Structure based inference of functional single nucleotide polymorphism and its role in TGFβ1 allied colorectal cancer (CRC)

Ankita Shukla and Tiratha Raj Singh*

Department of Biotechnology and Bioinformatics,
Jaypee University of Information Technology (JUIT),
Waknaghat, Solan, H.P., 173234, India
Email: shukla.ankita39@gmail.com
Email: tiratharaj@gmail.com
Website: http://www.juit.ac.in/Department/bio/PhDScholarsBio
Website: http://www.bioinfoindia.org/raj
*Corresponding author

Abstract: *Motivation*: Till date vast varieties of studies have given major attention to TGFβR1 and TGFβR2 receptors in colorectal cancer (CRC), however TGFβ1 remains to be poorly understood. It is still a major challenge to identify the functional SNPs in a CRC-related TGFβ1 gene. *Method*: In this study, total 136 mutations were retrieved for TGFβ1 out of which non-synonymous 37 mutations were considered. Initially sequence and structure based tools were used for damage prediction. The mutations that were predicted to be damaging by majority of the tools were then considered for the structure dynamics study. *Result*: In this paper we targeted only one mutation type, i.e., L28F to evaluate its effect on disease. Structure conservation studies have been performed to infer the effect of the mutation at the region with respect to its conservation profile. The study depicts the changes occurring to the overall structure due to a single amino acid variation (i.e. L28F).

Keywords: colorectal cancer; carcinogenesis; molecular dynamics; polymorphism.

Reference to this paper should be made as follows: Shukla, A. and Singh, T.R. (2021) 'Structure based inference of functional single nucleotide polymorphism and its role in TGFβ1 allied colorectal cancer (CRC)', *Int. J. Bioinformatics Research and Applications*, Vol. 17, No. 1, pp.80–99.

Biographical notes: Ankita Shukla is currently pursuing her PhD from Jaypee University of Information Technology, Waknaghat, Solan, India. She did her Master's degree from Panjab University, Chandigarh. Her area of interest is Systems biology, bioinformatics, structural biology and machine learning. She is working on DNA repair mechanisms and their role in associated malignancies.

Tiratha Raj Singh is currently working as an Associate Professor in Department of Biotechnology and Bioinformatics in Jaypee University of Information Technology, Solan, India. He did his PhD in Bioinformatics from MANIT, Bhopal and joint research work for PhD was also done from Tel-Aviv University, Tel-Aviv, Israel followed by post doctoral studies from TAU, Israel. He is working on various aspects of Genomics, and Proteomics towards their involvement in various human disease and also interested and working on the annotation of complex biological networks through

computational systems biology approach. More specifically, he is working on DNA repair mechanisms and their role in malignancies. He is also exploring Alzheimer's Disease (AD) through myriad of computational approaches.

This paper is a revised and expanded version of a paper entitled 'Structure based inference of functional single nucleotide polymorphism 'L28F' and to determine its role in TGFβ1 allied colorectal cancer (CRC)' presented at *Inbix' 2017*, BISR, Jaipur, India, 9 November, 2017.

1 Introduction

Colorectal cancer (CRC) sometime also called as the cancer of bowel or the colon; it initiates from the colon or sometimes from rectum (Haggar and Boushey, 2009). It is the third one majorly diagnosed malignancy and the fourth leading cause of cancer related deaths worldwide. It is expected to increase by the estimate of about 2.2 million incidences and 1.1 million cancer allied deaths by the year 2030 (Arnold et al., 2016). Mortality rate is prominent in the developed countries and risk is relatively higher in men than in women (Tariq and Ghias, 2016; Hisamuddin and Yang, 2006). There are many other factors also that can enhance the risk of cancer like chemical, environmental, lifestyle, etc., which are responsible for the CRC initiation and its metastasis (Haggar and Boushey, 2009; Grady and Markowitz, 2002).

Amongst the above mentioned factors DNA repair mechanisms play important role in maintaining inner biochemistry of the cells. It is the major player that take care of the damages that happens inside the cells, but if this mechanism becomes faulty it could lead to the major damages that remains unrepaired (Markowitz and Bertagnolli, 2009; Gavande et al., 2016; Dietlein et al., 2014). According to studies mismatch repair (MMR) mechanism act as a major player in regulating CRC (Peltomaki, 2001). Majority of the cancers were reported to be resistant to the growth inhibitory effects of TGFβ1 (Elliott and Blobe, 2005; Singh et al., 2007; Cui et al., 1996; Park et al., 2000). TGFβ1 induces and regulates apoptosis, through the Smad mediated pathway (Yamamura et al., 2000). However, in case of the Smad-independent pathway, Ras/Raf mediated mitogen-activated protein kinases (MAPK) pathway (Fink et al., 2001; Hanafusa et al., 1999) drives the proliferation of human colon and as well as that of prostate cancer cells (Park et al., 2000; Yan et al., 2001). Specifically wnt and TGFβ pathways are the major suppressors of the colon (Takaku et al., 1998), and pancreatic (Cullingworth et al., 2002) cancers, which they either execute individually or in a coordinated manner.

Transforming growth factor-beta (TGFβ) mediates several effects such as growth inhibition, apoptosis, and epithelial to mesenchymal transition (EMT) process, that augments cell migration and invasion (Akhurst and Hata, 2012). TGF-β1 is encoded on chromosome 19q; and is a precursor protein containing 391 amino acids. Out of which the C-terminal 112 amino acids constitute the mature protein having molecular weight of 25kDa (Rhyu et al., 2005). Inhibition of TGF-β1 signalling pathway has been reported to prevent progression and metastasis of certain advanced tumours (Ono et al., 2012). TGF-β1 is known to have a strong immunosuppressive effect including those on the tumour cells. In our study the TGFβ1 gene, A to G transition at amino acid number 28 changes leucine to phenylalanine and are termed Leu28Phe (L28F) mutation (Table 1) have been studied.

Table 1 Genotypic information for the studied SNP

SNP ID	*Mutation type*	*SNP region*	*AA change*	*Observed allele change*
rs199946261	Missense	Exon6	L28F	A/G

TGF-β1 has an important role in development, tissue repair, immune defense inflammation and tumorigenesis (Skeen et al., 2012). The studies have found to have involvement of the TGF-β1 in CRC in modulating the degree of angiogenesis (Ferrari et al., 2009). Drug resistance (oxaliplatin) has been noticed for the TGF-β1 via the promotion of EMT, cancer stem cell-like properties, crosstalk with interleukin 6, regulating mismatch repair system inducing cell cycle arrest and autophagy in several cancers (Asiedu et al., 2011). This shows that blocking the TGF-β1 pathway may enhance the efficacy of chemotherapy under certain circumstances. Pharmacological inhibition of TGFβ could be used as a therapeutic strategy to hinder tumour progression, improve drug delivery and efficacy of the treatment. The TGFβ targets (e.g., 1D11, AP12009, SD-208) as well as TGFβ inhibitory drugs (e.g., tranilast) have shown to reduce tumour progression and metastasis in vivo, mainly owing to augmentation of the immune response (Papageorgis and Stylianopoulos, 2015). It is anticipated that the studied mutational change in TGFβ will provide additional opportunity to scrutinise other agents for the therapeutic interaction of the CRC.

