TBS-T). The membrane pieces were then blocked for 1 hour in membrane-blocking buffer separately.

Primary Anti-FKRP antibody (Abcam, ab220059) was made up 1:1,000 in membrane-blocking buffer. Primary Anti-Vinculin antibody (Sigma Aldrich, V9131-100 µl) was made up 1:80,000 in membrane-blocking buffer. The membrane pieces were incubated in the corresponding primary antibodies overnight at 4°C whilst rocking to equally distribute antibody (BenchRocker 2D Genesee Scientific, 31-201).

The following day, the membrane pieces were washed 5 times for 5 minutes in TBS-T and incubated in corresponding secondary antibodies for 2 hours at room temperature. Anti-rabbit HRP conjugated secondary antibody (Cell Signaling Technology, 7074S) was used for FKRP at 1:2,000 in membrane-blocking buffer. Anti-mouse HRP conjugated secondary antibody (Cell Signaling Technology, 7076S) was used for vinculin at 1:2,000 in membrane-blocking buffer.

Membrane pieces were washed again 5 times for 5 mins at room temperature in TBS-T prior to developing with Clarity Max ECL substrate (BioRad, 1705062) or Clarity ECL substrate (BioRad, 1705060). Using the Chemidoc Imaging system (BioRad, 12005606), both chemiluminescent and colorimetric images were taken to observe protein expression and size of bands.

Method S10

Packaging ppVSV-LASV-GPC viral particles:

General information:

* The LASV-GPC plasmid and the ppVSVDG-VSV-G viral particles were obtained from Dr. Melinda Brindley's group.
* The ppVSVDG-VSV-G viral particles carry a GFP-coding sequence in their genomes, which can be used to determine the TCID50 (tissue culture infectious dose that will infect 50% of the cell monolayers).
* TCID50 of the viral particles was experimentally determined using a protocol provided by Dr. Melinda A. Brindley. The protocol was based on the Spearman-Karber method.
* To calculate the multiplicity of infection (MOI) based on TCID50, a working estimate was adopted:

$$plaque\ forming\ units\ (pfu)\ =\ 0.5\ *\ TCID50$$
$$MOI\ =\ pfu\ of\ virus\ /\ number\ of\ cells$$

* Exercise standard caution when handling viral particles!
* Biosafety Level 2 (BSL-2) for activities with materials and cultures known or reasonably expected to contain VSV.

Making ppVSV-LASV-GPC-Generation1:

Day 0	Seed HEK293T cells in a well of a 6-well plate.
Day 1	Perform Lipofectamine 3000 (Invitrogen, L3000001) transfection per manufacturer's instruction.
	*HEK293T should reach 70%-90% confluency.
	*4 µg LASV-GPC plasmid was used to transfect the cells.
Day 2	24 hrs after transfection, ppVSVDG-VSV-G (MOI=0.5) was added to the well.
	The cell number is estimated to be ~2 M. Use this number to decide the viral dose.
	1 hr after adding the ppVSVDG-VSV-G, the medium was removed, the well was washed with DPBS (Gibco, 14190144), and fresh medium was added to the well.
Day 3	Collect the newly generated viral particles, *i.e.,* ppVSV-LASV-GPC-Generation1, and titrate its titer.

Making ppVSV-LASV-GPC-Generation2:

Day 0	Seed 6M HEK293T cells in a 10-cm dish.
Day 1	Perform Lipofectamine 3000 (Invitrogen, L3000001) transfection per manufacturer's instruction.
	*30 µg LASV-GPC plasmid was used to transfect the cells.
Day 2	24 hrs after transfection, ppVSV-LASV-GPC-Generation1 (MOI=0.1) was added to the well.
	The cell number is estimated to be ~12 M. Use this number to decide the viral dose.
	1 hr after adding the ppVSV-LASV-GPC-Generation1, the medium was removed, the well was washed with DPBS (Gibco, 14190144), and fresh medium was added to the well.
Day 3	Collect the newly generated viral particles, *i.e.,* ppVSV-LASV-GPC-Generation2, and titrate its titer.

Using ppVSV-LASV-GPC-Generation2, ppVSV-LASV-GPC-Generation3 can be packaged, similar to ppVSV-LASV-GPC-Generation2.

NGS library construction for VSV-related experiments was performed using a procedure similar to that outlined in Method S8. The "high" primers were utilized for the infected groups, while the primers provided below were used for the non-infected groups.

FKRP-blk1-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCACTTGGTACCGGTTCTAGA
FKRP-blk2-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTGCCCGTGGTGGTGGCAGCCG
FKRP-blk3-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTGCCACGGCCAACCCTGCCAG
FKRP-blk4-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTGCCAGCCCCCGCTGGCCACG
FKRP-blk5-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCTATGTGGTGGGCGTGCTGG
FKRP-blk6-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTAACCACTTGCACGTGGACCT
LARGE1-blk1-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCACTTGGTACCGGTTCTAGA
LARGE1-blk2-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCGCGAGGTGGAGGAGGAGAA
LARGE1-blk3-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCGGATACAATGCCAGCCGGG
LARGE1-blk4-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTGTCTGATGAAGCTTGTCCTG
LARGE1-blk5-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTACAGGGGTGATCCTGTTACT

LARGE1-blk6-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTCGTGTCTGATCTAAAGGTCA
LARGE1-blk7-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTGAGAGCGCTTCACTGTCCAC
LARGE1-blk8-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTATGAGCCGCCACAACGTGGG
LARGE1-blk9-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTACTGCGCTACCGGCTGTCCT
LARGE1-blk10-noinf-F	CCCTACACGACGCTCTTCCGATCTCCTACGTTAGGCTTTGGCTGGAACAAA

Considerations for ppVSV application:
In our experiment, we noticed that the Lenti-*LARGE1* rescued *LARGE1*-KO MB135 appeared to be more sensitive to the ppVSV infection when compared to the Lenti-*FKRP* rescued *FKRP*-KO MB135. For *FKRP* application, the ppVSV MOI should be controlled to be approximated 2.5, while for *LARGE1* application, the MOI should be around 1. For other genes, we recommend testing the MOI in a range of 1-3. Lentiviral MOI, ppVSV MOI, and cell confluency can all affect the sensitivity of the ppVSV assay. We recommend starting the optimization with a low lentiviral MOI (<1) and a 20-40% confluency.

Method S11

Variants in the mini-libraries:

Gene	Variant	AA	ClinVar	Inclusion criteria
FKRP	c.135C>T	p.Ala45=	Benign	Highest AF in gnomAD v4 High SMuRF score: 0.17
FKRP	c.663C>A	p.Ser221Arg	Pathogenic	Low SMuRF score: -2.21
FKRP	c.1174T>G	p.Phe392Val	N/A	Previously unknown variants Observed in a patient with compound mutations Low SMuRF score: -1.13
FKRP	c.436G>T	p.Glu146Ter	N/A	3rd highest Nonsense SMuRF score SMuRF score: -1.08
FKRP	c.673C>T	p.Gln225Ter	N/A	2nd highest Nonsense SMuRF score SMuRF score: -0.76
FKRP	c.1037C>A	p.Ser346Ter	N/A	Low SMuRF score: -1.92
FKRP	c.1477G>T	p.Gly493Ter	N/A	Highest Nonsense SMuRF score SMuRF score: -0.44 The most 3' nonsense variants
FKRP	c.783A>T	p.Gly261=	N/A	Low SMuRF score: -3.42

				High confidence
FKRP	c.783A>G	p.Gly261=	N/A	High SMuRF score: 0.02 High confidence
FKRP	c.783A>C	p.Gly261=	N/A	Low SMuRF score: -2.62 Low confidence
LARGE1	c.435C>T	p.Ala145=	Benign	Highest AF in gnomAD v4 High SMuRF score: -0.43
LARGE1	c.992C>T	p.Ser331Phe	Pathogenic	Low SMuRF score: -2.58
LARGE1	c.639T>C	p.Asn213=	N/A	Disrupting a poly(A) motif Higher SMuRF score: 0.45
LARGE1	c.642A>G	p.Lys214=	N/A	Disrupting a poly(A) motif Higher SMuRF score: 0.64
LARGE1	c.690T>A	p.Thr230=	N/A	Low SMuRF score: -1.28 High confidence
LARGE1	c.693T>G	p.Leu231=	N/A	Low SMuRF score: -1.33 High confidence
LARGE1	c.699C>G	p.Ala233=	N/A	Low SMuRF score: -1.67 Low confidence
LARGE1	c.26G>C	p.Arg9Pro	N/A	SMuRF score: 0.18
LARGE1	c.1050C>G	p.Tyr350Ter	N/A	Low SMuRF score: -2.15
LARGE1	c.1720G>T	p.Glu574Ter	N/A	2nd highest Nonsense SMuRF score SMuRF score: -1.36
LARGE1	c.2257G>T	p.Glu753Ter	N/A	Highest Nonsense SMuRF score SMuRF score: 0.15 The most 3' nonsense variants

Analytical script example:
plasmidsaurus_enrichment_calculation_v2.py

Considerations for nanopore sequencing:
Reproducible high noise may arise at certain sites due to the nature of nanopore sequencing. It is recommended to ensure that such noise does not impact the enrichment calculations. If it does, consider conducting short-read sequencing instead.

Supplemental references

1. Pontiller, J., Gross, S., Thaisuchat, H., Hesse, F., and Ernst, W. (2008). Identification of CHO endogenous promoter elements based on a genomic library approach. Mol. Biotechnol. *39*, 135–139. 10.1007/s12033-008-9044-9.

2. Dhindsa, R.S., Wang, Q., Vitsios, D., Burren, O.S., Hu, F., DiCarlo, J.E., Kruglyak, L., MacArthur, D.G., Hurles, M.E., and Petrovski, S. (2022). A minimal role for synonymous variation in human disease. Am. J. Hum. Genet. *109*, 2105–2109. 10.1016/j.ajhg.2022.10.016.

