In silico approaches for pathogenicity prediction of missense variants of uncertain significance (VUS) in ADPKD

Parnthanutcha Wetchanien[1], Duangkamon Bunditworapoom[2], Manop Pithukpakorn[2], Chanin Limwongse[2,3], Kriengsak Vareesangthip[4], Anunchai Assawamakin[5], Wanna Thongnoppakhun[3*]

[1]Graduate Program in lmmunology, Department of lmmunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
[2]Division of Medical Genetics, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand
[3]Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand
[4]Division of Nephrology, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand
[5]Department of Pharmacology, Faculty of Pharmacy, Mahidol University, Bangkok 10400, Thailand.

**Corresponding author: wanna.tho@mahidol.ac.th*

ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) is the most common dominantly inherited kidney disease, being caused by mutations in *PKD1* (85%) *and PKD2* (15%). Approximately 25% of all *PKD1* mutations are pathogenic missense, but additional 12.8% of those are reported as indeterminate or missense variants of uncertain significance (VUS), of which the deleterious nature are unclear in clinical practice. Functional studies to assess the impact of missense variants, particularly for *PKD1*, are difficult due to the large size and ambiguous functions of the proteins, polycystins. A variety of *in silico* tools were developed to evaluate interspecies variations and biochemical impacts of amino acid substitutions of missense VUS. To evaluate the tool suitability for *PKD1 and PKD2*, we applied additional 12 state-of-the-art web-based tools with their claimed superior performances based on different algorithms for the prediction in parallel with the three benchmark programs (PolyPhen2, SIFT and MutationTaster). A total of 15 tools were assessed in a gene-specific manner with *PKD1 and PKD2* variants of known pathogenicity from the PKDB mutation database. We found that each of the genes had suitable predictions from different sets of tools. Combined uses of 3, 5, 8 and 12 tools with high performance in descending order (in which the benchmarks were excluded) gave consensus predictions for the selected nine VUS in *PKD1*, except one VUS which might have a mild pathogenicity that only the use of 12 tools predicted differently. In this situation, using more numbers of tools might prevent the misinterpretation of milder VUS. Classification of the pathogenicity of the VUS in *PKD1 and PKD2* became an essential part of molecular diagnosis of ADPKD and is useful for a clinical decision and would expand knowledge of mutation spectrum in ADPKD.

Keywords: ADPKD; variant of uncertain significance (VUS); *in silico*; pathogenicity; missense

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is the dominantly inherited and life-threatening kidney disease, being one of the most common monogenic disorders affecting 1:400-1,000 worldwide**.** The majority (85%) of ADPKD cases is caused by mutations in *PKD1* located on chromosome 16p13.3 and the rest 15% by *PKD2* at 4q21 (Tan *et al*., 2011). According to the Autosomal Dominant Polycystic Disease: Mutation Database (PKDB; http://pkdb.mayo.edu/; accessed March 1, 2015), mutation spectrum of *PKD1* roughly includes 1,272 pathogenic mutations and 857 polymorphisms. A total of 25.2% (320/1,272) are missense mutations and additional 12.8% (163/1,272) being missense indeterminate variants which cannot be definitely classified, so called "missense variant of uncertain significance (VUS)". *PKD2* mutations include 202 pathogenic mutations, 59 polymorphisms and 13 missense VUS.

The finding of VUS can complicate the diagnosis process and genetic counseling. Our previous molecular study of ADPKD in 70 unrelated Thai patients has identified a total of 22 and 4 missense VUS in *PKD1* and *PKD2*, respectively (Thongnoppakhun, 2012). The findings of significant numbers of missense variants are challenging to interpret, thus requiring further evaluation of their pathogenic effects before reporting for clinical use. To assess an effect of VUS using functional studies or laboratory experiments is not only expensive and time consuming, but also difficult, especially for *PKD1*. Computational analysis is a method of choice for primary finding of the impact of VUS whether it would be pathogenic, likely pathogenic, likely benign or benign. This may provide improved clinical diagnostics that help the clinicians to take better decisions for the patient management.

