Supplementary Table S3A: Equations and regression coefficients used in this study to predict disease severity

Predicting Phenotype								
Prediction for	**F-ratio[g]**	**Equation[e]**	**Variable**	***B***	**SE_B**	**β**	**Explained Variability**	**Correctly predicted[f]**
Allele	F(2,32) = 31.403	Predicted APV = 2.732 + 0.078 x %PAH – 0.349 x FoldX	Intercept	2.732	0.764		64.1 %	60.0 %
			%PAH	0.078	0.014	0.660		
			FoldX	-0.349	0.161	-0.254		
	F(2,38) = 22.762	Predicted APV = 2.595 + 0.073 x %PAH + 0.526 x SNP3D	Intercept	2.595	0.785		52.1 %	65.7 %
			%PAH	0.073	0.016	0.585		
			SNPs3D	0.526	0.279	0.241		
Deleterious genotypes[a]								≈ 100 %
Functionally hemizygous Genotypes[b]	F(1,152) = 504.706	Predicted phenotype = 0.877 + 0.219 x APV	Intercept	0.877	0.043		76.7 %	83.1 – 94.8 %
			APV	0.219	0.01	0.877		
	F(1,143) = 57.739	Predicted phenotype = 0.951 + 0.02 x %PAH	Intercept	0.951	0.1		32.5 %	
			%PAH	0.02	0.002	0.574		
	F(2,134) = 10.639	Predicted phenotype = 2.208 + 0.162 x SNP3D – 0.136 x FoldX	Intercept	2.208	0.113		12.4 %	
			SNPs3D	0.162	0.058	0.228		
			FoldX	-0.136	0.043	-0.257		
Compound heterozygous genotype[c]	F(1,93) = 183.498	Predicted phenotype = 1.062 + 0.301 x $\bar{x}_{APV}^{h}$	Intercept	1.062	0.90		66.0 %	77.9 %
			$\bar{x}_{APV}^{h}$	0.301	0.022	0.815		
	F(2,84) = 28.920	Predicted phenotype = 1.518+ 0.025 x $\bar{x}_{\%PAH}^{i}$ + 0.233 x $\bar{x}_{SNP3D}^{k}$	Intercept	1.518	0.207		39.4 %	
			$\bar{x}_{\%PAH}^{i}$	0.025	0.004	0.536		
			$\bar{x}_{SNP3D}^{k}$	0.233	0.081	0.247		
Compound homozygous genotype[d]	F(1,26) = 117.001	Predicted phenotype = 0.998 + 0.252 x APV	Intercept	0.998	0.091		81.1 %	92.9 %
			APV	0.252	0.023	0.905		
	F(1,29) = 27	Predicted phenotype = 1.252 + 0.019 x %PAH	Intercept	1.252	0.274		27.5 %	
			%PAH	0.019	0.007	0.561		

g) F-ratio = mean sum of squares for regression / mean sum of squares for the residuals
h) $\bar{x}_{AVP}$ = average allelic phenotype = {APV (allele 1) + APV (allele 2)} / 2
i) $\bar{x}_{\%PAH}$ = average allelic %PAH = {%PAH (allele 1) + %PAH (allele 2)} / 2
k) $\bar{x}_{SNP3D}$ = average allelic SNPs3D value = {SNP3D (allele 1) + SNPs3D (allele 2)} / 2

Supplementary Table S3B: Equations and regression coefficients used in this study to predict BH4 response

Predicting BH4 Response								
Prediction for	**F-ratio[g]**	**Equation[e]**	**Variable**	***B***	**SE_B**	**β**	**Explained Variability**	**Correctly predicted[f]**
Deleterious genotypes[a]								≈ 100 %
Functionally hemizygous genotypes[b]	F(2,75) = 51.144 F(1,63) = 42.998	Predicted BH_4 response = 0.91 + 0.2 x APV + 0.008 x %PAH	Intercept	0.91	0.14		58.4 %	68.4 %
			APV	0.20	0.03	0.64		
			%PAH	0.008	0.004	0.201		
Compound heterozygous genotypes[c]	F(1,63) = 42.998 F(1,14) = 10.404	Predicted BH_4 response = 1.44 + 0.3 x $\overline{x}_{APV}$ [h]	Intercept	1.44	0.17		39.6 %	50.0 %
			$\overline{x}_{AVP}$ [h]	0.30	0.05	0.60		
Homozygous genotypes[d]	F(1,14) = 10.404 F(1,18) = 8.377	Predicted BH_4 response = 1.79 + 0.021 x %PAH	Intercept	1.79	0.28		38.5 %	81.3 %
			%PAH	0.021	0.006	0.653		
	F(1,18) = 8.377	Predicted BH_4 response = 1.63 + 0.19 x APV	Intercept	1.63	0.27		28.0 %	
			APV	0.19	0.07	0.56		

Multiple regression analysis was run to predict APV from FoldX, PolyPhen-2, SNPs3D, SIFT Blist and %PAH values. For prediction of disease phenotype, we determined the average phenotype ($\overline{x}_{PHENO}$) for each genotype. $\overline{x}_{PHENO}$ was a linear variable defined as (3 x MHP patients + 2 x mild PKU patients + 1 x classic PKU patients) / (all patients within genotype) ranging from 1 (classic PKU) to 3 (MHP) representing the genotype's phenotype tendency. Similarly, average BH_4 response ($\overline{x}_{BH4}$) was determined in all genotypes with at least two patients where response was known. $\overline{x}_{BH4}$ = (3x responders + 2x slow responders + 1x non–responders)/ (all patients within genotype) was a linear variable ranging from 1 (non-responding) to 3 (fast responding) representing the genotype's response tendency. Potentially predictive variables were (a) genotype in deleterious null-null cases, (b) allelic values of the non-null allele in hemizygous patients, (c) averaged allelic values in compound heterozygotes and (d) values of either allele in homozygous genotypes. (e) Equations from multiple regression analysis were applied to the collective. (f) Predicted phenotype and BH_4 response were considered correct if they were less than 0.5 away from de facto $\overline{x}_{PHENO}$ and $\overline{x}_{BH4}$. In the case of functionally hemizygous genotypes the difference between predicted phenotype and $\overline{x}_{PHENO}$ was <0.50 in 128 cases (83.1 %), ≈ 0.50 in 18 cases (11.7 %) and > 0.50 in 8 cases (5.2 %). Therefore the equation correctly predicted $\overline{x}_{PHENO}$ in 83.1 – 94.8 % of all hemizygous genotypes.