1   **Multiplexed measurement of variant abundance and activity reveals VKOR topology,**

2   **active site and human variant impact**

3

4   Melissa A. Chiasson[1], Nathan J. Rollins[2], Jason J. Stephany[1], Katherine A. Sitko[1], Kenneth A.

5   Matreyek[1], Marta Verby[3], Song Sun[3], Frederick P. Roth[3], Daniel DeSloover[4], Debora Marks[2],

6   Allan E. Rettie[5], Douglas M. Fowler[1,6*]

7

8   [1]Department of Genome Sciences, University of Washington, Seattle, WA, USA.

9   [2]Department of Systems Biology, Harvard Medical School, Boston, MA, USA.

10   [3]Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of

11   Toronto, and Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON,

12   Canada

13   [4]Color Genomics, Burlingame, CA, USA.

14   [5]Department of Medicinal Chemistry, University of Washington, Seattle, WA, USA.

15   [6]Department of Bioengineering, University of Washington, Seattle, WA, USA.

16

17   * Correspondence should be addressed to D.M.F. (dfowler@uw.edu).

18 **ABSTRACT**

19 Vitamin K epoxide reductase (VKOR) drives the vitamin K cycle, activating vitamin K-dependent

20 blood clotting factors. VKOR is also the target of the widely used anticoagulant drug, warfarin

21 Despite VKOR's pivotal role in coagulation, its structure and active site remain poorly understood.

22 In addition, VKOR variants can cause vitamin K-dependent clotting factor deficiency 2 or alter

23 warfarin response. Here, we used multiplexed, sequencing-based assays to measure the effects

24 of 2,695 VKOR missense variants on abundance and 697 variants on activity in cultured human

25 cells. The large-scale functional data, along with an evolutionary coupling analysis, supports a

26 four transmembrane domain topology, with variants in transmembrane domains exhibiting

27 strongly deleterious effects on abundance and activity. Functionally constrained regions of the

28 protein define the active site, and we find that, of four conserved cysteines putatively critical for

29 function, only three are absolutely required. Finally, 25% of human VKOR missense variants show

30 reduced abundance or activity, possibly conferring warfarin sensitivity or causing disease.

31 **INTRODUCTION**

32   The enzyme vitamin K epoxide reductase (VKOR) drives the vitamin K cycle, which

33   activates blood coagulation factors. VKOR, an endoplasmic reticulum (ER) localized

34   transmembrane protein encoded by the gene *VKORC1*, reduces vitamin K quinone and vitamin

35   K epoxide to vitamin K hydroquinone (Li et al., 2004; Rost et al., 2004). Vitamin K hydroquinone

36   is required to enable gamma-glutamyl carboxylase (GGCX) to carboxylate Gla domains on

37   vitamin K-dependent blood clotting factors. VKOR is inhibited by the anticoagulant drug warfarin

38   (Czogalla et al., 2016; Zimmermann and Matschiner, 1974), and *VKORC1* polymorphisms

39   contribute to an estimated ~25% of warfarin dosing variability (Owen et al., 2010). For example,

40   variation in *VKORC1* noncoding and coding sequence can cause warfarin resistance (weekly

41   warfarin dose > 105 mg) or warfarin sensitivity (weekly warfarin dose < ~10 mg) (Osinbowale et

42   al., 2009; Yuan et al., 2005).

43   Though 15 million prescriptions are written for warfarin each year

44   (https://www.clincalc.com), fundamental questions remain regarding its target, VKOR. For

45   example, the structure of human VKOR is unsolved, though a bacterial homolog has been

46   crystallized (Li et al., 2010). A homology model based on bacterial VKOR has four transmembrane

47   domains, but the quality of the homology model is unclear, as human VKOR has only 12%

48   sequence identity to bacterial VKOR. Moreover, experimental validation of VKOR topology

49   yielded mixed results: similar biochemical assays suggested either three- or four-

50   transmembrane- domain topologies (Schulman et al., 2010; Tie et al., 2012; Wu et al., 2018).

51   Topology informs basic aspects of VKOR function including where vitamin K and warfarin

52   bind, so determining the correct topology and validating the homology model is critical. In

53   particular, VKOR has four functionally important, absolutely conserved cysteines at positions 43,

54   51, 132, and 135, the orientation of which differs between the two proposed topologies. In the

55   four transmembrane domain topology, all four cysteines are located on the ER lumenal side of

56   the enzyme. In this topology, cysteines 43 and 51 are hypothesized to be "loop cysteines" that

57   pass electrons from an ER-anchored reductase, possibly TMX (Schulman et al., 2010), to the

58   active site (Rishavy et al., 2011). However, in the three transmembrane domain topology, these

59   cysteines are located in the cytoplasm and other pathways would be required to convey electrons

60   to the redox center. Even for non-catalytic residues, topology plays an important role. For

61   example, vitamin K presumably binds near the redox center, and topology dictates which residues

62   make up the substrate binding site.

63   To understand the effect of human variants and to define the vitamin K and warfarin

64   binding sites, VKOR variant activity has been extensively studied in cell-based assays (Czogalla

65   et al., 2016; Shen et al., 2017; Tie et al., 2013). In addition to activity, VKOR protein abundance

66   has also been studied because abundance is an important driver of disease and warfarin

67   response. For example, VKOR R98W is a decreased- abundance variant that, in homozygous

68   carriers, causes vitamin K-dependent clotting factor deficiency 2 (Rost et al., 2004). A 5' UTR

69   polymorphism reduces VKOR abundance and can be used to predict warfarin sensitivity (Gong

70   et al., 2011). However, so far, the activity and abundance of only a handful of VKOR variants has

71   been tested.

72   Here, we used multiplexed, sequencing-based assays (Gasperini et al., 2016) to measure

73   the effects of 2,695 VKOR missense variants on abundance and 697 variants on activity. Our

74   analysis of the large-scale functional data supports a four transmembrane domain topology, which

75   an orthogonal evolutionary coupling analysis confirmed. Next, we identified distinct mutational

76   tolerance groups, which are concordant with a four transmembrane homology model. Combining

77   this homology model with variant abundance and activity effects, we identified an active site that

78   contains the catalytic residues C132 and C135 and shares six positions with a previously

79   proposed vitamin K binding site (Czogalla et al., 2016). We found that of four conserved cysteines

80   putatively critical for function, only three are absolutely required, and analyzed the mutational

81   signatures of two putative ER retention motifs. Human *VKORC1* variants present in genetic

82   databases and contributed by a commercial genetic testing laboratory were each classified based

4

83   on abundance and activity. While most variants show wild type-like activity, 25% show low

84   abundance or activity, which could confer warfarin sensitivity or cause disease in a homozygous

85   context. Finally, we analyzed warfarin resistance variants and found that they span a range of

86   abundances, indicating that increased abundance is an uncommon mechanism of warfarin

87   resistance.

88

89   **RESULTS**

90   **Multiplexed measurement of VKOR variant abundance using VAMP-seq**

91   To measure the abundance of VKOR variants, we applied Variant Abundance by

92   Massively Parallel sequencing (VAMP-seq), an assay we recently developed (Matreyek et al.,

93   2018). In VAMP-seq, a protein variant is fused to eGFP with a short amino acid linker. If the

94   variant is stable and properly folded, then the eGFP fusion will not be degraded, and cells will

95   have high eGFP fluorescence. In contrast, if the variant causes the protein to misfold, protein

96   quality control machinery will detect and degrade the eGFP fusion, leading to a decrease in eGFP

97   signal (Fig. 1a). mCherry is also expressed from an internal ribosomal entry site (IRES) to control

98   for expression. Differences in abundance are measured on a flow cytometer using the ratio of

99   eGFP to mCherry signal. To determine whether VAMP-seq could be applied to VKOR, we fused

100   eGFP to VKOR N- or C-terminally and found that both orientations had high eGFP signal (Figure

101   1—figure supplement 1). We compared N-terminally tagged wild type (WT) VKOR to R98W, a

102   variant that ablates a putative ER retention motif and reduces abundance (Czogalla et al., 2014),

103   and to TMD1Δ, a deletion of residues 10-30 which comprise the putative first transmembrane

104   domain (TMD1; Fig. 1b). Both reduced- abundance variants exhibited much lower eGFP:mCherry

105   ratios than WT, demonstrating that VAMP-seq could be applied to VKOR.

106   We constructed a barcoded site-saturation mutagenesis VKOR library that covered 92.5%

107   of all 3,240 possible missense variants. To express this library in HEK293T cells we used a Bxb1

108   recombinase landing pad system we previously developed (Matreyek et al., 2017). In this system,

109   each cell expresses a single VKOR variant. Recombined, VKOR variant-expressing cells were

110   then sorted into quartile bins based on their eGFP:mCherry ratios. Each bin was deeply

111   sequenced, and abundance scores were calculated based on each variant's distribution across

112   bins. Raw abundance scores were normalized such that WT-like variants had a score of one and

113   total loss of abundance variants had a score of zero (Fig. 1c). We performed seven replicates,

114   which were well correlated (Figure 1—figure supplement 2, mean Pearson's r = 0.73; mean

115   Spearman's ρ = 0.7, Supplementary Table 1). Abundance score means and confidence intervals

116   for each variant were calculated from the replicates.

117   The final dataset describes the effect of 2,695 of the 3,240 possible missense VKOR

118   variants on abundance (Fig. 1d and 1e). Validation of 10 randomly selected variants spanning the

119   abundance score range showed high concordance between individual eGFP:mCherry ratios

120   assessed by flow cytometry and VAMP-seq derived abundance scores (Fig. 1f, Pearson's r =

121   0.96, Spearman's ρ = 0.97).
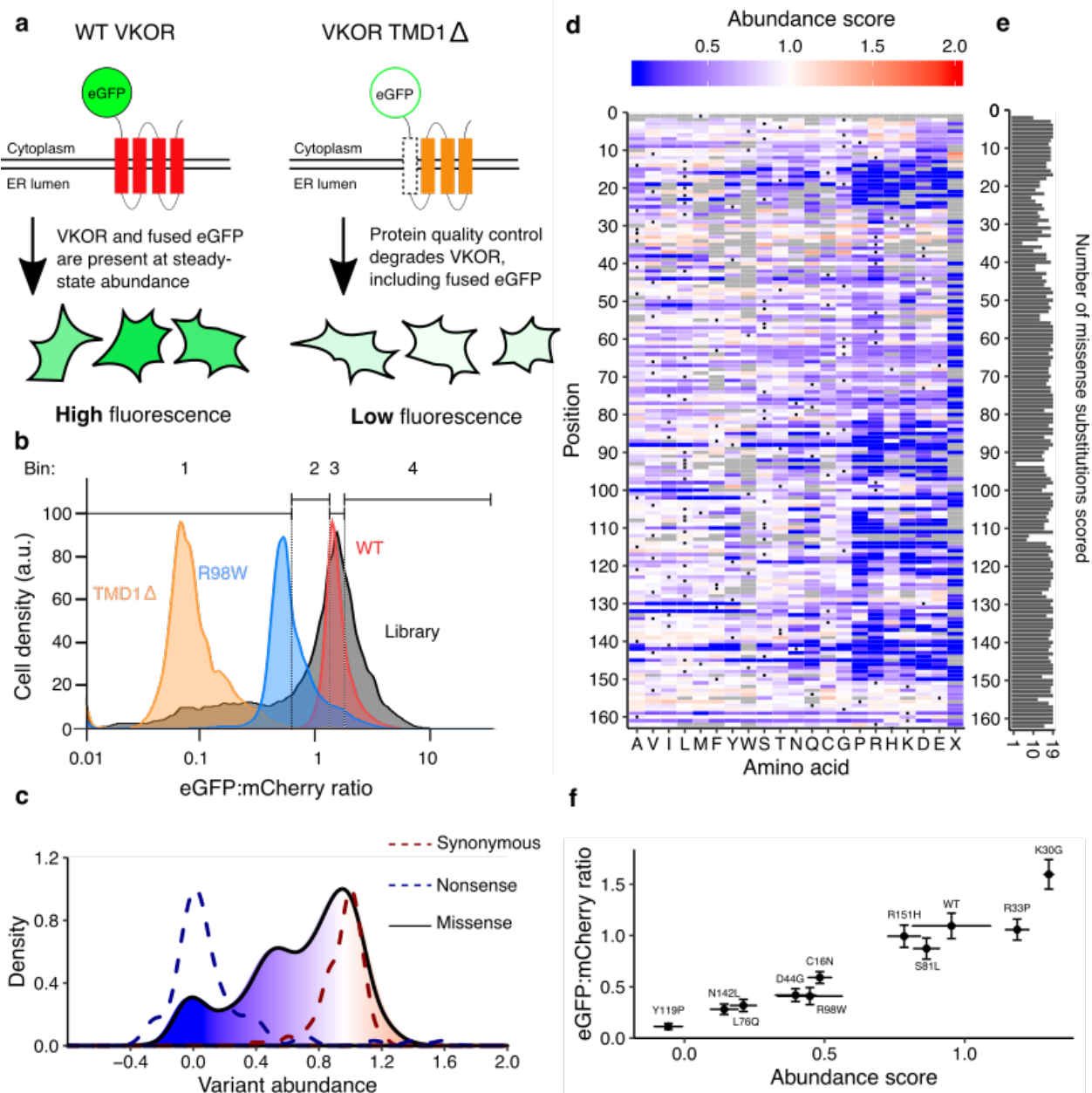
Figure 1. Multiplexed measurement of VKOR variant abundance using VAMP-seq. a, To measure abundance, an eGFP reporter is fused to VKOR. eGFP-tagged WT VKOR is folded correctly, leading to high eGFP fluorescence. However, a destabilized variant is degraded by protein quality control machinery, leading to low eGFP fluorescence. **b**, Flow cytometry is used to bin cells based on their eGFP:mCherry fluorescence intensity. Density plots of VKOR library expressing cells (grey, n = 12,109) relative to three controls: WT VKOR (red, n = 4,756), VKOR

7

129 98W (blue, n = 2,453) , and VKOR TMD1Δ (orange, n = 2,204) are shown. Quartile bins for FACS

130 of the library are marked. **c**, Abundance score density plots of nonsense variants (dashed blue

131 line, n = 88), synonymous variants (dashed red line, n = 127**)**, and missense variants (filled, solid

132 line, n = 2,695). The missense variant density is colored as a gradient between the lowest 10%

133 of abundance scores (blue), the WT abundance score (white) and abundance scores above WT

134 (red). **d**, Heatmap showing abundance scores for each substitution at every position within VKOR.

135 Heatmap color indicates abundance scores scaled as a gradient between the lowest 10% of

136 abundance scores (blue), the WT abundance score (white), and abundance scores above WT

137 (red). Grey bars indicate missing variants. Black dots indicate WT amino acids. **e**, Number of

138 substitutions scored at each position for abundance. **f**, Scatterplot comparing VAMP-seq derived

139 abundance scores to mean eGFP:mCherry (n = 1 replicate) ratios measured individually by flow

140 cytometry. Variants were selected at random to span the abundance score range.

141 **Figure 1-source data 1.** VKOR variant abundance and activity scores.

142 **Figure 1-source data 2.** Flow cytometry for monoclonal validation of variants. 11 variants were

143 run individually, values show mean and error for VAMP-seq score and eGFP:mcherry intensity.

144 **Multiplexed measurement of VKOR variant activity using a gamma-glutamyl**

145 **carboxylation reporter**

146 We also measured VKOR variant activity, adapting a HEK293 cell assay based on vitamin

147 K- dependent gamma-glutamyl carboxylation of a cell-surface reporter protein (Haque et al.,

148 2014). In this assay, if VKOR is active, a Factor IX domain reporter is carboxylated, secreted and

149 retained on the cell surface where it is detected with a carboxylation-specific, fluorophore-labeled

150 antibody. However, if VKOR is inactive, the reporter is not carboxylated and the antibody cannot

151 bind (Fig. 2a). We modified the HEK293 activity reporter cell line to eliminate endogenous VKOR

152 activity by knocking out both *VKORC1* and its paralog, VKORC1-like 1 (*VKORC1L1*) (Tie et al.,

153 2013) (Figure 2—figure supplement 1). We also installed a Bxb1 landing pad to facilitate

154 expression of individual VKOR variants or libraries (Figure 2—figure supplement 1).