The computational evaluation for the functional significance of missense VUS can be performed using several freely available online web-based programs based on various algorithms such as interspecies sequence variations, biochemical differences of resulting amino acid substitutions, and the location and context within the protein sequence. PolyPhen2, SIFT and MutationTaster (Adzhubei *et al*., 2009; Kumar *et al*., 2009; Schwarz *et al*., 2010) are the most commonly used prediction tools for missense variant interpretation in clinical laboratories (Richards *et al*., 2015). We hypothesized that the 9 selected missense VUS in *PKD1* identified by our group would be pathogenic. This study thus aimed to assess the performance of several up-to-date *in silico* tools for missense VUS prediction using *PKD1* and *PKD2* genes as models. Subsequently, a number of informative *in silico* tools were utilized to evaluate the pathogenicity of the *PKD1* missense VUS.

MATERIALS AND METHODS

In this study, PolyPhen2, SIFT, MutationTaster and additional 12 *in silico* tools (Reva *et al*., 2011; Shihab *et al*., 2013; Li *et al*., 2009; Choi *et al*., 2012; Bromberg *et al*., 2007; Capriotti *et al*., 2006; 2011; Bendl *et al*., 2014; Yates *et al*., 2014; Pappalardo *et al*., 2014; Capriotti *et al*., 2013; Zeng *et al*., 2014), as listed in Table 1, were chosen for prediction of missense variants. These web-based programs with high accuracy (around 80%) use different training datasets for prediction and different classification approaches (algorithms) based on either sequence and evolutionary conservation, protein sequence and structure, or machine learning methods (MLM) like support vector machine (SVM), naive Bayes (NB), random forest (RF) and neural networks (NN).

Dataset used and performance evaluation of the 15 *in silico* tools

In the PKDB database for *PKD1* and *PKD2* respectively, there are 144 and 13 base substitutions (missense variants) which are classified as "highly likely pathogenic", while an opposite class of 295 and 25 substitutions are "likely neutral". The former missense variants were used as the 'deleterious' dataset, while the latter being the 'neutral' one. To evaluate the performance of the 15 tools, four parameters: sensitivity, specificity, accuracy and precision are calculated based on the results of *in silico* analysis of known classes of missense variants in *PKD1* and *PKD2*, according to the following formulas:

Sensitivity (Sens.) = TP /(TP + FN)
(Number of true positive prediction)/(Number of all pathogenic variants)
Specificity (Spec.) = TN /(TN + FP)
(Number of true negative prediction)/(Number of all neutral variants)
Accuracy (Acc.) = (TN + TP) /(TN+TP+FN+FP)
(Number of correct predictions)/(Number of all predictions)
Precision (Prec.) = TP /(TP+FP)
(Number of true positive prediction)/(Number of all positive predictions)

Where:

True positive (TP) = Disease is predicted as Disease;
True negative (TN) = Neutral is predicted as Neutral;
False positive (FP) = Neutral is predicted as Disease;
False negative (FN) = Disease is predicted as Neutral

Table 1 *In silico* tools used for pathogenic prediction of known missense variants in *PKD1* and *PKD2.*

