Applying Multiplex Assays to Understand Variation in Pharmacogenes

Melissa Chiasson[1], Maitreya J. Dunham[1,2], Allan E. Rettie[3] and Douglas M. Fowler[1,2,4,*]

Genome sequencing has enabled the detection of unprecedented numbers of new pharmacogene variants. However, interpreting how these variants affect pharmacogene biology and ultimately drug response is difficult. Multiplexed assays for variant effects (MAVEs) leverage high throughput DNA sequencing to assess the functional consequences of thousands of variants simultaneously. We discuss the utility of large-scale functional data in pharmacogene variant interpretation and suggest that implementing MAVEs could empower pharmacogenetics and improve patient care.

Genomes can now be sequenced with ease, but understanding the effect of the variants found therein poses a major challenge. Each uninterpreted variant represents a missed opportunity to improve patient outcomes. For example, the Clinical Pharmacogenetics Implementation Consortium (CPIC) lists 358 gene–drug pairs in which variation can change drug response. For 63 of these 358 pairs, CPIC has issued guidelines regarding clinical interventions that may improve patient care. These guidelines focus on common variants (minor allele frequencies, typically >5%) whose clinical consequences are most clearly documented. However, understanding the effects of rare variants (minor allele frequency <0.5–1%) is also essential, and this goal is far from realized.

The magnitude of the unmet need requires consideration of the totality of rare variation that will be identified as sequencing becomes more common. As of February 2019, the Genome Aggregation Database contained ~ 125,000 exomes and ~ 15,000 genomes, which included 404 rare coding single-nucleotide variants in cytochrome P450 (CYP)2C9 alone, 212 of which were singletons. Only 55 of these variants were in the PharmVar database (accessed February 10, 2019), and only about a dozen have been functionally annotated. Undoubtedly, as sequencing continues, many more *CYP2C9* variants will be identified. This issue is not confined to *CYP2C9*: 731 novel nonsynonymous variants in 12 CYP genes were discovered in the exomes of ~ 6,500 individuals.[1] Approximately 10% of individuals carried at least one of these potentially deleterious novel variants. These results, obtained from a handful of genes in a few individuals relative to the number that will ultimately be sequenced, illustrate that an onslaught of new and potentially important variants is coming.

THE CHALLENGE OF VARIANT FUNCTIONAL ANALYSIS

Current methods for determining the impact of pharmacogene variants fall into two categories. Biochemical assays using known substrates for drug disposition genes can reveal variant functional consequences. However, this approach is limited in scale to tens or hundreds of missense variants. Computational predictions can scale to all possible variants of a gene of interest but are of limited value as they often produce incorrect or conflicting results. For example, the *CYP2C9*3* variant, present in ~ 7% of people of European ancestry, confers ~ 90% loss of function according to experimental data[2] but is predicted computationally to be benign. To overcome the limitations of biochemical assays and computational predictions, an experimental approach to assess pharmacogene variants on a massive scale is needed.