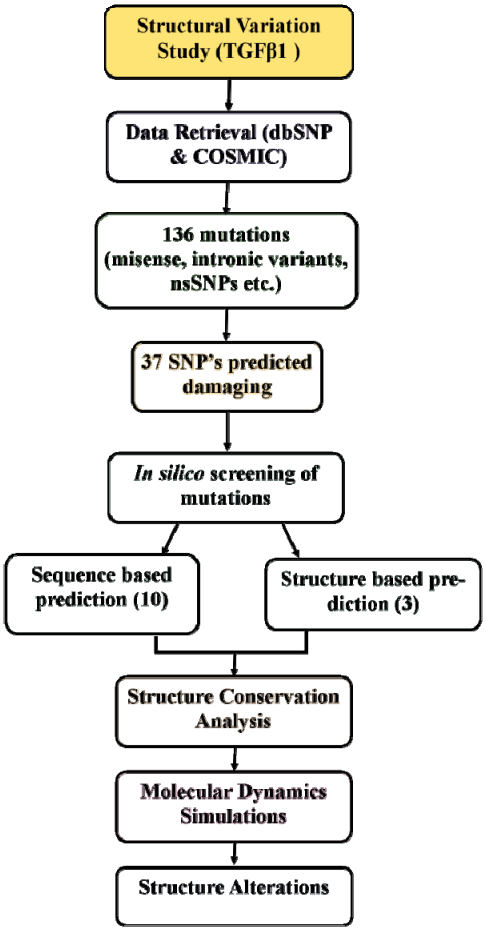
Several evidences support the significant role of TGFβ1, TGFβR1, and TGFβR2 in signalling which in case of deletion can cause angiogenesis (Dickson et al., 1995; Oshima et al., 1996; Larsson et al., 2001). In normal condition TGFβ1 hamper the propagation of normal intestinal epithelial cells (Kurokowa et al., 1987), and control the proliferation and differentiation of normal colonic epithelium (Avery et al., 1993). TGFβ1 is known to be the most abundant and ubiquitously expressed isoform of TGFβ (Xu and Pasche, 2007). Therefore, the study of TGFβ1 could prove to be clinically beneficial for diagnosis of tumours. Based upon the established and essential role of TGFβ1 we targeted our study in the site specific mutation analysis. The pipeline of the analysis performed for the structural studies are mentioned in Figure 1.

2.1 SNP's collection and damaging ns-SNPs prediction

Data collection for the SNPs was done using dbSNP (Sherry et al., 2001) and COSMIC (Forbes et al., 2011) databases using keyword 'TGFβ1' and disease 'colorectal cancer'. This in total resulted in 136 mutations including misense, intronic variants, and non-synonymous (ns-SNPs). By considering the focus only on ns-SNPs, the data has left only to 37, final analysis were performed only on these ns-SNPs. The nsSNP mutations were predicted to be damaging or not using an array of sequence based (PolyPhen (Adzhubei et al., 2013), I-Mutant Suite (Capriotti et al., 2005), PROVEAN (Choi and Chan, 2015), MutPred (Li et al., 2009), SNP & GO (Calabrese et al., 2009), PredictSNP (Bendl et al., 2014), MAPP (Stone and Sidow, 2005), SNAP (Bromberg and Rost, 2007), SIFT (Sim et al., 2012), Mutation Accessor (Reva et al., 2011)) (Supplementary Table 1) and structure based (SNP Effect 4.0 (De Baets et al., 2012), Eris (Yin et al., 2007), SNP&GO3D (Capriotti et al., 2013)) SNP damage prediction tools (Supplementary Table 2). Out of these 37 ns-SNPs, we considered L28F for the analysis as it was found to be on the interface and therefore thought to be the significant ns-SNP that makes the

higher chances of altering the cavity site which could therefore hinder the binding of the ligand. Also, this SNP were predicted to be damaging by maximum of the tools, and therefore taken further for structural analysis and simulations.

Figure 1 Flowchart for the structure analysis (see online version for colours)



2 Materials and methods

2.2 Structure selection for TGFβ1

For performing structure level analysis, the structure search was done using the keyword 'TGFβ1' in Protein DataBank (PDB) (Sussman et al., 1998), the search resulted in a five structures, i.e., three NMR spectroscopy (1KLA, 1KLC, 1KLD) and two X-ray crystallography structures at a resolution of 3Å (3KFD, 4KV5). Now the question arouse that what structure could be used for the analysis, therefore we decided to make use of the Theseus-3D tool (Theobald and Wuttke, 2008). Theseus-3D is a maximum likelihood (ML) method that performs the PCA (principal components analysis) for examining the

complex correlations among atoms within a structural ensemble. From this 4KV5 was found to be most suitable as it comes out be median structure among all. Therefore, its PDB structure was obtained from the PDB with both A and B chains as the structure was homodimer. PyMOL's mutagenesis tool was then used to create the mutations on the structure (DeLano, 2002), where the rotamers for each mutant amino acid with least steric clashes were selected.

2.3 Molecular dynamics (MD)

Molecular dynamics helps to evaluate the structural stability of the native and mutant proteins. In our study molecular dynamics simulations were performed using GROMACS package v5.1.2 package (Abraham et al., 2015). The dynamics was run using optimised potential for liquid simulations (OPLS) force field on all-atom (Siu et al., 2012). The MD simulation was executed for the L28F variant including the native one for the time period of 100 ns each. Solvation (aiding the water molecules) was done using water model TIP3P inside a triclinic box (Jorgensen et al., 1983). The counter ions were added using the genion tool so as to replace the solvent molecules by monoatomic ions (i.e., Cl), four chloride ions were added each for L28F, and for the wild protein with most favourable electrostatic potential. Particle-based method was used to calculate the electrostatic potential on all atoms. After then structure was relaxed through energy minimisation steps of 50,000 which is then followed by solvent equilibration and the equilibration of the system for 100 ps. 100 ns run of production step was performed for 100 ns using 100 ps time step for the integration of the equation of motion in the NPT at 1 atmospheric pressure. The MD simulation coordinates were saved at 100 ps time period for carrying further analysis. Further processing and analysis were carried out using GROMACS analysis tools and trajectories were visualised using PyMOL (DeLano, 2002) and VMD (Humphrey et al., 1996) and plots were plotted using gnuplot (Williams et al., 2012).

2.4 Determining polar and non-polar Interactions

The polar and non-polar interactions play a key role in determining the structural impact on the protein. For determining the interaction types we have used the Ligplot; it is a program that make use of a standard PDB file that generates schematic 2D representations of protein-ligand complexes (Wallace et al., 1995). The program is helps to show instructive representation regarding the intermolecular interactions like hydrogen bonds, hydrophobic interactions etc. The polar interactions obtained through Ligplot were cross validated using the H-bond command of the GROMACS, and the interactions that were actually existed have shown in Table 2.

Table 2 STRIDE residue conformations

	Phi-angle (Φ)	*Psi-angle* (Ψ)	*Solvent accessible surface area (SASA)*
Leu28	–109.84	–18.68	6.8
Phe28	–119.06	–32.42	37.2

3 Results and discussion

3.1 Structural conservation studies

CONSURF is a server that has been widely used for determining the functional regions in proteins (Ashkenazy et al., 2016). It uses empirical Bayesian inference to calculate the evolutionary conservation, in sequence as well as the structure of the proteins and nucleic acids. Generally the degree to which an amino acid position is conserved strongly depends on its structural and functional importance. In our study, when structural conservation was accessed for the mutant (Leu28) form of the protein; it attained high level of conservation score (Figure 2(a)). It suggests the significance level of the residue for the structural and functional impact. Similar procedure was applied to the mutant one (Phe28), this also shown to have good conservation score (lightpink colour) (Figure 2(b)). Therefore the mutation in this region is not favourably accepted as it may lead to the damaging effect.