3. Mueller, W.F., Larsen, L.S.Z., Garibaldi, A., Hatfield, G.W., and Hertel, K.J. (2015). The silent sway of splicing by synonymous substitutions. J. Biol. Chem. *290*, 27700–27711. 10.1074/jbc.M115.684035.

4. Gaither, J.B.S., Lammi, G.E., Li, J.L., Gordon, D.M., Kuck, H.C., Kelly, B.J., Fitch, J.R., and White, P. (2021). Synonymous variants that disrupt messenger RNA structure are significantly constrained in the human population. Gigascience *10*. 10.1093/gigascience/giab023.

5. Zeng, Z., Aptekmann, A.A., and Bromberg, Y. (2021). Decoding the effects of synonymous variants. Nucleic Acids Res. *49*, 12673–12691. 10.1093/nar/gkab1159.

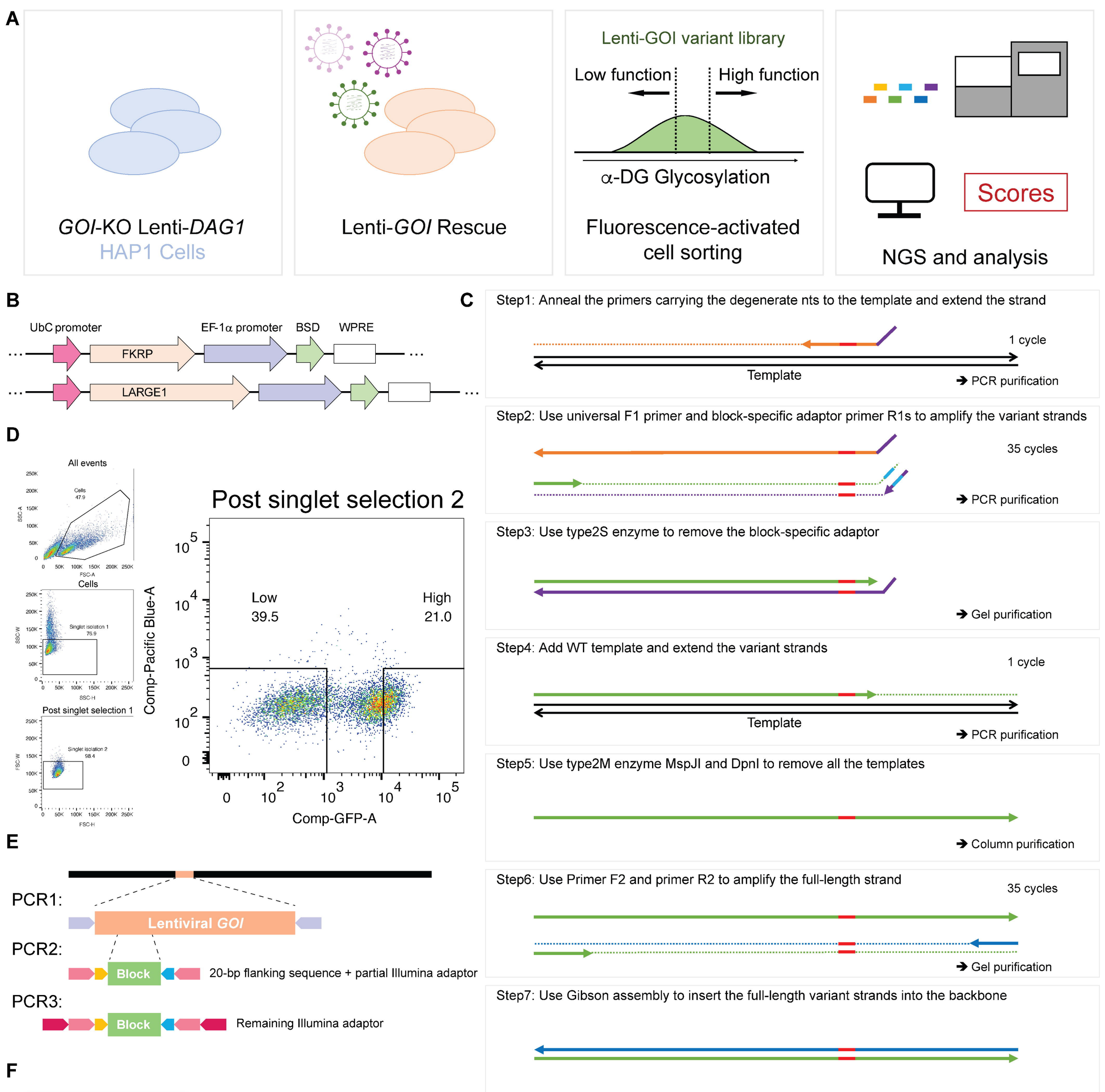
6. Zhang, Y., Liu, L., Qiu, Q., Zhou, Q., Ding, J., Lu, Y., and Liu, P. (2021). Alternative polyadenylation: methods, mechanism, function, and role in cancer. J. Exp. Clin. Cancer Res. *40*, 51. 10.1186/s13046-021-01852-7.

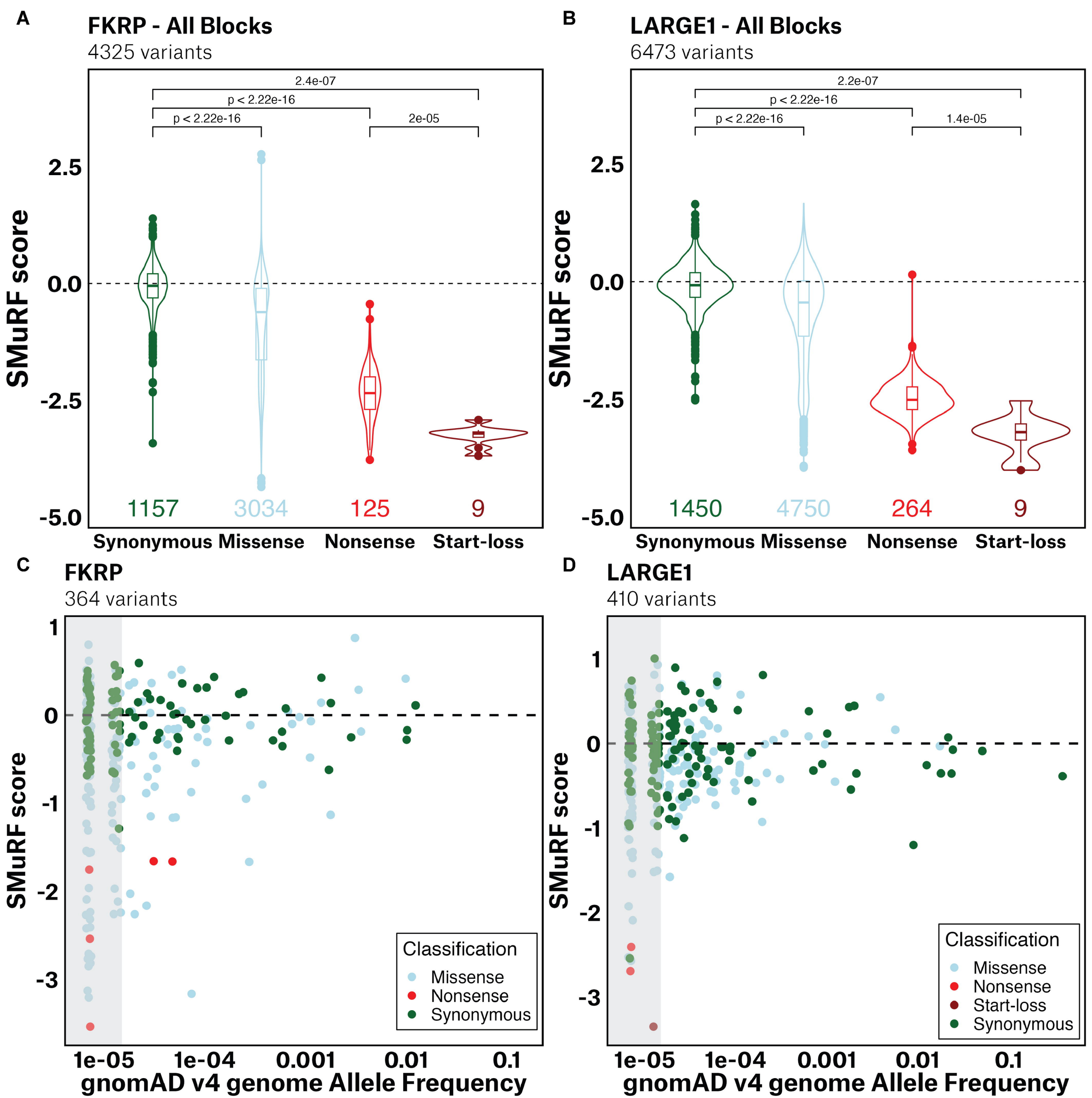
7. Stein, D., Bayrak, Ç.S., Wu, Y., Stenson, P.D., Cooper, D.N., Schlessinger, A., and Itan, Y. (2022). Genome-wide prediction of pathogenic gain- and loss-of-function variants from ensemble learning of diverse feature set. BioRxiv. 10.1101/2022.06.08.495288.

