

Insilico analysis for DPYD gene and the effect of the mutation DPD enzyme

ABSTRACT

Aim: The aim of this study was to implement an insilico bioinformatics analysis for clinically observed missense mutations in human DPYD gene to investigate the effect these mutations on DPD protein's structure and function.

Methods: The human DPYD gene was investigated in dbSNP/NCBI, 273238 SNPs were found; 99645 SNPs were Homo sapiens; of which 534 were missense SNPs. Missense SNPs were selected for in silico analysis; SIFT, Polyphen2, SNPs & GO, I-mutant 2.0, Mutation 3D, UCSF Chimera and HOPE were used to investigate the effect of SNPs on DPD protein's structure and function.

Results: 69 SNPs were found to be highly damaging for the protein by SIFT and Polyphen, of which 4 SNPs were observed to be associated to clinical presentations (M166V, V335L, I560S, D949V). This 69 SNPs were further analyzed by SNPs & GO, one SNP (D949V) was observed to be associated to clinical presentations. The 4 nsSNPs that observed to be associated to clinical presentations were further analyzed by I-Mutant, HOPE and chimera tools to predict their stability index and visualize wide and mutant residues in their protein 3D structure.

Conclusions: M166V SNP was located at the interface of the DPD protein that is important for binding of other molecules. I560S located at the interface of the DPD protein interfering with its activity and contact with residues in another domain. M166V SNP and D949V make the mutant residue smaller than the wild type residue, this will cause a possible loss of external interactions. I560S and V335L make the mutant residue bigger than the wild-type residue, probably will not fit in the core of the protein affecting its stability. DPYD variants Screening before the administration of 5-FU is crucial to identify the patients how suspected to develop an adverse 5-FU drug reaction or toxicity toward reducing the early severe life-threatening toxicities associated with 5-FU and reduce the incidence of toxic death due to this drug.

KEYWORDS: DPYD, 5 FLUOROURACIL, DPD, IN SILICO ANALYSIS, SIFT, POLYPHEN2, SNPS & GO, IMUTANT2.0, MUTATION 3D, UCSF CHIMERA, HOPE.

1. INTRODUCTION

The human DPYD gene located on chromosome 1p21.3 present as a single copy consists of 23 exons, cover approximately 950 kb in length with 3 kb of coding sequence and an average intron size of about 43kb [1]. It encode for dihydropyrimidine dehydrogenase (DPD) enzyme, an important regulatory enzyme in the metabolism of pyrimidines; thymine and uracil that occur naturally in the cell, beside cancer chemotherapy fluoropyrimidine 5-FU drug [2].

fluoropyrimidine 5-FU drug (5-FU) is a standard anticancer treatment used in the management of colorectal cancer [3]. The mode of action of 5-fluorouracil (5-FU) its conversion to 5-fluoro-2'deoxy-5' monophosphate (FdUMP) leading to inhibition of thymidylate synthase (TS) and then of DNA synthesis and this will lead to DNA replication inhibition [4]. The fluoropyrimidines (5-FU) degradation pathway is carry out by (DPD) enzyme, DPD transformed 5-FU to 5-fluoro-5,6-dihydrouracil (5-FDHU), which is metabolized to α -fluoro- β -alanine (FBAL) by two others additional enzymes and lastly excreted by kidneys [5].

Regarding DPYD gene and 5-FU there is a direct link or relationship between the individual's DPYD activity and the metabolism of 5-FU its self [6]. In 5-FU-based cancer chemotherapy, drug resistance highly vary among colorectal cancer patients limiting its clinical usage. Genetic differences among patients has been confirmed in many research to be contribute to variability drug response [7-10]. Patients who have mutant allele in DPYD gene suffering from severe (lethal) toxicity after using 5FU drug [7,8]. Moreover, the adverse effects of 5-FU are often lethal especially when this mutant reduce the DPD efficiency [9,10]. An estimated 10%–20% of treated patients develop serious, sometimes life-threatening, 5-FU toxicity [11].

