Table S2. Composition of PMD testing dataset.

<table>
<thead>
<tr>
<th>Amino acid residues</th>
<th>Pathogenic variants</th>
<th>Wild-type</th>
<th>Neutral variants</th>
<th>All variants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>O/E ratio $^a$</td>
<td>Observed</td>
</tr>
<tr>
<td>Ala</td>
<td>45</td>
<td>168</td>
<td>27%</td>
<td>42</td>
</tr>
<tr>
<td>Arg</td>
<td>198</td>
<td>116</td>
<td>171%</td>
<td>109</td>
</tr>
<tr>
<td>Asn</td>
<td>59</td>
<td>103</td>
<td>57%</td>
<td>72</td>
</tr>
<tr>
<td>Asp</td>
<td>199</td>
<td>116</td>
<td>172%</td>
<td>81</td>
</tr>
<tr>
<td>Cys</td>
<td>137</td>
<td>40</td>
<td>343%</td>
<td>72</td>
</tr>
<tr>
<td>Gln</td>
<td>59</td>
<td>141</td>
<td>42%</td>
<td>40</td>
</tr>
<tr>
<td>Glu</td>
<td>113</td>
<td>92</td>
<td>123%</td>
<td>116</td>
</tr>
<tr>
<td>Gly</td>
<td>99</td>
<td>159</td>
<td>62%</td>
<td>24</td>
</tr>
<tr>
<td>His</td>
<td>131</td>
<td>49</td>
<td>267%</td>
<td>63</td>
</tr>
<tr>
<td>Ile</td>
<td>94</td>
<td>123</td>
<td>76%</td>
<td>38</td>
</tr>
<tr>
<td>Leu</td>
<td>137</td>
<td>203</td>
<td>67%</td>
<td>53</td>
</tr>
<tr>
<td>Lys</td>
<td>203</td>
<td>130</td>
<td>156%</td>
<td>109</td>
</tr>
<tr>
<td>Met</td>
<td>23</td>
<td>63</td>
<td>37%</td>
<td>26</td>
</tr>
<tr>
<td>Phe</td>
<td>86</td>
<td>87</td>
<td>99%</td>
<td>34</td>
</tr>
<tr>
<td>Pro</td>
<td>69</td>
<td>114</td>
<td>61%</td>
<td>38</td>
</tr>
<tr>
<td>Ser</td>
<td>143</td>
<td>165</td>
<td>87%</td>
<td>124</td>
</tr>
<tr>
<td>Thr</td>
<td>96</td>
<td>134</td>
<td>72%</td>
<td>60</td>
</tr>
<tr>
<td>Trp</td>
<td>90</td>
<td>29</td>
<td>310%</td>
<td>23</td>
</tr>
<tr>
<td>Tyr</td>
<td>188</td>
<td>74</td>
<td>254%</td>
<td>77</td>
</tr>
<tr>
<td>Val</td>
<td>80</td>
<td>145</td>
<td>55%</td>
<td>47</td>
</tr>
<tr>
<td>All</td>
<td>2,249</td>
<td>2,249</td>
<td></td>
<td>1,248</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino acid residues</th>
<th>Pathogenic variants</th>
<th>Wild-type</th>
<th>Neutral variants</th>
<th>All variants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>O/E ratio $^a$</td>
<td>Observed</td>
</tr>
<tr>
<td>Ala</td>
<td>719</td>
<td>168</td>
<td>428%</td>
<td>575</td>
</tr>
<tr>
<td>Arg</td>
<td>138</td>
<td>116</td>
<td>119%</td>
<td>78</td>
</tr>
<tr>
<td>Asn</td>
<td>125</td>
<td>103</td>
<td>121%</td>
<td>29</td>
</tr>
<tr>
<td>Asp</td>
<td>66</td>
<td>116</td>
<td>57%</td>
<td>34</td>
</tr>
<tr>
<td>Cys</td>
<td>48</td>
<td>40</td>
<td>120%</td>
<td>21</td>
</tr>
<tr>
<td>Gln</td>
<td>86</td>
<td>141</td>
<td>61%</td>
<td>49</td>
</tr>
<tr>
<td>Glu</td>
<td>109</td>
<td>92</td>
<td>118%</td>
<td>42</td>
</tr>
<tr>
<td>Gly</td>
<td>74</td>
<td>159</td>
<td>47%</td>
<td>42</td>
</tr>
<tr>
<td>His</td>
<td>50</td>
<td>49</td>
<td>102%</td>
<td>11</td>
</tr>
<tr>
<td>Ile</td>
<td>51</td>
<td>123</td>
<td>41%</td>
<td>31</td>
</tr>
<tr>
<td>Leu</td>
<td>103</td>
<td>203</td>
<td>51%</td>
<td>39</td>
</tr>
<tr>
<td>Lys</td>
<td>97</td>
<td>130</td>
<td>75%</td>
<td>45</td>
</tr>
<tr>
<td>Met</td>
<td>31</td>
<td>63</td>
<td>49%</td>
<td>19</td>
</tr>
<tr>
<td>Phe</td>
<td>128</td>
<td>87</td>
<td>147%</td>
<td>61</td>
</tr>
<tr>
<td>Pro</td>
<td>49</td>
<td>114</td>
<td>43%</td>
<td>12</td>
</tr>
<tr>
<td>Ser</td>
<td>164</td>
<td>165</td>
<td>99%</td>
<td>73</td>
</tr>
<tr>
<td>Thr</td>
<td>65</td>
<td>134</td>
<td>49%</td>
<td>26</td>
</tr>
<tr>
<td>Trp</td>
<td>20</td>
<td>29</td>
<td>69%</td>
<td>12</td>
</tr>
<tr>
<td>Tyr</td>
<td>45</td>
<td>74</td>
<td>61%</td>
<td>18</td>
</tr>
<tr>
<td>Val</td>
<td>81</td>
<td>145</td>
<td>56%</td>
<td>31</td>
</tr>
<tr>
<td>All</td>
<td>2,249</td>
<td>2,249</td>
<td></td>
<td>1,248</td>
</tr>
</tbody>
</table>

$^a$ – Expected numbers of amino acid residues were extracted from 105,990 sequences in the non-redundant OWL protein database (release 26.0)

$^b$ – O/E ratio – observed to expected ratio