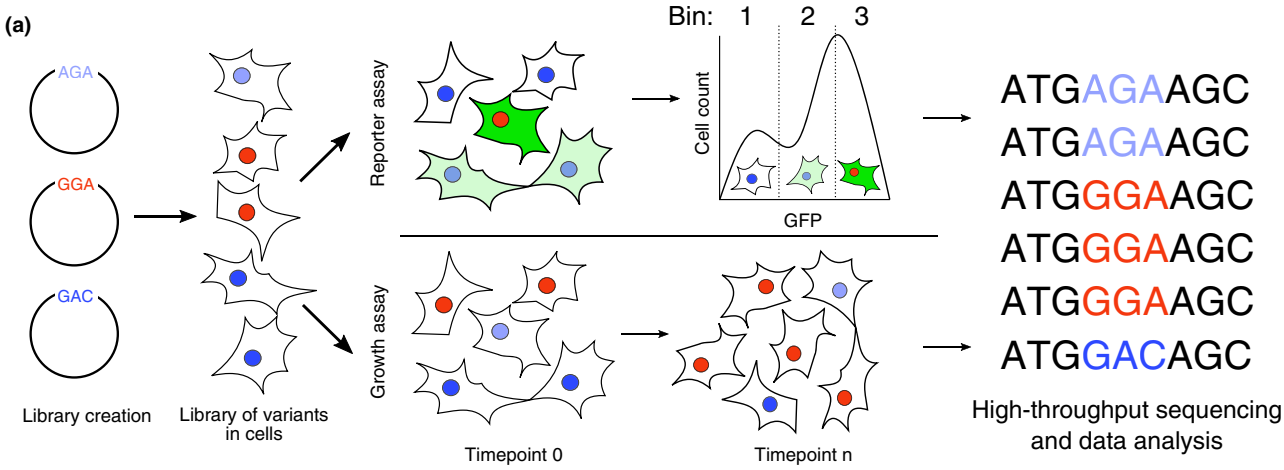
MAVES CAN CHARACTERIZE TENS OF THOUSANDS OF VARIANTS SIMULTANEOUSLY

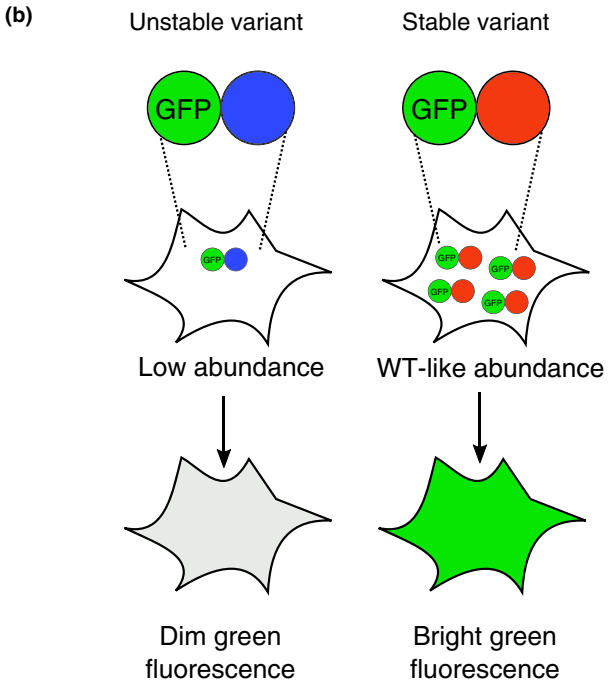
A MAVE measures the functional consequences of a large library of genetic variants simultaneously.[3,4] MAVEs can be applied to a wide range of genetic elements, including mRNA untranslated regions (UTRs), promoters, enhancers, splice sites, and proteins. The result of a MAVE is a variant effect map that reveals the functional consequences of all possible single variants in the genetic element.

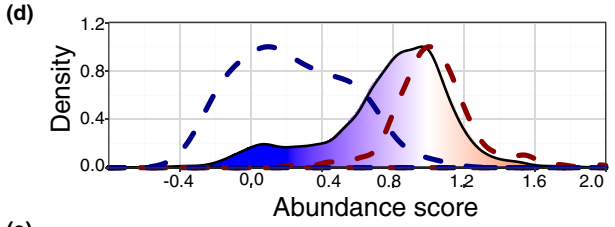
All MAVEs share the same basic design (**Figure 1a**, reviewed in refs. 3,4). First, a pooled library of variants is constructed either by polymerase chain reaction-based mutagenesis or synthesized oligo arrays programmed with mutations of interest. The library is then introduced into an experimental system, typically yeast or cultured human cells. Each cell must express a single variant to maintain the link between variant sequence and phenotype. For example, in human cells, expression of a single variant is typically achieved using lentiviral transduction or recombinase-based systems. Cells expressing the library of interest are then assayed for a phenotype of interest, like growth or reporter activation. These assays stratify variants based on their phenotypic effect. For example, in a growth assay, cells expressing wild type (WT)-like variants grow rapidly whereas