155 Recombination of WT *VKORC1* into the landing pad of the HEK293 VKOR activity reporter cell

156 line yielded robust reporter activation, demonstrating that the reporter line could be used to assess

157 the activity of a library of VKOR variants (Fig. 2b).

158 We recombined a library of *VKORC1* variants into the HEK293 activity reporter cell line,

159 and sorted recombinant cells into quartile bins based on carboxylation-specific antibody binding.

160 Each bin was deeply sequenced and, as for VAMP-seq, an activity score was computed for each

161 variant. Final activity scores and confidence intervals were computed from six replicates for a total

162 of 697 missense variants, 21.5% of those possible (Figure 2—figure supplement 2, mean

163 Pearson's r = 0.62 and mean Spearman's ρ = 0.56, Supplementary Table 2). Our activity score

164 density plot showed that most variants had WT-like activity scores (Fig. 2c).

165

166
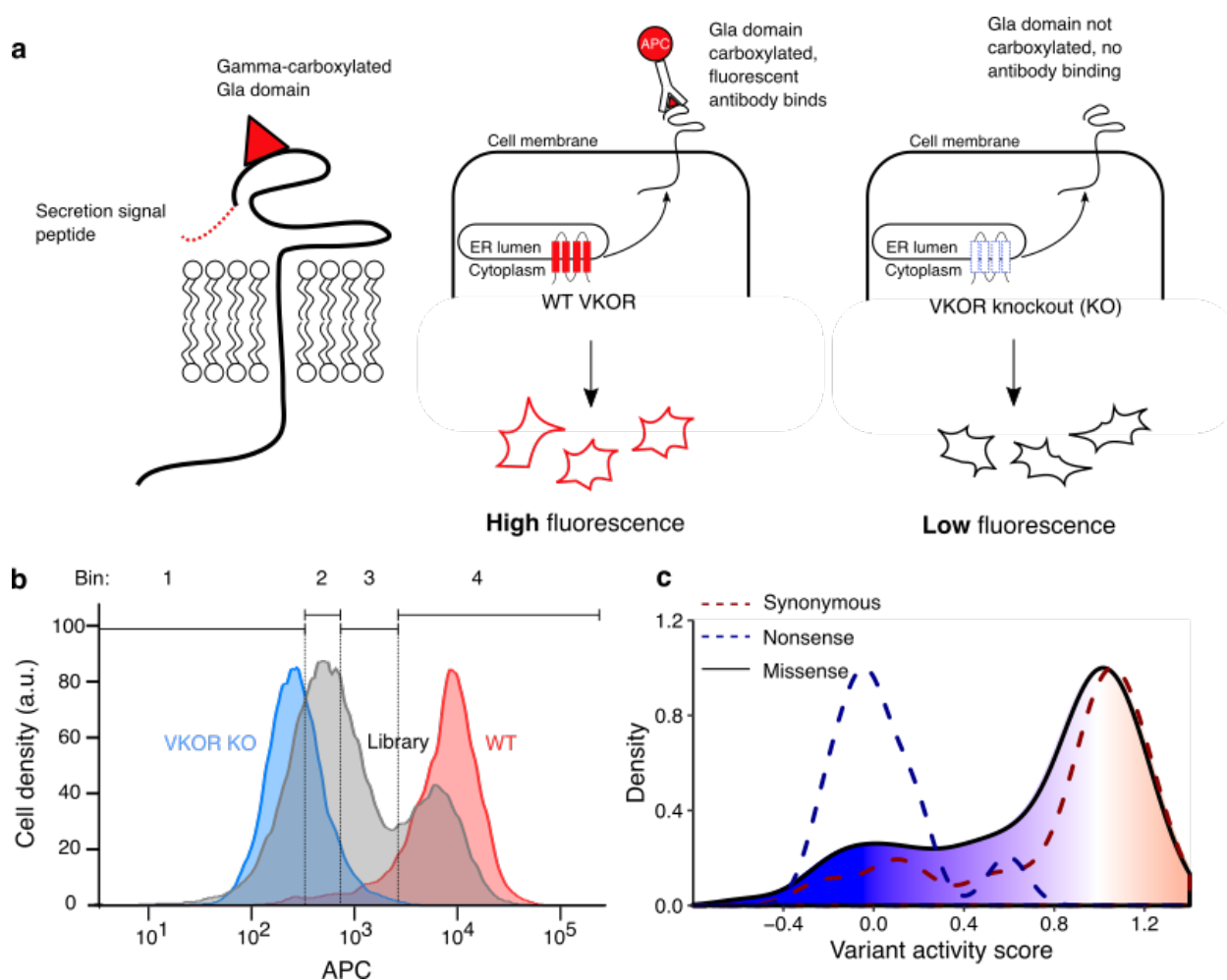
Figure 2. Multiplexed measurement of VKOR variant activity using a gamma-glutamyl carboxylation reporter. a, left panel, a Factor IX Gla domain reporter is expressed inHEK293 cells and consists of a prothrombin pre-pro-peptide which allows for processing and secretion, a Factor IX Gla domain, and Proline rich Gla protein 2 (PRGP2) transmembrane and cytoplasmic domains. **middle panel,** Cells expressing WT VKOR carboxylate the reporter Gla domain, which, upon trafficking to the cell surface, can be stained using a carboxylation-specific antibody conjugated to the fluorophore APC. VKOR knockout cells do not carboxylate the reporter, so the fluorescent antibody does not bind. **b**, Density plots of HEK293 activity reporter cells stained with APC-labeled carboxylation-specific antibody expressing no VKOR (blue, n = 7,188), WT VKOR (red, n = 4,107), or the VKOR variant library (grey, n = 41,418). Quartile bins

177   for FACS of the library are marked. **c**, Activity score density plots of nonsense variants (dashed

178   blue line, n = 14), synonymous variants (dashed red line, n = 35), and missense variants (filled,

179   solid line, n = 697). The missense variant density is colored as a gradient between the lowest

180   10% of activity scores (blue), the WT activity score (white) and activity scores above WT (red).

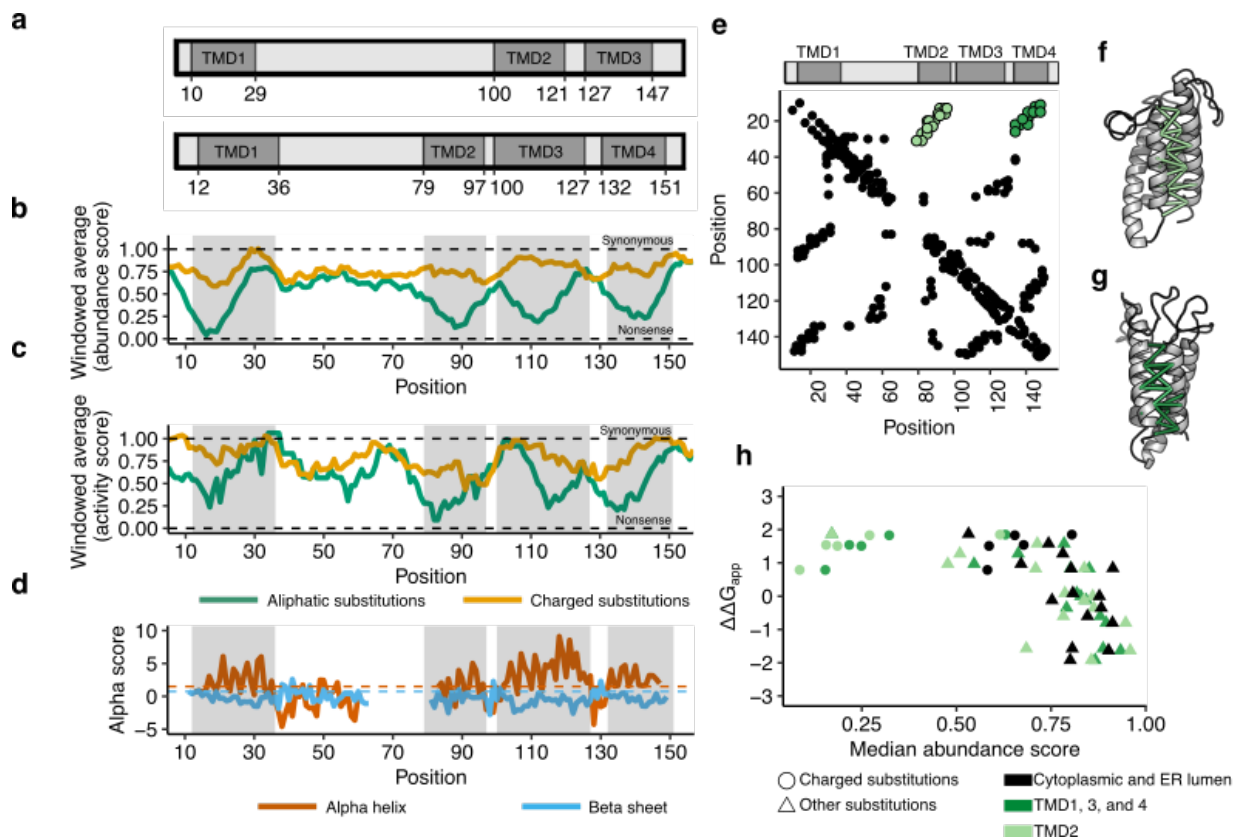181 **Human VKOR has four transmembrane domains**

182   Two different domain models, one with three transmembrane domains and another with

183 four, have been proposed for human VKOR (Li et al., 2010; Tie et al., 2012) (Fig. 3a). Because

184 charged amino acids occur infrequently in transmembrane domains and should be less tolerated,

185 we reasoned we could discriminate between these two models using a sliding window average of

186 the effect of charged substitutions on VKOR abundance (A. Elazar et al., 2016; Sharpe et al.,

187 2010). We found four clearly demarcated regions where charged substitutions profoundly reduced

188 VKOR abundance, relative to aliphatic substitutions (Fig. 3b). To exclude the possibility that the

189 eGFP tag used in our VAMP-seq assay somehow affected topology, we also analyzed the activity

190 score data. The activity data, derived using native, untagged VKOR, revealed the same four

191 minima as the abundance data (Fig. 3c). In addition to these four minima, we also observed an

192 activity score minimum at position 57, corresponding to a conserved serine at this position. This

193 serine occurs at the end of the lumenal half-helix hypothesized to shield the active site from non-

194 specific oxidation, so it is likely this signal is the result of disruption of that half helix. Together,

195 these results strongly support the hypothesis that, like its distant bacterial homolog, human VKOR

196 has four transmembrane domains.

197   To validate these findings, we performed evolutionary coupling analysis to infer the three-

198 dimensional structure suggested by co-evolution. We aligned 2,770 VKOR sequences from both

199 eukaryotes and prokaryotes and identified coupled residues using the EVcouplings software

200 (Hopf et al., 2012; Marks et al., 2011). Local patterns of evolutionary couplings (i.e. between

201 nearby positions, i to i+4) supported a four-helix topology. The helices predicted by these local

202 evolutionary couplings overlapped 70 of the 82 residues in alpha-helices of the bacterial structure

203 (PDB 4NV5) (Shen et al., 2017) and included in our alignment, non-gapped in >70% of aligned

204 VKOR sequences (hyper-geometric test p-value = 3.26^{-23}, Fig. 3d).

205   We identified non-local evolutionary coupling patterns characteristic of three-dimensional

206 contacts, which also strongly supported the four transmembrane domain model. Using these

207   contacts, we computationally folded human VKOR, yielding a modeled structure similar to the

208   bacterial structure (RMSD = 2.58 ÅÅ over 97/143 C_{alpha}, Figure 3—figure supplement 1). The

209   predicted tertiary structure hasd a four-helix topology, with antiparallel contacts between

210   transmembrane domains 1 and 2 (Fig. 3e, Fig. 3f) and between transmembrane domains 1 and

211   4 (Fig. 3e, Fig. 3g), which are only possible in a four-helix topology.

212   Comparison of our abundance data to the energy required to insert different amino acids

213   into the membrane yields additional evidence for the four transmembrane domain model. The

214   apparent change in free energy ($\Delta\Delta G_{app}$) of insertion relative to wild type for every amino acid has

215   been determined experimentally using deep mutational scanning of bacterial membrane proteins

216   (A. A. Elazar et al., 2016). Median abundance score and $\Delta\Delta G_{app}$ for each amino acid are

217   correlated (Fig. 3h). In particular, the large energetic cost of insertion of transmembrane domains

218   with charged amino acids is apparent, including within the second transmembrane domain TMD2.

219   Beyond insertion energies of individual amino acids, the overall hydrophobicity of transmembrane

220   helices contributes to membrane protein insertion (A. A. Elazar et al., 2016), as well as topology

221   (A. Elazar et al., 2016) and degradation (Guerriero et al., 2017). To determine whether overall

222   helix hydrophobicity was a large factor contributing to abundance scores, we calculated the free

223   energy for insertion (ΔG_{helix}) of each helix in the four transmembrane domain model using the ΔG

224   prediction server v1.0 (Hessa et al., 2007). The four helices of VKOR have different ΔG_{helix}, with

225   only transmembrane domain 3 having favorable ΔG_{helix} for insertion (TMD1: 0.435, TMD2: 1.551,

226   TMD3: -1.749, and TMD4: 1.734). Interestingly, we observed that TMD3 has a high density of

227   substitutions with WT-like scores (Figure 3—figure supplement 2), suggesting that TMD3's

228   favorable insertion energy might explain its mutational tolerance.

229

Figure 3. Abundance, activity, and evolutionary data support four transmembrane domains. a, Three and four transmembrane domain (TMD) models of VKOR, with TMDs in dark grey (Li et al., 2010; Tie et al., 2012). **b**, Windowed abundance score means (width = 10 positions) for charged substitutions (green) and aliphatic substitutions (gold). Dark grey boxes correspond to TMDs proposed in the four domain model. Dashed lines show median synonymous and the nonsense abundance scores. **c**, Windowed activity score means (width = 10 positions) for charged substitutions (green) and aliphatic substitutions (gold). Boxes and dashed lines as described in **b**. **d**, Secondary structure classification from local evolutionary couplings shown as alpha scores calculated for alpha helices (red) and beta sheets (blue). Dashed lines show significance cut-offs for alpha helices (1.5, red) and beta sheets (0.75, blue) (Toth-Petroczy et al., 2016). **e**, A contact map derived from evolutionary couplings. Black points show pairs of positions with significant coupling. Light green points show predicted contacts between TMD1 and TMD2. Dark green points show predicted contacts between TMD1 and TMD4. **f**, Predicted tertiary

14

243   contacts between TMD1-TMD2 (shown in light green in **e**) and **g**, TMD1-TMD4 (shown in dark

244   green in **e**) shown on the evolutionary couplings-derived hVKOR structural model. **h**, Scatterplot

245   comparing change in free energy for membrane insertion (A. A. Elazar et al., 2016) ($\Delta\Delta G_{app}$) to

246   median abundance score for each amino acid substitution. Cytoplasmic and lumenal positions

247   shown in black, TMD2 in light green, and TMDs 1, 3, and 4 in dark green. Charged substitutions

248   shown as circles, all other substitutions as triangles.

249   **Figure 3-source data 1.** Evolutionary couplings secondary structure predictions. Rows show

250   position, with columns showing alpha helix or beta sheet values and predictions.

251   **Figure 3-source data 2.** Evolutionary couplings 3D contact predictions. Rows show pairs of

252   residues with contact probabilities.

253   **Figure 3-source data 3.** Insertion energies from Elazar et al., 2016. Amino acids with

254   calculated insertion energy.

Detailed structural context of VKOR variant abundance effects

Having confirmed that human VKOR has four transmembrane domains, we next explored the detailed pattern of mutational effects we observed in the context of a four transmembrane domain homology model. We generated a homology model of human VKOR with I-TASSER using the bacterial VKOR structure (Shen et al., 2017; Yang et al., 2015). We performed hierarchical clustering of positions based on abundance scores, which yielded four groups of positions with characteristic mutational patterns (Fig. 4a). In Group 1, most substitutions were neutral or increased abundance; in Group 2, charged amino acid and proline substitutions decreased abundance; in Group 3, all substitutions decreased abundance; and in Group 4, all substitutions decreased abundance profoundly. Each group corresponded to a spatially distinct region of the homology model structure (Fig. 4b).