Figure 2 CONSURF results (see online version for colours)

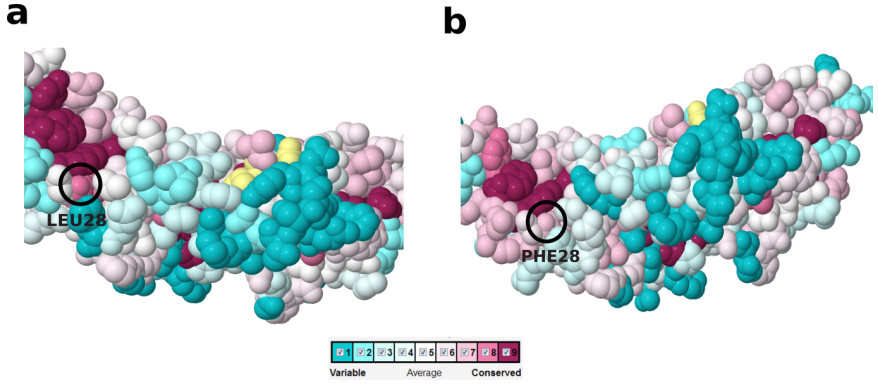
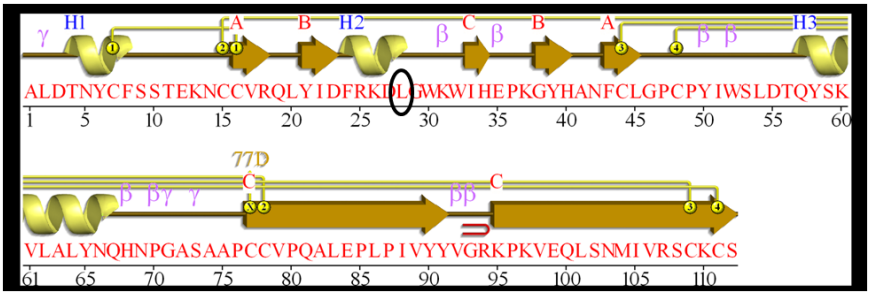


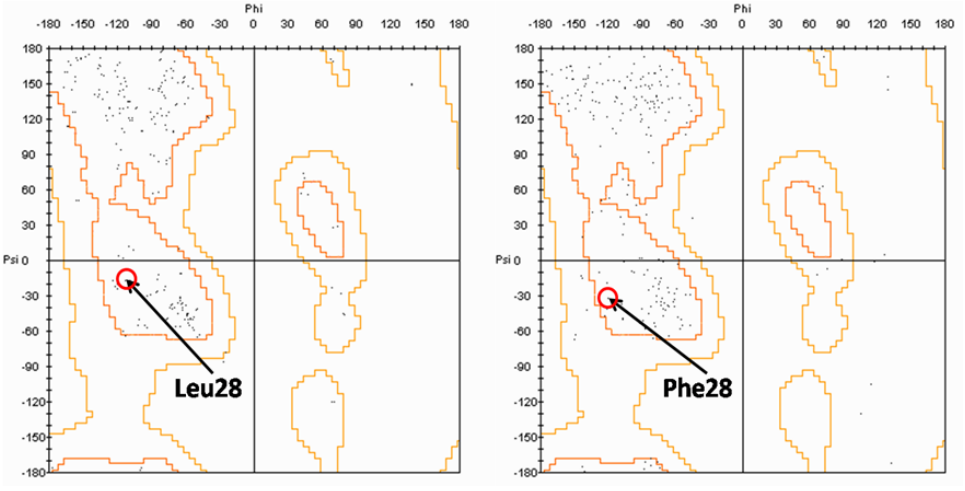
Figure 3 Secondary structure prediction through PDBsum. Prediction showing the position number 28 tends to be in a helix before the mutation inclusion (see online version for colours)



To determine the position of the Leu28 at the structure level, structure prediction was captured from the PDBSum (Laskowski et al., 2017). Through PDBSum it was found Leu28 lies at helix conformation before mutation (Figure 3). For confirming these results we have used STRIDE web server. STRIDE (Heinig and Frishman, 2004) provide the secondary structure assignment for the given structure. Figure 4 have shown the

position of the Leu28 (present in α-helix region) and Phe28 position (present in turn region), both of them fall in allowed region it has also shown the phi-psi angles along with the solvent accessible surface area (SASA) (Table 2).

Figure 4 Ramachandran Plot signifies the position of Leu28 and the mutated Phe28 (see online version for colours)



3.2 MD analysis

Through the MD simulation it is possible to observe and predict the structural effects occurring due to the mutational events that could be proved to provide possible insights in CRC. The MD simulation was completed for 100 ns time period for the each native and mutant structure. Both the global and local factors were analysed from the molecular dynamics studies. Globally, the root mean square deviation (RMSD) (Figure 5), root mean square fluctuation (RMSF) (Figure 6), radius of gyration (Rg) (Figure 7), solubility accessible surface area (SASA) (Figure 8), were analysed to check the stability of MD simulation and to determine when does the time of equilibration attained.

Significant changes were observed within the native and the mutant form of the protein 4KV5. It was observed that the mutant L28F (RMSD of 3.1Å) was similar to the native's (final RMSD of 2.9Å) trajectory until 50 ns, however after 50 ns these mutants were observed to be different in comparison to the native trajectory that reflected the deviation of the structural coordinates from its actual position that could result into the structure alterations (specifically at secondary structure level). By looking at the individual residue's fluctuations throughout the 100 ns simulation, there is relatively higher residue fluctuation in the mutant in comparison to the native structure specifically for the Arg94, and Tyr50 residues (Figure 6).

The radius of gyration was found to be 2.12 nm for the mutant (L28F) which was almost equivalent to the wild (2.11 nm). Radius of gyration tells that a globular protein becomes a linear molecule if undergone denaturation. Therefore from the analysed radius of gyration data we found that the structure remains intact and globular throughout the simulation and thus confirms that the MD trajectories are stable (Figure 7).

Figure 5 Graph displaying the root mean square deviation (RMSD) (see online version for colours)

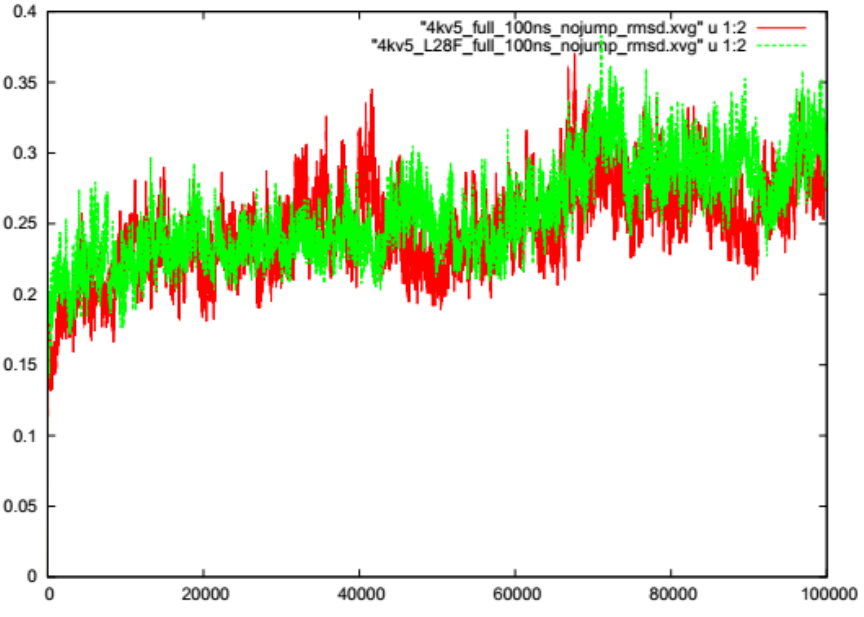


Figure 6 Graph displaying the root mean square fluctuation (RMSF) (see online version for colours)

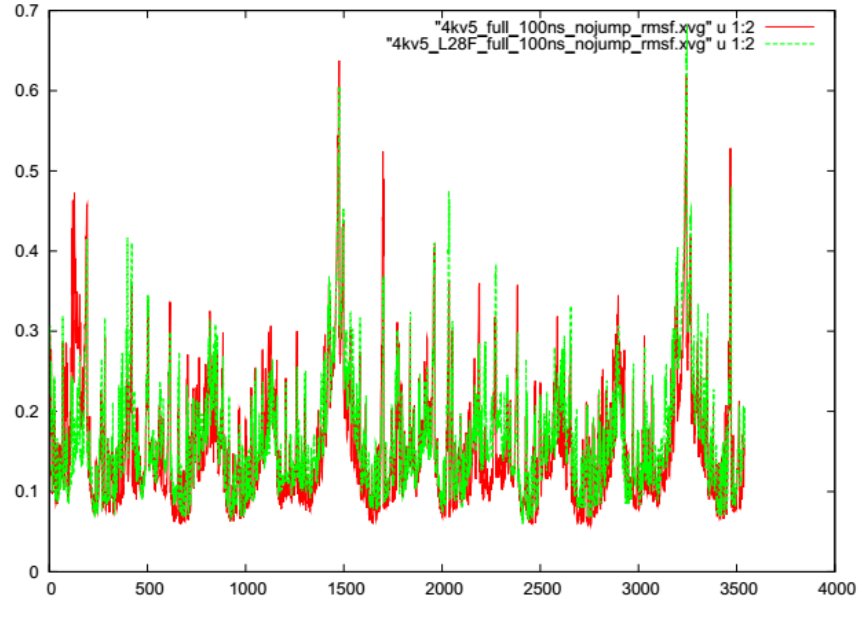


Figure 7 Graph displaying the radius of gyration (Rg) (see online version for colours)