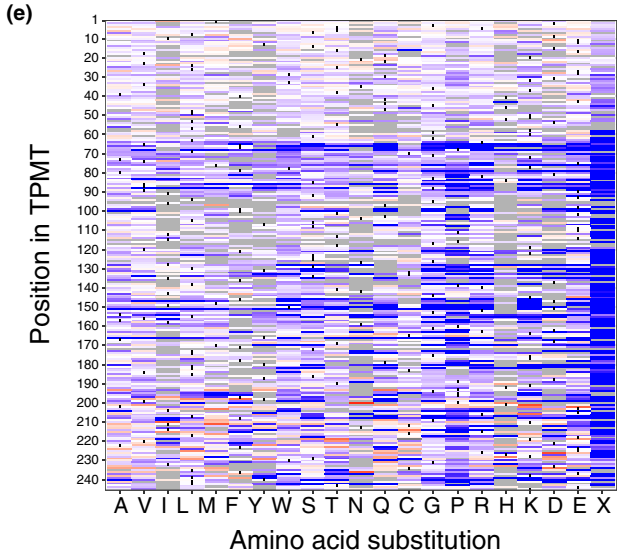
[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA; [2]Genetic Networks Program, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario, Canada; [3]Department of Medicinal Chemistry, University of Washington, Seattle, Washington, USA; [4]Department of Bioengineering, University of Washington, Seattle, Washington, USA. *Correspondence: Douglas M. Fowler (dfowler@uw.edu)

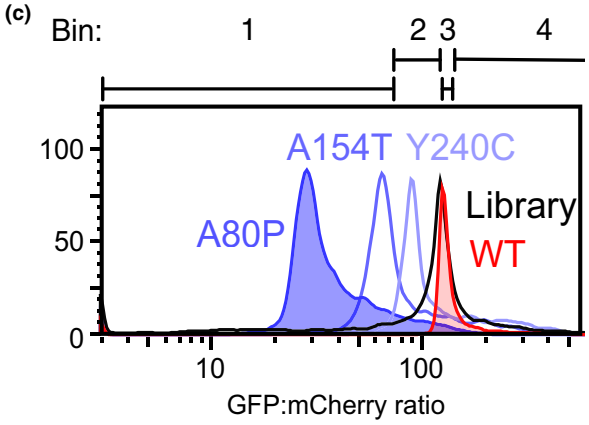
Received February 13, 2019; accepted April 3, 2019. doi:10.1002/cpt.1468


(a)
AGA
GGA
GAC
Library creation
Library of variants in cells
Reporter assay
Growth assay
Bin: 1 2 3
Cell count
GFP
Timepoint 0
Timepoint n
ATGAGAAGC
ATGAGAAGC
ATGGGAAGC
ATGGGAAGC
ATGGGAAGC
ATGGACAGC
High-throughput sequencing and data analysis


(b)
Unstable variant
Stable variant
GFP
Low abundance
WT-like abundance
Dim green fluorescence
Bright green fluorescence


(d)
Density
1.2
0.8
0.4
0.0
-0.4
0.0
0.4
0.8
1.2
1.6
2.0
Abundance score


(e)
Position in TPMT
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
A V I L M F Y W S T N Q C G P R H K D E X
Amino acid substitution


(c)
Bin: 1 2 3 4
100
50
0
A154T Y240C
A80P
Library
WT
10
100
GFP:mCherry ratio

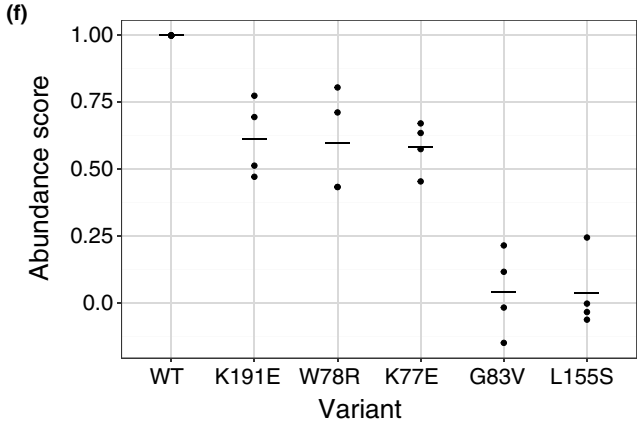

(f)
Abundance score
1.00
0.75
0.50
0.25
0.0
WT
K191E
W78R
K77E
G83V
L155S
Variant