Out of 30 potentially relevant SNPs reported in the literature [12,13], we could determine the following four missense SNPs, (M166V, V335L, I560S, D949V) as most often identified in clinical DPD deficiency. Studies suggested when aspartic acid change to Valine in 949 position that causing a directly interferes with electron transport or cofactor binding and strongly associate with grade III and

IV toxicity in a cohort of patients with colorectal cancer treated with 5-FU-based therapy [10], whereas Isoleucine change to serine in position 560, that point mutation due to destabilize the DPD protein [10]. When Isoleucine change to serine at position 560 then the enzyme activity will be (12.5%–25%), and will be more than 25% when aspartic acid change to valine at 949 position but still less than wild type [14].

Bioinformatics is an important role in almost all aspects of drug discovery, development and assessment. As bioinformatics resources (data bases and software) facilitate the understanding and the prediction of the drug metabolism, elimination and also toxicity of drugs [15,16]. Identification of SNPs responsible for adverse drug reaction or toxicity, provide a basic information for physicians to setup a therapy toward reducing the risk of complications and improving patient quality of life. In this study we implemented an insilico analysis for clinically observed missense mutations in human DPYD gene using different bioinformatics databases and softwares to investigate the effect these mutations on DPD protein's structure and function and whether these variations can affect its activity or not.

2. MATERIAL AND METHODS

Different soft-wares; SIFT, polyphen-2, I-mutant3.0, SNPs & GO and chimera were used to investigate the effect of SNPs mutations on DPD protein structure and function. Prediction of deleterious effect of non synonymous SNPs was done by SIFT and Polyphen-2 soft-wares. The association of missense SNPs with disease was done by SNPs & GO software. Prediction of stability changes was investigated in I mutant. The structural changes in 3D structure were analyzed using HOPE and Chimera software.

2.1 Datasets

SNPs located in DPYD gene were obtained from the database of SNPs (dbSNP); it is, a genetic variation database established by the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/SNP/>). In this study Homosapie missense SNPs had been selected and submitted to bioinformatics tools for further investigation.

2.2 Prediction of Structural Impact of nsSNPs on Protein by SIFT software

Sorting Tolerant from Intolerant (SIFT) is an online bioinformatics software that uses an algorithm to predict the effect of amino acid substitutions, resulting from Non-synonymous SNPs (nsSNPs), on protein function [17]. SIFT uses sequence homology to predict the effects of all possible substitutions at each position in the protein sequence. The threshold intolerance score for SNPs is 0.05 or less [18]. nsSNPs within dbSNP retrieved data were selected as an input for SIFT. Available at: (<http://sift.jcvi.org/>).

2.3 Prediction of functional modification of coding nsSNPs by Polyphen-2

Polymorphism Phenotyping v2 (PolyPhen-2) is another bioinformatics server tool that predicts the damaging effects of missense mutations. PolyPhen searches for protein 3D structures and make multiple alignments of homologous sequences and amino acid contact in several protein databases and calculate position-specific independent count scores (PSIC) for each of two variants and then computes the PSIC scores difference between two variants, where the higher PSIC score difference indicates that the functional impact of amino acid substitution is likely to occur. PolyPhen-2 outcome can be one of the following: probably damaging, possibly damaging, or benign, with score range from 0 to 1 [19].

Available at: (<http://genetics.bwh.harvard.edu/pph2/>).

The input FASTA sequence of protein with the position of interest and the new residue were submitted to Polyphen to predict functional impact of mutations.

2.4 Prediction of Disease Associated Variations by SNPs&GO

Single Nucleotide Polymorphism Database (SNPs) & Gene Ontology (GO) is a support vector machine (SVM) based on the method to accurately predict the disease related mutations from protein sequence. FASTA sequence of whole protein is considered to be an input option and output will be the prediction results based on the discrimination among disease related and neutral variations of protein sequence. The probability score higher than 0.5 reveals the disease related effect of mutation on the parent protein function [20].

