Pakistan J. Zool., vol. 47(5), pp. 1305-1318, 2015,

In Silico Analysis of SNPs in Coding Region of Human *c-Myc* Gene

Mamoona Noreen,* Sheeba Murad and Hayat Khan
Atta-ur-Rehman School of Applied Biosciences, National University of Sciences & Technology, H-12 Sector, Islamabad, Pakistan

Abstract.- Deregulation and over expression of human c-Myc proto-oncogene protein can induce cancer. Presence of genetic variations is a key player among many others which affect susceptibility and progression of disease. Single nucleotide polymorphisms (SNPs) are the most frequent variations in human genome. Non synonymous single nucleotide polymorphisms (nsSNPs) leading to change in amino acids may result in altered protein structure, function and molecular morphology. We used SIFT, PolyPhen and SNPeffect algorithms to predict tolerant and deleterious SNPs in coding region of *c-Myc* gene and their possible structural, molecular and phenotypic effects on c-Myc protein. Effect of four nsSNPs of coding region; rs114570780 (Tyr47His), rs150308400 (Cys148Tyr), rs137906262 (Leu159Ile) and rs200431478 (Ser362Phe) is predicted to be damaging. These results may form the basis for further large-scale population based association studies.

Key Words: SNPs; *c-Myc* gene, single nucleotide polymorphism, SIFT, PolyPhen.

INTRODUCTION

Human Myc proto-oncogene protein is a basic dimerization motif helix-loop-helix leucine zipper (HLH-LZ) containing transcription factor that was initially discovered to be the cellular homologue of v-myc myelocytomatosis viral oncogene (Vennstrom *et al.*, 1982). In mammals three proteins with reasonable structural resemblance; *c-Myc*, N-Myc and L-Myc are encoded by three different genes and their over expression is reported to lead to tumor induction (Oster *et al.*, 2002). In humans the *c-Myc* gene is located on 8q24.21 chromosomal position and is reported to have three exons. N-terminal region contains four virtually identical and highly conserved domains, identified as Myc Box I (MBI), MBII, MBIII and MBIV (Oster *et al*., 2002; Cowling *et al*., 2006). N-terminal region contains several well characterized phosphorylation and ubiquitination sites (Fig. 1). The basic region (BR) of *c-Myc* mediates the DNA while the helix-loop-helix (HLH) and leucine zipper (LZ) domains mediate heterodimerization with other proteins (Hann, 2006).

* Corresponding author:
mamoona.10.trfphd.33@asab.nust.edu.pk,
mamoona786@hotmail.com
0030-9923/2015/0005-1305 $ 8.00/0

On the basis of discrete translational initiation sites, *c-Myc* gene has been reported to encode three distinct isoforms; *c-Myc*1, *c-Myc2* (Hann *et al.*, 1988) and *c-MycS* (Hann, 1995) with identical carboxy terminal domain having BR-HLH-LZ and different N-terminal regions. Major isoform *i.e. c-Myc2* is a 439 amino acids long 64 kDa protein and is encoded when translation starts at AUG start site in the second exon. Translational start at a non canonical CUG codon located 15 codons upstream of *c-Myc2* start site adds a short extension of additional 15 amino acids at N-terminal region of protein yielding a longer isoform of 454 amino acids and 67 kDa called *c-Myc1* or p67 (Hann *et al.*, 1988). Considerably shorter, N-terminally-truncated *c-MycS* isoform requires translational initiation at AUG site located 100 codons downstream of *c-Myc2* start site. Such translation deprives the protein of well conserved motif MBI lying in 100 amino acids region of the trans-activation domain at N-terminus (Spotts *et al.*, 1997).