Figure 1 (**a**) Overview of multiplexed assays for variant effects. A library of variants of the genetic element of interest is created and introduced into cells. The cells are subjected to a growth-based or fluorescent reporter–based assay. High-throughput sequencing is used to determine the frequency of variants before and after the assay, and variant frequencies are used to calculate functional scores. (**b**) Variant Abundance by Massively Parallel sequencing uses a green fluorescent protein (GFP) fusion reporter to measure steady-state variant abundance. GFP was fused N-terminally to a library of thiopurine methyltransferase (TPMT) variants; mCherry was used as a transcriptional control. This library was introduced into HEK293T cells using a serine integrase landing pad system such that only one variant is expressed per cell. Cells were sorted based on their fluorescence into four bins. High-throughput sequencing was used to determine the frequency of every variant in each bin. Frequencies were then converted to abundance scores. (**c**) A flow cytometry plot of wild-type (WT) TPMT (red) and three high-frequency variants known to be low abundance (blue): A80P, A154T, and Y240C. The library of TPMT variants and bins used for sorting are shown (gray). (**d**) Density plot of abundance scores, with dotted blue line showing distribution of nonsense variants and red dotted line showing synonymous variants. The missense variant distribution is shaded from blue (low abundance) to red (high abundance). (**e**) Heatmap of TPMT abundance scores shaded from blue (low abundance) to red (high abundance); gray indicates missing data. (**f**) Abundance scores from four replicates for five new TPMT variants found in gnomAD.

cells expressing loss-of-function variants grow slowly. In a fluorescent reporter assay, WT-like variants drive high fluorescence, whereas loss-of-function variants drive low fluorescence. Cells are sorted into bins according to fluorescence. High throughput sequencing is used to measure a variant's frequency in the assay, either before and after growth or across bins. Variant frequencies are then used to compute effect scores.

MAVEs for coding and noncoding variants differ in the type of assays used. For example, noncoding MAVEs generally measure how variants affect expression, often by quantifying mRNA transcripts or using a fluorescent reporter. Coding MAVEs measure different aspects of a protein's function. For example, reporter assays can measure specific protein properties like abundance or substrate binding using fluorescent protein tags or fluorophore-labeled antibodies. Growth-based assays measure each variant's ability to drive cell growth, either in the context of a deletion of the genomic copy of the protein or by using a metabolic reporter.

MAVEs have the power to functionally annotate variants in many, if not most, pharmacogenes. However, achieving this goal will take time and effort, requiring the implementation of existing MAVEs and the development of new assays. To illustrate these issues, we first discuss the recent application of an MAVE to thiopurine methyltransferase (TPMT) and then consider other pharmacogenes that could benefit most from MAVEs.

ANALYZING TPMT ABUNDANCE REVEALS NEW VARIANTS THAT CONFER THIOPURINE TOXICITY RISK

TPMT inactivates thiopurine drugs commonly used to treat cancer and autoimmune diseases, including 6-thioguanine and 6-mercaptopurine. Thus, TPMT reduces the quantity of drug available for transformation into thioguanine nucleotides, which inhibit *de novo* purine synthesis. During routine dosing with thiopurines, TPMT deficiency results in high levels of thioguanine nucleotides and, ultimately, hematopoietic toxicity. Three variants, A80P, A154T, and Y240C, are known to lead to decreased TPMT function. CPIC recommends testing for these three variants, enabling patients to be classified as normal, intermediate, or poor metabolizers based on diplotype, with doses adjusted accordingly.

Previously, we applied Variant Abundance by Massively Parallel sequencing (VAMP-seq), a generalizable, multiplex assay for measuring protein abundance inside cells, to TPMT[5] (**Figure 1b**). We generated abundance scores for 3,689 of the 4,655 possible variants (**Figure 1c–e**). A80P, A154T, and Y240C were all low-abundance variants, in accordance with their poor metabolizer status. In contrast, four rare variants from a clinical study of acute lymphoblastic leukemia (S125L, Q179H, R215H, and R226Q) were all WT-like in abundance, and patients with these variants tolerated higher doses of 6-mercaptopurine better than those with A80P, A154T, or Y240C. We then identified 31 reduced-abundance variants in gnomAD and suggested that patients with these variants could have increased risk for thiopurine toxicity. Since our publication of the TPMT variant abundance map, seven new TPMT variants have been added to gnomAD: K77E, W78R, G83V, L155S, P160A, K191E, and C216Y. VAMP-seq data indicate that K77E, W78R, G83V, L155S, and K191E are of low abundance relative to WT (**Figure 1f**). Accordingly, these variants might confer drug sensitivity in patients who carry them.