***In silico* tools**	**Training Dataset**	**Algorithm**			**Accuracy (%)**	**Reference**
		Sequence Conservation (matric)	**Machine Learning Methods (MLM) & Others**	**Protein Analysis**		
1. Polyphen 2	HumDiv, HumVar	+ (PSIC)	MLM: NB	+	92, 73	Adzhubei *et al*., 2009
2. SIFT	None	+	Multi-step alignment score, SIFT score	-	69	Kumar *et al*., 2009
3. MutationTaster	HGMD, dbSNP, ClinVar	+	MLM: NB	-	87.5	Schwarz *et al*., 2010
4. Mutation Assessor	UniProt, HUMSAVAR, COSMIC	+	FIS score	+	79	Reva *et al*., 2011
5. FATHMM	HGMD, UniProt, VariBench, SwissVar	+	Hidden Markov model (HMM)	-	86	Shihab *et al*., 2013
6. Mutpred	SwissProt, HGMD, Somatic cancer variants, The kinase data set	+	MLM: RF	+	83.5	Li *et al*., 2009
7. PROVEAN	UniProt human protein variants	+	Delta alignment score (PROVEAN score)	-	79.19	Choi *et al*., 2012
8. SNAP	Protein Mutant Database (PMD)	+	MLM: NN	+	77	Bromberg *et al*., 2007
9. PhD-SNP	HumVar Swiss-Prot	+	MLM: SVM	-	74	Capriotti *et al*., 2006
10. SNPs&GO	SwissVar database	+ (subSPEC)	MLM: SVM	-	81	Capriotti *et al*., 2011
11. PredictSNP* – MAPP – nsSNPAnalyzer – PANTHER – PhD-SNP – PolyPhen-1 – PolyPhen-2 – SIFT – SNAP	SwissVar, Swiss-Prot, HGMD, HUMSAVAR, PON-P, HumVar	+	MLM	+ (by SNAP)	71.6, 78.4	Bendl *et al*., 2014
12. SuSPect	HUMSAVAR / VariBench from dbSNP	Sequence features	MLM: SVM	-	82	Yates *et al*., 2014
13. VarMod	VariBench	Sequence features	MLM: SVM	-	-	Pappalardo 2014
14. Meta-SNP* – PANTHER – PhD-SNP – SIFT – SNAP	SwissVar	+	MLM: RF	+ (by SNAP)	79	Capriotti *et al*., 2013
15. EFIN	Human SwissProt, HumDiv	+	MLM: RF	-	83.7	Zeng *et al*., 2014

* PredictSNP and Meta-SNP are metaservers that integrate the predicted results from 8 and 4 *in silico* tools as listed, respectively to form a consensus prediction for a particular variant.

Pathogenicity prediction of missense VUS in *PKD1*

Nine missense VUS in *PKD1* found by our group (Thongnoppakhun, 2012) in 9 unrelated ADPKD patients including P676L, S1863Y, R3274C, P3551L, W3726R, L3749P, P3788R, I4105L and V4236F were selected to be studied because of their absence in relevant SNP databases e.g. PubMed, HGMD, OMIM, dbSNP, 1000 Genomes, ThaiSNP, ClinVar and Exome Variant Server. These variants were determined for their pathogenic potentials using the 15 selected *in silico* tools mentioned above.

RESULTS AND DISCUSSION

Pathogenic prediction of known missense variants in *PKD1* and *PKD2* by the 15 *in silico* tools (with two analysis results from EFIN, thus would be recounted as 16 tools hereafter) showed high performance in terms of sensitivity (0.88-1.00) for both genes in the three benchmark programs (1-3 in Table 2 i.e. PolyPhen2, SIFT and MutationTaster). This could reflect minimal false negative results obtained from them. However, other values (specificity, accuracy and precision) among the three benchmark tools were varied (0.40-0.84). Four programs i.e. VarMod, FATHMM, SNPs&GO and SuSPect had low concordance (20.5-65.3%) for *PKD1* pathogenic missense mutations. The other eight tools were comparable to the benchmarks, having superior performances in terms of accuracy, precision, sensitivity and specificity for *PKD1* (MutPred, PROVEAN, PhD-SNP and EFIN, Swiss-Prot score) and for *PKD2* (MutationAssessor, PROVEAN, PredictSNP and Meta-SNP), with only exceptions in sensitivity of PROVEAN for *PKD1*, and PROVEAN and Meta-SNP for *PKD2*.