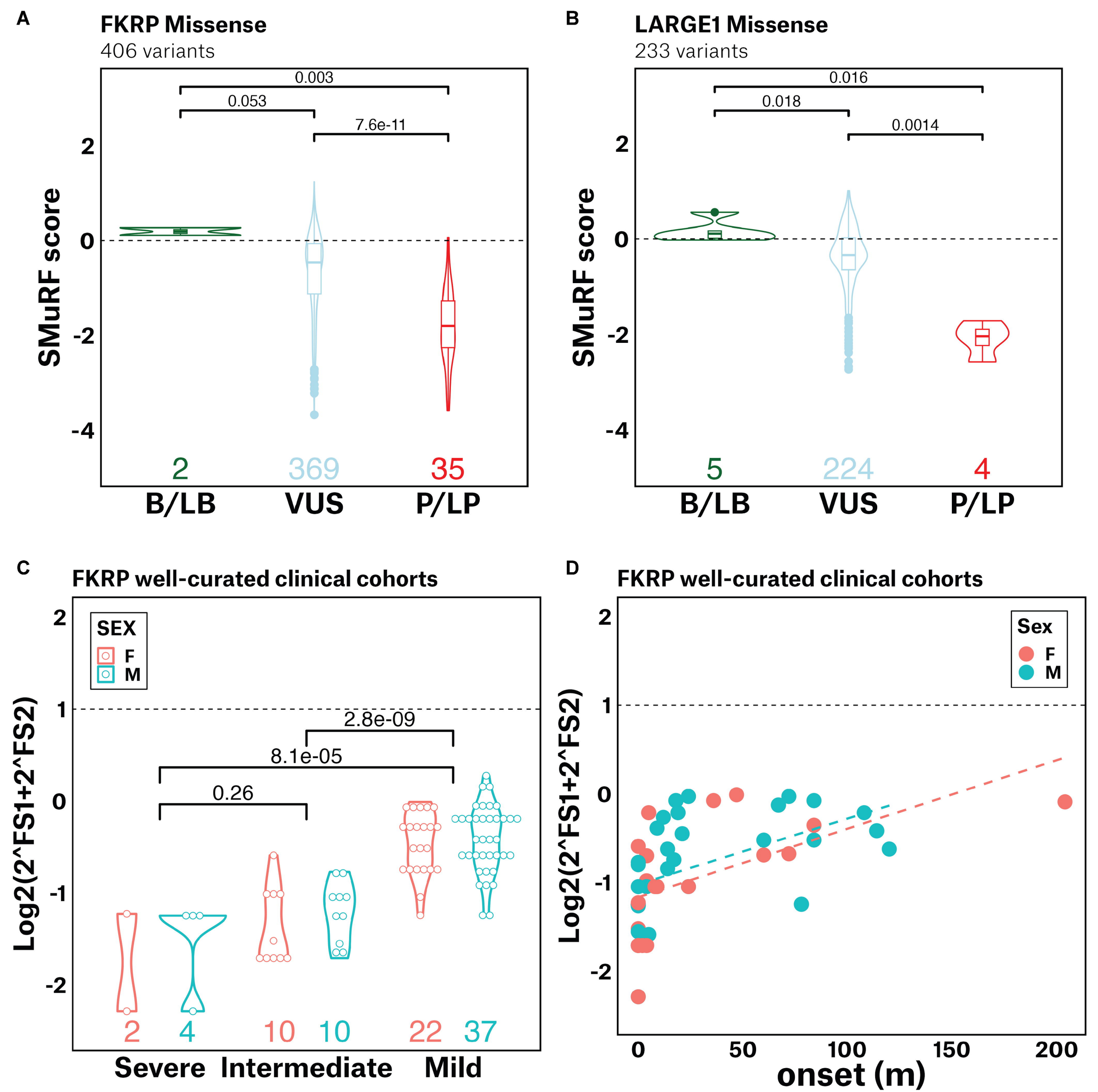
8. Nishihara, R., Kobayashi, K., Imae, R., Tsumoto, H., Manya, H., Mizuno, M., Kanagawa, M., Endo, T., and Toda, T. (2018). Cell endogenous activities of fukutin and FKRP coexist with the ribitol xylosyltransferase, TMEM5. Biochem. Biophys. Res. Commun. *497*, 1025–1030. 10.1016/j.bbrc.2018.02.162.

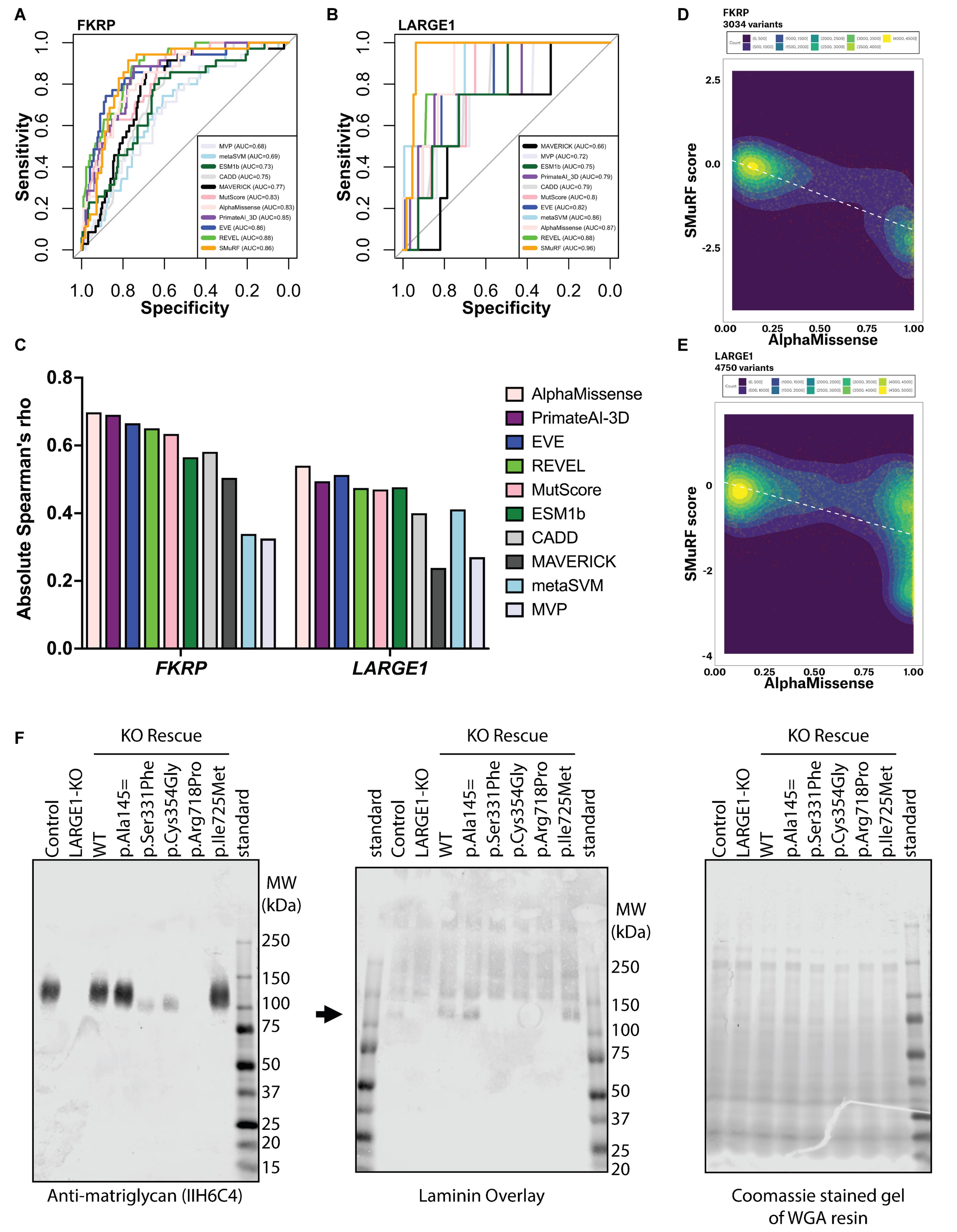
9. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. 10.1038/s41592-022-01488-1.

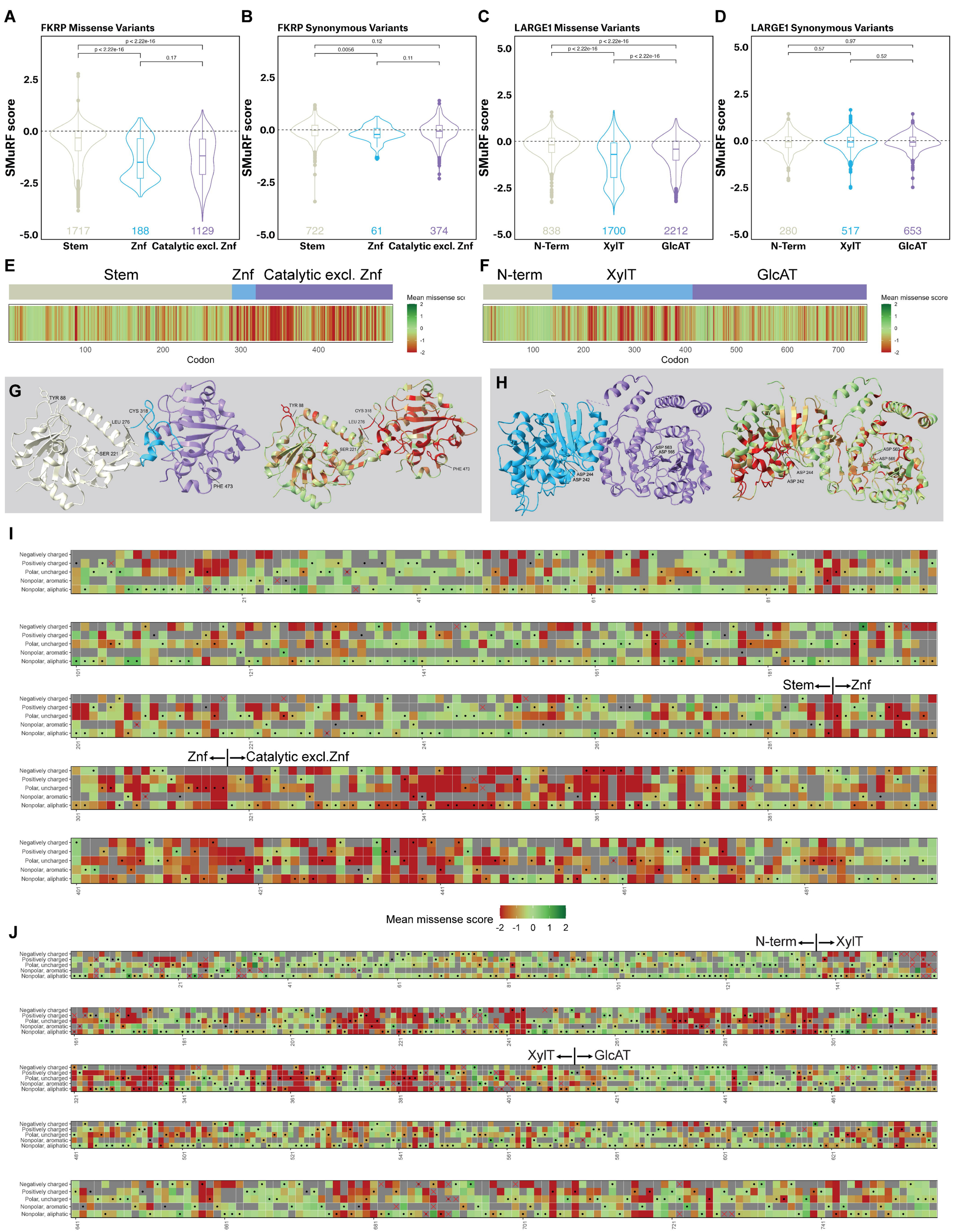
10. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A.W., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. BioRxiv. 10.1101/2021.10.04.463034.