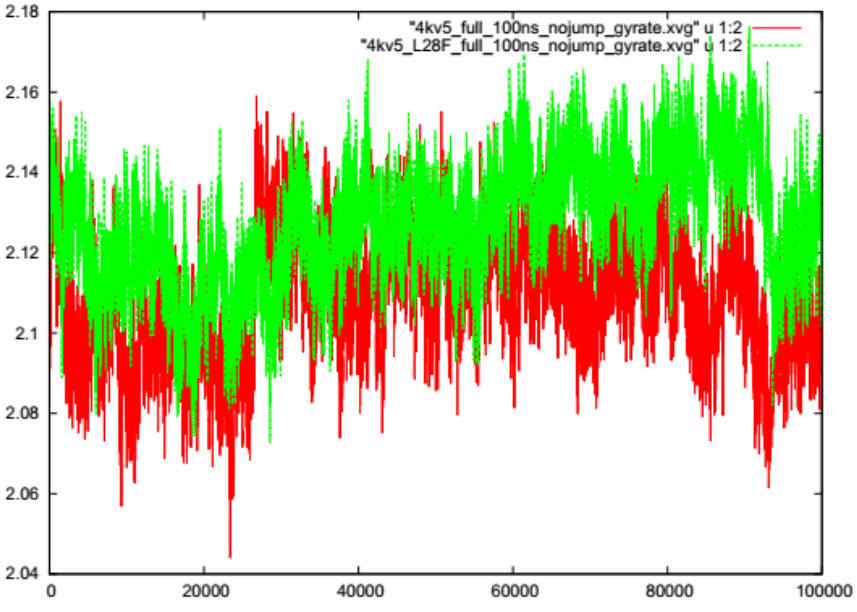
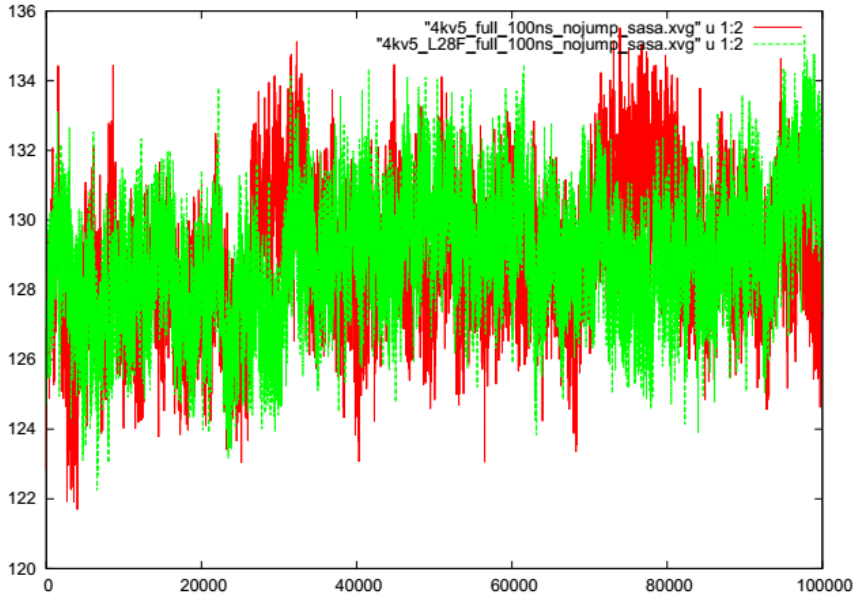


Figure 8 Graph displaying the solvent accessible surface area (SASA) (see online version for colours)



One of the key factor in measuring the protein stability is the solvent accessible surface area (SASA); that predicts the protein conformational changes when binds to its substrate or ligand (Marsh and Teichmann, 2011). Therefore from the SASA analysis not much exposure has been found for the L28F variant (Figure 8). The superposition of the first

timeframe (0 ns) to the last timeframe (100 ns) reflected the minor conformational changes in terms of secondary structure conformation, i.e., at the loop region, helix region and also at the beta-sheet region (Figure 9(a)). Also, residues around the 5Å regions of the Phe28 have been captured to find alterations in structure. As shown in Figure 9(b) significant structural differences has been found in the residues within 5Å radius.

Also we looked at the longevity of the hydrogen bonds that stabilise the dimer, and the new hydrogen bonds formed during the simulation. Along with the hydrogen bonds polar interactions and hydrophobic interactions were also studied through the Ligplot (Figure 10). Hydrogen bonds give information regarding most of the directional interactions that fortify protein folding, protein structure and molecular recognition. New inter-chain hydrogen bonds were found to be formed (Table 3) in the final snapshot of the simulation trajectory (Figure 11). Some bonds were formed at the beginning of the simulation and maintained throughout the 100 ns trajectory, while others were formed at the end of the simulation only. We hypothesise that former bonds are created due to the internal structural adjustments to compensate the entropic cost of mutating a residue and they are maintained to make the dimer structure stable. The bonds not present throughout the trajectory, but only in the last timeframe of the trajectory (at 100 ns) are transient bonds formed as result of the minor fluctuations and are not contributing to the stability of the mutant structure. Interestingly, these new bonds are not observed in the native structure at the end of its 100 ns simulation, thus emphasises the possibility of bonds created to stabilise the structure.

Also, trajectory visualisation has been performed for all 10,000 frames (100 ns) for the native and for mutant (L28F) using VMD (Humphrey et al., 1996). Figure 12 shown structural variations through localised RMSD differences, where the red regions shows high deviation, green regions with intermediate deviations, and blue regions is showing low deviation. We observed that no major changes are happening to the overall protein structure. However, in the secondary structure regions (i.e., loop region, helix region) changes were observed and such changes might cause conformational change and hence can lead to the functional alterations.

Our study is not limited to the single mutation study, in future we will come up with analysis of other damaging mutations particularly for TGFβ1 gene that could be targeted in colorectal cancer study.

Figure 9 Mapping the respective frames around 5Å region of the mutant residues (see online version for colours)