Group 1 positions were located in or adjacent to cytoplasmic and ER lumenal loops, which were more tolerant of substitutions than the transmembrane domains. At four Group 1 positions, K30, R33, R35, and R37, almost every substitution increased abundance. These positively charged positions are positioned either at the edge of TMD1 (K30) or in the ER lumen directly abutting the top of TMD1 (R33, R35, and R37). The "positive inside rule" (von Heijne, 1989), suggests that positive charges in membrane proteins generally reside in the cytoplasm, and this phenomenon is important for driving topology and membrane insertion (A. Elazar et al., 2016; Nilsson and von Heijne, 1990; von Heijne, 1989). K30, R33, R35, and R37 violate the positive inside rule, and substitutions at these positions may increase abundance by reducing charge inside the ER, reducing topological frustration or increasing membrane insertion efficiency. Compared to the other 12 arginine and lysine positions in WT VKOR, K30, R33, R35, and R37 are the only ones where substitutions generally increased abundance (Figure 4—figure supplement 1). Our observations are consistent with a screen of rat VKOR variants intended to improve protein expression in *E. coli* where deletion of positions 31 to 33 increased protein levels (Hatahet et al., 2015).

281   In Group 2, charged amino acids or proline substitutions generally decreased abundance.

282   Group 2 consisted mostly of transmembrane positions that had side chains projecting into the

283   lipid bilayer. Such transmembrane positions usually have hydrophobic, nonpolar side chains

284   (Ulmschneider and Sansom, 2001). Proline has poor helix forming propensity, explaining why

285   proline substitutions decreased abundance at these positions. Group 3 consisted of a mixture of

286   cytoplasmic, ER lumenal and transmembrane positions where most substitutions decreased

287   abundance. The cytoplasmic positions in this group included the putative dilysine ER localization

288   motif at positions 159 and 161. Also in this group were R98, part of another putative ER retention

289   motif at positions 98 and 100, and a glycine adjacent to TMD1 at position nine. The

290   transmembrane positions had side chains projecting towards neighboring transmembrane

291   helices, suggesting that, as for other membrane proteins (Fleming and Engelman, 2001; Mravic

292   et al., 2019), intramolecular sidechain packing is important for abundance.

293   Finally, substitutions in Group 4, consisting of positions G19, Y88, I141, and L145, resulted

294   in catastrophic loss of abundance. These positions are all in transmembrane domains with side

295   chains projecting into the interior of the protein. On the basis of strict mutational intolerance of

296   these positions, we hypothesized that their coordinated side chain packing comprises the core of

297   the VKOR four helix bundle. Indeed, Group 4 residues had dramatically lower relative solvent

298   accessibility than Groups 1-3 (Fig. 4c).

299   The four transmembrane domain homology models also allowed us to explain VKOR's

300   unusual trimodal distribution of variant abundance scores. Previous VAMP-seq derived

301   abundance score distributions for the cytosolic proteins TPMT and PTEN were bimodal (Figure

302   4—figure supplement 2) (Matreyek et al., 2018), and 15 of 16 deep mutational scans of other

303   soluble proteins using a variety of other assays also exhibited bimodal functional score

304   distributions (Gray et al., 2017). Because VKOR is an ER resident, transmembrane protein, we

305   hypothesized that its unusual trimodal abundance score distribution resulted from transmembrane

306   domain substitutions. Indeed, the lowest mode of the distribution was composed almost

17

307   exclusively of deleterious transmembrane domain substitutions (Fig. 4d). In contrast, the

308   intermediate mode consisted of substitutions in the ER lumen, cytoplasm, and transmembrane

309   domains. Similarly, substitutions that profoundly decreased activity occurred in transmembrane

310   domains (Fig. 4e).

311
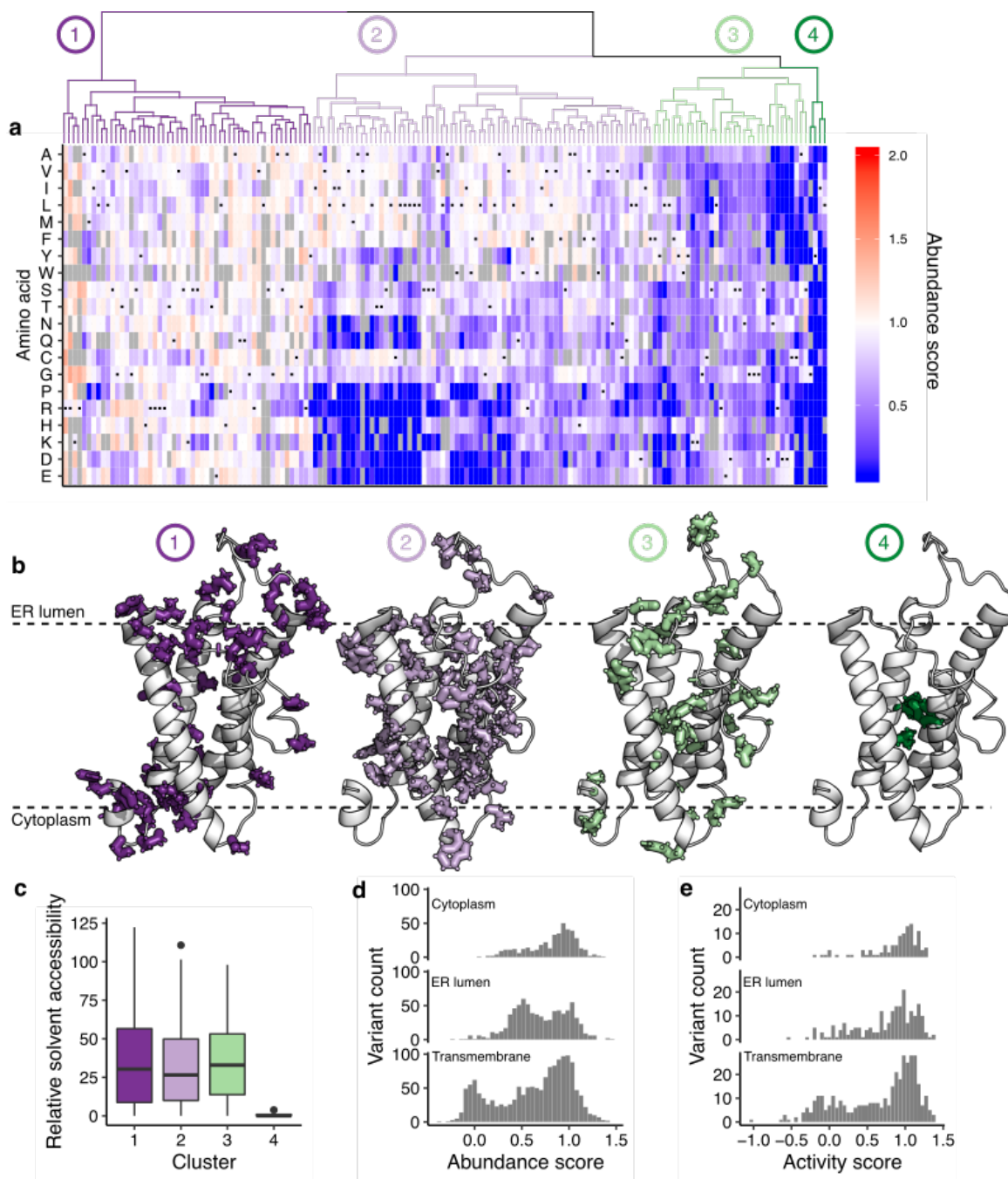
Figure 4. Hierarchical clustering of abundance scores and distributions of abundance and activity scores by domain. a, A heatmap showing hierarchical clustering of positions based on abundance score vectors, with the dendrogram above. Groups of positions, chosen based on the dendrogram, are numbered and colored. Heatmap color indicates abundance scores scaled as a gradient between the lowest 10% of abundance scores (blue), the WT abundance score (white)

317   and abundance scores above WT (red). Grey bars indicate missing variants. Black dots indicate

318   WT amino acids. **b**, Positions in groups 1-4 shown on the VKOR homology model, with numbers

319   and colors corresponding to panel **a**. **c**, Boxplot showing relative solvent accessibility of positions

320   in each cluster determined using DSSP (Kabsch and Sander, 1983; Touw et al., 2015) and

321   colored as in **b**. Bold black line shows median, box shows 25th and 75th percentile. Line shows

322   1.5 interquartile range above and below percentiles, and outliers are shown as black points. **d**,

323   Histograms of abundance scores for missense variants in the cytoplasmic, ER lumenal, or

324   transmembrane domains. **e**, Histograms of activity scores for missense variants in the

325   cytoplasmic, ER lumenal, or transmembrane domains.

Variant activity and abundance identify functionally constrained regions of VKOR

326

327   We reasoned that our activity and abundance data could reveal the location of functionally

328   important positions in VKOR, including the active site, since functionally important positions

329   should have many loss-of-activity but few loss-of-abundance variants. Thus, we calculated the

330   specific activity for each variant by taking the ratio of its rescaled activity score and abundance

331   scores (see Methods) ratio. We computed the median specific activity for each position;

332   substitutions at positions with low median specific activity generally have low activity relative to

333   their abundance. We set a specific activity threshold based on two absolutely conserved cysteines

334   that form VKOR's redox center, C132 and C135. Using this threshold, positions with the lowest

335   12.5% of specific activity scores and with at least four variants scored for activity were deemed

336   functionally constrained and mapped on the homology model of VKOR (Fig. 5a, Figure 5—figure

337   supplement 1). These 11 functionally constrained positions are organized around C132 and C135

338   and define, at least in part, the VKOR active site (Fig. 5b,c, Figure 5—figure supplements 1).

339   Among the functionally constrained positions are six positions previously identified in vitamin K

340   docking simulations (Czogalla et al., 2016) (Figure 5—figure supplement 1), including F55, which

341   is hypothesized to bind vitamin K. Three functionally constrained positions, G60, R61, and A121,

342   did not match any position in the active site predicted by docking, but were immediate neighbors

343   of W59 and L120, positions that were.

344   Besides C132 and C135, VKOR has two additional absolutely conserved cysteines, C43

345   and C51. In the four transmembrane domain model, C43 and C51 are postulated to be loop

346   cysteines that relay electrons to the C132/C135 redox center (Liu et al., 2014). We classified C43

347   as having low specific activity, but we only observed one variant at this position, so it was not

348   included in our set of functionally constrained positions (Figure 5—figure supplement 2). In

349   contrast, substitutions at C51 resulted in only modest activity loss, a phenomenon that has been

350   observed previously (Shen et al., 2017). Interestingly, every substitution at C51 and 15 of 19 at

351   C132 decreased VKOR abundance (Figure 5—figure supplement 2). Inside cells, the majority of

352   VKOR molecules have a C51-C132 disulfide bond, and warfarin binds to this redox state of VKOR

353   (Shen et al., 2017). Since disruption of this disulfide bond apparently impacts abundance as well

354   as activity, this bond may be important for VKOR folding and stability.

355   VKOR is thought to contain two sequences important for ER localization. The first is a

356   diarginine motif (RxR) at positions 98-100, and the second is a dilysine motif (KXKXX) at positions

357   159-163. While we did not directly measure localization, we found that only six of 19 R98 variants

358   and seven of 14 R100 variants resulted in low abundance (Figure 5—figure supplement 3). In

359   contrast, nearly all variants at K159 (14 of 18) and K161 (17 of 19) resulted in low abundance

360   (Figure 5—figure supplement 3). A histidine substitution was tolerated at position 161, which

361   mimics the KXHXX motif commonly found in coronaviruses and a small number of human proteins

362   (Ma and Goldberg, 2013). Because protein localization and degradation are coupled (Hessa et

363   al., 2011), we suggest that the reductions in abundance we observe are the result of degradation

364   caused by mislocalization, and that the dilysine motif at positions 159-163 is essential for VKOR

365   ER localization. Overall, comparison of VKOR variant activity and abundance revealed

366   functionally important regions, refining our understanding of the active site, redox-active

367   cysteines, and ER retention motifs.

368
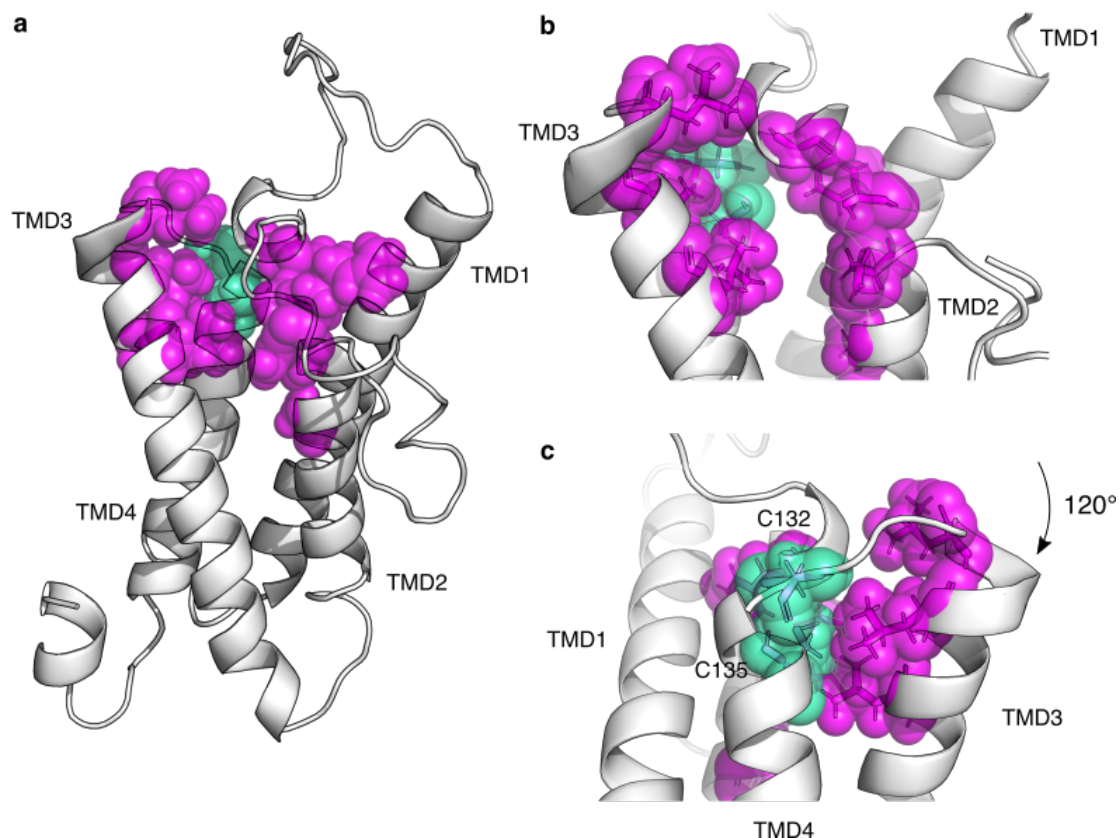
Figure 5. Functionally constrained positions reveal VKOR active site and critical cysteines.

a, Positions with the lowest 12.5% of median specific activity scores and at least four variants scored for activity are shown as magenta spheres on the VKOR homology model. Cysteines C132, and C135, also in the bottom 12.5% of median specific activity scores, are shown in green spheres. **b**, Magnified view of the redox center cysteines (positions 132, and 135, green spheres) and surrounding residues that define the active site (magenta spheres). Residues shown in transparent spheres, with side chains also shown in sticks. **c**, panel **b** rotated 120°C.

Figure 5-source data 1. VKOR positional abundance and activity scores. Rows show positions, with columns showing median abundance score, median activity score, rescaled scores, and specific activity score.

23

Functional consequences of VKOR variants observed in humans

Variation in VKOR is linked to both disease and warfarin response, but the overwhelming majority of VKOR variants found in humans so far have unknown effects. Thus, we curated a total of 215 variants that had either been previously reported in the literature as affecting warfarin response (Supplementary Table 3), were in ClinVar (Landrum et al., 2014), were in gnomAD v2 or v3 (Karczewski et al., 2019), or were present in individuals whose healthcare provider had ordered a multi-gene panel test from a commercial testing laboratory (Color Genomics) (Supplementary Table 4). Of eight variants present in ClinVar, we included only one (D36Y) in our analysis as it was the only variant reviewed by an expert panel (Kurnik et al., 2012). 159 variants were present in gnomAD, and all but one missense variant (D36Y) had population frequencies less than 0.2%. 28 were literature-curated warfarin response variants, only 12 variants of which were in one of the databases surveyed. D36Y was the only warfarin response variant present in all databases, ClinVar, gnomAD, and Color (Figure 6—figure supplement 1).