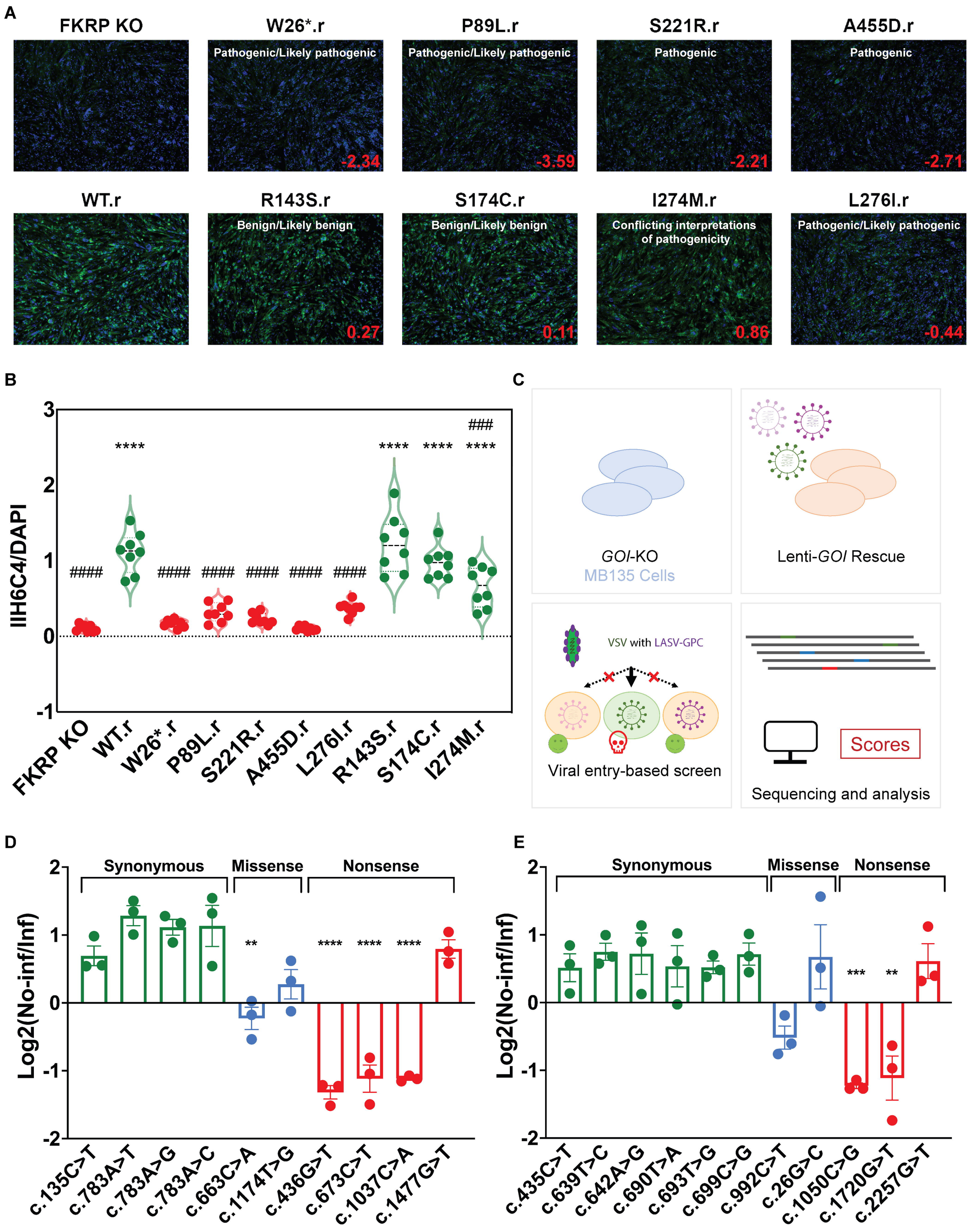

A
GOI-KO Lenti-DAG1
HAP1 Cells
Lenti-GOI Rescue
Lenti-GOI variant library
Low function
High function
α-DG Glycosylation
Fluorescence-activated cell sorting
Scores
NGS and analysis
B
UbC promoter
EF-1α promoter
BSD
WPRE
FKRP
LARGE1
C
Step1: Anneal the primers carrying the degenerate nts to the template and extend the strand
1 cycle
Template
PCR purification
Step2: Use universal F1 primer and block-specific adaptor primer R1s to amplify the variant strands
35 cycles
PCR purification
Step3: Use type2S enzyme to remove the block-specific adaptor
Gel purification
Step4: Add WT template and extend the variant strands
1 cycle
Template
PCR purification
Step5: Use type2M enzyme MspJI and DpnI to remove all the templates
Column purification
Step6: Use Primer F2 and primer R2 to amplify the full-length strand
35 cycles
Gel purification
Step7: Use Gibson assembly to insert the full-length variant strands into the backbone
D
All events
Cells
Singlet isolation 1
Post singlet selection 1
Singlet isolation 2
Post singlet selection 2
Low
39.5
High
21.0
Comp-Pacific Blue-A
Comp-GFP-A
E
PCR1:
Lentiviral GOI
PCR2:
Block
20-bp flanking sequence + partial Illumina adaptor
PCR3:
Block
Remaining Illumina adaptor
F
Demultiplexing
Alignment
Pre-cleaning
Cleaning
Counting
Raw
Reads QC
SMuRF scores
Normalization
DiMSum-empowered replicate combination