Thus, protein abundance is a useful phenotype for identifying loss-of-function variants. We also anticipate that measurement of protein activity will be necessary for many pharmacogenes. Fortunately, in some cases, existing low-throughput activity assays can be adapted. For example, a reporter cell line developed to measure vitamin K oxidoreductase (VKOR) activity[6] could be combined with a variant library to assess activity of all VKOR missense variants. Because some VKOR variants confer resistance to warfarin, cells could also be treated with warfarin to reveal the relationship between activity and resistance. Ultimately, the activity and resistance scores from such an assay could be used to help predict a patient's warfarin dose based on their VKOR sequence.

MAVES COULD AID PHARMACOGENE VARIANT INTERPRETATION

Including *TPMT*, CPIC lists 127 genes that have differing levels of evidence for identification as an actionable pharmacogene. There are 5,132,280 possible single-nucleotide variants that exist among these genes. Assaying such a large number of variants is possible but daunting. Thus, we suggest prioritization of the most promising pharmacogenes.

We focused solely on missense variants for this analysis; however, many pharmacogenes have noncoding variants that contribute to drug response and could be assayed with an appropriately designed noncoding MAVE. First, we restricted our analysis to the 31 genes that are designated as CPIC level A or B where genetic information can be used to guide drug therapy. We annotated each gene according to the localization of the protein it encodes, length,

Table 1 Thirty-one genes designated by CPIC as A or B level genes, along with factors to consider when designing MAVEs

Gene	Drug(s)	Length (amino acids)	Total possible single amino acid variants	Missense variants in gnomAD	Missense variants in PharmVar	Localization
MT-RNR1	Aminoglycoside antibacterials	16	300	0	0	Secreted
NUDT15	Azathioprine, mercaptopurine, thioguanine	164	3,260	83	12	Cytoplasm
IFNL3	Peginterferon alfa-2a, Peginterferon alfa-2b, Ribavirin	196	3,900	152	0	Secreted
HPRT1	Mycophenolic acid	218	4,340	20	0	Cytoplasm
TPMT	Azathioprine, mercaptopurine, thioguanine	245	4,880	119	0	Cytoplasm
OTC	Valproic acid	354	7,060	88	0	Mitochondrion matrix
HLA-B	Abacavir, allopurinol, carbamazepine, oxcarbazepine	362	7,220	180	0	Membrane; single-pass type I membrane protein
HLA-A	Carbamazepine, allopurinol	365	7,280	194	0	Membrane; single-pass type I membrane protein
ASS1	Valproic acid	412	8,220	211	0	Cytoplasm
ASL	Valproic acid	464	9,260	242	0	Cytoplasm, extracellular exosome
CYP2C9	Phenytoin, warfarin, acenocoumarol	490	9,780	381	55	Endoplasmic reticulum membrane, peripheral membrane
CYP2C19	Amitriptyline, clopidogrel, citalopram, voriconazole	490	9,780	375	5	Endoplasmic reticulum membrane, peripheral membrane
CYP2B6	Efavirenz, methadone	491	9,800	331	0	Endoplasmic reticulum membrane, peripheral membrane
CYP2D6	Codeine, oxycodone, tamoxifen, tramadol	497	9,920	374	30	Endoplasmic reticulum membrane, peripheral membrane
CYP3A5	Tacrolimus	502	10,020	215	11	Endoplasmic reticulum membrane, peripheral membrane
G6PD	Rasburicase, chloramphenicol, chloroquine, ciprofloxacin	515	10,280	171	0	Cytoplasm, extracellular exosome, nucleus
CYP4F2	Warfarin, acenocoumarol	520	10,380	344	2	Endoplasmic reticulum membrane, peripheral membrane
UGT1A1	Atazanavir, irinotecan, belinostat	533	10,640	308	0	Endoplasmic reticulum membrane; single-pass membrane protein
NAGS	Carglumic acid	534	10,660	216	0	Mitochondrion matrix
GBA	Velaglucerase alfa	536	10,700	247	0	Lysosome membrane, peripheral membrane protein
SLCO1B1	Simvastatin, cerivastatin	691	13,800	399	0	Basolateral cell membrane, multi pass membrane protein
DPYD	Capecitabine, fluorouracil	1,025	20,480	566	0	Cytoplasm
ABL2	Valproic acid	1,182	23,620	509	0	Cytoplasm, cytoskeleton
POLG	Valproic acid	1,239	24,760	762	0	Mitochondrion, mitochondrion matrix, mitochondrion nucleoid
ABCB1	Antidepressants, digoxin	1,280	25,580	578	0	Cell membrane, multi pass membrane protein
CFTR	Ivacaftor	1,480	29,580	991	0	Apical cell membrane
CPS1	Valproic acid	1,500	29,980	679	0	Mitochondrion, nucleus, nucleolus