The four programs with low sensitivity mentioned above were excluded from the *PKD1* VUS analysis due to the resulting under-prediction that will impede a consensus prediction. Thus, the rest 12 tools were used for prediction of pathogenic potentials of 9 missense VUS in *PKD1* (Table 3). As no single tool seemed to be the best program (Table 2), the use of multiple tools for VUS interpretation of *PKD1* is preferable for accuracy improvement due to their different strengths and weaknesses based on the algorithms (Adzhubei *et al*., 2010). The performance of the tools in Table 2 was ranked due to the prediction for *PKD1* gene, being determined based on how well the tool was able to identify both pathogenic and neutral variants, thus reducing the number of false positive and false negative calls. The first three tools (Polyphen2, SIFT and MutationTaster) were ranked according to an acceptance as the combined benchmark tools, regardless of their priorities. MutPred and PROVEAN were ranked as the 4th and 5th tool due to their overall best performance in terms of accuracy, precision, sensitivity and specificity (top average value). PhD-SNP and EFIN were almost as good compared with each other, but EFIN has 2 analysis methods that gave some different results, thus their orders as the 6th, 7th and 8th are not significant, since they were in the same grouping (all of them were added to the '5 tools' to be '8 tools'). In addition, whenever EFIN was used, both Swiss-Prot score and HumDiv score were analyzed. Therefore, the two methods of EFIN were intended to be adjacently ordered as the 7th and 8th. PredictSNP, Meta-SNP, MutationAssessor and SNAP were ranked as the 9th to 12th tools according to their overall descending performance.

The performance for *PKD2* prediction in Table 2 was also performed in addition to that for *PKD1* in order to demonstrate that the performance of each individual tool as well as of combined sets of tools was different between *PKD1* and *PKD2*, thus supporting the assessment of tools in a gene-specific manner. This information could be useful for future evaluation of VUS which have been already found in *PKD2* as well.

Combinations of tools starting from 3 (benchmark tools) to 5, 8 and 12 tools according to the descending performance (shown at the bottom of Table 3) showed consensus predictions for all of the nine VUS being predicted in *PKD1*, except only one, S1863Y, of which the use of 12 tools gave a different output to be "likely pathogenic" as opposed to "likely neutral" predicted by the 3, 5 and 8 tools. This could reflect awareness of prediction for VUS with mild effects that could be missed by the use of fewer tools. Thus, more tools should be considered to ensure that all likely pathogenic variants would not be missed. From our prediction of the 141 known *PKD1* missense mutations (*PKD1*-"highly likely pathogenic" in Table 2) performed by prior combination of 3, 5, 8 and 12 tools (data not shown), true positive predictions (consensus in ≥ 50% of tools) were found in 136 (>66.7%), 140 (>60%), 140 (>62.5%) and 140 (>66.7%) mutations, respectively. Prediction by the combined 5, 8 and 12 tools was shown to have the same outcome (140/141, 99.3%), being better than that of the 3 benchmark tools (136/141, 96.5%) which was even worse than those of some individual tools such as PhD-SNP and EFIN (Swiss-Prot score) (137/141, 97.2%) (Table 2). Therefore, the combined

5 tools seemed to be mostly practical since they used minimal tools while giving comparably best performance with the other more redundant sets of 8 and 12 tools. However, a slightly difference in consensus of predictions among the combined 5, 8 and 12 tools as >60%, >62.5% and >66.7%, respectively was observed. This information more or less may support the highest reliability of the 12 tools used, especially in cases of mildly pathogenic variants.

Table 2 Performance of the 16 *in silico* tools for pathogenic prediction of known missense variants in PKD1 and PKD2 (values ≥ 80% are bold to stress the high performance).