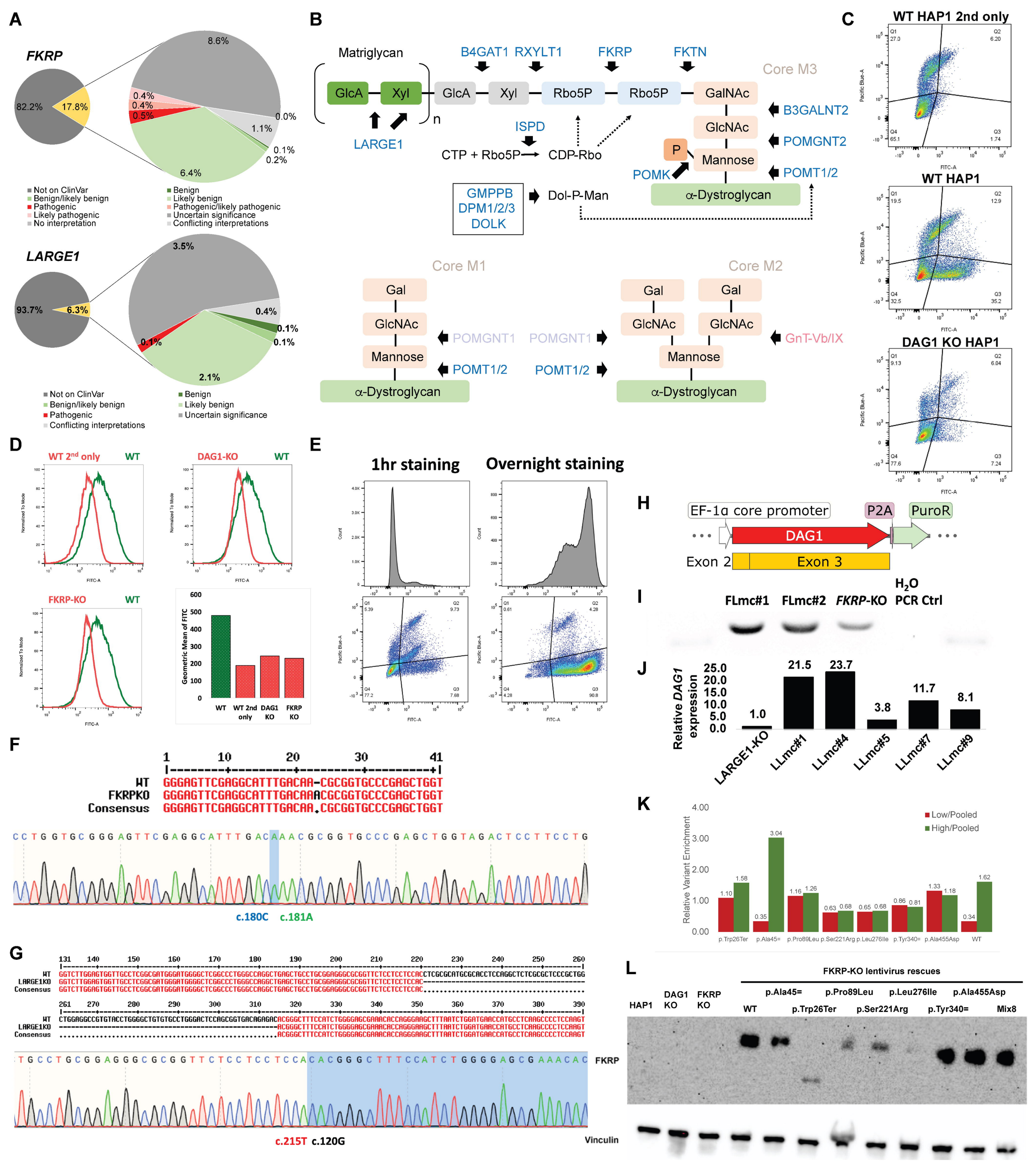

A
FKRP - All Blocks
4325 variants
2.4e-07
p < 2.22e-16
p < 2.22e-16
2e-05
SMuRF score
2.5
0.0
-2.5
-5.0
1157
3034
125
9
Synonymous
Missense
Nonsense
Start-loss
B
LARGE1 - All Blocks
6473 variants
2.2e-07
p < 2.22e-16
p < 2.22e-16
1.4e-05
SMuRF score
2.5
0.0
-2.5
-5.0
1450
4750
264
9
Synonymous
Missense
Nonsense
Start-loss
C
FKRP
364 variants
SMuRF score
1
0
-1
-2
-3
Classification
Missense
Nonsense
Synonymous
1e-05
1e-04
0.001
0.01
0.1
gnomAD v4 genome Allele Frequency
D
LARGE1
410 variants
SMuRF score
1
0
-1
-2
-3
Classification
Missense
Nonsense
Start-loss
Synonymous
1e-05
1e-04
0.001
0.01
0.1
gnomAD v4 genome Allele Frequency

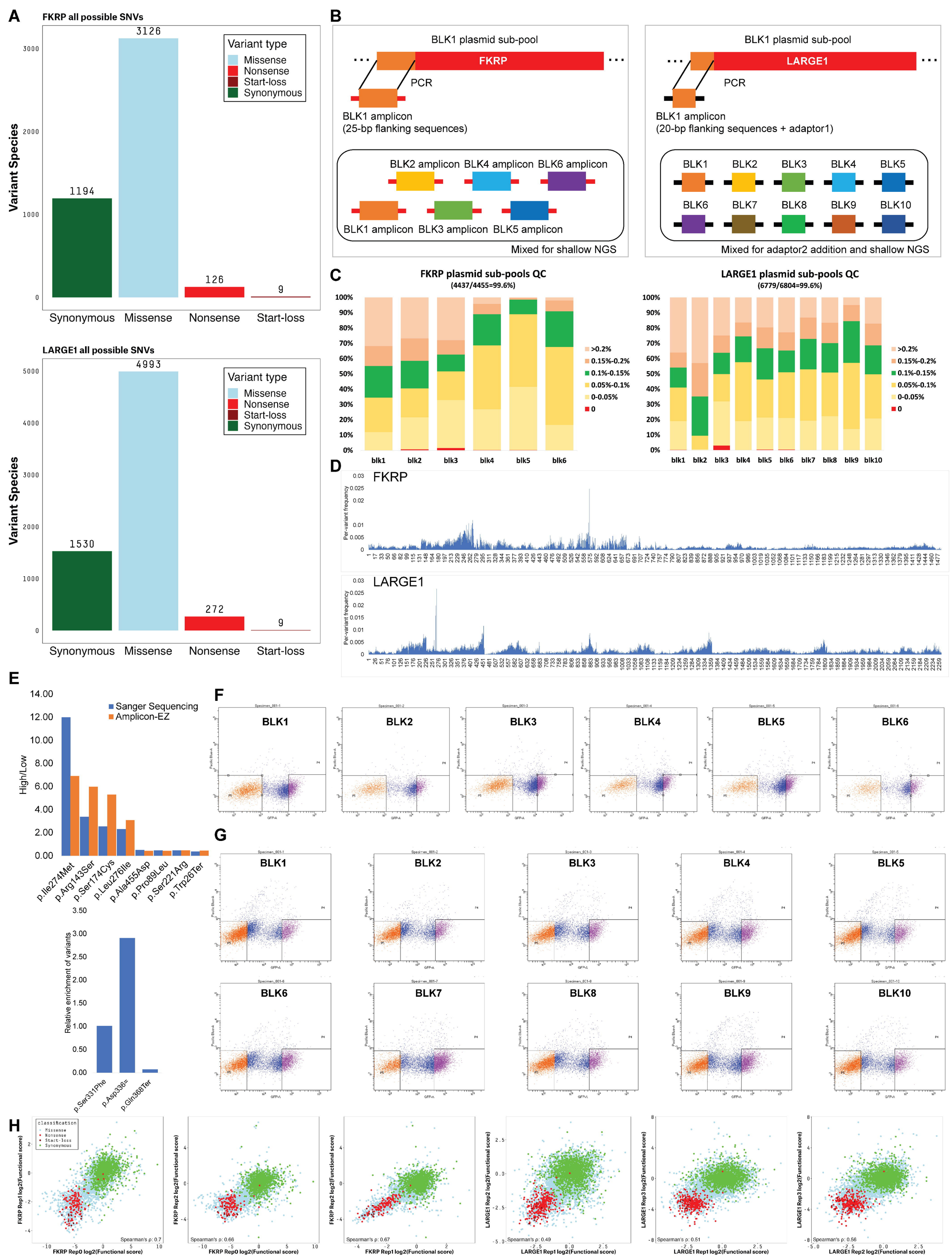

A
FKRP Missense
406 variants
0.003
0.053
7.6e-11
SMuRF score
2
0
-2
-4
2
369
35
B/LB
VUS
P/LP
B
LARGE1 Missense
233 variants
0.016
0.018
0.0014
SMuRF score
2
0
-2
-4
5
224
4
B/LB
VUS
P/LP
C
FKRP well-curated clinical cohorts
SEX
F
M
Log2(2^FS1+2^FS2)
2
1
0
-1
-2
2.8e-09
8.1e-05
0.26
2
4
10
10
22
37
Severe
Intermediate
Mild
D
FKRP well-curated clinical cohorts
Sex
F
M
Log2(2^FS1+2^FS2)
2
1
0
-1
-2
0
50
100
150
200
onset (m)

