Prediction of deleterious nsSNPs in human UGT1A1 gene by web available algorithm tools

Rodrigues C. [a,b] *, Costa E.[c,e], VieiraE.[d], Santos R.[d], Santos-Silva A.[a,e],Bronze-da-Rocha E.[a,e]

a. Departamento de Ciências Biológicas, Laboratório de Bioquímica, FFUP; b. Departamento das Tecnologias de Diagnóstico e Terapêutica, ESS Instituto Politécnico de Bragança. c. Departamento de Ciências da Saúde, UCP. d. Unidade de Genética Molecular, IGMJM, INSARJ. e. Instituto de Biologia Molecular e Celular , UP.

1 – INTRODUCTION

The uridine diphosphate glucuronosyltransferase (UGT1A1) belongs to the class of phase II enzymes involved in the metabolism and detoxification of numerous xenobiotic and endogenous compounds[1] (e.g. bilirubin). Genotyping data lead to the discovery of over 100 single nucleotide polymorphisms (SNPs) within the UGT1A1 gene[2]. Some of the non-synonymous (ns) SNPs (nsSNPs) of the human UGT1A1 gene variants have been associated to hyperbilirubinemia in Gilbert's and Crigler-Najjar syndromes[3], as well as altered drug clearance and/or drug response[4].

In UGT1A1, and other genes, there are many nsSNPs which genotype-phenotype correlations were not established, since the study of the functional impact of all SNPs is time consuming and expensive. Alternatively, bioinformatics tools have gained an increased importance with the prospect of reducing the totality of detailed studies at protein level. The aim of this study was to investigate the potential of bioinformatics approaches, using five web available to predict the phenotype of 28 human UGT1A1 nsSNPs, previously characterized at protein level by in vivo and in vitro studies.

2 – MATERIAL AND METHODS

Information describing the *UGT1A1* gene variants was obtained from mutation database websites:

- http://www.polydoms.cchmc.org/polydoms;
- http://www.mutdb.org;
- http://www.ncbi.nlm.nih.gov/sites/entrez.

To predict the phenotype of 28 human UGT1A1 nsSNPs previously characterized at protein level by in vivo and in vitro studies we used five web available algorithms:

- http://sift.jcvi.org/ - Sorting Intolerant from Tolerant (SIFT);
- http://genetics.bwh.harvard.edu/pph2/ - Polymorphism Phenotyping-2 (PolyPhen-2);
- http://agvgd.iarc.fr/ - Align Grantham Variance/Grantham Difference (A-GVGD);
- http:// mendel.standford.edu/ - Multivariate Analysis of Protein Polymorphism (MAPP);
- http://www.ebi.ac.uk/help/matrix.html/ Block Substitution Matrix score 62 (BLOSUM62).

3 – RESULTS

Fig 1. Correct prediction rate of each *in silico* analysis tool.

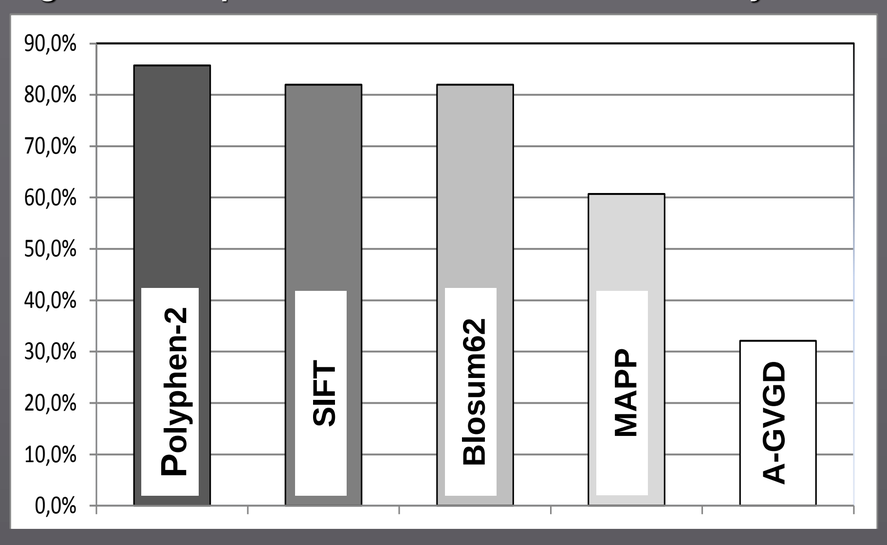

Table 1. shows the phenotype of 28 human UGT1A1 nsSNPs, previously characterized at protein level by in vivo and in vitro studies. From those, 24 SNPs were confirmed as responsible for changes in protein function and in 4 there were no detected impact. Results from in silico analysis (fig 1.) showed a correct prediction rate of 85.7% for Polyphen-2, 82.0% for both BLOSUM62 and SIFT, 60.7% for MAPP and 32.1% for Align-GVGD. The five computational methods had concordant results using Polyphen-2 and SIFT algorithms in 78.6% (n=23) of variants. Concordance in variants prediction, between the five used methods and with results obtained at protein levels, was observed in 14.3% (n=6) nsSNPs.