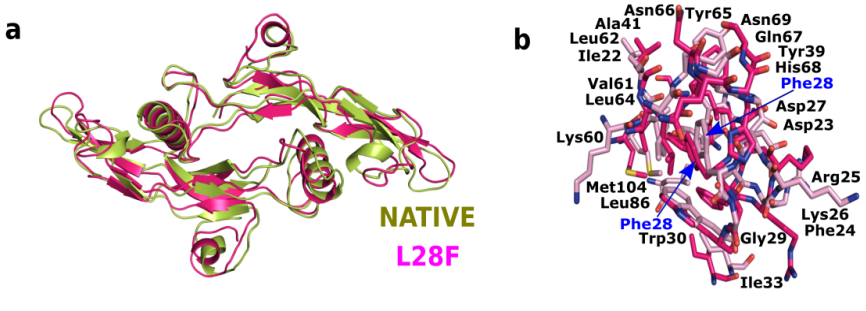
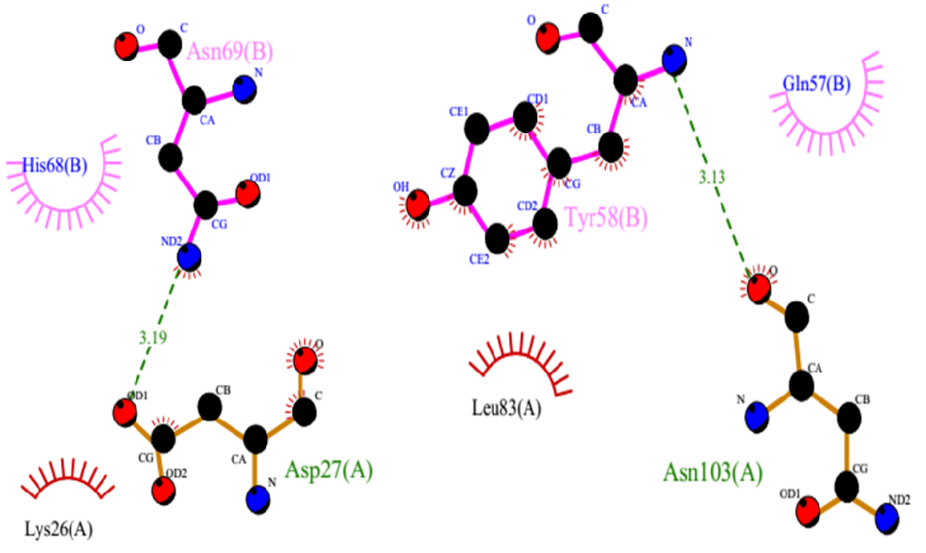
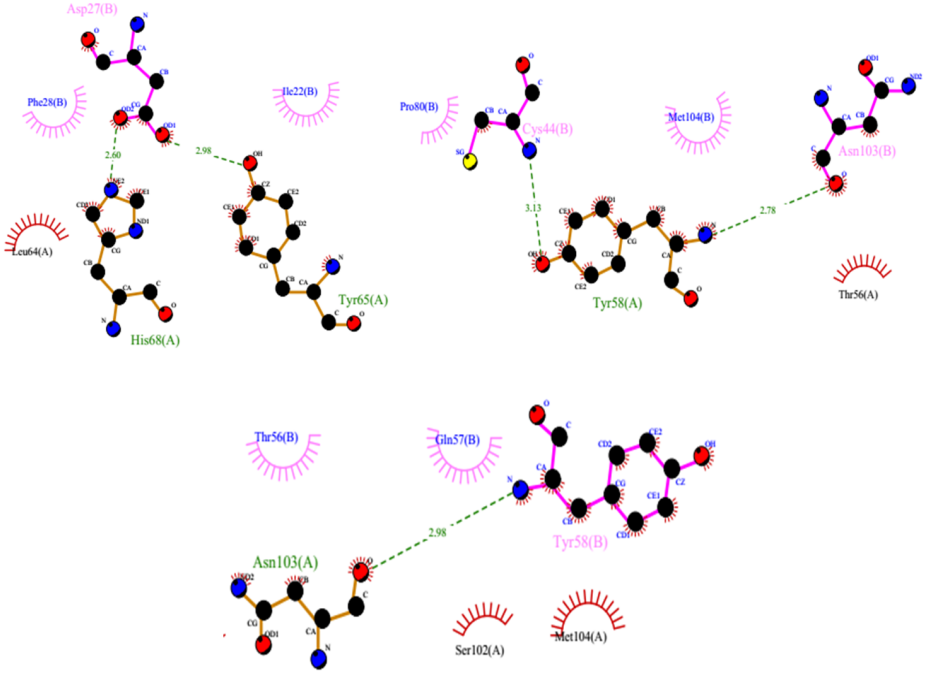


Figure 10 Ligplot interactions: (a) for the first frame of the mutant L28F (0 ns) and (b) for the last frame of the mutant L28F (100 ns) (see online version for colours)



(a)



(b)

Figure 11 Hydrogen – bond interactions with respect to time (see online version for colours)

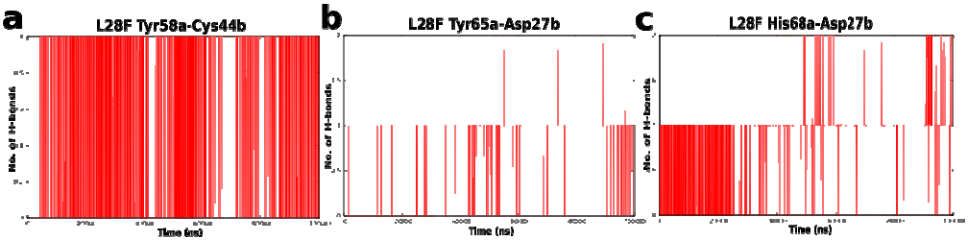


Figure 12 Superposed snapshots of 100 ns simulation of TGFβ1 dimer site mutants (see online version for colours)

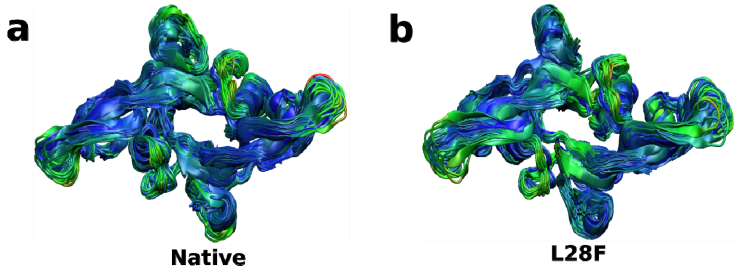


Table 3 Polar interactions obtained through the Ligplot

4kv5_Native_t0		*4kv5_Native_t100*		*4kv5_L28F_t0*		*4kv5_L28F_t100*	
ChainA-ChainB	*H-Bond*	*ChainA-ChainB*	*H-Bond*	*ChainA-ChainB*	*H-Bond*	*ChainA-ChainB*	*H-Bond*
Gln57-Ser102	2.98	Asp27-Asn69	2.74	Asp27-Asn69	3.19	His68-Asp27	2.6
Tyr58-Asn103	3.01	Asn103-Tyr58	2.87	Asn103-Tyr58	3.13	Tyr65-Asp27	2.98
Asn69-Asp27	2.83	Asn42-Tyr58	2.74			Tyr58-Cys44	3.13
Asp27-Asn69	3.08	Gln57-Ser102	3.31			Tyr58-Asn103	2.78
Asn103-Tyr58	2.82	Gln57-Asn103	2.99			Asn103-Tyr58	2.98
Asn42-Tyr58	3.03	Tyr58-Asn103	2.89				
		Tyr58-Asn42	2.84				

4 Conclusion

The study characterises the association of deleterious mutation resulting in colorectal cancer progression by interfering the functionality of the TGFβ1. In this study rs199946261 was identified to be a damaging SNP that can hinder the functionality of the involved protein. Using MD analysis, we observed that the changes occurring at structure level due to a single amino acid variation at position (L28F) can cause damage to the structure by altering its conformation and thus possibly inhibit its activity. Also, in future we will consider some other TGFβ1 related damaging SNPs to determine their damaging

effects at structural level. We hope that this computational analysis will provide broad view of the disease to the biological researchers so that effective methods can be developed for the disease eradication thus enhancing the survival rates. The results presented here will be helpful for experimental biologist to test the effect of SNP and its subsequent phenotypical responses.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) 'GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, Vol. 1, pp.19–25.

Adzhubei, I., Jordan, D.M. and Sunyaev, S.R. (2013) 'Predicting functional effect of human missense mutations using PolyPhen-2', *Curr. Protoc. Hum. Genet*., Chapter 7, p. Unit7 20.

Akhurst, R.J. and Hata, A. (2012) 'Targeting the TGFβ signalling pathway in disease', *Nature Reviews Drug Discovery*, Vol. 11, No. 10, p.790.

Arnold, M., Sierra, M.S., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2016) 'Global patterns and trends in colorectal cancer incidence and mortality', *Gut*, gutjnl-2015-310912.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules', *Nucleic Acids Res*., Vol. 44, pp.W344–50.

Asiedu, M.K., Ingle, J.N., Behrens, M.D., Radisky, D.C. and Knutson, K.L. (2011) 'TGFβ/TNFα-mediated epithelial–mesenchymal transition generates breast cancer stem cells with a claudin-low phenotype', *Cancer Research*, Vol. 71, pp.4707–4719.

Avery, A., Paraskeva, C., Hall, P., Flanders, K., Sporn, M. and Moorghen, M. (1993) 'TGF-beta expression in the human colon: differential immunostaining along crypt epithelium', *British Journal of Cancer*, Vol. 68, p.137.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J. and Damborsky, J. (2014) 'PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations', *PLoS Computational Biology*, Vol. 10, p.e1003440.