We classified 193 of the 215 variants we curated according to their abundance (Supplementary Table 4). All synonymous variants with the exception of two were WT-like or possibly WT-like, while the three nonsense variants scored had as having low abundance (Fig. 6a). Missense variants spanned all abundance categories, with 129 (60%) having WT-like or possibly WT-like abundance. 30 missense variants were low abundance, and 12 were high abundance. The single known pathogenic variant R98W was low abundance (Fig. 6b). We also classified 54 variants according to their activity (Supplementary Table 4). Only one variant, A115V, exhibited low activity. It had WT-like abundance, indicating that the loss of activity is not due to loss of abundance.

We examined warfarin response variants including W5X, the only variant observed so far linked to human warfarin sensitivity (Oldenburg et al., 2004). As expected, W5X was low abundance, reinforcing that heterozygous loss of VKOR is the cause of warfarin sensitivity in carriers of this variant. Warfarin resistance variants, on the other hand, are predicted to abrogate

24

405   warfarin binding (Li et al., 2010), but it is unclear whether these variants have appreciable effects

406   on abundance or activity. We found that warfarin resistance variants span a range of abundances

407   and that the distribution of warfarin resistant variant abundance was not different from missense

408   variants generally (Fig. 6c, two-sided Kolmogorov-Smirnov test p= 0.438). Five warfarin-

409   resistance variants had low abundance, suggesting that these variants must block drug binding

410   or increase activity to confer resistance. One variant, A26T, had high abundance, a possible

411   mechanism of warfarin resistance. The five warfarin resistance variants, R58G, W59L, V66M,

412   G71A, and N77S, whose activity we scored, were all WT-like. Thus, our abundance and activity

413   data are consistent with warfarin resistance arising largely from variants that block warfarin

414   binding.
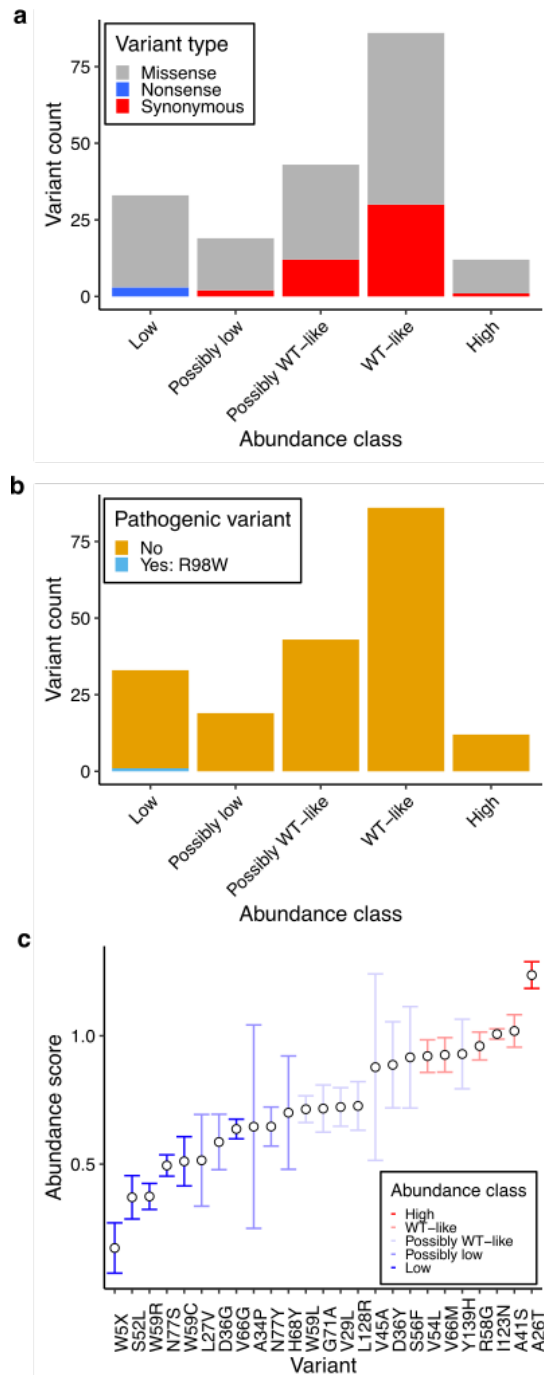
Figure 6. Characterization of human variants using abundance and activity data. a, Histogram of abundance classifications for variants from gnomAD, ClinVar, and Color Genomics. Nonsense variants colored in blue, synonymous in red, and missense in grey. **b**, Histogram of abundance classifications for same variants in **a**, colored by pathogenicity. The only variant known to cause disease, R98W, is colored in blue. All other variants shown in yellow. **c**, Scatterplot

26

421   showing abundance scores for literature-curated warfarin resistance variants. Bars show

422   standard error and are colored by abundance class. Variants are arranged in order of abundance

423   score.

424   **DISCUSSION**

425   We conducted multiplexed assays to measure the effects of 2,695 VKOR variants on

426   abundance and 697 variants on activity. Both abundance and activity data provided evidence for

427   a four transmembrane topology, which was further supported by evolutionary couplings analysis.

428   We evaluated a VKOR homology model in the context of the patterns of variant effects on

429   abundance we measured, and found that the homology model could explain these patterns. Low

430   specific activity residues mapped onto this homology model identify, at least in part, the active

431   site, which largely overlaps with the results of a vitamin K docking simulation (Czogalla et al.,

432   2016). Our active site is shallower than what the docking simulation predicts; this is the result of

433   low abundance scores at some of the deeper, transmembrane positions predicted by docking to

434   bind the isoprenoid chain of vitamin K (F87, Y88), and poor coverage of activity scores for other

435   positions (V112, S113). In light of the fact that substitutions at F87 and Y88 resulted in low

436   abundance, we note that the modeled vitamin K binding mode would disrupt packing of VKOR

437   core residues and require repacking of helices to maintain protein stability (Merkle et al., 2018).

438   In addition to the active site, substitutions at the dilysine and, to a lesser extent, the diarginine ER

439   localization motifs caused abundance loss.

440   We also used our large-scale functional data to analyze 215 VKOR variants found in

441   humans. 16% of these variants affect neither activity nor abundance; we identified 54 previously

442   uncharacterized low abundance or low activity variants that could be pathogenic or alter warfarin

443   response. We found that only one warfarin resistance variant had increased abundance,

444   indicating that increased abundance is not a pervasive warfarin resistance mechanism. All five of

445   the warfarin resistance variants whose activity we scored were WT-like. Taken together these

446   data support the notion that warfarin resistance generally involves alterations to warfarin binding

447   rather than abundance or activity. We analyzed one known warfarin sensitivity variant, W5X, and

448   found that it is low abundance, suggesting that one should not exclude the possibility that any of

449   the 52 other low abundance variants, if found in a person, also confer warfarin sensitivity.

28

450   While our VKOR variant abundance and activity data illuminates various aspects of

451   VKOR's structure and function, the data have limitations. For example, neither assay captures

452   variant effects on mRNA splicing. Both assays have limited dynamic ranges, meaning that subtle

453   effects on abundance or activity cannot be discerned. In addition, both assays have inherent

454   noise, largely arising from the limited number of cells we can sample due to the bottleneck of cell

455   sorting. We account for this noise by filtering each dataset based on variant frequency and

456   presenting a confidence interval for each abundance and activity score.
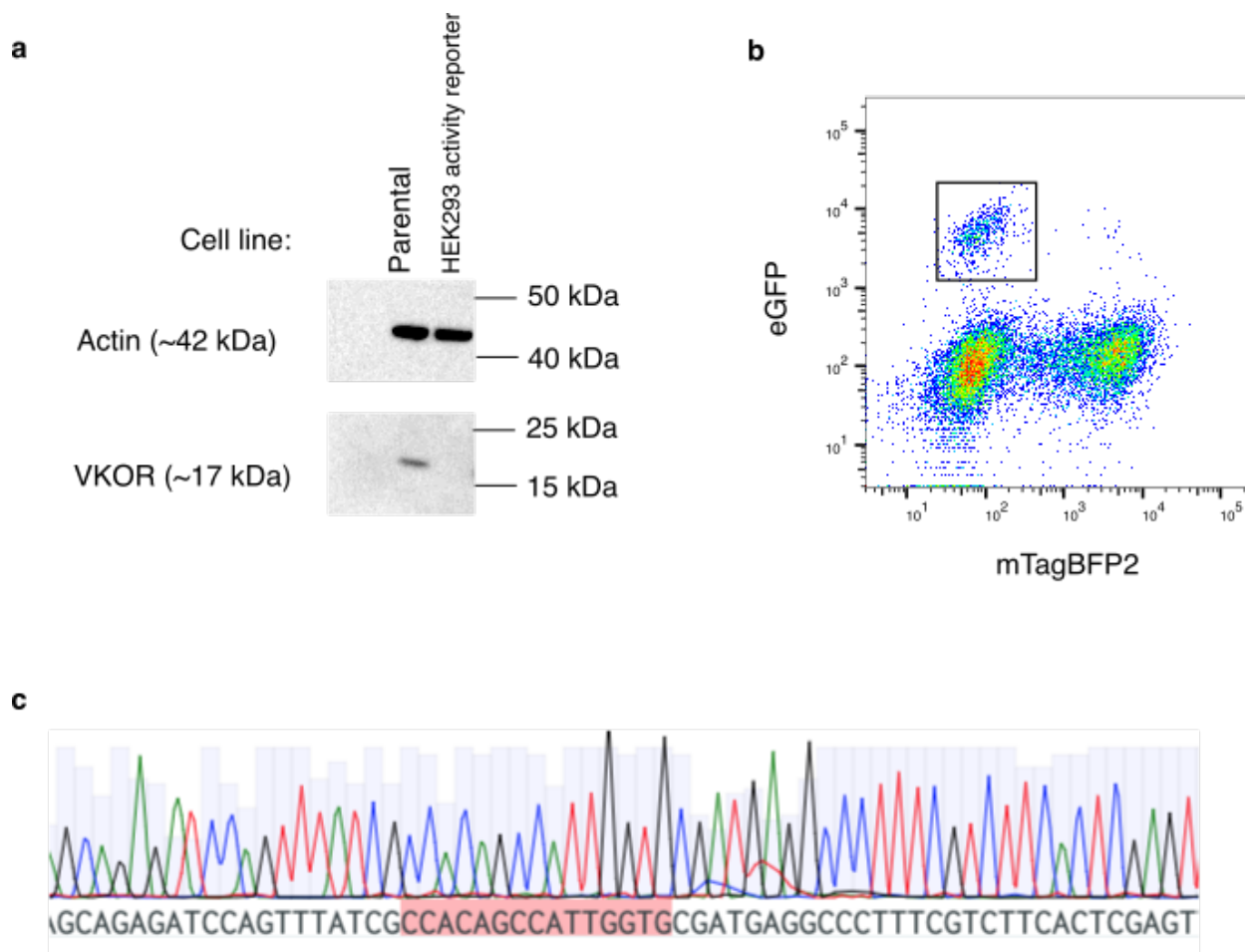
457   In the future, we envision that the assays we used could be employed to better understand

458   VKOR's interaction with warfarin. Here, we could measure warfarin's effect on both variant

459   abundance and activity, mapping the warfarin binding site more finely. In addition, we could

460   identify warfarin resistance mutations that have not yet been observed in the clinic and group

461   variants by their putative resistance mechanism. Overall, our work highlights the value of

462   multiplexed assays of variant effect for better understanding protein structure, function and human

463   variant effects.
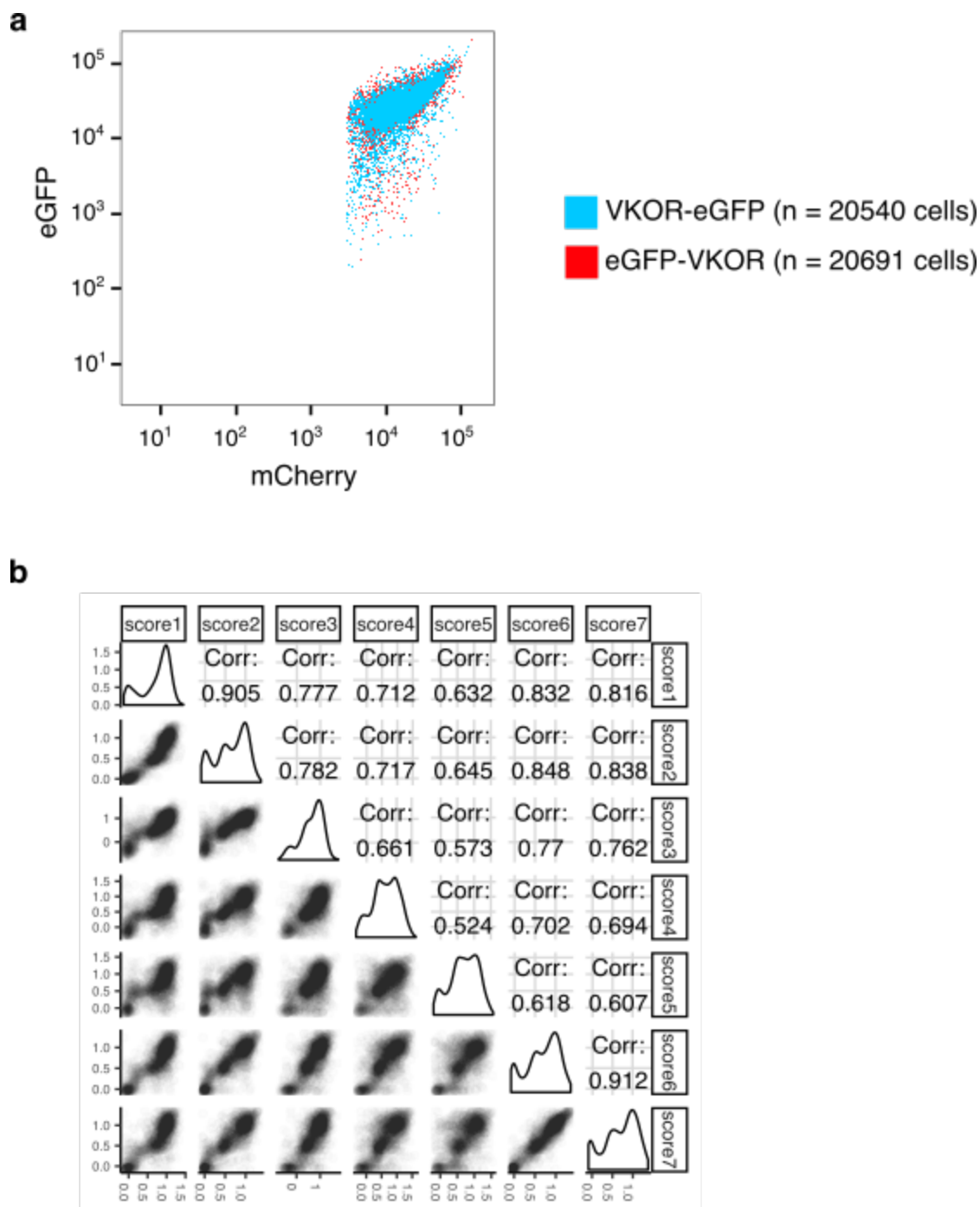
Figure 1—figure supplement 1. **VKOR abundance assay pilot experiment and replicate correlations. a**, Scatterplot of eGFP vs. mCherry fluorescence for cells expressing either C-terminally eGFP-tagged VKOR (VKOR-eGFP, blue) or N-terminally eGFP-tagged VKOR (eGFP-VKOR, red). **b**, Pairwise abundance score correlations between replicate sorting experiments.