Available at: (<http://snps.biofold.org/snps-and-go/pages/method.html>).

2.5 Analysis of nsSNPs' Impact on Protein Stability using I-Mutant 2.0 Server:

Change in protein stability disturbs both protein structure and protein function [21]. I-Mutant is a suite of support vector machine, based predictors integrated in a unique web server. It offers the opportunity to predict the protein stability changes upon single-site mutations. from the protein structure or sequence. [22]. The FASTA sequence of protein retrieved from UniProt is used as an input to predict the mutational effect on protein and stability RI value (reliability index) computed.

Available at: (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant2.0.cgi>).

2.6. Automatic Protein Structural Analysis and Information Using HOPE Server

Automatic mutant analysis server can provide insight into the structural effects of a mutation. HOPE collects information from a wide range of information sources including calculations on the 3D coordinates of the protein by using sequence annotations from the UniProt database, and predictions by DAS services. Input method of Project HOPE carries the protein sequence and selection of Mutant variants. HOPE server predicts the output in the form of structural variation between mutant and wild type residues. HOPE builds a report with text, figures, and animations that is easy to use and understandable for medical researchers [23]

Available at: (<http://www.cmbi.ru.nl/hope/>)

2.7 Modeling amino acid substitution by UCSF Chimera Model Software

CPHmodels-3.2 server is a protein homology modeling prediction server, used to predict the 3D structure of proteins with an unknown 3D structure model. Modeling SNPs on the 3D structure of the proteins is a very helpful action in order to predict the effect that SNPs may cause on the structural level. The template recognition based on profile-profile alignment guided by secondary structure and exposure predictions [24]. Protein sequences requirements were submitted to CPH server to get the model as PDB file.

Available at: <http://www.cbs.dtu.dk/services/CPHmodels/>.

The resultant PDB files were opened using Chimera program which was used to visualize the PDB structure.

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data. Chimera software was used to scan the 3D (three-dimensional) structure of specific protein, and hence

modifies the original amino acid with the mutated one to see the impact that can be produced. The outcome is then a graphic model depicting the mutation. Chimera (version 1.8) currently. Available at ([http:// www.cgl.ucsf.edu/chimera](http://www.cgl.ucsf.edu/chimera))

3.Results and discussion

Mutations of DPYD gene had been reported to be associated with 5-fluorouracil toxicities, in the present study we presented a computational analysis for the DPYD SNPs that observed to be associated to clinical presentations using several bioinformatics softwares and databases to understand how do these mutations affect the DPD protein structure and function and promote the 5-fluorouracil toxicities.

3.1 Retrieval of SNPs

The human DPYD gene investigated in this work were retrieved from the NCBI dbSNP database (<http://ncbi.nlm.nih.gov/>). DPYD gene was containing a total of 273238 SNPs; 99645 SNPs were Homosapins ; of which 534 were missense SNPs. Initially missense coding SNPs were selected for our investigation.

3.2 Predicted Results by SIFT, PolyPhen, and SNAP& GO

Coding SNPs were analyzed using SIFT and Polyphen soft-wares. Batch nsSNPs (rs-IDs) were submitted to SIFT and Polyphen server. Out of 534 , 69 SNPs were predicted to be damaging by both servers. Out this 69 SNPs, four SNPs were observed to be associated with clinical presentations ; rs2297595 (M166V), rs55886062 (I560S), rs67376798 (D949V), rs72549306 (V335L).

SNPs that predicted to be deleterious with SIFT, Polyphen; were submitted to SNAP&GO, 20 nsSNP (K→F, I→T, K→N, S→R, Y→C, R→W, L→S,R→C, S→L, T→I, L→W, A→D, P→S, G→R, D→N, I→N, D→V, K→E, H→Q, V→F) were predicted to be associated with disease. One SNP, rs67376798 (D949V) was observed to be associated to clinical presentations (Table1).