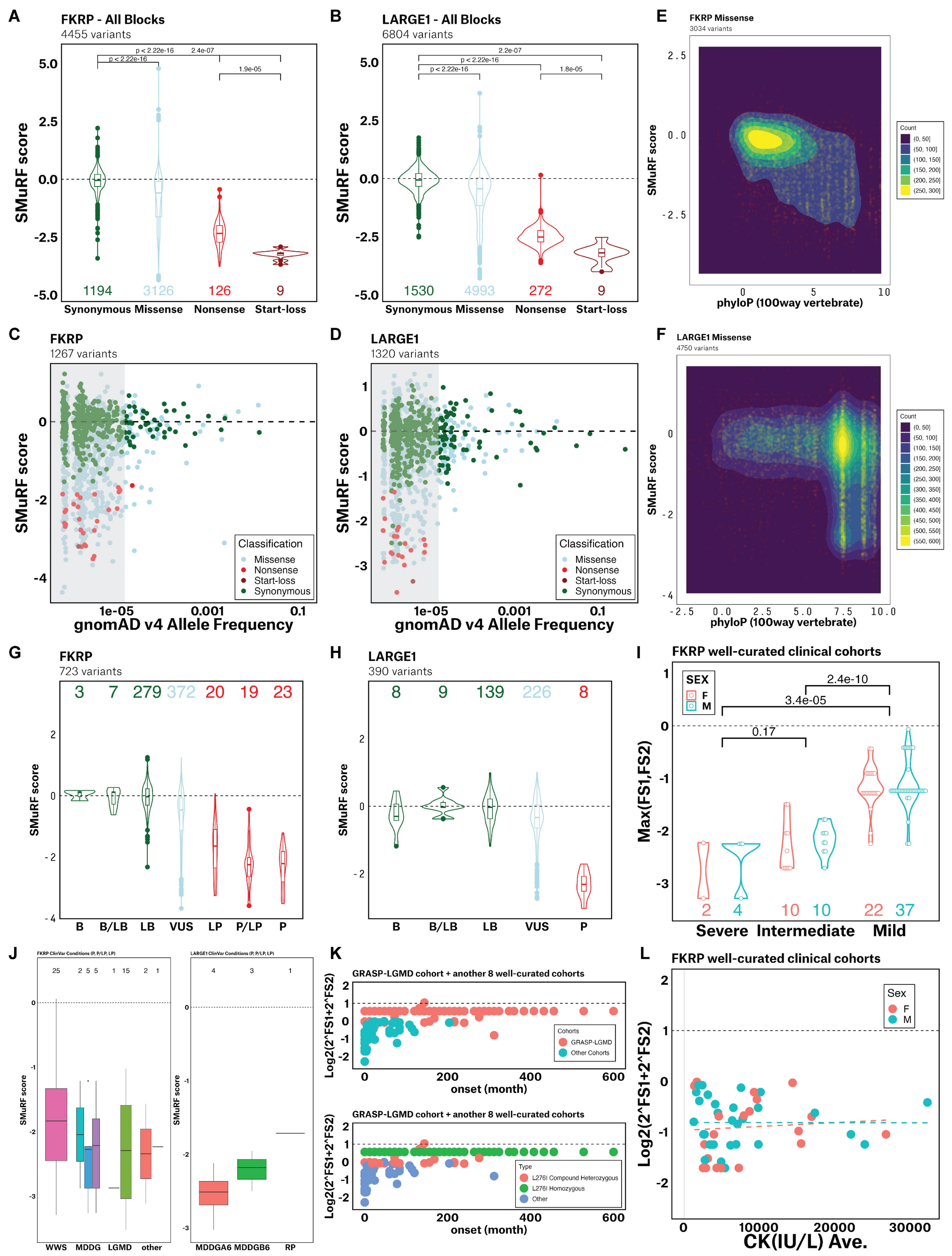

A
FKRP
Sensitivity
Specificity
MVP (AUC=0.68)
metaSVM (AUC=0.69)
ESM1b (AUC=0.73)
CADD (AUC=0.75)
MAVERICK (AUC=0.77)
MutScore (AUC=0.83)
AlphaMissense (AUC=0.83)
PrimateAI_3D (AUC=0.85)
EVE (AUC=0.86)
REVEL (AUC=0.88)
SMuRF (AUC=0.86)
B
LARGE1
Sensitivity
Specificity
MAVERICK (AUC=0.66)
MVP (AUC=0.72)
ESM1b (AUC=0.75)
PrimateAI_3D (AUC=0.79)
CADD (AUC=0.79)
MutScore (AUC=0.8)
EVE (AUC=0.82)
metaSVM (AUC=0.86)
AlphaMissense (AUC=0.87)
REVEL (AUC=0.88)
SMuRF (AUC=0.96)
C
Absolute Spearman's rho
FKRP
LARGE1
AlphaMissense
PrimateAI-3D
EVE
REVEL
MutScore
ESM1b
CADD
MAVERICK
metaSVM
MVP
D
FKRP
3034 variants
SMuRF score
AlphaMissense
E
LARGE1
4750 variants
SMuRF score
AlphaMissense
F
KO Rescue
Control
LARGE1-KO
WT
p.Ala145=
p.Ser331Phe
p.Cys354Gly
p.Arg718Pro
p.Ile725Met
standard
MW (kDa)
250
150
100
75
50
37
25
20
15
Anti-matriglycan (IIH6C4)
Laminin Overlay
Coomassie stained gel of WGA resin

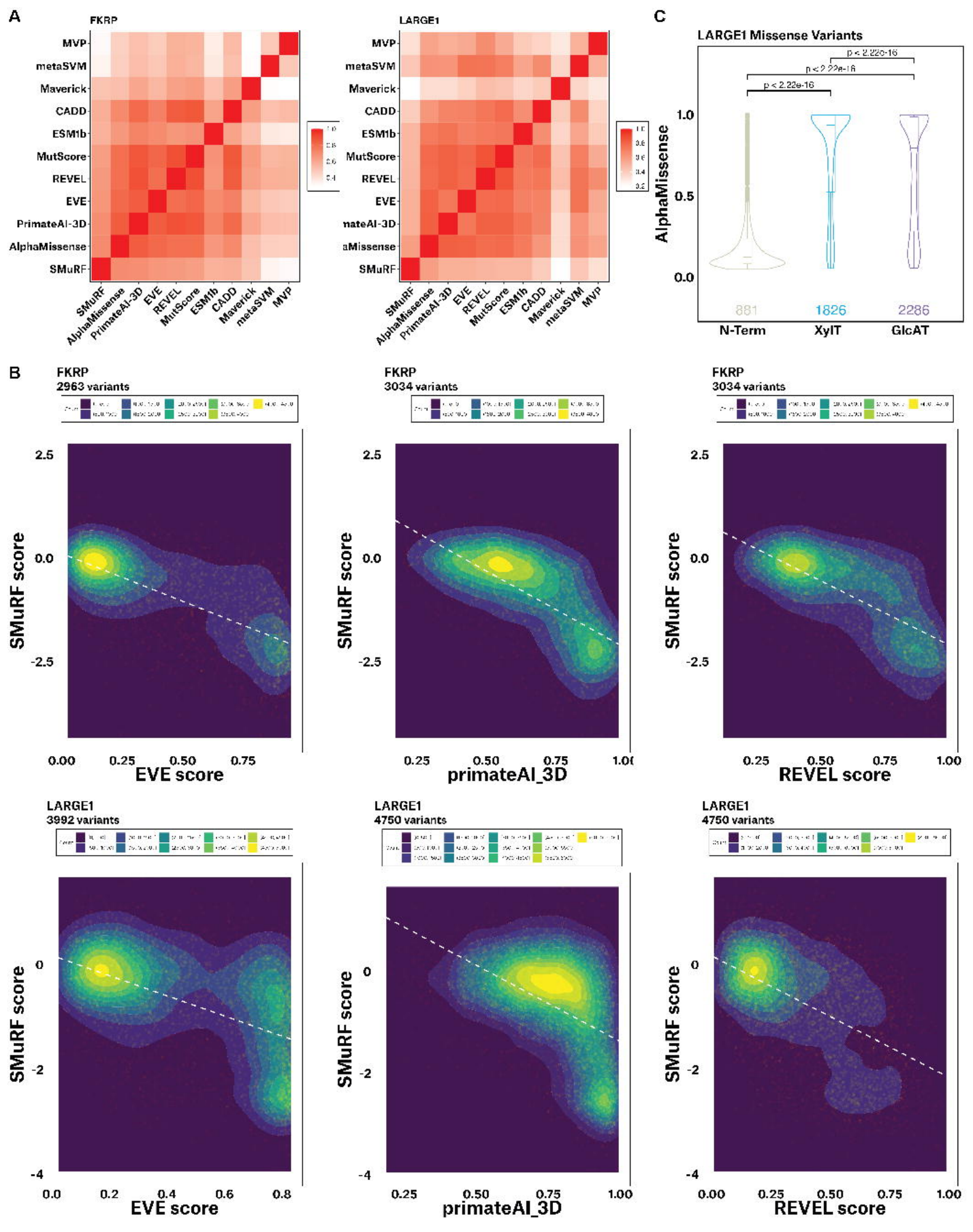

A
FKRP Missense Variants
p < 2.22e-16
p < 2.22e-16
0.17
SMuRF score
1717
188
1129
Stem
Znf
Catalytic excl. Znf
B
FKRP Synonymous Variants
0.12
0.0056
0.11
722
61
374
C
LARGE1 Missense Variants
p < 2.22e-16
p < 2.22e-16
p < 2.22e-16
838
1700
2212
N-Term
XylT
GlcAT
D
LARGE1 Synonymous Variants
0.97
0.57
0.52
280
517
653
E
Stem
Znf
Catalytic excl. Znf
Mean missense sc
Codon
F
N-term
XylT
GlcAT
Mean missense score
Codon
G
TYR 88
CYS 318
LEU 276
SER 221
PHE 473
H
ASP 563
ASP 565
ASP 244
ASP 242
I
Negatively charged
Positively charged
Polar, uncharged
Nonpolar, aromatic
Nonpolar, aliphatic
Stem
Znf
Znf
Catalytic excl.Znf
Mean missense score
-2 -1 0 1 2
J
N-term
XylT
XylT
GlcAT