469   Seven VAMP-seq replicates were performed. Pearson's correlation coefficients are shown. Score

470   numbers in this figure correspond to replicate numbers shown in Supplementary Table 1.

a, Western blot of parental cell line vs. HEK293 activity reporter cell line. Loading control is actin (42 kDa). VKOR was probed using an antibody generated against a peptide from the C-terminal of VKOR (FRKVQEPQGKAKRH)(Hallgren et al., 2006). The band for VKOR at 17 kDA is visible in the parental cell line but is not present in the HEK293 activity reporter cell line. **b**, Scatterplot showing mTagBFP2 vs. eGFP mean fluorescence intensities for HEK293 activity reporter cells recombined with a construct encoding WT VKOR followed by internal ribosomal entry sequence and eGFP. The emergence of a distinct recombined population that is eGFP positive and mTagBFP2 negative (black outline, n = 768 cells) supports the presence of a single landing pad into the cell genome, and not multiple insertions. **c**, A chromatogram showing the barcode sequence of the landing pad inserted at the *AAVS1* locus in the HEK293 activity reporter cell line.

471

472 **Figure 2—figure supplement 1. HEK293 VKOR activity reporter cell line characterization**.

483   The presence of a single barcode, highlighted in red, instead of mixed peaks, supports insertion

484   of one landing pad rather than multiple landing pads.
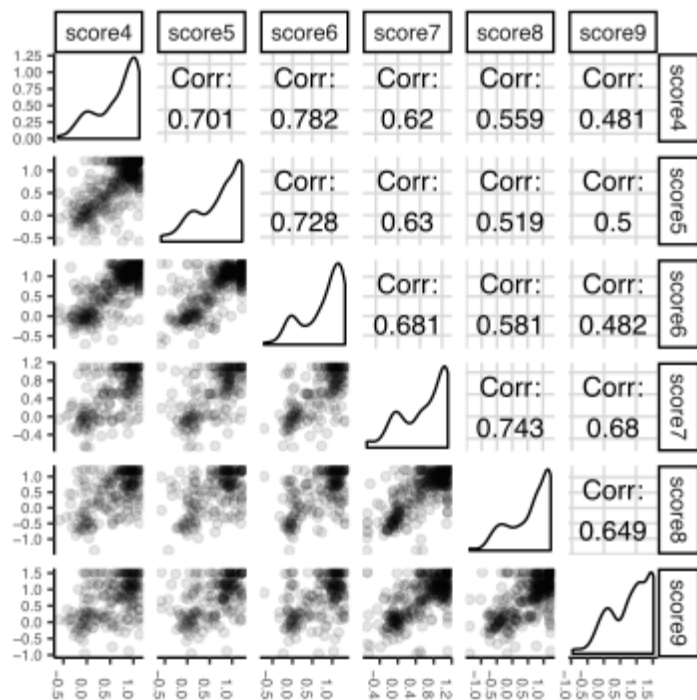
485

486 **Figure 2—figure supplement 2. Correlations of activity assay replicates.** Pairwise score

487 correlations between replicate sorting experiments of VKOR activity. Six replicates of the activity

488 assay were performed. Pearson's correlation coefficients are shown. Score numbers in this panel

489 correspond to replicate numbers shown in Supplementary Table 2.

490
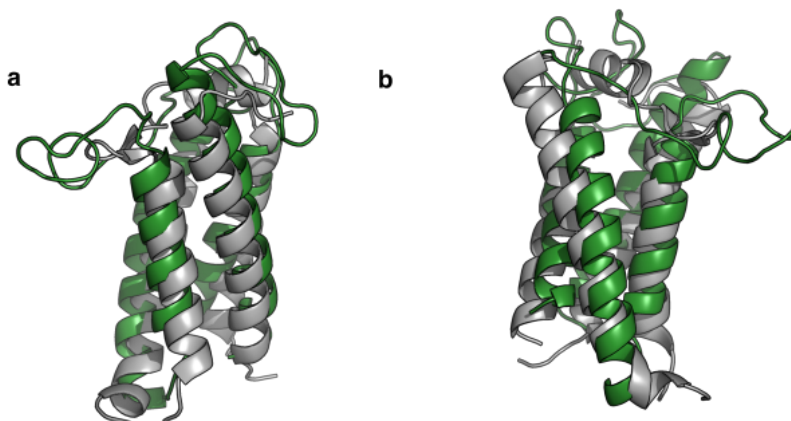
Figure 3—figure supplement 1. **Bacterial VKOR structure and EV-couplings folded model**

are highly similar. a, Pymol graphic showing overlap between EVcouplings-folded model of

VKOR (shown as a cartoon in green) compared to the bacterial structure (PDB: 4NV5, shown

as a cartoon in grey). **b**, shows the same two structures, rotated 120°C.
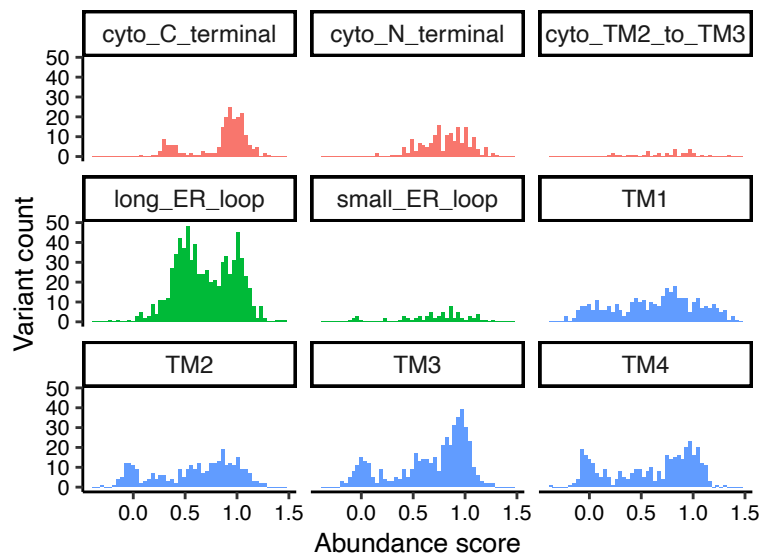
Figure 3—figure supplement 2. Specific domain abundance scores. Histograms of abundance scores for missense variants, grouped by domain and colored by cytoplasmic, ER lumenal, or transmembrane localization.

499

Figure 4—figure supplement 1. **TMD1-adjacent positive residues show pattern of**
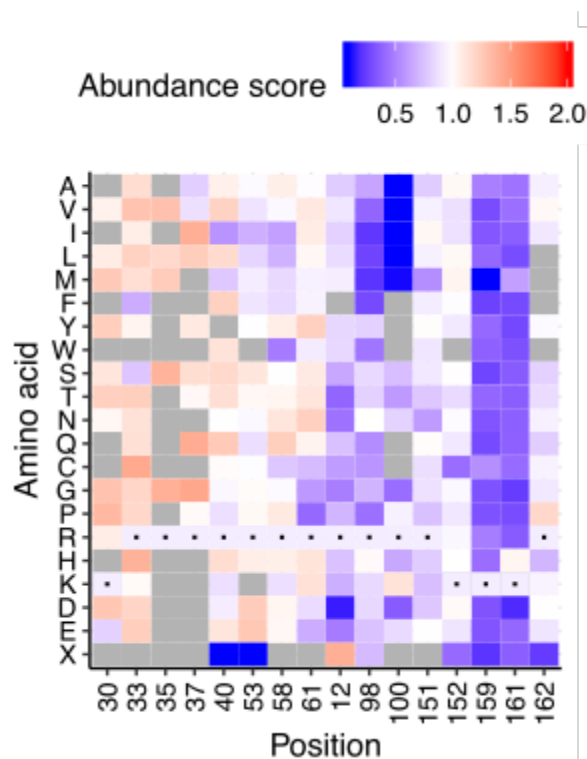
increased abundance. Heatmap of abundance scores for all arginines and lysines in VKOR.

First four positions (K30, K33, K35, K37) are in or proximal to transmembrane domain 1.

Heatmap color indicates abundance scores scaled as a gradient between the lowest 10% of

abundance scores (blue), the WT abundance score (white) and abundance scores above WT

(red). Grey bars indicate missing variants. Black dots indicate WT amino acids.

506
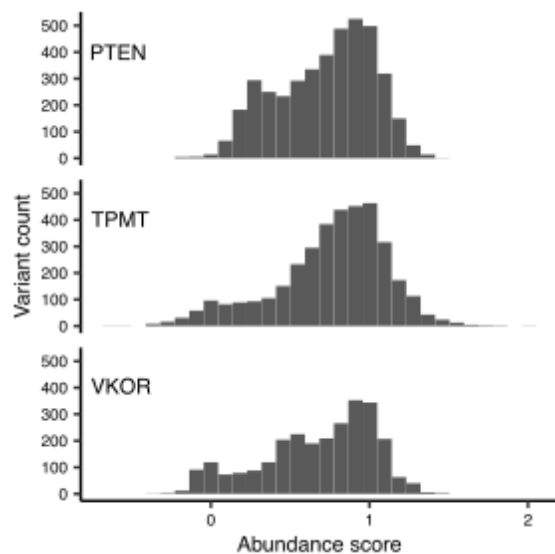
Figure 4—figure supplement 2. **Trimodality of missense variant abundance scores is unique to VKOR.** Histograms of abundance scores for missense variants for three proteins: PTEN, TPMT, and VKOR.

510
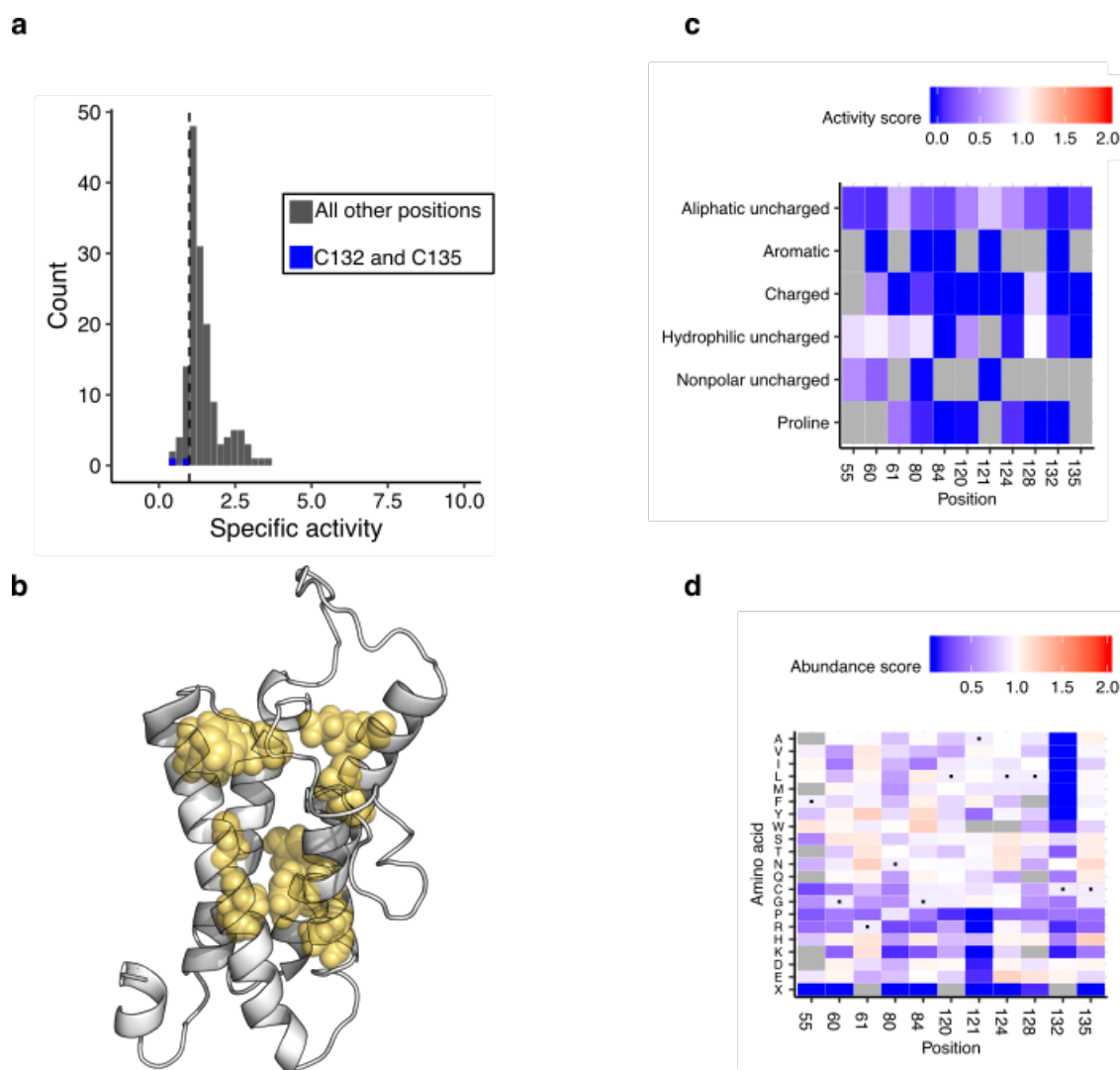
Figure 5—figure supplement 1. **VKOR active site analysis.** Histogram of specific activity,

with catalytic cysteines C132 and C135 labeled in blue. Dashed line demarcates bottom 12.5%.

b, Active site positions as defined by computational docking, shown on the homology model as

yellow spheres(Czogalla et al., 2016). **c**, Heatmap of activity scores for residues with lowest

12.5% of specific activity scores, collapsed by amino acid class. Color indicates activity scores

scaled as a gradient between the lowest 10% of activity scores (blue), the WT activity score

(white) and activity scores above WT (red). Grey indicates missing data. **d**, Heatmap of

39

518   abundance scores for residues with lowest 12.5% of specific activity scores. Color legend same

519   as described in **c,** applied to abundance scores.

520

521
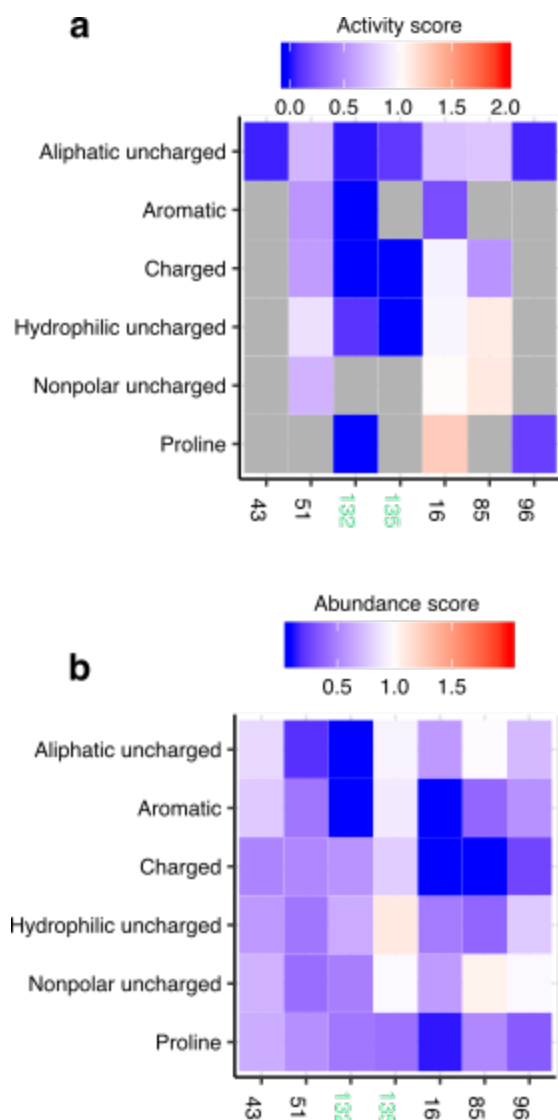
Figure 5—figure supplement 2. **Conserved cysteine analysis. a**, Heatmap of activity scores for cysteines. Catalytic cysteines C132 and C135 labeled in green. Color indicates activity scores scaled as a gradient between the lowest 10% of activity scores (blue), the WT activity score (white) and activity scores above WT (red). Grey indicates missing data. **b**, Heatmap of abundance scores for cysteines. Catalytic cysteines C132 and C135 labeled in green. Color legend same as described in **a,** applied to abundance scores.
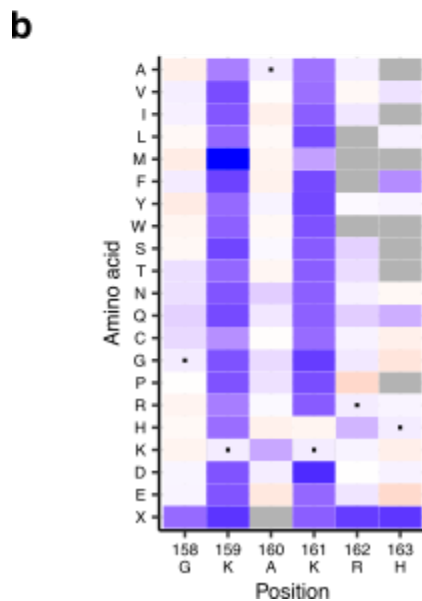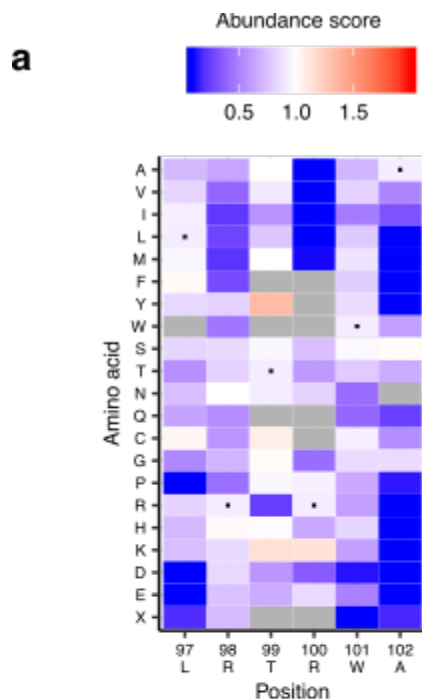
Figure 5—figure supplement 3. **VKOR localization motif analysis. a**, Heatmap of

abundance scores for diarginine ER retention motif. X-axis shows residues and position. Color

indicates abundance scores scaled as a gradient between the lowest 10% of abundance scores

(blue), the WT abundance score (white) and abundance scores above WT (red). Grey indicates

missing data. **b**, Heatmap of abundance score for dilysine ER retention motif. X-axis shows

residues and position.