The differences in prediction capabilities refer to the fact that every prediction algorithm uses different sets of sequences and alignments.

The 4 nsSNPs that observed to be associated to clinical presentations were selected to predict their stability index and visualize wide and mutant residues in their protein 3D structure .

3.3 Prediction of of nsSNPs Impact on the Protein Stability by I-Mutant 2.0 server

We submitted independently the four proteins sequence containing missense SNPs which observed to be associated to clinical presentations due to effect of mutation on DPD activity to I-Mutant 2.0 server , all 4 SNPs (rs2297595, rs55886062, rs67376798, rs72549306). were predicted to decrease protein stability, this results demonstrated that all proteins stability were changed due to SNPs alteration (Table 2).

This finding reveal that, the DPYD mutation associated with decrease of DPD capacity to degrade 5FU resulting in excess drug accumulation and toxicity and death, therefore the genetic screening for the presence of these mutations in cancer patients is essential before the administration of 5-FU provide a basic information for physicians to setup a therapy toward reducing the risk of

complications and improving patient quality of life. through dose-adaptation needs to be considered or prescribed non-fluoropyrimidine-based chemotherapy.

3.4 Protein 3D structure Modeling by Chimera and HOPE

Protein sequences containing the nsSNPs were again submitted to CPH 3.2 server and HOPE to get the protein 3D structure model, then the CPH 3.2 resultant PDB files were opened using Chimera program to browse and respectively locate the 3D structure of the each protein and to alter the native amino acid with a mutated one, the final outcome was a graphic model depicting the mutation.

rs2297595 (M166V) change in the amino acid Methionine into Valine at position 166. The mutant residue is smaller than the wild-type residue. The mutant will cause an empty space in the core of the protein. This mutant is located in a domain that is important for binding of other molecules (Fig1-a/1-c) and (Fig1-b/1-d).

rs55886062 (I560S) change in the amino acid Isoleucine into Serine at position 560. The mutant is bigger than the wild-type. the charge of wild-type was neutral, the mutant residue charge is negative. The wild-type residue is more hydrophobic than the mutant residue. The mutated residue is located in a domain that is important for the activity of the protein and in contact with residues in another domain (Fig2-a/2-c) and (Fig2-b/2-d).

rs67376798 (D949V) change in the amino acid Aspartic acid into Valine at position 949. There a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type is lost by this mutation, this can cause loss of interaction with other molecules. The mutant residue is smaller than the wild-type this will cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant differs. (Fig3-a/3-c) and (Fig3-b/3-d)

rs72549306 (V335L) change in the amino acid Valine into leucine at position 335. The mutant residue is bigger than the wild-type residue. The mutated residue is located in a domain that is important for the activity of the protein and in contact with another domain that is also affect protein function. The wild-type residue was buried in the core of the protein, The mutant residue is bigger and probably will not fit. (Fig4-a/4-c) and (Fig4-b/4-d)

Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties. In addition the effect of mutant residue on the binding site between different substrates or cofactors and protein also affect in the stability of the protein.

The mutant site for protein plays an important role in how this mutation affects the protein stability. mutant residue can located on the surface of the protein so mutation can disturb interactions with other parts of the protein, or located in a domain that is important for binding of other molecules, like M166V SNP, and important for activity of the protein and contact with residues in another domain like I560S SNP. Also the wild-type residue can be located in preferred secondary structure of protein, a β -strand. The mutant residue prefers to be in another secondary structure, therefore the local conformation will be slightly destabilized Like V335L SNP.

When mutant residue is bigger than the wild-type residue, makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did, like I560S SNP, this mutated residue probably will not fit in the core of the protein. this affect in the stability of the protein too. Sometimes mutant residue is smaller than the wild-type residue like M166V and D949V SNPs, thus mutant will cause an empty space in the core of the protein and causing a possible loss of external interactions.