Table 1. UGT1A1 snSNPs and predicted effect of on protein function by the 5 web tools.

SNP Position	Amino acid Position	Enzyme Activity		Computational Methods				
		In vivo	*In vitro*	SIFT[1]	PolyPhen-2[2]	Blosum62[3]	MAPP[4]	A-GVGD[5]
44T>G	L15R	Reduced	Reduced	Deleteriuos	Possibly Damaging	Deleterious	Neutral	Neutral
211G>A	G71R	Reduced	Reduced	Tolerated	Possibly Damaging	Deleterious	Neutral	Neutral
476T>C	I159T	Normal	Normal	Tolerated	Benign	Deleterious	Neutral	Neutral
524T>A	L175Q	38.4%	Reduced	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
529T>C	C177R	Inactive	Inactive	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
625C>T	R209W	Reduced	Reduced	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Deleteriuos
686C>A	P229Q	Reduced	Reduced	Tolerated	Probably Damaging	Deleterious	Neutral	Neutral
826G>C	G276R	Inactive	Inactive	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Deleteriuos
872C>T	E291V	Absent	Absent	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
881T>C	I294T	40-55%	-	Deleteriuos	Benign	Deleterious	Deleteriuos	Neutral
923G>A	G308E	Inactive	Absent	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Deleteriuos
928A>G	M310V	26-51%	-	Deleteriuos	Probably Damaging	Tolerated	Neutral	Neutral
962C>G	A321G	Normal	Normal	Tolerated	Benign	Tolerated	Neutral	Neutral
964A>G	I322V	Normal	Normal	Deleteriuos	Probably Damaging	Tolerated	Neutral	Neutral
991C>T	Q331R	Reduced	Reduced	Deleteriuos	Probably Damaging	Tolerated	Neutral	Neutral
1006C>T	R336W	0-10%	Absent	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Deleteriuos
1069C>T	Q357R	Inactive	Inactive	Deleteriuos	Probably Damaging	Tolerated	Deleteriuos	Neutral
1075G>A	D359N	Normal	Normal	Deleteriuos	Probably Damaging	Tolerated	Neutral	Neutral
1091C>T	P364L	-	Reduced	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
1099C>G	R367G	Reduced	Reduced	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
1102G>A	A368T	Absent	Absent	Deleteriuos	Probably Damaging	Tolerated	Deleteriuos	Neutral
1124C>T	S375F	Inactive	Inactive	Deleteriuos	Probably Damaging	Deleterious	Neutral	Neutral
1143C>G	S381R	Absent	Absent	Deleteriuos	Benign	Deleterious	Deleteriuos	Neutral
1159C>T	P387S	Absent	Absent	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Deleteriuos
1201G>C	A401P	Absent	Absent	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Neutral
1292T>C	I431T	Reduced	-	Tolerated	Probably Damaging	Deleterious	Deleteriuos	Neutral
1381T>C	W461R	-	Inactive	Deleteriuos	Probably Damaging	Deleterious	Deleteriuos	Neutral
1456T>G	Y486D	Reduced	Reduced	Deleteriuos	Probably Damaging	Deleterious	Deleterious	Neutral

[1]Qualitative classification given by Polyphen: Probably damaging, Possibly damaging and Benign. [2] Sort Intolerant from Tolerant, [3]Align Grantham Variance/ Grantham Distance;[4]. BLOSUM62 (Blok Substitution Matrix); [5] Multivariate Analysis of Protein Polymorphism.

4 - CONCLUSIONS

- SIFT and Polyphen-2 together, were the best predictor methods of nsSNPs phenotype in human UGT1A1 gene. These tools have the advantage of directing and complement functional assays.
- A limitation of this approach could be the lack of the 3D structure of UGT1A1 and some of this tools rely on structural databases.
- The observed discrepancy in variants prediction phenotype may be eliminated with a method combining all the currently available criterions.

REFERENCES: (1) Strasburg CP. *et al*., *Arch Toxicol* (2008) 82:415–433; (2) Strasburg CP. *et al*., *Pharmacogenomics* (2008) 9:703-15; Costa E. *Blood Cell Mol Dis*. (2006) 36:77-80; (4)Sinafi SB. *et al*.,Biochem J. (1994) 303:233-40.

ACKNOWLEDGEMENTS: This work was supported by PhD grant (SFRH/BD/42791/2007) from Fundação para a Ciência e Tecnologia (FCT) attributed to Carina Rodrigues and Fundo Social Europeu (FSE).