Bromberg, Y. and Rost, B. (2007) 'SNAP: predict effect of non-synonymous polymorphisms on function', *Nucleic Acids Research*, Vol. 35, No. 11, pp.3823–3835.

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P.L. and Casadio, R. (2009) 'Functional annotations improve the predictive score of human disease-related mutations in proteins', *Hum. Mutat*., Vol. 30, pp.1237–1244.

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P.L., Altman, R.B. and Casadio, R. (2013) 'WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation', *BMC Genomics*, Vol. 14, p.S6.

Capriotti, E., Fariselli, P. and Casadio, R. (2005) 'I-mutant2.and 0: predicting stability changes upon mutation from the protein sequence or structure', *Nucleic Acids Research*, Vol. 33, Suppl. 2, pp.W306–W310.

Choi, Y. and Chan, A.P. (2015) 'Provean web server: a tool to predict the functional effect of amino acid substitutions and indels', *Bioinformatics*, p. btv195.

Cui, W., Fowlis, D. J., Bryson, S., Duffie, E., Ireland, H., Balmain, A. and Akhurst, R.J. (1996) 'TGFbeta1 inhibits the formation of benign skin tumors, but enhances progression to invasive spindle carcinomas in transgenic mice', *Cell*, Vol. 86, pp.531–542.

Cullingworth, J., Hooper, M.L., Harrison, D.J., Mason, J.O., Sirard, C., Patek, C.E. and Clarke, A.R. (2002) 'Carcinogen-induced pancreatic lesions in the mouse: effect of Smad4 and Apc genotypes', *Oncogene*, Vol. 21, p.4696.

De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., Schymkowitz, J. and Rousseau, F. (2012) 'SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants', *Nucleic Acids Research*, Vol. 40, pp.D935–D939.

DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*.

Dickson, M.C., Martin, J.S., Cousins, F.M., Kulkarni, A.B., Karlsson, S. and Akhurst, R.J. (1995) 'Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice', *Development*, Vol. 121, pp.1845–1854.

Dietlein, F., Thelen, L. and Reinhardt, H.C. (2014) 'Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches', *Trends in Genetics*, Vol. 30, No. 8, pp.326–339.

Elliott, R.L. and Blobe, G.C. (2005) 'Role of transforming growth factor beta in human cancer', *J. Clin. Oncol.*, Vol. 23, No. 9, pp.2078–2093.

Ferrari, G., Cook, B.D., Terushkin, V., Pintucci, G. and Mignatti, P. (2009) 'Transforming growth factor-beta 1 (TGF-β1) induces angiogenesis through vascular endothelial growth factor (VEGF)-mediated apoptosis', *Journal of Cellular Physiology*, Vol. 219, pp.449–458.

Fink, S.P., Swinler, S.E., Lutterbaugh, J.D., Massagué, J., Thiagalingam, S., Kinzler, K.W., Vogelstein, B., Willson, J.K. and Markowitz, S. (2001) 'Transforming growth factor-β-induced growth inhibition in a Smad4 mutant colon adenoma cell line', *Cancer Research*, Vol. 61, pp.256–260.

Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K. and Menzies, A. (2011) 'COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer', *Nucleic Acids Research*, Vol. 39, pp.D945–D950.

Gavande, N.S., Vandervere-Carozza, P.S., Hinshaw, H.D., Jalal, S.I., Sears, C.R., Pawelczak, K.S. and Turchi, J.J. (2016) 'DNA repair targeted therapy: the past or future of cancer treatment?', *Pharmacol. Ther.*, Vol. 160, pp.65–83.

Grady, W.M. and Markowitz, S.D. (2002) 'Genetic, and epigenetic alterations in colon cancer', *Annual Review of Genomics and Human Genetics*, Vol. 3, No. 1, pp.pp.101–128.

Haggar, F.A. and Boushey, R.P. (2009) 'Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors', *Clin. Colon. Rectal. Surg.*, Vol. 22, No. 4, pp.p.191–197.

Hanafusa, H., Ninomiya-Tsuji, J., Masuyama, N., Nishita, M., Fujisawa, J-I., Shibuya, H., Matsumoto, K. and Nishida, E. (1999) 'Involvement of the p38 mitogen-activated protein kinase pathway in transforming growth factor-β-induced gene expression', *Journal of Biological Chemistry*, Vol. 274, pp.27161–27167.

Heinig, M. and Frishman, D. (2004) 'Stride: a web server for secondary structure assignment from known atomic coordinates of proteins', *Nucleic Acids Res*, Vol. 32, Web Server issue, pp.W500–W502.

Hisamuddin, I.M. and Yang, V.W. (2006) 'Molecular genetics of colorectal cancer: an overview', *Current Colorectal Cancer Reports*, Vol. 2, No. 2, pp.53–59.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', *J. Mol. Graph*, Vol. 14, No. 1, pp.33–38, 27–28.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) 'Comparison of simple potential functions for simulating liquid water', *The Journal of Chemical Physics*, Vol. 79, pp.926–935.

Kurokowa, M., Lynch, K. and Podolsky, D.K. (1987) 'Effects of growth factors on an intestinal epithelial cell line: transforming growth factor β inhibits proliferation and stimulates differentiation', *Biochemical and Biophysical Research Communications*, Vol. 142, No. 3, pp.775–782.

Larsson, J., Goumans, M.J., Sjöstrand, L.J., Van Rooijen, M.A., Ward, D., Levéen, P., Xu, X., Ten Dijke, P., Mummery, C.L. and Karlsson, S. (2001) 'Abnormal angiogenesis but intact hematopoietic potential in TGF-β type I receptor-deficient mice', *The EMBO Journal*, Vol. 20, pp.1663–1673.

Laskowski, R.A., Jabonska, J., Pravda, L., Varekova, R.S. and Thornton, J.M. (2017) 'PDBsum: structural summaries of PDB entries', *Protein Sci*.

Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D. N., Mooney, S.D. and Radivojac, P. (2009) 'Automated inference of molecular mechanisms of disease from amino acid substitutions', *Bioinformatics*, Vol. 25, pp.2744–2750.

Markowitz, S.D. and Bertagnolli, M.M. (2009) 'Molecular origins of cancer: molecular basis of colorectal cancer', *N. Engl. J., Med.*, Vol. 361, No. 25, pp.2449–2460.

Marsh, J.A. and Teichmann, S.A. (2011) 'Relative solvent accessible surface area predicts protein conformational changes upon binding', *Structure*, Vol. 19, No. 6, pp.859–867.

Ono, Y., Hayashida, T., Konagai, A., Okazaki, H., Miyao, K., Kawachi, S., Tanabe, M., Shinoda, M., Jinno, H. and Hasegawa, H. (2012) 'Direct inhibition of the transforming growth factor-β pathway by protein-bound polysaccharide through inactivation of Smad2 signaling', *Cancer Science*, Vol. 103, pp.317–324.

Oshima, M., Oshima, H. and Taketo, M.M. (1996) 'TGF-β receptor Type, I.I. deficiency results in defects of yolk sac hematopoiesis and vasculogenesis', *Developmental Biology*, Vol. 179, No. 1, pp.297–302.

Papageorgis, P. and Stylianopoulos, T. (2015) 'Role of TGFβ in regulation of the tumor microenvironment, and drug delivery', *International Journal of Oncology*, Vol. 46, No. 3, pp.933–943.

Park, B.J., Park, J.I., Byun, D.S., Park, J.H. and Chi, S.G. (2000) 'Mitogenic conversion of transforming growth factor-beta1 effect by oncogenic Ha-Ras-induced activation of the mitogen-activated protein kinase signaling pathway in human prostate cancer', *Cancer Res*., Vol. 60, pp.3031–3038.

Peltomaki, P. (2001) 'Deficient DNA mismatch repair: a common etiologic factor for colon cancer', *Hum. Mol. Genet.*, Vol. 10, No. 7, pp.735–740.

Reva, B., Antipin, Y. and Sander, C. (2011) 'Predicting the functional impact of protein mutations: application to cancer genomics', *Nucleic Acids Research*, p. gkr407.