% Concordance of the specified variants in each class as predicted by each tool					**Performance of each tool for *PKD1* (upper) and *PKD2* (lower)**			
***In silico* Tools**	***PKD1***		***PKD2***		**Acc.**	**Prec.**	**Sens.**	**Spec.**
	Highly likely pathogenic	**Likely Neutral**	**Highly likely pathogenic**	**Likely Neutral**				
1. Polyphen2	**87.9**	51.2	**100.0**	76.0	0.63	0.46	**0.88**	0.51
	(124/141)	(151/295)	**(13/13)**	(19/25)	**0.84**	0.68	**1.00**	0.76
2. SIFT	**92.9**	77.6	**100.0**	72.0	**0.83**	0.66	**0.93**	0.78
	(131/141)	(229/295)	**(13/13)**	(18/25)	**0.82**	0.65	**1.00**	0.72
3. MutationTaster	**92.9**	74.9	**100.0**	40.0	**0.81**	0.64	**0.93**	0.75
	(131/141)	(221/295)	**(13/13)**	(10/25)	0.61	0.46	**1.00**	0.40
4. MutPred	**93.6**	**89.8**	**100.0**	60.0	**0.91**	**0.81**	**0.94**	**0.90**
	(132/141)	**(265/295)**	**(13/13)**	(15/25)	0.74	0.57	**1.00**	0.60
5. PROVEAN	**89.4**	**89.2**	**92.3**	**92.0**	**0.89**	**0.80**	**0.89**	**0.89**
	(126/141)	**(263/295)**	**(12/13)**	**(23/25)**	**0.92**	**0.86**	**0.92**	**0.92**
6. PhD-SNP	**97.2**	**80.3**	**100.0**	48.0	**0.86**	0.70	**0.97**	**0.80**
	(137/141)	**(237/295**)	**(13/13)**	(12/25)	0.66	0.50	**1.00**	0.48
7. EFIN (Swiss-Prot score)	**97.2**	**82.0**	**92.3**	76.0	**0.87**	0.72	**0.97**	**0.82**
	(137/141)	**(242/295**)	**(12/13)**	(19/25)	**0.82**	0.67	**0.92**	0.76
8. EFIN (HumDiv score)	**96.5**	64.4	**100.0**	44.0	0.75	0.56	**0.96**	0.64
	(136/141)	(190/295)	**(13/13)**	(11/25)	0.63	0.48	**1.00**	0.44
9. PredictSNP	**95.7**	79.3	**100.0**	**88.0**	**0.85**	0.69	**0.96**	0.79
	(135/141)	(234/295)	**(13/13)**	**(22/25)**	**0.92**	**0.81**	**1.00**	**0.88**
10. Meta-SNP	**95.7**	76.9	**92.3**	**80.0**	**0.83**	0.67	**0.96**	0.77
	(135/141)	(227/295)	**(12/13)**	**(20/25)**	**0.84**	0.71	**0.92**	**0.80**
11. MutationAssessor	**88.7**	76.6	**100.0**	**92.0**	**0.81**	0.64	**0.89**	0.77
	(125/141)	(226/295)	**(13/13)**	**(23/25)**	**0.95**	**0.87**	**1.00**	**0.92**
12. SNAP	**92.2**	57.6	**100.0**	72.0	0.69	0.51	**0.92**	0.58
	(130/141)	(170/295)	**(13/13)**	(18/25)	**0.82**	0.65	**1.00**	0.72
13. SuSPect	65.3	**95.9**	**84.6**	**96.0**	**0.86**	**0.88**	0.65	**0.96**
	(92/141)	**(283/295)**	**(11/13)**	**(24/25)**	**0.92**	**0.92**	**0.85**	**0.96**
14. SNPs&GO	59.6	**98.3**	**100.0**	76.0	**0.86**	**0.94**	0.60	**0.98**
	(84/141)	**(290/295)**	**(13/13)**	(19/25)	**0.84**	0.68	**1.00**	0.76
15. FATHMM	23.4	**98.9**	38.5	**100.0**	0.75	**0.92**	0.23	**0.99**
	(33/141)	**(292/295)**	(5/13)	**(25/25)**	0.79	**1.00**	0.38	**1.00**
16.VarMod	20.5	**96.3**	69.2	**92.0**	0.72	0.73	0.21	**0.96**
	(29/141)	**(284/295)**	(9/13)	**(23/25)**	**0.84**	**0.82**	0.69	**0.92**

Table 3 Pathogenic prediction of unknown missense *PKD1* VUS using 12 acceptable *in silico* tools.