The hydrophobicity of the Wild-type and mutant residue can differs like D949V, The charge of the wild-type may lost by mutation, this can cause loss of interaction with other molecules and this happens when There is a difference in charge between the wild-type and mutant amino acid. In return when the mutant residue charge is negative and wild-type residue was neutral, The wild-type residue then more hydrophobic than the mutant residue because the new amino acid charge is negative instead of neutral, like I560S SNP.

Table 1: Predicted Results by SIFT, PolyPhen, and SNAP& GO

SNPs RS Number	Amino acids change	SIFT Result	POLYPHEN Result	SNP& GO Result	SIFT Score	PLOYPHEN Score	SNP & GO Score
rs367619008	K63E	DELETERIOUS	probably damaging	Disease	0	1	0.657
rs375990187	I97T	DELETERIOUS	probably damaging	Disease	0.003	1	0.596
rs370615432	K114N	DELETERIOUS	probably damaging	Disease	0.002	1	0.506

rs72549308	S201R	DELETERIOUS	probably damaging	Disease	0.007	0.999	0.724
rs72549307	Y211C	DELETERIOUS	probably damaging	Disease	0.026	0.999	0.689
rs1801266	R235W	DELETERIOUS	probably damaging	Disease	0.003	1	0.654
rs150437414	L310S	DELETERIOUS	probably damaging	Disease	0.009	1	0.689
rs143154602	R353C	DELETERIOUS	probably damaging	Disease	0.001	1	0.815
rs72549304	S492L	DELETERIOUS	probably damaging	Disease	0.004	1	0.511
rs112766203	T760I	DELETERIOUS	probably damaging	Disease	0.004	1	0.563
rs200643089	L775W	DELETERIOUS	probably damaging	Disease	0.022	0.987	0.657
rs374825099	A777D	DELETERIOUS	probably damaging	Disease	0.005	0.999	0.897
rs267598785	P789S	DELETERIOUS	probably damaging	Disease	0.03	1	0.601
rs147800606	G795R	DELETERIOUS	probably damaging	Disease	0	0.998	0.83
rs199777072	D829N	DELETERIOUS	probably damaging	Disease	0.004		0.637
rs372307932	I948N	DELETERIOUS	probably damaging	Disease	0.01	0.968	0.693
rs67376798	D949V	DELETERIOUS	probably damaging	Disease	0.004	0.827	0.839
rs141044036	K958E	DELETERIOUS	probably damaging	Disease	0.001	0.998	0.753
rs201268750	H978Q	DELETERIOUS	probably damaging	Disease	0.039	0.923	0.541
rs1801268	V995F	DELETERIOUS	probably damaging	Disease	0.011	0.591	0.726

Table 2: Prediction result of I-Mutant software

SNP ID	Amino acid Position	WT	MT	pH	Temp (°C)	Stability	RI
rs2297595	166	M	V	7.0	25	Decrease	7
rs55886062	560	I	S	7.0	25	Decrease	9
rs67376798	949	D	V	7.0	25	Decrease	1
rs72549306	335	V	L	7.0	25	Decrease	8