(Continues)

Table 1 (Continued)

Gene	Drug(s)	Length (amino acids)	Total possible single amino acid variants	Missense variants in gnomAD	Missense variants in PharmVar	Localization
CACNA1S	Desflurane, enflurane, isoflurane, halothane	1,873	37,440	1,071	0	Cell membrane, sarcolemma, T-tubule, multi pass membrane protein
SCN1A	Carbamazepine	2,009	40,160	587	0	Cell membrane, multi pass membrane protein
RYR1	Desflurane, enflurane, isoflurane, halothane	5,038	100,740	2,663	0	Sarcoplasmic reticulum membrane, multi pass membrane protein

CPIC, Clinical Pharmacogenetics Implementation Consortium; CYP, cytochrome P450; MAVEs, multiplexed assays for variant effects; TPMT, thiopurine methyltransferase.

number of missense variants already in gnomAD, and number of variants registered in PharmVar (**Table 1**).

Among this list, small proteins should be given high priority, because they have fewer possible variants and are, thus, easier to assay. Larger proteins affecting dosing of multiple, widely prescribed drugs should also be prioritized, as they impact many patients. For these, we suggest focusing initial efforts on functionally important domains. All the genes have tens to thousands of variants deposited in gnomAD; however, most genes do not have any variants deposited yet in PharmVar. Therefore, concentrating on the genes that have the greatest number of rare variants in gnomAD but no information in PharmVar would yield new insight. Two pharmacogenes, *IFN3* and *MT-RNR1*, encode secreted proteins requiring new assays that maintain the sequence-phenotype link. In addition to these factors, analyzing published clustered regularly interspaced short palindromic repeats (CRISPR) screen data will identify which of these genes cause growth defects in a relevant cell line; growth-based MAVEs would be an attractive starting point for these. For the remainder, we suggest applying reporter-based assays, such as VAMP-seq.

Despite their promise, MAVEs also have limitations. MAVEs often take the genetic element of interest out of its endogenous genomic or cellular context and, thus, demand careful validation of results. Data generated from MAVEs, although comprehensive, can be noisy. Thus, adequate replication is required to improve measurement accuracy and facilitate error estimation. Finally, MAVEs generally focus on one or a few experimental conditions and so may not fully capture condition-dependent effects. For pharmacogenes, therefore, it will be critical to evaluate variants in physiologically relevant concentration, time, and drug contexts.

In summary, a community-wide effort to apply MAVEs to high-priority pharmacogenes would result in variant-effect maps that could aid in the interpretation of variants seen in the clinic. As pharmacogene variant-effect maps are produced, they will yield a better understanding of pharmacogene biology and create opportunities for more rigorous, data-driven customization of patient treatment.

FUNDING

This work was supported by the National Institute of General Medical Sciences (5R24GM115277 to D.M.F., A.R., and M.D., R01GM132162 to D.M.F., A.R., and M.D., P01 GM116691 for A.R.). D.M.F. is a Canadian Institute for Advanced Research Azrieli Global Scholar. M.D. is a Senior Fellow in the Genetic Networks program at the Canadian Institute for Advanced Research. M.D. is supported in part by a Faculty Scholars grant from the Howard Hughes Medical Institute.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

1. Gordon, A.S. *et al.* Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum. Mol. Genet.* **23**, 1957–1963 (2014).
2. Rettie, A.E. & Jones, J.P. Clinical and toxicological relevance of CYP2C9: drug-drug interactions and pharmacogenetics. *Annu. Rev. Pharmacol. Toxicol.* **45**, 477–494 (2005).
3. Starita, L.M. *et al.* Variant interpretation: functional assays to the rescue. *Am. J. Hum. Genet.* **101**, 315–325 (2017).
4. Weile, J. & Roth, F.P. Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. *Hum. Genet.* **137**, 665–678 (2018).
5. Matreyek, K.A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
6. Haque, J.A., McDonald, M.G., Kulman, J.D. & Rettie, A.E. A cellular system for quantitation of vitamin K cycle activity: structure-activity effects on vitamin K antagonism by warfarin metabolites. *Blood* **123**, 582–589 (2014).