Myc family of proto-oncogenes has been considered to be one of the most studied oncogenes to date (Meyer and Penn, 2008). Crucial involvement of Myc proto-oncogene protein in the regulation of other gene's expression (Lee and Dang, 2006), cellular proliferation (Lemaitre *et al.*, 1996), growth (Henriksson and Luscher, 1996), apoptosis (Thompson, 1998), metabolic transformation and oncogenesis (Miller *et al.*, 2012) has become evident during last two decades. Lower

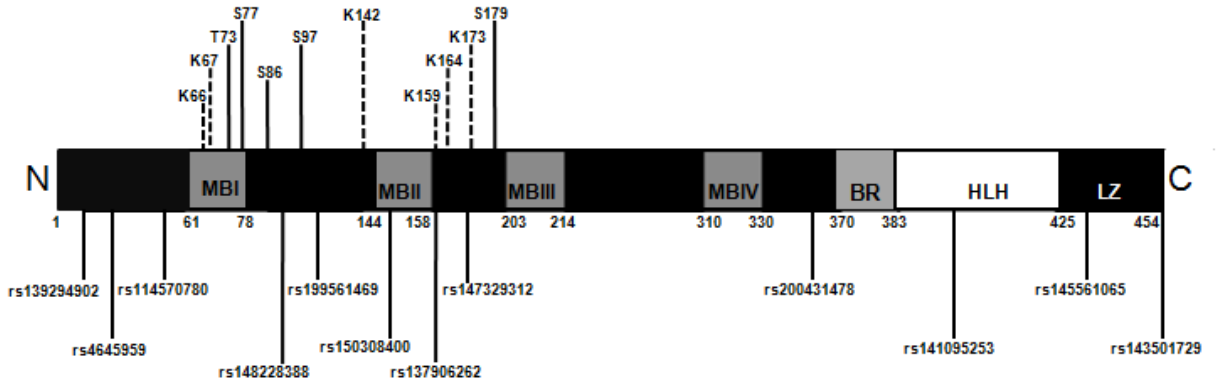

Fig. 1. Schematic diagram of the significant domains and position of phosphorylation, ubiquitination and nsSNPs in c-Myc1 protein. It details the location of highly conserved Myc homology boxes (MBI, MBII,MBIII and MBIV), basic region (BR), helix-loop-helix (HLH) and leucine zipper (LZ) domains. The numbers below represent the amino acid that borders each significant domain of protein. Specific N-terminal features, including the location of phosphorylation and ubiquitination (lysine residues) are indicated on upper side while location of nsSNPs (predicted to be damaging by SIFT, PolyPhen or SNPeffect algorithms) are indicated below.

expression of *c-Myc* in normal resting cell becomes enormously high in the transformed cell stimulated via a broad range of signaling pathways, designating it to be an early response gene. It is not clear whether *c-Myc* over expression is the cause or the consequence of cell transformation (Miller *et al.*, 2012).

Commonly, disease risk is influenced by the existing polymorphisms in the genome. One of the most frequent types of polymorphisms found in human genome is single nucleotide polymorphisms (SNP) (Noreen *et al.*, 2012). Usually the SNPs associated with disease risk are missense or non-synonymous SNPs (nsSNPs) involving substitution of amino acid (Gul *et al.*, 2015). These SNPs are likely to bring structural as well as functional changes in the protein. Genetic variations in cis-regulators of transcription at 8q24 can considerably change the germline expression levels of *c-Myc* thus contribute to cancer susceptibility (Sole *et al.*, 2008). Many studies are conducted to find association of a well documented SNP; rs13281615 harboring the non-coding chromosomal region 8q24 near *c-Myc* and susceptibility of breast cancer (Easton *et al.*, 2007; Pei *et al.*, 2013), prostate cancer (Meyer *et al.*, 2009; Haiman *et al.*, 2007) and colorectal cancer (Tomlinson *et al.*, 2007). However, the results are inconsistent. A meta-analysis including more than 100,000 subjects was performed to evaluate the role of this SNP in risk of breast cancer development. Increased risk of breast cancer development was found to be significantly associated with presence of allele G (Odds Ratio (OR): 1.10, 95% Confidence interval (CI): 1.06–1.14, p value: 0.001) and genotype "GG" (OR: 1.20, 95% CI: 1.12–1.29, p value: 0.001), while the genotype AA was found to be a protective factor against breast cancer development (OR: 0.89, 95% CI: 0.84–0.93, p value: 0.001) (Gong *et al.*, 2013).