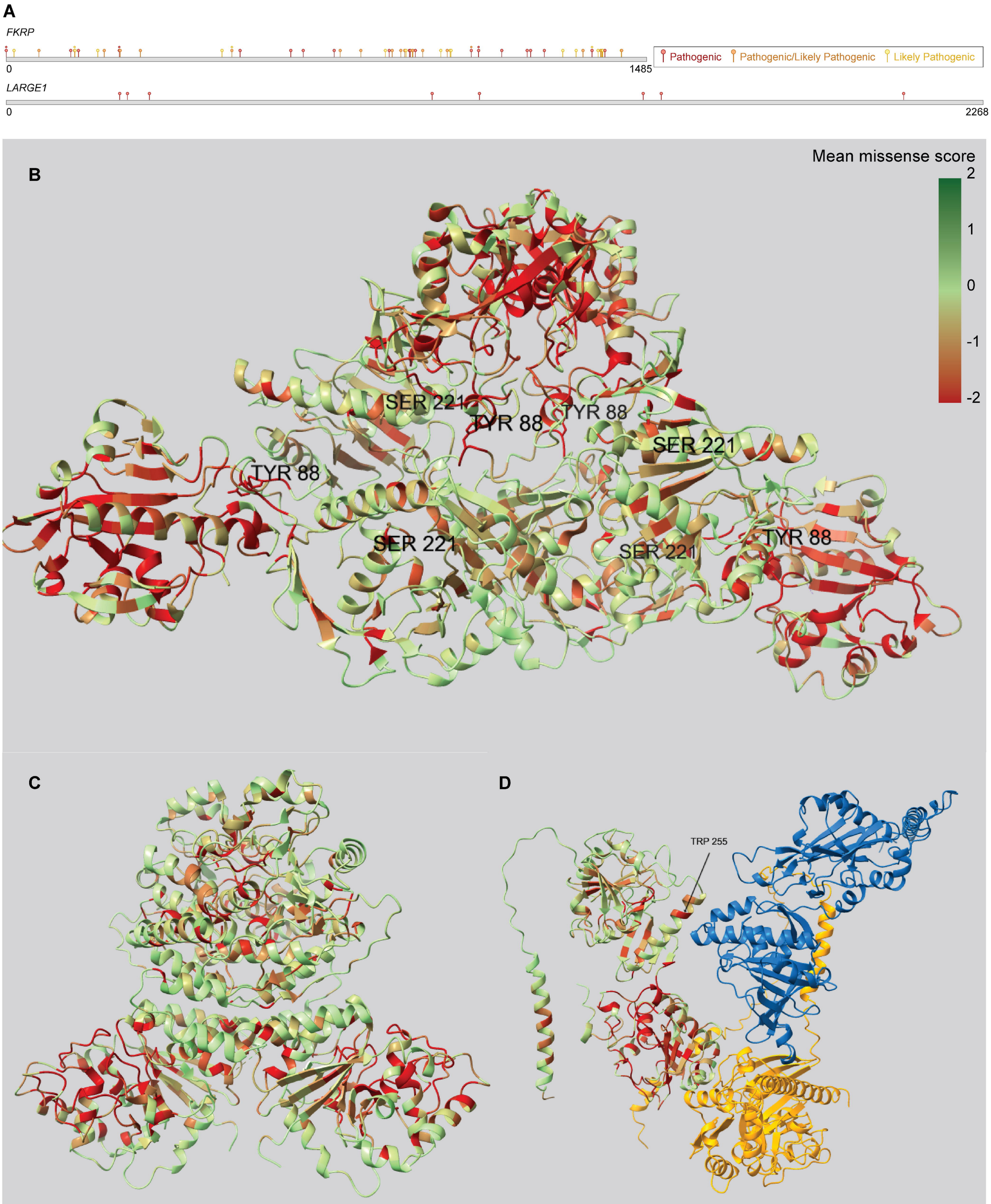

A
FKRP KO
W26*.r
Pathogenic/Likely pathogenic
-2.34
P89L.r
Pathogenic/Likely pathogenic
-3.59
S221R.r
Pathogenic
-2.21
A455D.r
Pathogenic
-2.71
WT.r
R143S.r
Benign/Likely benign
0.27
S174C.r
Benign/Likely benign
0.11
I274M.r
Conflicting interpretations of pathogenicity
0.86
L276I.r
Pathogenic/Likely pathogenic
-0.44
B
IIH6C4/DAPI
FKRP KO
WT.r
W26* .r
P89L.r
S221R.r
A455D.r
L276I.r
R143S.r
S174C.r
I274M.r
C
GOI-KO
MB135 Cells
Lenti-GOI Rescue
VSV with LASV-GPC
Viral entry-based screen
Scores
Sequencing and analysis
D
Synonymous
Missense
Nonsense
Log2(No-inf/Inf)
c.135C>T
c.783A>T
c.783A>G
c.783A>C
c.663C>A
c.1174T>G
c.436G>T
c.673C>T
c.1037C>A
c.1477G>T
E
Synonymous
Missense
Nonsense
Log2(No-inf/Inf)
c.435C>T
c.639T>C
c.642A>G
c.690T>A
c.693T>G
c.699C>G
c.992C>T
c.26G>C
c.1050C>G
c.1720G>T
c.2257G>T

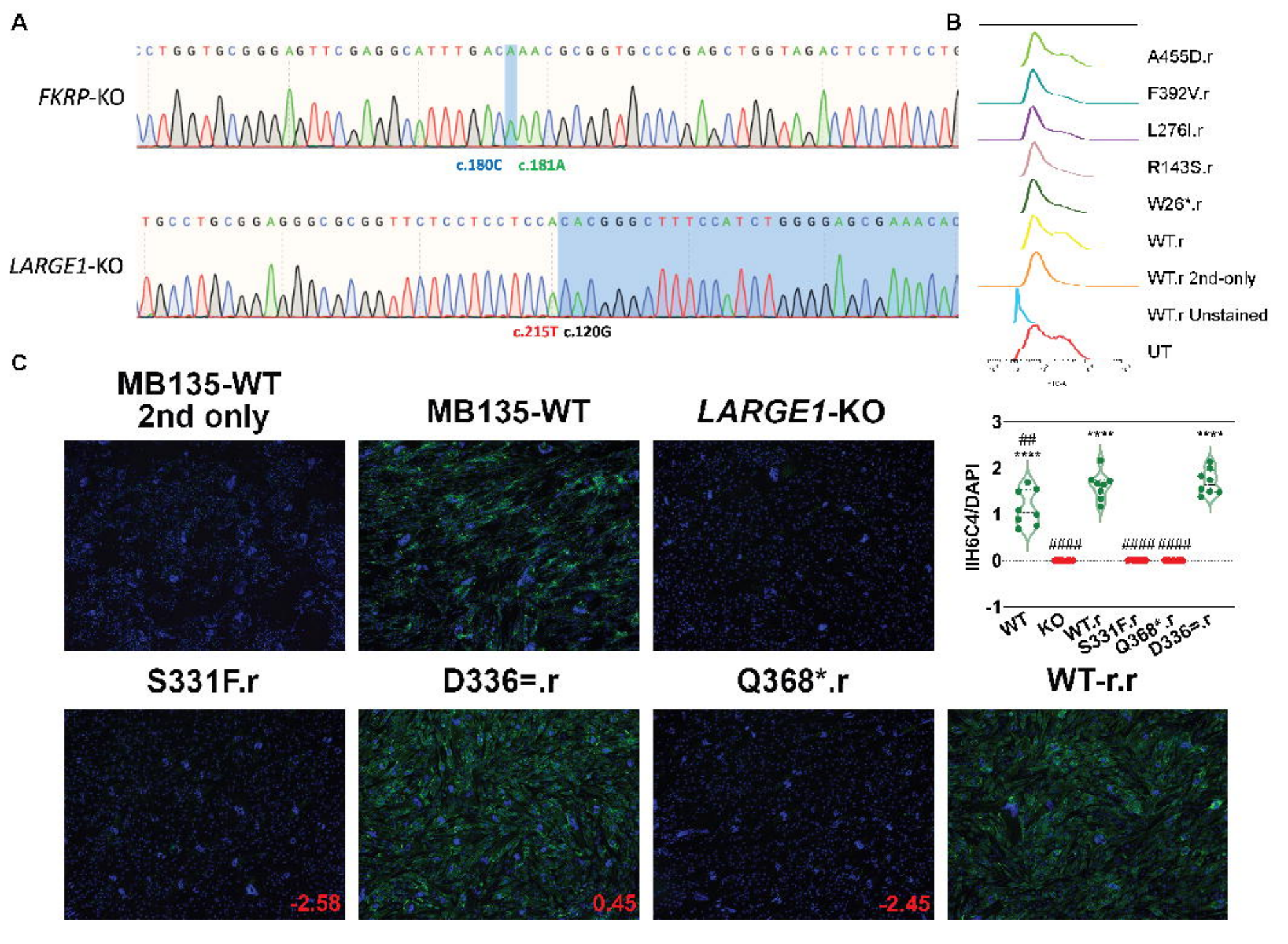

A
FKRP
82.2%
17.8%
8.6%
0.4%
0.4%
0.5%
0.0%
1.1%
0.1%
0.2%
6.4%
Not on ClinVar
Benign/likely benign
Pathogenic
Likely pathogenic
No interpretation
Benign
Likely benign
Pathogenic/likely pathogenic
Uncertain significance
Conflicting interpretations
LARGE1
93.7%
6.3%
3.5%
0.4%
0.1%
0.1%
0.1%
2.1%
Not on ClinVar
Benign/likely benign
Pathogenic
Conflicting interpretations
Benign
Likely benign
Uncertain significance
B
Matriglycan
GlcA
Xyl
n
LARGE1
B4GAT1
RXYLT1
FKRP
FKTN
GlcA
Xyl
Rbo5P
Rbo5P
GalNAc
Core M3
B3GALNT2
GlcNAc
POMGNT2
ISPD
CTP + Rbo5P
CDP-Rbo
P
Mannose
POMK
POMT1/2
GMPPB
DPM1/2/3
DOLK
Dol-P-Man
α-Dystroglycan
Core M1
Gal
GlcNAc
Mannose
α-Dystroglycan
POMGNT1
POMT1/2
Core M2
Gal
GlcNAc
Mannose
α-Dystroglycan
POMGNT1
POMT1/2
GnT-Vb/IX
C
WT HAP1 2nd only
WT HAP1
DAG1 KO HAP1
Pacific Blue-A
FITC-A
D
WT 2nd only
WT
DAG1-KO
FKRP-KO
Normalized To Mode
Geometric Mean of FITC
WT
WT 2nd only
DAG1 KO
FKRP KO
E
1hr staining
Overnight staining
Count
Pacific Blue-A
FITC-A
H
EF-1α core promoter
P2A
PuroR
DAG1
Exon 2
Exon 3
I
FLmc#1
FLmc#2
FKRP-KO
H2O PCR Ctrl
J
Relative DAG1 expression
1.0
21.5
23.7
3.8
11.7
8.1
LARGE1-KO
LLmc#1
LLmc#4
LLmc#5
LLmc#7
LLmc#9
F
WT
FKRPKO
Consensus
c.180C
c.181A
G
WT
LARGE1KO
Consensus
c.215T
c.120G
K
Relative Variant Enrichment
Low/Pooled
High/Pooled
1.10
1.58
0.35
3.04
1.16
1.26
0.63
0.68
0.65
0.68
0.86
0.81
1.33
1.18
0.34
1.62
p.Trp26Ter
p.Ala45=
p.Pro89Leu
p.Ser221Arg
p.Leu276Ile
p.Tyr340=
p.Ala455Asp
WT
L
FKRP-KO lentivirus rescues
HAP1
DAG1 KO
FKRP KO
WT
p.Ala45=
p.Trp26Ter
p.Pro89Leu
p.Ser221Arg
p.Leu276Ile
p.Tyr340=
p.Ala455Asp
Mix8
FKRP
Vinculin