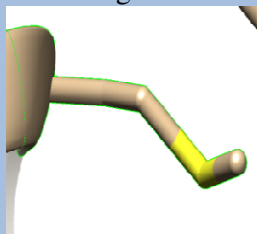
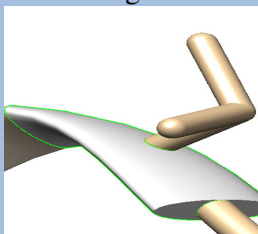
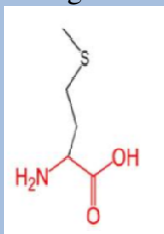
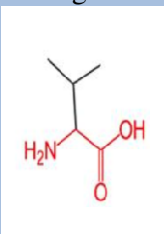
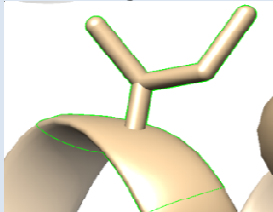
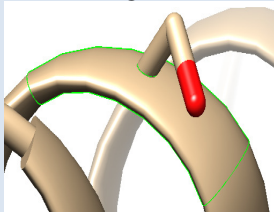
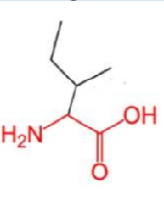
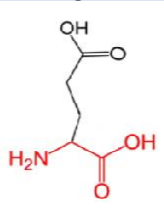
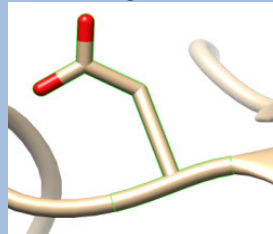
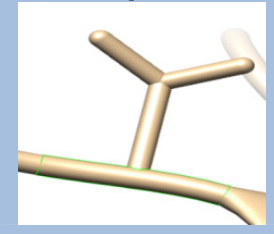
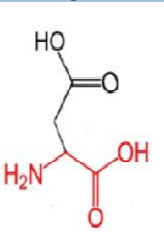
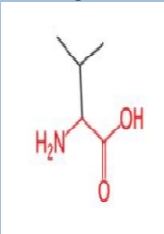
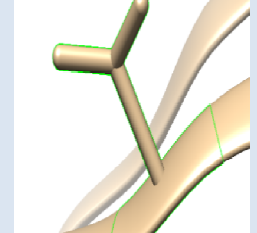
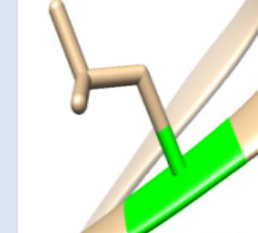
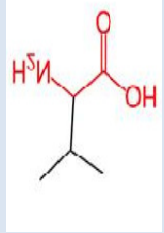
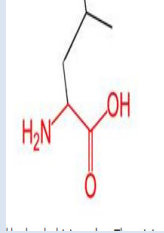
SNP ID	CHIMERA ORIGINAL	CHIMERA MUTANT	HOPE ORIGINAL	HOPE MUTANT
rs2297595 (M166V)	Fig1-a 	Fig1-b 	Fig1-c 	Fig1-d 
rs55886062 (I560S)	Fig2-a 	Fig2-b 	Fig2-c 	Fig2-d 
rs67376798 (D949V)	Fig3-a 	Fig3-b 	Fig3-c 	Fig3-d 
rs72549306 (V335L)	Fig4-a 	Fig4-b 		

Figure1-4: Schematic structures of the original (a) and the mutant (b) amino acid by Chimera and HOPE

4. Conclusion :

The current study shows the in silico analysis of genetic single variants within the coding human DPYD gene. These polymorphisms could influence the intermolecular and intramolecular interactions of amino acid residues and affect the stability of the protein result in loss of external interactions. By analyzing the conformational changes and interactions of amino acid residues within DPD proteins, we have identified significant structural and functional changes that can explain the activity deviations, caused by several mutations ((M166V, V335L, I560S, D949V). Finally we hope that our comprehensive investigation provide a guideline to set up a portecol for DPYD variants Screening before the administration of 5-FU to identify the patients how suspected to develop an adverse 5-FU drug reaction or toxicity toward reducing the early severe life-threatening toxicities associated with 5-FU and reduce the incidence of toxic death due to this drug.