535
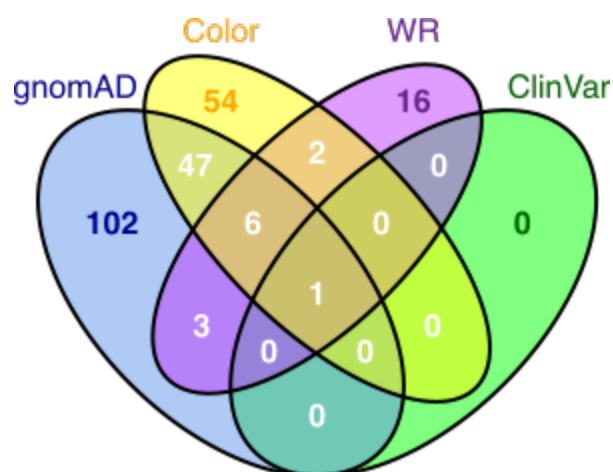
Figure 6—figure supplement 1. Human VKOR variant curation summary. Venn diagram of VKOR missense variants present in gnomAD v2 and v3, ClinVar, Color Genomics, a commercial genetic testing company, and literature-reported warfarin resistant variants.

Replicate	Cells recombined	Cells sorted in four-way sort
1	256,324	200,000
2	256,324	200,000
3	111,570	183,000
4	155,169	100,000
5	105,000	103,000
6	54,045	120,000
7	54,045	125,000

539

540   Supplementary Table 1. The seven replicates of VAMP-seq performed with cells recombined

541   and sorted for each.

Replicate	Cells recombined	Cells sorted in four-way sort
1	85,492	200,000
2	85,492	150,000
3	85,492	100,000
4	165,000	90,000
5	165,000	90,000
6	165,000	100,000

542 Supplementary Table 2. The six replicates of the activity assay performed with cells recombined

543 and sorted for each.

Position	Variant	Described in
5	W5*	(Oldenburg et al., 2004)
26	A26T	(Watzka et al., 2011)
26	A26P	(Bodin et al., 2008)
27	L27V	(Peoc'h et al., 2009)
28	H28Q	(Watzka et al., 2011)
29	V29L	(Oldenburg et al., 2004)
34	A34P	(Harrington et al., 2011)
36	D36Y	(Loebstein et al., 2007)
36	D36G	(Watzka et al., 2011)
41	A41S	(Rieder et al., 2005)
45	V45A	(Oldenburg et al., 2004)
52	S52W	(Watzka et al., 2011)
52	S52L	(Schmeits et al., 2010)
54	V54L	(Bodin et al., 2008; Harrington et al., 2008)
56	S56F	(Watzka et al., 2011)
58	R58G	(Oldenburg et al., 2004)
59	W59R	(Wilms et al., 2008)
59	W59L	(Watzka et al., 2011)
59	W59C	(Watzka et al., 2011)

66	V66M	(Oldenburg et al., 2004)
66	V66G	(Watzka et al., 2011)
68	H68Y	(Osman et al., 2006)
71	G71A	(Watzka et al., 2011)
77	N77Y	(Watzka et al., 2011)
77	N77S	(Watzka et al., 2011)
123	I123N	(Watzka et al., 2011)
128	L128R	(Oldenburg et al., 2004)
139	Y139H	(Watzka et al., 2011)

544

545   **Supplementary Table 3**. Variants found in humans that cause warfarin sensitivity or resistance,

546   and references in which they were first reported.

547

548   **Additional supplementary files and source data**

549   **Supplementary Table 4**. Abundance and activity data for human variants found in ClinVar,

550   gnomAD v2 and v3, and Color Genomics dataset.

551   **Supplementary Table 5.** Reagents and resources table.

552   **Supplementary Table 6**. Names and sequences for oligos used in this paper.

553   **Figure 1-source data 1. VKOR variant abundance and activity scores.**

554   **Figure 1-source data 2. Flow cytometry for monoclonal validation of variants.**

555   **Figure 3-source data 1. Evolutionary couplings secondary structure predictions.**

556   **Figure 3-source data 2. Evolutionary couplings 3D contact predictions.**

557   **Figure 3-source data 3. Insertion energies from Elazar et al., 2016.**

558   **Figure 5-source data 1. VKOR positional abundance and activity scores.**

47

559　**METHODS**

560　**General reagents, DNA oligonucleotides, and plasmids.**

561　　Details on general reagents can be found in Supplementary Table 5. Unless otherwise

562　noted, all chemicals were obtained from Sigma and all enzymes were obtained from New

563　England Biolabs. *E. coli* were cultured at 37°C in Luria broth. All cell culture reagents were

564　purchased from ThermoFisher Scientific unless otherwise noted. HEK 293T cells (ATCC CRL-

565　3216) and derivatives thereof were cultured in Dulbecco's modified Eagle's medium

566　supplemented with 10% fetal bovine serum, 100 U ml^{-1} penicillin, and 0.1 mg ml^{-1} streptomycin.

567　Cells were induced with 2.5 ug mL^{-1} doxycycline. Cells were passaged by detachment with

568　trypsin-EDTA 0.25%, and cells were prepared for sorting by detachment with versene. All cell

569　lines tested negative for mycoplasma. Because our activity assay is vitamin-K dependent, all

570　activity assays were done with the same lot of FBS to ensure similar concentrations of vitamin K

571　in each replicate.

572　　All synthetic oligonucleotides were obtained from IDT and can be found in

573　Supplementary Table 6. All non-library-related plasmid modifications were performed with

574　Gibson assembly(Gibson et al., 2009).

575

576　**Library construction**

577　　A gBLOCK with a codon-optimized sequence for human VKOR was ordered from IDT. It

578　was then cloned into the vector pHSG298 (Clontech). Saturation mutagenesis primers were

579　designed for each codon in VKOR from positions 2 to 163(Jain and Varadarajan, 2014) and

580　ordered resuspended from IDT. Forward and reverse primers for each position were mixed at 2.5

581　mM, and used in a PCR reaction with 125 pg of pHSG298-VKOR, 5% DMSO, and 5 uL of KAPA

582　Hifi Hotstart 2X ReadyMix. PCR products were visualized on a 0.7% agarose gel to confirm

583　amplification of the correct product.

584

585 PCR products were then quantified using the Quant-iT PicoGreen dsDNA Assay kit

586 (Invitrogen) using DNA control curves done in triplicate. To pool, a total amount of DNA for

587 each reaction was calculated that maximized the volume to be drawn from the lowest

588 concentration PCR product. Pooled PCR products were cleaned and concentrated using

589 Zymogen Clean and Concentrate kit and then gel extracted. The pooled library was

590 phosphorylated with T4 PNK (NEB), incubated at 37°C for 30 minutes, 65°C for 20 minutes, and

591 then 4° indefinitely. 8.5 uL of this phosphorylated product was combined with 1 uL of 10X T4

592 ligase buffer (NEB) and 0.5 uL of T4 DNA ligase (NEB) to make a 10 uL overnight ligation reaction.

593 This reaction was incubated at 16°C overnight.

594 The overnight ligation was then cleaned and concentrated (Zymogen) and eluted in 6 uL

595 of ddH2O. 1 uL of this ligation was then transformed into high efficiency *E. coli* using

596 electroporation at 2 kV. Each reaction contained 1 uL of ligation (or ligation control or pUC19 10

597 pg/uL) and 25 uL of E. coli. 975 uL of pre-warmed SOC media was added to each cuvette after

598 electroporation, transferred to a culture tube, and recovered at 37°C, shaking for 1 hour. At 1

599 hour, 1 and 10 uL samples from all cultures were taken and plated on appropriate media (LB +

600 kanamycin for ligation and ligation control; LB + ampicillin for pUC19), the remaining 989 uL was

601 used to inoculate a 50 mL culture (+ kanamycin). Plates and 50 mL culture were incubated at

602 37°C overnight (shaking for 50 mL culture). Colonies on plates were then counted, and counts

603 were used to calculate how many unique molecules were transformed to gauge coverage of the

604 library. 50 mL culture was spun down and midiprepped.

605 To transfer the library from pKan to the recombination vector, the pKan library and

606 recombination vector were digested with XbaI and AflIII for 1 hour at 65°C. The library and cut

607 vector were then gel extracted. The library was then ligated with the cut vector at 5:1 using T4

608 ligase, overnight at 16°C. The ligation was heat inactivated the next morning, clean and

609 concentrated. Another high efficiency transformation was performed the same as described

610 above, except this ligation was plated on LB + ampicillin (antibiotic switching strategy). Plates and

49

611   50 mL culture were incubated at 37°C overnight (shaking for 50 mL culture). Colonies on plates

612   were then counted, and counts were used to calculate how many unique molecules had been

613   transformed to gauge coverage of the library. A 50 mL culture was spun down and midiprepped.

614   To barcode individual variants, plasmid library harvested from midiprep was digested with

615   EcoRI-HF and NdeI at 37°C for 1 hour, 65°C for 20 minutes. Barcode oligos were ordered from

616   IDT, resuspended at 100 uM, and then annealed by combining 1 uL each of primer with 4 uL

617   CutSmart Buffer and 34 uL ddH2O and running at 98°C for 3 minutes followed by ramping down

618   to 25°C at -0.1°C/second. After annealing, 0.8 uL of Klenow polymerase (exonuclease negative,

619   NEB) and 1.35 uL of 1 mM dNTPS was then combined with the 40 uL of product to fill in the

620   barcode oligo (cycling conditions: 25°C for 15:00, 70°C for 20:00, ramp down to 37°C at -0.1°C/s).

621   Digested vector and barcode oligo were then ligated overnight at 16°C.

622   The overnight ligation was then cleaned and concentrated and eluted in 6 uL of ddH2O. 1

623   uL of this ligation was then transformed into high efficiency *E. coli* using electroporation at 2 kV.

624   Each reaction contained 1 uL of ligation (or ligation control or pUC19 10 pg/uL) and 25 uL of *E.*

625   *coli*. 975 uL of pre-warmed SOC media was added to each cuvette after electroporation,

626   transferred to a culture tube, and recovered at 37°C, shaking for 1 hour. At 1 hour, 1 and 10 uL

627   samples from water and pUC19 cultures were taken and plated on LB supplemented with

628   ampicillin. For ligation and ligation control, four flasks were prepared with 50 mLs of LB and

629   ampicillin, and then 500 uL, 250 uL, 125 uL, 62.5 uL was sample from the 1 mL of recovery and

630   transferred into a corresponding flask. From those flasks, 1 uL, 10 uL, and 100 uL, were sampled

631   and plated onto LB ampicillin plates. Plates and 50 mL culture were incubated at 37°C overnight.

632   Colonies on plates were then counted, and counts were used to calculate how many unique

633   molecules were transformed to gauge number of barcodes. Flask with the target number of

634   barcodes was then spun down and midiprepped.

635

636   **Cell line description**

637   *VAMP-seq assay cell line*

638   HEK293T cells with a serine integrase landing pad integrated at the AAVS1 locus were

639   used (Matreyek et al., 2017).

640

641   *Activity assay cell line*

642   We used a previously published reporter cell line(Haque et al., 2014) and inserted a

643   recombinase-based landing pad at the *AAVS1* safe harbor locus using a previously published

644   strategy (Matreyek et al., 2017). Single cell clones were transfected with TALENs for AAVS1 and

645   the landing pad plasmid, and single cell clones were sorted. Presence of one landing pad was

646   confirmed by 1) barcode sequencing of the landing pad and 2) co-transfection experiment with

647   GFP and mCherry. From this, we moved forward with one clone demonstrated to have only one

648   landing pad present (clone 45).

649   gRNAs to delete portions of the first exon of both *VKORC1* and *VKORC1L1* were ordered

650   and cloned into pSpCas9(BB)-2A-GFP (PX458), which was a gift from Feng Zhang (Addgene

651   plasmid #48138 ; http://n2t.net/addgene:48138 ; RRID:Addgene_48138). Clone 45 was then

652   transfected with these four plasmids, and single cells were sorted based on GFP positivity.

653   Disruption of *VKORC1* and *VKORC1L1* was confirmed by performing nested PCR, TA cloning,

654   and then sequencing of products. We detected three alleles for both *VKORC1* and *VKORC1L1*,

655   indicating that these loci are triploid in HEK293 cells.

656   A Western blot was also used to confirm absence of VKOR protein product in our activity

657   reporter cell line. Protein lysates were harvested from ~1 million cells using 100 uL NP40 lysis

658   buffer with freshly prepared protease inhibitor cocktail and 1 mM PMSF. Protein lysates were

659   Qubited for concentration, and 20 ug of each protein lysate was loaded. 4-12% BisTris NuPage

660   gel (Thermo Fisher) was used with MES buffer + 500µl of antioxidant added to the inner chamber.

661   The gel ran at 150V for 90 min. Gel was then transferred to Nitrocellulose using 1X transfer buffer

662   20% EtOH at 24V for 1 hour on ice. The blot was washed for 5 minutes with 1X TBS-T 0.1%

663   Tween 3 times. Blot was then blocked for overnight 1X TBS-T 0.1% Tween + 5% Milk. Blot was

664   then washed for 5 minutes with 1X TBS-T 0.1% Tween 3 times. Blot was then cut in half at the

665   between the 25kDa and 35kDa molecular weight markers. The bottom blot was incubated with:

666   αVKOR 1:1000 + 1X TBS-T 0.1% Tween + 5% Milk. The top blot was incubated with αbeta-actin

667   dHRP 1:1000+ 1X TBS-T 0.1% Tween + 5% Milk. Both blots were incubated with their primary

668   antibodies overnight at 4ºC. The αVKOR blot was washed for 5 minutes with 1X TBS-T 0.1%

669   Tween 3 times. The αVKOR blot was then incubated with 1:10,000 secondary anti-mouse-HRP

670   (GE Healthcare NA931V) + 1X TBS-T 0.1% Tween + 5% Milk for one hour. The αbeta-actin dHRP

671   blot remained in primary antibody during this time, as no secondary antibody is needed for a direct

672   HRP conjugate. Both blots were then washed for 5 minutes with 1X TBS-T 0.1% Tween 3 times.

673   Blots were then incubated with Supersignal West Dura Extended Duration Substrate (Thermo

674   Fisher). 500μl of both substrates incubated on blot for 5 min. Blots were then dried by kimwipe

675   and exposed using the colorimetric and chemiluminescence functions on the BioRad ChemiDoc

676   MP (Biorad).

677

678   *Recombination of variants in cell lines: Abundance assay*

679   Cells were transfected in six well plates, 250,000 cells per well (12-24 wells transfected

680   total for each experiment). Sequential transfections were performed. On day 1, 3 ug of pCAG-

681   NLS-Bxb1 was diluted in 250 uL of OptiMEM and 6 uL of Fugene (Promega). On day 2, 3 ug of

682   barcoded library was diluted in 250 uL of OptiMEM and 6 uL of Fugene6 and transfected. 48

683   hours after this second transfection, cells were induced with doxycycline at a final concentration

684   of 2.5 ug/mL.