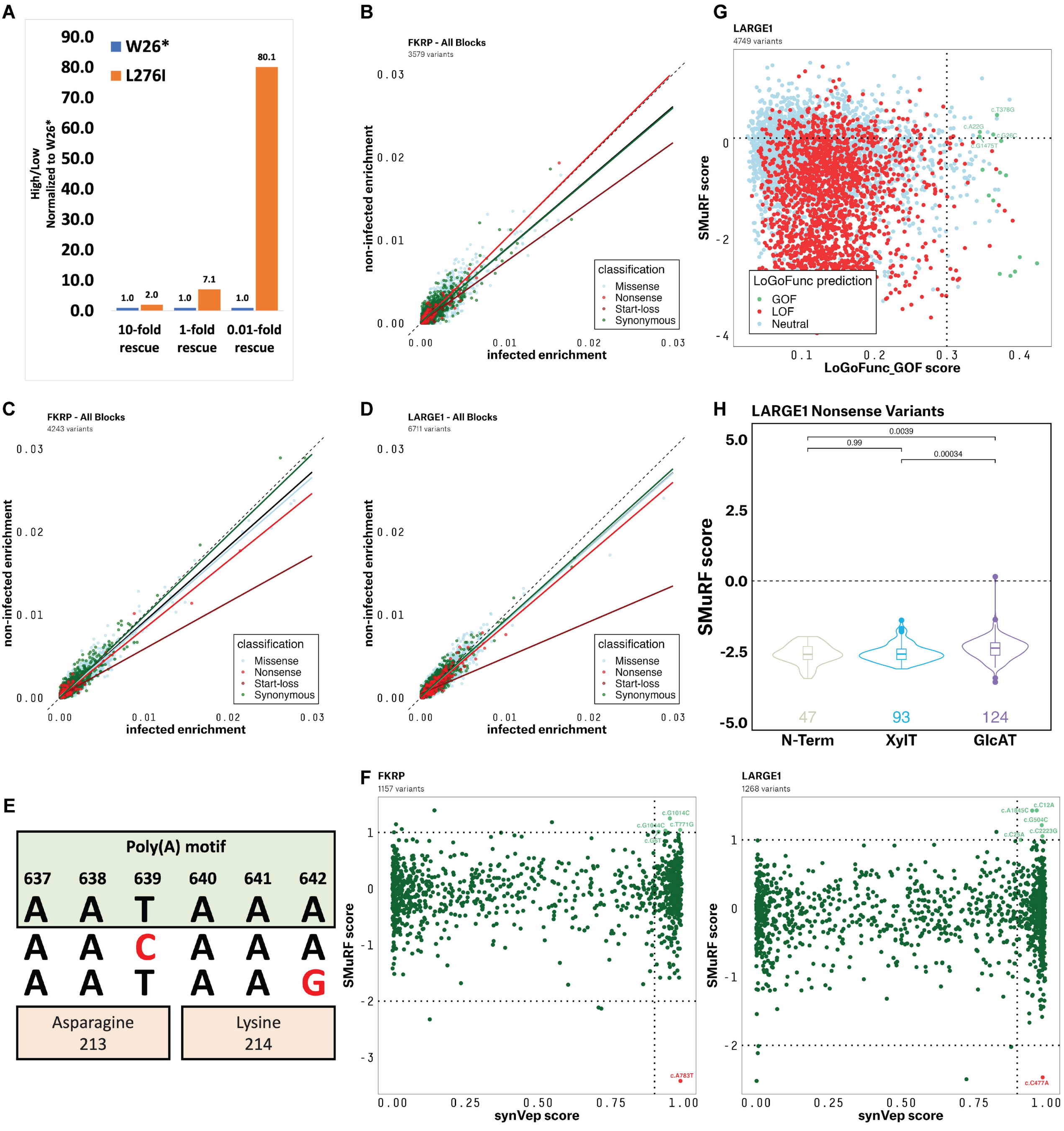

A
FKRP all possible SNVs
Variant Species
Variant type
Missense
Nonsense
Start-loss
Synonymous
1194
3126
126
9
Synonymous
Missense
Nonsense
Start-loss
LARGE1 all possible SNVs
1530
4993
272
9
B
BLK1 plasmid sub-pool
FKRP
PCR
BLK1 amplicon
(25-bp flanking sequences)
BLK2 amplicon
BLK4 amplicon
BLK6 amplicon
BLK1 amplicon
BLK3 amplicon
BLK5 amplicon
Mixed for shallow NGS
BLK1 plasmid sub-pool
LARGE1
PCR
BLK1 amplicon
(20-bp flanking sequences + adaptor1)
BLK1
BLK2
BLK3
BLK4
BLK5
BLK6
BLK7
BLK8
BLK9
BLK10
Mixed for adaptor2 addition and shallow NGS
C
FKRP plasmid sub-pools QC
(4437/4455=99.6%)
LARGE1 plasmid sub-pools QC
(6779/6804=99.6%)
>0.2%
0.15%-0.2%
0.1%-0.15%
0.05%-0.1%
0-0.05%
0
D
FKRP
LARGE1
Per-variant frequency
E
Sanger Sequencing
Amplicon-EZ
High/Low
p.Ile274Met
p.Arg143Ser
p.Ser174Cys
p.Leu276Ile
p.Ala455Asp
p.Pro89Leu
p.Ser221Arg
p.Trp26Ter
Relative enrichment of variants
p.Ser331Phe
p.Asp336=
p.Gln368Ter
F
BLK1
BLK2
BLK3
BLK4
BLK5
BLK6
G
BLK1
BLK2
BLK3
BLK4
BLK5
BLK6
BLK7
BLK8
BLK9
BLK10
H
classification
Missense
Nonsense
Start-loss
Synonymous
FKRP Rep1 log2(Functional score)
FKRP Rep0 log2(Functional score)
Spearman's ρ: 0.7
FKRP Rep2 log2(Functional score)
FKRP Rep0 log2(Functional score)
Spearman's ρ: 0.66
FKRP Rep2 log2(Functional score)
FKRP Rep1 log2(Functional score)
Spearman's ρ: 0.67
LARGE1 Rep2 log2(Functional score)
LARGE1 Rep1 log2(Functional score)
Spearman's ρ: 0.49
LARGE1 Rep3 log2(Functional score)
LARGE1 Rep1 log2(Functional score)
Spearman's ρ: 0.51
LARGE1 Rep3 log2(Functional score)
LARGE1 Rep2 log2(Functional score)
Spearman's ρ: 0.56


A
FKRP - All Blocks
4455 variants
SMuRF score
p < 2.22e-16
p < 2.22e-16
2.4e-07
1.9e-05
1194
3126
126
9
Synonymous
Missense
Nonsense
Start-loss
B
LARGE1 - All Blocks
6804 variants
2.2e-07
p < 2.22e-16
p < 2.22e-16
1.8e-05
1530
4993
272
9
E
FKRP Missense
3034 variants
phyloP (100way vertebrate)
Count
(0, 50]
(50, 100]
(100, 150]
(150, 200]
(200, 250]
(250, 300]
C
FKRP
1267 variants
gnomAD v4 Allele Frequency
Classification
Missense
Nonsense
Start-loss
Synonymous
D
LARGE1
1320 variants
F
LARGE1 Missense
4750 variants
(300, 350]
(350, 400]
(400, 450]
(450, 500]
(500, 550]
(550, 600]
G
FKRP
723 variants
3
7
279
372
20
19
23
B
B/LB
LB
VUS
LP
P/LP
P
H
LARGE1
390 variants
8
9
139
226
8
I
FKRP well-curated clinical cohorts
SEX
F
M
2.4e-10
3.4e-05
0.17
Max(FS1,FS2)
2
4
10
10
22
37
Severe
Intermediate
Mild
J
FKRP ClinVar Conditions (P, P/LP, LP)
LARGE1 ClinVar Conditions (P, P/LP, LP)
WWS
MDDG
LGMD
other
MDDGA6
MDDGB6
RP
K
GRASP-LGMD cohort + another 8 well-curated cohorts
Log2(2^FS1+2^FS2)
onset (month)
Cohorts
GRASP-LGMD
Other Cohorts
Type
L276I Compound Heterozygous
L276I Homozygous
Other
L
FKRP well-curated clinical cohorts
Sex
F
M
CK(IU/L) Ave.


A
FKRP
LARGE1
MVP
metaSVM
Maverick
CADD
ESM1b
MutScore
REVEL
EVE
PrimateAI-3D
AlphaMissense
SMuRF
C
LARGE1 Missense Variants
p < 2.22e-16
AlphaMissense
1.0
0.5
0.0
881
1826
2286
N-Term
XylT
GlcAT
B
FKRP
2963 variants
3034 variants
SMuRF score
EVE score
primateAI_3D
REVEL score
LARGE1
3992 variants
4750 variants


A
FKRP
0
1485
Pathogenic
Pathogenic/Likely Pathogenic
Likely Pathogenic
LARGE1
0
2268
B
Mean missense score
2
1
0
-1
-2
SER 221
TYR 88
TYR 88
SER 221
TYR 88
SER 221
SER 221
TYR 88
C
D
TRP 255


A
FKRP-KO
c.180C c.181A
LARGE1-KO
c.215T c.120G
B
A455D.r
F392V.r
L276I.r
R143S.r
W26*.r
WT.r
WT.r 2nd-only
WT.r Unstained
UT
C
MB135-WT 2nd only
MB135-WT
LARGE1-KO
IIH6C4/DAPI
WT
KO
WT.r
S331F.r
Q368*.r
D336=.r
S331F.r
D336=.r
Q368*.r
WT-r.r
-2.58
0.45
-2.45


A
W26*
L276I
High/Low Normalized to W26*
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1.0
2.0
1.0
7.1
1.0
80.1
10-fold rescue
1-fold rescue
0.01-fold rescue
B
FKRP - All Blocks
3579 variants
non-infected enrichment
infected enrichment
classification
Missense
Nonsense
Start-loss
Synonymous
G
LARGE1
4749 variants
SMuRF score
LoGoFunc_GOF score
LoGoFunc prediction
GOF
LOF
Neutral
c.T378G
c.A22G
c.G26C
c.G1475T
C
FKRP - All Blocks
4243 variants
non-infected enrichment
infected enrichment
classification
Missense
Nonsense
Start-loss
Synonymous
D
LARGE1 - All Blocks
6711 variants
non-infected enrichment
infected enrichment
classification
Missense
Nonsense
Start-loss
Synonymous
H
LARGE1 Nonsense Variants
0.0039
0.99
0.00034
SMuRF score
47
93
124
N-Term
XylT
GlcAT
E
Poly(A) motif
637 638 639 640 641 642
A A T A A A
A A C A A A
A A T A A G
Asparagine 213
Lysine 214
F
FKRP
1157 variants
c.G1014C
c.G1044C
c.T771G
c.A783T
SMuRF score
synVep score
LARGE1
1268 variants
c.C12A
c.A1845C
c.G504C
c.C2223G
c.C25A
c.C477A
SMuRF score
synVep score