Rhyu, D.Y., Yang, Y., Ha, H., Lee, G.T., Song, J.S., Uh, S-T. and Lee, H.B. (2005) 'Role of reactive oxygen species in TGF-β1-induced mitogen-activated protein kinase activation and epithelial-mesenchymal transition in renal tubular epithelial cells', *Journal of the American Society of Nephrology*, Vol. 16, pp.667–675.

Sherry, S.T., Ward, M-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic Acids Research*, Vol. 29, pp.308–311.

Sim, N-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P.C. (2012) 'SIFT web server: predicting effects of amino acid substitutions on proteins', *Nucleic Acids Research*, Vol. 40, pp.W452–W457.

Singh, B., Murphy, R.F., Ding, X.Z., Roginsky, A.B., Bell Jr., R.H. and Adrian, T.E. (2007) 'On the role of transforming growth factor-beta in the growth inhibitory effects of retinoic acid in human pancreatic cancer cells', *Mol. Cancer*, Vol. 6, p.82.

Siu, S.W., Pluhackova, K. and Böckmann, R.A. (2012) 'Optimization of the OPLS-AA force field for long hydrocarbons', *Journal of Chemical Theory and Computation*, Vol. 8, No. 4, pp.1459–1470.

Skeen, R.V., Paterson, I., Paraskeva, C. and Williams, C.A. (2012) 'TGF-β1 signalling, connecting aberrant inflammation and colorectal tumorigenesis', *Current Pharmaceutical Design*, Vol. 18, pp.3874–3888.

Stone, E.A. and Sidow, A. (2005) 'Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity', *Genome Research*, Vol. 15, No. 7, pp.pp.978–986.

Sussman, J.L., Lin, D., Jiang, J., Manning, N.O., Prilusky, J., Ritter, O. and Abola, E. (1998) 'Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules', *Acta Crystallographica Section D: Biological Crystallography*, Vol. 54, pp.1078–1084.

Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M.F. and Taketo, M.M. (1998) 'Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes', *Cell*, Vol. 92, pp.645–656.

Tariq, K. and Ghias, K. (2016) 'Colorectal cancer carcinogenesis: a review of mechanisms', *Cancer Biol Med*, Vol. 13, No. 1, pp.120–135.

Theobald, D.L. and Wuttke, D.S. (2008) 'Accurate structural correlations from maximum likelihood superpositions', *PLoS Comput. Biol.*, Vol. 4, No. 2, p.e43.

Wallace, A.C., Laskowski, R.A. and Thornton, J.M. (1995) 'LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions', *Protein Eng*, Vol. 8, No. 2, pp.127–134.

Williams, T., Kelley, C., Bröker, H., Campbell, J., Cunningham, R., Denholm, D., Elber, E., Fearick, R., Grammes, C. and Hart, L. (2012) *Gnuplot 4.6: An Interactive Plotting Program*, URL: http://www.gnuplot.info

Xu, Y. and Pasche, B. (2007) 'TGF-β signaling alterations, and susceptibility to colorectal cancer', *Human Molecular Genetics*, Vol. 16, No. R1, pp.R14–R20.

Yamamura, Y., Hua, X., Bergelson, S. and Lodish, H.F. (2000) 'Critical role of Smads and AP-1 complex in transforming growth factor-β-dependent apoptosis', *Journal of Biological Chemistry*, Vol. 275, pp.36295–36302.

Yan, Z., Deng, X. and Friedman, E. (2001) 'Oncogenic Ki-Ras confers a more aggressive colon cancer phenotype through modification of transforming growth factor-β Receptor III', *Journal of Biological Chemistry*, Vol. 276, No. 2, pp.1555–1563.

Yin, S., Ding, F. and Dokholyan, N.V. (2007) 'Eris: an automated estimator of protein stability', *Nature Methods*, Vol. 4, No. 6, pp.466–467.

Supplementary Table 1 Damaging nsSNPs prediction through sequence prediction tools

		PolyPhen	*I-Mutant Suite*	*Provean*	*MutPred*		*SNP&GO*		*PredictSNP*	*MAPP*	*SNAP*	*SIFT*	*Mutation Accessor*
SNP_ID	*AA change*	*Prediction*	*Prediction*	*Prediction*	*Score*	*PhD-SNP*	*PANTHER*	*SNPs&GO*	*Prediction*	*Prediction*	*Prediction*	*Prediction*	*Functional_ Impact*
rs201700967	R107H	PD	D	N	0.573	N	U	N	D	N	N	D	L
rs747563417	V106L	PD	N	N	0.855	D	D	N	D	D	N	D	M
rs781566009	N103T	PD	N	D	0.891	D	D	D	D	D	D	D	H
rs770505137	R94C	PD	D	D	0.448	D	D	D	D	D	D	D	M
rs758966134	P85R	PD	D	D	0.913	D	D	N	D	D	D	D	H
rs541829714	L83R	PD	D	D	0.918	D	D	D	D	D	D	D	M
rs199849225	V79G	PD	N	D	0.83	D	D	D	D	D	D	D	M
rs775550741	A74V	PD	N	D	0.685	N	D	N	N	D	N	D	L
rs763943753	N69Y	PD	D	D	0.776	D	D	D	D	D	D	D	M
rs199699574	S59N	PD	N	N	0.798	D	N	N	D	D	N	D	L
rs199699574	S59T	PD	N	N	0.747	N	N	N	N	D	N	N	N
rs753287325	T56M	PD	D	D	0.612	D	D	D	D	D	D	D	M
rs200230083	D55N	PD	D	D	0.572	N	D	D	N	D	N	D	L
rs199713772	S53N	B	N	N	0.597	N	D	D	N	D	N	N	N
rs199713772	S53T	B	N	N	0.571	N	D	N	N	D	N	D	L
rs745482429	S53G	PD	N	D	0.58	N	D	N	N	D	D	D	L
rs56361919	P49T	PD	D	D	0.862	D	D	D	N	D	N	N	M
rs200209614	P47A	B	D	N	0.486	N	N	N	N	N	N	N	N

Supplementary Table 1 Damaging nsSNPs prediction through sequence prediction tools (continued)

		PolyPhen	*I-Mutant Suite*	*Provean*	*MutPred*		*SNP&GO*		*PredictSNP*	*MAPP*	*SNAP*	*SIFT*	*Mutation Accessor*
SNP_ID	*AA change*	*Prediction*	*Prediction*	*Prediction*	*Score*	*PhD-SNP*	*PANTHER*	*SNPs&GO*	*Prediction*	*Prediction*	*Prediction*	*Prediction*	*Functional_ Impact*
rs768250306	G46R	PD	D	D	0.962	D	D	D	D	D	D	D	H
rs748135261	L45F	PD	D	N	0.429	N	N	N	N	N	N	N	N
rs769434404	F43L	PD	D	D	0.861	D	D	D	D	D	D	D	L
rs763101036	H40R	PD	D	D	0.561	N	N	N	N	N	D	D	L
rs766572720	K37N	PD	D	D	0.577	N	D	D	D	N	D	D	M
rs199516461	K37R	PD	D	N	0.441	D	D	D	N	N	N	D	L
rs200763912	E35D	PD	N	N	0.749	D	D	D	D	D	D	D	M
rs767685429	H34Q	PD	D	D	0.595	D	D	D	D	D	D	D	M
rs369182751	K31R	PD	D	N	0.621	N	N	N	N	N	N	N	L
rs761279576	W30R	PD	D	D	0.823	D	D	D	D	D	D	D	H
rs199946261	L28F	PD	D	D	0.757	N	D	N	D	D	N	D	M
rs775861573	R25H	PD	D	D	0.729	D	U	N	D	D	D	D	M
rs200527282	R25C	PD	D	D	0.673	D	U	D	D	D	D	D	M
rs201635147	I22T	PD	D	D	0.773	D	N	D	D	D	D	D	M
rs200164212	R18Q	PD	D	D	0.75	N	D	D	D	D	N	D	L
rs747968669	K13T	PD	D	N	0.523	N	U	N	N	N	N	D	N
rs770754343	T11M	PD	D	N	0.405	N	D	N	N	D	N	N	M
rs749275172	Y6C	PD	D	D	0.432	N	D	N	N	D	N	D	M
rs778711968	L2P	PD	D	D	0.677	D	D	D	D	D	N	D	L

*PD – Probably Damaging, B – Benign, D – Disease, N – Neutral, U – Unclassified, L – Low, M – Medium, H – High.