REFERENCES

References

1. Wei X, Elizondo G, Sapone A, McLeod HL, Raunio H, FernandezSalguero P, et al. Characterization of the human dihydropyrimidine dehydrogenase gene. *Genomics* 1998; 51: 391-400.
2. Diasio RB. The role of dihydropyrimidine dehydrogenase (DPD) modulation in 5-FU pharmacology. *Oncology (Williston Park, NY)*. 1998;12(10 Suppl 7):23-7.
3. Longley DB, Harkin DP, Johnston PG. ('5-fluorouracil: mechanisms of action and clinical strategies).', *Nat Rev Cancer*, may,2003.5(3), p. 330-8.
4. Van den Bosch, B., Evolution of Dihydropyrimidine Dehydrogenase (DPD) Diagnostics in A Single Center in A Time-Period of Seven Years. *Clinical Therapeutics*, 2017. 39(8): p. 11.
5. Del Re, M., Dihydropyrimidine dehydrogenase polymorphisms and fluoropyrimidine toxicity: ready for routine clinical application within personalized medicine? *EPMA Journal*, 2010. 1(3): p. 495-502.
6. Salonga, D., Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clinical Cancer Research*, 2000. 6(4): p. 1322-1327.
7. Wei, X., et al., Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *Journal of Clinical Investigation*, 1996. 98(3): p. 610.
8. Raida, M., et al., Prevalence of a common point mutation in the dihydropyrimidine dehydrogenase (DPD) gene within the 5'-splice donor site of intron 14 in patients with severe 5-fluorouracil (5-FU)-related toxicity compared with controls. *Clinical Cancer Research*, 2001. 7(9): p. 2832-2839.
9. Van Kuilenburg, A., et al., Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Human genetics*, 1999. 104(1): p. 1-9.
10. Van Kuilenburg, A.B., et al., Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. *Biochemical Journal*, 2002. 364(1): p. 157-163.
11. Andre, T., Colin, P., Louvet, C. et al. Semimonthly versus monthly regimen of fluorouracil and leucovorin administered for 24 or 36 weeks as adjuvant therapy in stage II and III colon cancer: results of a randomized trial. *J Clin Oncol*. 2003; 21: 2896–2903
12. Parker WB, Cheng YC 'Metabolism and mechanism of action of 5-fluorouracil', *Pharmacol. Ther.*, 1990;48(3), p. 381-95.
13. Morel, A., Boisdron-Celle, M., Fey, L., Soulié, P., Craipeau, M.C., and Gamelin, E. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms (SNP) upon 5-fluorouracil tolerance in a prospective clinical study in a french caucasian population. *Mol Cancer Ther*. 2006; 5: 2895–2904.
14. Offer SM, Wegner NJ, Fossum C, Wang K, Diasio RB. Phenotypic profiling of DPYD variations relevant to 5 fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer Res* 2013;73: 58–68.
15. Wishart DS. Bioinformatics in drug development and assessment. *Drug metabolism reviews*. 2005;37(2):279-310.
16. Alanazi M., Abduljaleel Z., Khan W., et al. In silico analysis of single nucleotide polymorphism (SNPs) in human β -globin gene. *PLoS ONE*. 2011;6(10)e2587.
17. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research*. 2001;11(5):863-74.

295 18. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids*
296 *Research*. 2003;31(13):3812-4.
297 19. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS,
298 Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat*
299 *Methods*. 2010;7:248–249,
300 20. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the
301 predictive score of human disease-related mutations in proteins. *Hum Mutat*. 2009;30(8):1237–1244.
302 21. Daggett V, Fersht AR. Is there a unifying mechanism for protein folding? *Trends in biochemical*
303 *sciences*. 2003;28(1):18-25
304 22. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from
305 sequences using support vector machines. *Bioinformatics (Oxford, England)*. 2005;21 Suppl 2:ii54-8.
306 23. Smigielski, E.M., et al., dbSNP: a database of single nucleotide polymorphisms. *Nucleic acids*
307 *research*, 2000. 28(1): p. 352-355.
308 24. Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels-3.0--remote homology modeling
309 using structure-guided sequence profiles. *Nucleic Acids Research*. 2010;38(Web Server issue):W576-
310 81.