In Silico* Tools Suitable for *PKD1		***PKD1* missense VUS identified by our group which are selected for pathogenicity prediction**								
		P676L	**S1863Y**	**R3274C**	**P3551L**	**W3726R**	**L3749P**	**P3788R**	**I4105L**	**V4236F**
1. Polyphen2		+	+	+	+	+	+	-	+	+
2. SIFT		+	-	-	+	+	+	-	-	+
3. MutationTaster		-	-	+	+	+	+	-	-	-
4. MutPred		-	+	+	+	+	+	+	+	-
5. PROVEAN		+	-	+	+	+	+	-	-	-
6. PhD-SNP		+	+	+	+	+	+	+	-	+
7. EFIN (Swiss-Prot score)		-	-	-	+	+	+	-	-	+
8. EFIN (HumDiv score)		+	-	+	+	+	+	-	-	+
9. Predict SNP		-	+	-	+	+	+	-	-	+
10. Meta-SNP		-	+	-	+	+	+	+	-	+
11. Mutation Assessor		+	+	+	+	+	+	-	-	+
12. SNAP		+	+	+	-	+	+	+	+	-
Scores obtained from the combined tools	**1-3**	**2/3 (67%)**	1/3 (33%)	**2/3 (67%)**	**3/3 (100%)**	**3/3 (100%)**	**3/3 (100%)**	**0/3 (0%)**	**1/3 (33%)**	**2/3 (67%)**
	1-5	**3/5 (60%)**	2/5 (40%)	**4/5 (80%)**	**5/5 (100%)**	**5/5 (100%)**	**5/5 (100%)**	**2/5 (40%)**	**1/5 (20%)**	**3/5 (60%)**
	1-8	**5/8 (63%)**	3/8 (38%)	**6/8 (75%)**	**8/8 (100%)**	**8/8 (100%)**	**8/8 (100%)**	**2/8 (25%)**	**2/8 (25%)**	**5/8 (63%)**
	1-12	**7/12 (58%)**	**7/12 (58%)**	**8/12 (67%)**	**11/12 (92%)**	**12/12 (100%)**	**12/12 (100%)**	**4/12 (33%)**	**3/12 (25%)**	**8/12 (67%)**
Summarized Prediction		**Likely Pathogenic**	**Likely Pathogenic**	**Likely Pathogenic**	**Pathogenic**	**Pathogenic**	**Pathogenic**	**Likely Benign**	**Likely Benign**	**Likely Pathogenic**

Remarks:

+ represents Probably/Possibly Damaging (1), Damaging (2), Disease causing (3), Deleterious (4, 5, 9), Disease (6, 7, 8, 10), Medium (11), Non-neutral (12)

- represents Benign (1), Tolerated (2), Polymorphism (3), Neutral (4-12)

Therefore, combination of 12 different tools seemed to give more reliable predictions (Table 3). The frequent inconsistency among the 12 *in silico* tools were caused by the different training datasets and algorithms they are based on. Both Polyphen2 (1st tool in Table 3) and MutationTaster (3rd tool) have used a machine learning method (MLM) based on naive Bayes (NB), but the former predicted S1863Y to be 'pathogenic' while the latter assigning it as 'neutral'. The different results may be resulted from the other algorithms of Polyphen2 such as PSIC and an additional protein analysis. Moreover, the training datasets of both tools were also dissimilar (see details in Table 1). Mutpred (4thtool), Meta-SNP (10th tool) and EFIN (7th and 8th tool) used the same random forest (RF) based MLM, but the prediction for S1863Y was 'pathogenic' only by Mutpred and Meta-SNP. An explanation for this set is analogous to the above comparison. It was noticed that an additional protein analysis as used in Polyphen2, Mutpred, Meta-SNP (as a metaserver including SNAP which uses an additional protein analysis) may have an important role in the prediction for S1863Y as 'pathogenic'. This observation was also supported by the positive prediction for S1863Y in the other three tools which also applied an additional protein analysis regardless of their MLM, i.e. PredictSNP (9th tool, as a metaserver including SNAP which uses an additional protein analysis), Mutation Assessor (11th tool) and SNAP (12th tool).