Supplementary Table 2 Damaging nsSNPs prediction through structure prediction tools

	SNP_Effect_4.0(A-chain)					*SNP_Effect_4.0(B-chain)*					*ERIS*		*SNP&GO$^{3D\ (A\text{-}CHAIN)}$*					*SNP&GO$^{3D\ (B\text{-}CHAIN)}$*				
SNP_ID	*FOLDX* (kcal/mol)	*dLIMBO*	*WALTZ*	*TANGO*	*Prediction*	*FOLDX* (kcal/mol)	*dLIMBO*	*WALTZ*	*TANGO*	*Prediction*	*Prediction (ΔΔG* (kcal/mol))	*S3D-PROF*	*PANTHER*	*SNPs&GO*	*S3Ds&GO*	*S3D-PROF*	*PANTHER*	*SNPs&GO*	*S3Ds&GO*
rs201700967	2.14	0	0.03	−3.89	RS	1.52	0	0.03	−3.89	RS	0.34	D	U	N	N	D	U	N	N
rs747563417	3.84	0	−0.07	−2.09	RS	3.12	0	−0.07	−2.09	RS	1.86	D	D	N	D	D	D	N	D
rs781566009	2.28	0	−0.29	57.05	RS	1.57	0	−0.29	57.05	RS	1.27	D	D	N	D	D	D	N	D
rs770505137	0.24	−6.17	−119.79	56.03	NE	0.41	−6.17	−119.79	56.03	NE	−1.59	D	D	D	D	D	D	D	D
rs758966134	1.63	0.01	1.13	−1.42	RS	1.75	0.01	1.13	−1.42	RS	−3.85	D	D	D	D	D	D	D	D
rs541829714	2.41	0	0	0	RS	1.61	0	0	0	RS	2.32	D	D	D	D	D	D	D	D
rs199849225	3.09	0	0	0	RS	3.16	0	0	0	RS	5.28	D	D	D	D	D	D	D	D
rs775550741	2.24	0	0	0	RS	1.96	0	0	0	RS	−1.63	D	D	N	D	D	D	N	D
rs763943753	1.83	0	0.04	0.08	RS	1.68	0	0.04	0.08	RS	0.28	D	D	D	D	D	D	D	D
rs199699574	6.05	−2.01	0.52	−4.45	SRS	2.52	−2.01	0.52	−4.45	RS	2.83	D	N	N	D	D	N	N	D
rs199699574	1.99	−1.15	−2.56	29.64	RS	1.84	−1.15	−2.56	29.64	RS	4.34	N	N	N	N	N	N	N	N
rs753287325	−1.09	0	0.37	9.8	NE	0.3	0	0.37	9.8	NE	0.83	N	D	D	D	N	D	D	D
rs200230083	0.28	0	3.73	32.43	NE	0.28	0	3.73	32.43	NE	−0.18	N	D	D	N	N	D	D	N
rs199713772	−0.74	0	11.7	−19.45	NE	−0.7	0	11.7	−19.45	RS	−3.31	N	D	N	D	N	D	N	D
rs199713772	0.41	0	−0.96	94.03	NE	0.21	0	−0.96	94.03	NE	−1.02	D	D	N	D	D	D	N	D
rs745482429	−0.03	0	0.1	−8.38	NE	−0.1	0	0.1	−8.38	NE	0.95	N	D	D	D	N	D	D	D
rs56361919	2.19	0	0.5	47.34	RS	2.3	0	958.82	47.34	RS	−1.51	D	D	D	D	D	D	D	D
rs200209614	1.26	0	0.05	−0.01	RS	0.74	0	958.82	−0.01	RS	−1.57	N	N	N	N	N	N	N	N
rs768250306	17.12	0	0	0.01	SRS	14.95	0	0	0.01	SRS	>10	D	D	D	D	D	D	D	D
rs748135261	0.22	0	0.22	−0.02	NE	0.15	0	0.22	−0.02	NE	2.51	N	N	N	N	N	N	N	N

Supplementary Table 2 Damaging nsSNPs prediction through structure prediction tools (continued)

	SNP_Effect_4.0(A-chain)					*SNP_Effect_4.0(B-chain)*					*ERIS*					*SNP&GO$^{3D\ (A\text{-}CHAIN)}$*					*SNP&GO$^{3D\ (B\text{-}CHAIN)}$*				
SNP_ID	*FOLDX* (kcal/mol)	*dLIMBO*	*WALTZ*	*TANGO*	*Prediction*	*FOLDX* (kcal/mol)	*dLIMBO*	*WALTZ*	*TANGO*	*Prediction*	*Prediction* (*ΔΔG* (kcal/mol))	*S3D-PROF*	*PANTHER*	*SNPs&GO*	*S3Ds&GO*	*S3D-PROF*	*PANTHER*	*SNPs&GO*	*S3Ds&GO*	*PANTHER*	*SNPs&GO*	*S3Ds&GO*
rs769434404	3.05	0	−0.7	0	RS	3.41	0	−0.7	0	RS	−0.5	D	D	D	D	D	D	D	D	D	D	D
rs763101036	0.04	0	−0.65	0	NE	0.1	0	−0.65	0	NE	−0.26	N	N	N	N	N	N	N	N	N	N	N
rs766572720	−0.14	0	−0.01	0	NE	0.36	0	−0.01	0	NE	−3.12	D	D	D	D	D	D	D	D	D	D	D
rs199516461	0.03	0	0	0	NE	−0.02	0	0	0	NE	−1.29	N	D	D	N	N	N	D	D	D	D	D
rs200763912	0.64	0.04	−0.01	0	SRS	1.1	0.04	−0.01	0	RS	0.98	D	D	D	D	D	D	D	D	D	D	D
rs767685429	−0.58	0.06	0.25	0	ES	−0.21	0.06	0.25	0	NE	0.26	D	D	D	D	D	D	D	D	D	D	D
rs369182751	0.12	0	−0.08	0	NE	0.28	0	−0.08	0	NE	−1.22	N	N	N	N	N	N	N	N	N	N	N
rs761279576	3.61	0.04	0.1	0	RS	3.45	0.04	0.1	0	RS	−0.96	D	D	D	D	D	D	D	D	D	D	D
rs199946261	2.89	0	0.27	0	RS	4.5	0	0.27	0	RS	1.9	N	D	N	D	N	D	D	N	D	N	D
rs775861573	0.22	−1.12	0.67	0	NE	1.93	−1.12	0.67	0	RS	0	D	U	N	N	D	U	N	N	U	N	N
rs200527282	0.16	−1.09	−0.11	6.62	NE	2.01	−1.09	−0.11	6.62	RS	−1.31	D	U	D	N	D	U	D	N	U	D	N
rs201635147	2.9	−0.13	−128.78	0	RS	3.17	−0.13	−128.78	0	RS	4.24	D	N	D	D	D	N	D	D	N	D	D
rs200164212	1.46	0	421.2	1.3	RS	1.46	0	421.2	1.3	RS	1.28	D	D	D	D	D	D	D	D	D	D	D
rs747968669	0.79	0	−0.14	0	SRS	0.67	0	−0.14	0	RS	−1.17	N	U	N	N	N	U	N	N	U	N	N
rs770754343	−0.08	0	−0.01	0	NE	−0.1	0	−0.01	0	NE	−2.44	N	D	N	N	N	D	N	N	D	N	N
rs749275172	3.02	0	−2.33	0.11	RS	2.92	0	−2.33	0.11	RS	3.77	D	D	N	D	D	D	D	D	D	N	D
rs778711968	2.2	0	−2.19	0	RS	2.26	0	−2.19	0	RS	>10	D	D	D	D	D	D	D	D	D	D	D

*RS – Reduces Stability, NE – No Effect, SRS – Severely Reduces Stability, ES – Enhances Stability, D – Disease, U – Unclassified, N – Neutral.