685

686   *Recombination of variants in cell lines: Activity assay*

687   Cells were transfected in six- well plates, 500,000 cells per well (18-24 wells transfected

688   total for each experiment). 272 ng of pCAG-NLS-Bxb1 was diluted in 125 uL of OptiMEM with 2.7

689   ug of barcoded library. 2.25 uL of Lipofectamine 3000 (Thermo Fisher) was diluted in 125 uL of

690   OptiMEM in a separate tube. The DNA mixture was then added to the Lipofectamine 3000 mixture

691   and incubated at room temperature for 15 minutes. Transfection mixture was then added

692   dropwise to one six-well plate. Cells were induced with doxycycline 48 hours after transfection,

693   with a final concentration of 2.5 ug/mL doxycycline .

694

695   *Enrichment sorting for recombined cells*

696   Cells were washed once with PBS, then dissociated with versene. Media was added to

697   dilute EDTA, and cells transferred to 15 mL conical and spun down at 300 x g for 4 minutes. Media

698   was aspirated off, and cells were resuspended in PBS, then filtered through a 35 um nylon mesh

699   filter. Cells were sorted on a BD Aria III FACS machine. mTagBFP2, expressed from the

700   unrecombined landing pad, was excited with a 405 nM laser. Recombined cells either expressed

701   mCherry (abundance) or eGFP (activity), and these were excited by a 561 nm laser and a 488

702   nm laser, respectively. Samples were gated for live cells using FSC-A and SSC-A, then singlets

703   using SSC-H vs. SSC-W, FSC-H vs. FSC-W. For activity assay reporter cell line, cells were then

704   sorted for DsRed positivity to ensure robust expression of reporter. Cells that had successfully

705   recombined a single VKOR variant were gated on recombinant mTagBFP2 negativity and either

706   mCherry positivity (abundance) or eGFP positivity (activity) (see Supplementary Fig. 2b for gating

707   example). Recombined cells were sorted on "Yield" mode in the BD Diva software and grown out

708   for 3-5 days.

709

710   *Abundance assay quartile sorting*

711   Recombined cells were run on a BD Aria III FACS machine. Cells were prepared for

712   sorting as described above, and were then gated for live, recombined singlets. A ratio of

713   eGFP/mCherry was created using the BD Diva software as a unique parameter, and the

714   histogram of this ratio was divided into four equal bins. Each quartile was sorted into a 5 mL tube

715   on "4-Way Purity" mode. Sorted cells were grown out for 2-4 days post sorting to ensure enough

716   DNA for sequencing. The details of replicate sorts for activity assay are in Supplementary Table

717   1.

718

719   *Antibody conjugation*

720   Factor IX Gla domain antibody specific for carboxylation was conjugated to APC following

721   LYNX Rapid APC Antibody Conjugation Kit instructions. Antibody was resuspended at 1 mg/mL

722   in nuclease-free water. 1 uL of Modifier reagent was then added for every 10 uL of antibody and

723   mixed by pipetting. That mixture was then pipetted directly onto the LYNX lyophilized mix and

724   gently mixed by pipetting up and down twice. The conjugation mixture was then capped and

725   incubated in the dark at room temperature overnight. After overnight incubation, 1 uL of Quencher

726   reagent was added for every 10 uL of antibody used and left to incubate for 30 minutes. At that

727   point, antibody was divided into 20 uL aliquots to be used for replicate experiments and stored at

728   -20°C.

729

730   *Activity assay antibody staining and quartile sorting*

731   Cells were plated in six-well plates at 500,000 cells per well with D10 media with no

732   doxycycline. All replicates were performed with 18-24 wells of cells total. After 24 hours,

733   doxycycline was added to cells to induce expression of reporter and VKOR variant. Cells were

734   then incubated with doxycycline for 48 hours. On day of cell sorting, each six well was washed

735   with cold PBS, dissociated with 200 uL of versene, and then resuspended in 1 mL of phenol red-

736   free DMEM + 1% FBS and transferred to a 5 mL FACS tube. Cells were spun at 300 x g, then

737   washed once with 1 mL of phenol red-free DMEM + 1% FBS. Cells were spun at 300 x g, and

738   after aspirating supernatant, cell pellet was resuspended in 100 uL of antibody diluted 1:100 in

739   phenol red-free DMEM + 1% FBS. Cells were incubated in antibody for 20 minutes at 4°C in the

740   dark, with vortexing at five minute intervals to ensure staining. After 20 minutes, 1 mL of staining

741 buffer was added to each tube to dilute out antibody. Cells were spun at 300 x g, washed twice

742 more similarly with staining buffer, then resuspended in 200 uL. At this point, all tubes were pooled

743 and filtered to remove clumps. Cells were then sorted using a FACSAria III (BD Biosciences) into

744 bins based on their APC intensity. First, live, single, recombinant cells were selected as described

745 above. A histogram of APC was created and gates dividing the library into four equally populated

746 bins based on APC fluorescence intensity were drawn. The details of replicate sorts for activity

747 assay are in Supplementary Table 2.

748

749 **gDNA prep, barcode amplification, and sequencing**

750   Cells were then collected, pelleted by centrifugation and stored at –20 °C. Genomic DNA

751 was prepared using a DNEasy kit, according to the manufacturer's instructions (Qiagen), with the

752 addition of a 30 min incubation at 37 °C with RNAse in the re-suspension step. Eight 50 µl first-

753 round PCR reactions were each prepared with a final concentration of ~50 ng µl–1 input genomic

754 DNA, 1×Q5 High-Fidelity Master Mix and 0.25 µM of the KAM499/VKORampR 1.1 primers. The

755 reaction conditions were 98 °C for 30 s, 98°C for 10 s, 65°C for 20 s, 72°C for 60 s, repeat 5 times,

756 72°C for 2 min, 4°C hold. Eight 50 µl reactions were combined, bound to AMPure XP (Beckman

757 Coulter), cleaned and eluted with 21 µl water. Forty percent of the eluted volume was mixed with

758 Q5 High-Fidelity Master Mix; VKOR_indexF_1.1 and one of the indexed reverse primers,

759 PTEN_seq_R1a through PTEN_seq_R2a, were added at 0.25 µM each. These reactions were

760 run with Sybr Green I on a BioRad MiniOpticon; reactions were denatured for 3 minutes at 95°C

761 and cycled 20 times at 95°C for 15s, 60°C for 15s, 72°C for 15s with a final 3 min extension at

762 72°C. The indexed amplicons were mixed based in relative fluorescence units and run on a 1%

763 agarose gel with Sybr Safe and gel extracted using a freeze and squeeze column (Bio-Rad). The

764 product was quantified using Kapa Illumina Quant kit.

765

Subassembly

766

767   Barcoded VKOR library was subassembled using a MiSeq 600 kit (Illumina). Two

768   amplicons were generated, one forward, one reverse. PCR reactions were each prepared with

769   ~500 ng input plasmid DNA, 1 × KAPA High-Fidelity Master Mix and 0.25 μM of the

770   VKOR_SA_amp_F/VKOR_SA_amp_R or VKOR_SA_for_amp_R2.0/VKOR_SA_rev_amp_F2.0

771   primers. PCR reactions were run at 95°C for five minutes, then cycled 15 times at 98°C for 0:20,

772   60°C for 0:15, 72°C for 0:30, with a final extension at 72°C for 2:00. Amplicons (741 bp) were gel

773   extracted on a 1.0% gel run at 130V for 35 mins. The product was quantified using Qubit and

774   Kapa Illumina quant kit. Read lengths were as follows: 289 bp forward read, 18 bp index1, 18 bp

775   index 2 (index = barcode forward and reverse). All reads sharing a common barcode sequence

776   were collapsed to form the consensus variant sequence, resulting in 175,052 barcodes after

777   filtering.

778

Barcode counting and variant calling

779

780   Enrich2 was used to quantify barcodes from bin sequencing, using a minimum quality filter

781   of 20 (Rubin et al., 2017). FASTQ files containing barcodes and the barcode map for VKOR were

782   used as input for Enrich2. Enrich2 configuration files for each experiment are available on the

783   GitHub repository. Barcodes assigned to variants containing insertion, deletions, or multiple

784   amino-acid alterations were removed from the analysis.

785

Calculating scores and classifications

786

787   Scores and classifications were assigned using previously published analysis pipeline

788   (Matreyek et al., 2018). Briefly, for each protein variant, frequencies in each bin were calculated

789   by dividing counts by total counts. From there, we filtered variants based on the number of

790   experiments in which it was observed (F_{expt} = 2) and their frequency (F_{freq} = 10^{-4}), after noticing

791   that low frequency variants introduced noise to the analysis. These frequencies were then each

56

792 weighted by multiplying by 0.25, 0.5, 0.75, and 1 in a bin-wise fashion. We generated a replicate

793 score for each variant by using min-max normalization: normalizing to the median weighted

794 average of the nonsense distribution set at 0 and the median weighted average of the

795 synonymous distribution set at 1. We then averaged those scores for a final, experiment-wide

796 variant score. Standard deviation and standard error were also calculated for each variant, and

797 95% confidence intervals were estimated using standard error, assuming a normal distribution.

798 Abundance and activity classifications were assigned by assessing variant score and confidence

799 intervals in relation to synonymous variant distribution. To do this, we established a cut-off that

800 separated the 5% of synonymous variants with the lowest abundance (or activity) scores from the

801 95% of synonymous variants with higher abundance (or activity) scores. Variants with both scores

802 and upper confidence intervals below this threshold were classified as "low," while those with

803 scores below but upper confidence above were classified as "possibly low." Variants with scores

804 and lower confidence intervals above the threshold were classified as "WT-like", while those with

805 scores above lower confidence intervals below the threshold were classified as "possibly WT-

806 like." Finally, another threshold was set that separated the 5% of synonymous variants with the

807 highest scores from the rest of the synonymous distribution. Variants that had scores above this

808 threshold, with lower confidence intervals above the lower threshold were classified as "high."

809

810 **Windowed abundance and activity analysis**

811 Windowed averages of abundance and activity scores were calculated using a window

812 length of 10 positions with center alignment. Scores were calculated for both charged amino acids

813 (R, K, H, D, E,) and aliphatic amino acids (G, A, V, I, L).

814

815 **Evolutionary couplings analysis**

816 EVcouplings extracts the constraints between pairs of residues, as evidenced in

817 alignments of homologous sequences: first homologous sequences must be collected and

57

818   aligned, and then a model of statistical energy costs and benefits between residues is fit to explain

819   the sequence variation in the alignment. We collected an alignment of 2770 sequences using

820   jackhammer (http://hmmer.org/) to query the human VKOR sequence against UniRef100

821   (https://www.uniprot.org/uniref/), with a bitscore per residue cutoffs of 0.4 and 7 search iterations.

822   We predicted secondary structures where the summed strength of couplings at would-be alpha

823   helix and beta strand contacts scored above 1.5 for alpha helices and 0.75 respectively, for two

824   or more consecutive residues. We extended the called helices and strands by one residue on

825   each side for a minimum structure size of four residues. All methods used for building alignments,

826   training the model, folding, and predicting secondary structure are part of the EVcouplings

827   software (https://evcouplings.org/) (Hopf et al., 2019).

828

829   **Homology modeling**

830   A homology model of human VKOR was made by accessing I-TASSER (Yang et al., 2015)

831   and using PDB structure 4NV5 as a template for threading. Model1 from results was used for all

832   figures in this paper.

833

834   **Hierarchical clustering**

835   Hierarchical clustering was performed on abundance score vectors for each position using

836   the hclust function in R. Dendrogram for hierarchically clustered heatmap was drawn using

837   dendextend package (version 1.12.0).

838

839   **Active site residue analysis**

840   Activity and abundance scores were rescaled so that the lowest score present in the

841   dataset was set at 0, and the highest score at 1. A ratio of rescaled activity to rescaled abundance

842   (specific activity) was then calculated for every variant. Using variant specific activity scores,

843   median specific activity was calculated for each position. Threshold for classification as an active

844   site position was drawn based on scores of known redox cysteines at positions 132 and 135,

845   resulting in lowest 12.5% of median specific activity scores being classified as active site residues.

846   We additionally required that any position within this group had been scored for at least four

847   variants to eliminate noise from poor sampling.

848

849   **Data availability**

850   All raw sequence data and function scores are freely available for all academic users by

851   non-exclusive license under reasonable terms to commercial entities that have committed to open

852   sharing of VKOR sequence variants and under a free non-exclusive license to non-profit entities.

853   The Illumina raw sequencing files and barcode–variant maps can be accessed at the NCBI Gene

854   Expression Omnibus (GEO) repository under accession number GSE149922. The data

855   presented in the manuscript are available as Supplementary Data files.

856

857   **Code availability**

858   Code for analysis is available at http://github.com/FowlerLab/VKOR.

859   The code used to train the evolutionary couplings model is available at the EVcouplings

860   GitHub repository (https://github.com/debbiemarkslab/EVcouplings). The data used to train the

861   model is publically available at uniprot (https://uniprot.org).

862

863   **ACKNOWLEDGEMENTS**

870   Foundation Graduate Research Fellowship. D.M. and N.J.R. thank the Chan Zuckerberg Initiative

871   DAF (grant number 2018-191853) for financial support.

872

873   **AUTHOR CONTRIBUTIONS**

874   M.A.C. carried out abundance and activity experiments and analyzed the data. N.R. and

875   D.M. performed evolutionary couplings analysis. J.S. prepared samples for next-generation

876   sequencing. K.S. cloned variants for abundance assay validation. K.A.M. provided abundance

877   analysis framework. D.D. provided human *VKORC1* variants from Color Genomics. A.R. provided

878   the parental activity reporter cell line. M.V., S.S. and F.P.R. provided a mutagenized VKOR library.

879   M.A.C. and D.M.F. wrote the manuscript with input from co-authors.

880

881   **COMPETING INTERESTS**

882   The authors declare that the variant functional data presented herein are copyrighted, and

883   may be freely used for non-commercial purposes. Licensing for commercial use may benefit the

884   authors. The authors declare no additional competing interests.

REFERENCES

885

886 Bodin L, Perdu J, Diry M, Horellou M-H, Loriot M-A. 2008. Multiple genetic alterations in vitamin

887   K epoxide reductase complex subunit 1 gene (VKORC1) can explain the high dose

888   requirement during oral anticoagulation in humans. *J Thromb Haemost* **6**:1436–1439.

889   doi:10.1111/j.1538-7836.2008.03049.x

890 Czogalla KJ, Biswas A, Höning K, Hornung V, Liphardt K, Watzka M, Oldenburg J. 2016.

891   Warfarin and vitamin K compete for binding to Phe55 in human VKOR. *Nat Struct Mol Biol*.

892   doi:10.1038/nsmb.3338

893 Czogalla KJ, Biswas A, Rost S, Watzka M, Oldenburg J. 2014. The Arg98Trp mutation in

894   human VKORC1 causing VKCFD2 disrupts a di-Arginine-based ER retention motif. *Blood*.

895   doi:10.1182/blood-2013-12-545988

896 Elazar AA, Weinstein J, Biran I, Fridman Y, Bibi E, Fleishman SJ. 2016. Mutational scanning

897   reveals the determinants of protein insertion and association energetics in the plasma

898   membrane. *eLife Sciences* e12125. doi:10.7554/eLife.12125

899 Elazar A, Weinstein JJ, Prilusky J, Fleishman SJ. 2016. Interplay between hydrophobicity and

900   the positive-inside rule in determining membrane-protein topology. *Proc Natl Acad Sci U S*

901   *A* **113**:10340–10345. doi:10.1073/pnas.1605888113

902 Fleming KG, Engelman DM. 2001. Specificity in transmembrane helix-helix interactions can

903   define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A* **98**:14340–

904   14344. doi:10.1073/pnas.251367498

905 Gasperini M, Starita L, Shendure J. 2016. The power of multiplexed functional analysis of

906   genetic variants. *Nat Protoc* **11**:1782–1787. doi:10.1038/nprot.2016.135

907 Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA 3rd, Smith HO. 2009. Enzymatic

908   assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343–345.