In this study, 3 out of 9 VUS in *PKD1* were predicted to be "pathogenic" by >90% of tools, while

the other 4 should be "likely pathogenic" as predicted by ≥50% of tools. Two VUS were "likely benign" as the positive prediction was from <50% of tools (Table 3). In case of no other approaches available, pathogenicity prediction of missense VUS using appropriate sets of *in silico* tools for *PKD1* and *PKD2* genes could at least provide improved clinicians' diagnosis for better decisions to manage the ADPKD patients. Finally, however, these predictions alone could not be used as the single source of evidence to assign a definite pathogenicity of the VUS. Genetic evidence from segregation analysis within affected families and population frequency of the VUS of interest should be further clarified to confirm these computational predictions.

CONCLUSION

The optimum tools for predicting VUS pathogenicity can be varied depending on the gene under investigation and parameters used. A total of 16 web-based tools were thus comparatively validated against missense variants of known effect in *PKD1* and *PKD2*. No one tool can be considered best performance. For *PKD1*, a combined benchmark tools (PolyPhen2, SIFT, MutationTaster) should be a minimum set of prediction tools. However, a combination of 12 tools (in Table 3) seemed to give more reliable prediction, especially for the VUS with mild effects. The ability to classify the pathogenicity of the selected 9 VUS in *PKD1* of this project will expand knowledge of *PKD1* mutation spectrum, providing the future applications in clinical practice.

ACKNOWLEDGEMENTS

This work was supported by the Thailand Research Fund (TRF) Mid-Carreer Research Grant # RMU5180045 (to WT) and the Chalermphrakiat Grant, Faculty of Medicine Siriraj Hospital, Mahidol University (MP, CL, KV, and WT).

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–9.

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 10: e1003440.

Bromberg Y, Rost B (2007) SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 35: 3823–35.

Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22: 2729–34.

Capriotti E, Altman RB (2011) Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics 12: 1–10.

Capriotti E, Altman RB, Bromberg Y (2013) Collective judgment predicts disease-associated single nucleotide variants. BMC Genomics 14: S2.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. de Brevern AG, editor. PLoS ONE 7: e46688.

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–81.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinforma Oxf Engl 25: 2744–50.

Pappalardo M, Wass MN (2014) VarMod: modelling the functional effects of non-synonymous variants. Nucleic Acids Res 42: W331–6.

Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39: e118.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–23.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7: 575–6.

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, Day IN, Gaunt TR (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 34: 57–65.

Tan YC, Blumenfeld J, Rennert H (2011) Autosomal dominant polycystic kidney disease: genetics, mutations and microRNAs. Biochim Biophys Acta 1812: 1202–12.

Thongnoppakhun W (2012) Improved molecular analysis of the genes causing autosomal dominant polycystic kidney disease (ADPKD): an efficient approach using denaturing high performance liquid chromatography (DHPLC). Final research report. Bangkok: The Thailand Research Fund (TRF); 106 p.

Yates CM, Filippis I, Kelley LA, Sternberg MJ (2014) SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. J Mol Biol 426: 2692–701.

Zeng S, Yang J, Chung BHY, Lau YL, Yang W (2014) EFIN: predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome. BMC Genomics 15: 455.

A Deep Catalogs of Human Genetic Variation (1000 Genomes), http://www.1000genomes.org/about (March 2015).

Autosomal Dominant Polycystic Disease: Mutation Database (PKDB), http://pkdb.mayo.edu/ (March 2015).

Human Gene Mutation Database (HGMD), http://www.hgmd.org/ (March 2015).

NCBI Database of Short Genetic Variations (dbSNP), http://www.ncbi.nlm.nih.gov/SNP/ (March 2015).

NHLBI GO Exome Sequencing Project (ESP) (Exome Variant Server), http://evs.gs.washington.edu/EVS/ (March 2015).

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/ (March 2015).

Public archive of relationships among sequence variation and human phenotype (ClinVar), http://www.ncbi.nlm.nih.gov/clinvar/ (March 2015).

Thailand SNP Discovery Project (ThaiSNP), http://www4a.biotec.or.th/GI/tools/thaisnp (March 2015).

US National Library of Medicine National Institutes of Health (PubMed), http://www.ncbi.nlm.nih.gov/pubmed (March 2015).