909   doi:10.1038/nmeth.1318

910   Gong IY, Schwarz UI, Crown N, Dresser GK, Lazo-Langner A, Zou G, Roden DM, Stein CM,

911   Rodger M, Wells PS, Kim RB, Tirona RG. 2011. Clinical and genetic determinants of

912   warfarin pharmacokinetics and pharmacodynamics during treatment initiation. *PLoS One*

913   **6**:e27808. doi:10.1371/journal.pone.0027808

914   Gray VE, Hause RJ, Fowler DM. 2017. Analysis of Large-Scale Mutagenesis Data To Assess

915   the Impact of Single Amino Acid Substitutions. *Genetics* **207**:53–61.

916   doi:10.1534/genetics.117.300064

917   Guerriero CJ, Reutter K-R, Augustine AA, Preston GM, Weiberth KF, Mackie TD, Cleveland-

918   Rubeor HC, Bethel NP, Callenberg KM, Nakatsukasa K, Grabe M, Brodsky JL. 2017.

919   Transmembrane helix hydrophobicity is an energetic barrier during the retrotranslocation of

920   integral membrane ERAD substrates. *Mol Biol Cell* **28**:2076–2090. doi:10.1091/mbc.E17-

921   03-0184

922   Hallgren KW, Qian W, Yakubenko AV, Runge KW, Berkner KL. 2006. r-VKORC1 Expression in

923   Factor IX BHK Cells Increases the Extent of Factor IX Carboxylation but Is Limited by

924   Saturation of Another Carboxylation Component or by a Shift in the Rate-Limiting Step†.

925   *Biochemistry* **45**:5587–5598. doi:10.1021/bi051986y

926   Haque JA, McDonald MG, Kulman JD, Rettie AE. 2014. A cellular system for quantitation of

927   vitamin K cycle activity: structure-activity effects on vitamin K antagonism by warfarin

928   metabolites. *Blood* **123**:582–589. doi:10.1182/blood-2013-05-505123

929   Harrington DJ, Gorska R, Wheeler R, Davidson S, Murden S, Morse C, Shearer MJ, Mumford

930   AD. 2008. Pharmacodynamic resistance to warfarin is associated with nucleotide

931   substitutions in VKORC1. *J Thromb Haemost* **6**:1663–1670. doi:10.1111/j.1538-

932   7836.2008.03116.x

933   Harrington DJ, Siddiq S, Allford SL, Shearer MJ, Mumford AD. 2011. More on: endoplasmic

934   reticulum loop VKORC1 substitutions cause warfarin resistance but do not diminish

935   gamma-carboxylation of the vitamin K-dependent coagulation factors. *J Thromb Haemost*

936   **9**:1093–1095. doi:10.1111/j.1538-7836.2011.04249.x

937   Hatahet F, Blazyk JL, Martineau E, Mandela E, Zhao Y, Campbell RE, Beckwith J, Boyd D.

938   2015. Altered Escherichia coli membrane protein assembly machinery allows proper

939   membrane assembly of eukaryotic protein vitamin K epoxide reductase. *Proc Natl Acad Sci*

940   *U S A* **112**:15184–15189. doi:10.1073/pnas.1521260112

941   Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH,

942   von Heijne G. 2007. Molecular code for transmembrane-helix recognition by the Sec61

943   translocon. *Nature* **450**:1026–1030. doi:10.1038/nature06387

944   Hessa T, Sharma A, Mariappan M, Eshleman HD, Gutierrez E, Hegde RS. 2011. Protein

945   targeting and degradation are coupled for elimination of mislocalized proteins. *Nature*

946   **475**:394–397. doi:10.1038/nature10181

947   Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. 2012. Three-Dimensional

948   Structures of Membrane Proteins from Genomic Sequencing. *Cell* **149**:1607–1621.

949   doi:10.1016/j.cell.2012.04.012

950   Hopf TA, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, Toth-Petroczy A,

951   Brock K, Riesselman AJ, Palmedo P, Kang C, Sheridan R, Draizen EJ, Dallago C, Sander

952   C, Marks DS. 2019. The EVcouplings Python framework for coevolutionary sequence

953   analysis. *Bioinformatics* **35**:1582–1584. doi:10.1093/bioinformatics/bty862

954   Jain PC, Varadarajan R. 2014. A rapid, efficient, and economical inverse polymerase chain

955   reaction-based method for generating a site saturation mutant library. *Anal Biochem*

956   **449**:90–98. doi:10.1016/j.ab.2013.12.002

957   Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of

958   hydrogen-bonded and geometrical features. *Biopolymers* **22**:2577–2637.

959   doi:10.1002/bip.360221211

960   Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia

961   KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D,

962   Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y,

963   Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman

964   E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C,

965   Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S,

966   Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S,

967   Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K,

968   Tolonen C, Wade G, Talkowski ME, The Genome Aggregation Database Consortium,

969   Neale BM, Daly MJ, MacArthur DG. 2019. Variation across 141,456 human exomes and

970   genomes reveals the spectrum of loss-of-function intolerance across human protein-coding

971   genes. *bioRxiv*. doi:10.1101/531210

972 Kurnik D, Qasim H, Sominsky S, Lubetsky A, Markovits N, Li C, Stein CM, Halkin H, Gak E,

973   Loebstein R. 2012. Effect of the VKORC1 D36Y variant on warfarin dose requirement and

974   pharmacogenetic dose prediction. *Thromb Haemost* **108**:781–788. doi:10.1160/TH12-03-

975   0151

976 Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014.

977   ClinVar: public archive of relationships among sequence variation and human phenotype.

978   *Nucleic Acids Res* **42**:D980–5. doi:10.1093/nar/gkt1113

979 Li T, Chang C-Y, Jin D-Y, Lin P-J, Khvorova A, Stafford DW. 2004. Identification of the gene for

980   vitamin K epoxide reductase. *Nature* **427**:541–544. doi:10.1038/nature02254

981 Liu S, Cheng W, Fowle Grider R, Shen G, Li W. 2014. Structures of an intramembrane vitamin

982   K epoxide reductase homolog reveal control mechanisms for electron transfer. *Nat*

983   *Commun* **5**. doi:10.1038/ncomms4110

984 Li W, Schulman S, Dutton RJ, Boyd D, Beckwith J, Rapoport TA. 2010. Structure of a bacterial

985   homologue of vitamin K epoxide reductase. *Nature* **463**:507–512. doi:10.1038/nature08720

986 Loebstein R, Dvoskin I, Halkin H, Vecsler M, Lubetsky A, Rechavi G, Amariglio N, Cohen Y,

987   Ken-Dror G, Almog S, Gak E. 2007. A coding VKORC1 Asp36Tyr polymorphism

988   predisposes to warfarin resistance. *Blood* **109**:2477–2480. doi:10.1182/blood-2006-08-

989   038984

990   Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. 2011. Protein

991   3D structure computed from evolutionary sequence variation. *PLoS One* **6**:e28766.

992   doi:10.1371/journal.pone.0028766

993   Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, Kircher M, Khechaduri

994   A, Dines JN, Hause RJ, Bhatia S, Evans WE, Relling MV, Yang W, Shendure J, Fowler

995   DM. 2018. Multiplex assessment of protein variant abundance by massively parallel

996   sequencing. *Nat Genet* **50**:874–882. doi:10.1038/s41588-018-0122-z

997   Matreyek KA, Stephany JJ, Fowler DM. 2017. A platform for functional assessment of large

998   variant libraries in mammalian cells. *Nucleic Acids Res*. doi:10.1093/nar/gkx183

999   Ma W, Goldberg J. 2013. Rules for the recognition of dilysine retrieval motifs by coatomer.

1000   *EMBO J* **32**:926–937. doi:10.1038/emboj.2013.41

1001   Merkle PS, Gotfryd K, Cuendet MA, Leth-Espensen KZ, Gether U, Loland CJ, Rand KD. 2018.

1002   Substrate-modulated unwinding of transmembrane helices in the NSS transporter LeuT. *Sci*

1003   *Adv* **4**:eaar6179. doi:10.1126/sciadv.aar6179

1004   Mravic M, Thomaston JL, Tucker M, Solomon PE, Liu L, DeGrado WF. 2019. Packing of apolar

1005   side chains enables accurate design of highly stable membrane proteins. *Science*

1006   **363**:1418–1423. doi:10.1126/science.aav7541

1007   Nilsson I, von Heijne G. 1990. Fine-tuning the topology of a polytopic membrane protein: role of

1008   positively and negatively charged amino acids. *Cell* **62**:1135–1141. doi:10.1016/0092-

1009   8674(90)90390-z

1010   Oldenburg J, Rost S, Fregin A, Geisen C, Ivaskevicius V, Seifried E, Scharrer I, Heistinger M,

1011   Tuddenham E, Muller-Reible C, Zieger B. 2004. Mutations in the VKORC1 Gene Cause

1012   Warfarin Resistance, Warfarin Sensitivity and Combined Deficiency of Vitamin K

1013   Dependent Coagulation Factors. *Blood* **104**:277–277.

1014   Osinbowale O, Al Malki M, Schade A, Bartholomew JR. 2009. An algorithm for managing

1015   warfarin resistance. *Cleve Clin J Med* **76**:724–730. doi:10.3949/ccjm.76a.09062

1016   Osman A, Enström C, Arbring K, Söderkvist P, Lindahl TL. 2006. Main haplotypes and

1017   mutational analysis of vitamin K epoxide reductase (VKORC1) in a Swedish population: a

1018   retrospective analysis of case records. *J Thromb Haemost* **4**:1723–1729.

1019   doi:10.1111/j.1538-7836.2006.02039.x

1020   Owen RP, Gong L, Sagreiya H, Klein TE, Altman RB. 2010. VKORC1 Pharmacogenomics

1021   Summary. *Pharmacogenet Genomics* **20**:642–644. doi:10.1097/FPC.0b013e32833433b6

1022   Peoc'h K, Pruvot S, Gourmel C, dit Sollier CB, Drouet L. 2009. A new VKORC1 mutation

1023   leading to an isolated resistance to fluindione. *Br J Haematol* **145**:841–843.

1024   doi:10.1111/j.1365-2141.2009.07687.x

1025   Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, McLeod HL, Blough DK, Thummel KE,

1026   Veenstra DL, Rettie AE. 2005. Effect of VKORC1 Haplotypes on Transcriptional Regulation

1027   and Warfarin Dose. *N Engl J Med* **352**:2285–2293. doi:10.1056/NEJMoa044503

1028   Rishavy MA, Usubalieva A, Hallgren KW, Berkner KL. 2011. Novel Insight into the Mechanism

1029   of the Vitamin K Oxidoreductase (VKOR) ELECTRON RELAY THROUGH Cys43 AND

1030   Cys51 REDUCES VKOR TO ALLOW VITAMIN K REDUCTION AND FACILITATION OF

1031   VITAMIN K-DEPENDENT PROTEIN CARBOXYLATION. *J Biol Chem* **286**:7267–7278.

1032   doi:10.1074/jbc.M110.172213

1033   Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K, Pelz H-J, Lappegard K, Seifried

1034   E, Scharrer I, Tuddenham EGD, Müller CR, Strom TM, Oldenburg J. 2004. Mutations in

1035   VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2.

1036   *Nature* **427**:537–541. doi:10.1038/nature02214

1037   Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM. 2017. A

1038   statistical framework for analyzing deep mutational scanning data. *Genome Biol* **18**:150.

1039   doi:10.1186/s13059-017-1272-5

1040   Schmeits PCJ, Hermans MHA, van Geest-Daalderop JHH, Poodt J, de Sauvage Nolting PRW,

1041   Conemans JMH. 2010. VKORC1 mutations in patients with partial resistance to

1042   phenprocoumon. *Br J Haematol* **148**:955–957. doi:10.1111/j.1365-2141.2009.08017.x

1043   Schulman S, Wang B, Li W, Rapoport TA. 2010. Vitamin K epoxide reductase prefers ER

1044   membrane-anchored thioredoxin-like redox partners. *Proc Natl Acad Sci U S A* **107**:15027–

1045   15032. doi:10.1073/pnas.1009972107

1046   Sharpe HJ, Stevens TJ, Munro S. 2010. A comprehensive comparison of transmembrane

1047   domains reveals organelle-specific properties. *Cell* **142**:158–169.

1048   doi:10.1016/j.cell.2010.05.037

1049   Shen G, Cui W, Zhang H, Zhou F, Huang W, Liu Q, Yang Y, Li S, Bowman GR, Sadler JE,

1050   Gross ML, Li W. 2017. Warfarin traps human vitamin K epoxide reductase in an

1051   intermediate state during electron transfer. *Nat Struct Mol Biol* **24**:69.

1052   doi:10.1038/nsmb.3333

1053   Tie J-K, Jin D-Y, Stafford DW. 2012. Human Vitamin K Epoxide Reductase and Its Bacterial

1054   Homologue Have Different Membrane Topologies and Reaction Mechanisms. *J Biol Chem*

1055   **287**:33945–33955. doi:10.1074/jbc.M112.402941

1056   Tie J-K, Jin D-Y, Tie K, Stafford DW. 2013. Evaluation of warfarin resistance using transcription

1057   activator-like effector nucleases-mediated vitamin K epoxide reductase knockout HEK293

1058   cells. *J Thromb Haemost* **11**:1556–1564. doi:10.1111/jth.12306

1059   Toth-Petroczy A, Palmedo P, Ingraham J, Hopf TA, Berger B, Sander C, Marks DS. 2016.

1060   Structured States of Disordered Proteins from Genomic Sequences. *Cell* **167**:158–170.e12.

1061   doi:10.1016/j.cell.2016.09.010

1062   Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, Vriend G. 2015. A series

1063   of PDB-related databanks for everyday needs. *Nucleic Acids Res* **43**:D364–8.

1064   doi:10.1093/nar/gku1028

1065   Ulmschneider MB, Sansom MS. 2001. Amino acid distributions in integral membrane protein

1066   structures. *Biochim Biophys Acta* **1512**:1–14. doi:10.1016/s0005-2736(01)00299-1

1067   von Heijne G. 1989. Control of topology and mode of assembly of a polytopic membrane protein

1068   by positively charged residues. *Nature* **341**:456–458. doi:10.1038/341456a0

1069   Watzka M, Geisen C, Bevans CG, Sittinger K, Spohn G, Rost S, Seifried E, Müller CR,

1070   Oldenburg J. 2011. Thirteen novel VKORC1 mutations associated with oral anticoagulant

1071   resistance: insights into improved patient diagnosis and treatment. *J Thromb Haemost*

1072   **9**:109–118. doi:10.1111/j.1538-7836.2010.04095.x

1073   Wilms EB, Touw DJ, Conemans JMH, Veldkamp R, Hermans M. 2008. A new VKORC1 allelic

1074   variant (p.Trp59Arg) in a patient with partial resistance to acenocoumarol and

1075   phenprocoumon. *J Thromb Haemost* **6**:1224–1226. doi:10.1111/j.1538-7836.2008.02975.x

1076   Wu S, Chen X, Jin D-Y, Stafford DW, Pedersen LG, Tie J-K. 2018. Warfarin and vitamin K

1077   epoxide reductase: a molecular accounting for observed inhibition. *Blood*.

1078   doi:10.1182/blood-2018-01-830901

1079   Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure

1080   and function prediction. *Nat Methods* **12**:7–8. doi:10.1038/nmeth.3213

1081   Yuan H-Y, Chen J-J, Lee MTM, Wung J-C, Chen Y-F, Charng M-J, Lu M-J, Hung C-R, Wei C-Y,

1082   Chen C-H, Wu J-Y, Chen Y-T. 2005. A novel functional VKORC1 promoter polymorphism is

1083   associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Hum Mol*

1084   *Genet* **14**:1745–1751. doi:10.1093/hmg/ddi180

1085   Zimmermann A, Matschiner JT. 1974. Biochemical basis of hereditary resistance to warfarin in

1086   the rat. *Biochem Pharmacol* **23**:1033–1040. doi:10.1016/0006-2952(74)90002-1