Chromosomal region 8q24.21 contains another discussable cancer risk SNP; rs6983267. Presence of allele G is variably frequent among all the populations ranging from 31% in Native Hawaiians to 85% in African American population. In Europeans the allele frequency is ~50% (Wang *et al.*, 2014). The rs6983267 is reported to up-regulate *c-Myc* transcription (Takatsuno *et al.*, 2013) and alter the physical interactions of *c-Myc* by affecting binding of transcription factor 7-like 2 (TCF7L2) and TCF4 (Pomerantz *et al.*, 2009). Presence of rs6983267 is associated with susceptibility to inflammatory breast cancer (Bertucci *et al.*, 2012), gastric cancer (Guo *et al.*, 2011), prostate cancer (Os'kina *et al.*, 2012), colorectal cancer and colorectal adenoma (Wang *et al.*, 2014). A cluster of four pathogenetically valuable nsSNPs; VAR_063384 (Glu39Asp), VAR_063385 (Pro57Ser), VAR_063386 (Pro59Ala) and

VAR_063387 (Asn86Thr) was reported to harbor 2nd and 3rd exon of *c-Myc2* trans-activation domain in Burkitt lymphoma samples (Rabbitts *et al.*, 1983; Bhatia *et al.*, 1993). A rare nsSNP (rs4645959) lying in N-terminal trans-activation domain of *c-Myc2* was reported to be putatively functional. Heterozygous carriers of this variant among Polish and German subjects were observed to have an increased risk of familial breast cancer (Wirtenberger *et al.*, 2005). Later on Figueiredo *et al.* (2007) conducted a large population based centralized pathology review revealing no biological relevance of rs4645959 with risk, tumor characteristics or endurance of breast cancer (Figueiredo *et al.*, 2007).

Molecular epidemiological association studies mainly focused on the nsSNPs lying in coding region of the gene (Ramensky *et al.*, 2002; Savas *et al.*, 2004; Zhu *et al.*, 2004). Studying the structural and the functional impact of the nsSNPs on the protein can help in the selection of functionally important nsSNPs. We have used some sequence and structural homology based algorithms to find those nsSNPs which can play potential role in structural and functional alteration of *c-Myc*.

MATERIALS AND METHODS

Dataset compilation

Neucleotide and protein sequence in FASTA format were obtained from NCBI (http://www.ncbi.nlm.nih.gov). Longest isoform of *c-Myc* was selected for analysis. Public SNP databases named dbSNP (http://www.ncbi.nlm.nih.gov/SNP), SNP500cancer (http://snp500cancer.nci.nih.gov/home.cfm), GeneCards (http://www.genecards.org/) and UniProt (http://www.uniprot.org) were used to identify SNPs related to longest isoform of human *c-Myc* proto-oncogene protein. Being in gene region, 165 SNPs were found to be related to *c-Myc* (gene ID: 4609, NCBI Reference Sequence: NP_002458.2). We kept 27 nsSNPs into analytical consideration, out of 43 lying in coding region.

Prediction of tolerated and deleterious SNPs and functional consequences of nsSNPs using SIFT

We used Sorting Intolerant from Tolerant (SIFT version 2) algorithm which is able to predict if an amino acid substitution can influence protein function. SIFT is known to make a distinction between tolerated and deleterious amino acid changes in proteins (http://blocks.fhcrc.org/sift/SIFT.html) (Ng and Henikoff, 2001). It works on basic principle that substitution of some fundamentally conserved amino acids in a protein can have a potential impact on protein function and can be deleterious (Ng and Henikoff, 2003).

Protein sequence in FASTA format was submitted as input to SIFT in order to predict possible effect of any of the other amino acids substitution at every position. For this analysis FIST searches for homologous sequences using SWISS-PORT version 51.3 and TrEMBL 34.3 databases and performs multiple alignments of available sequences until a median sequence conservation score (MSCS) for the protein sequence is reached to 3.00. Tolerant and intolerant effect of nsSNPs of coding region was predicted as tolerance index (TI) score which is normalized probability that the amino acid substitution is tolerated or not. It ranges from 0.0 to 1.0 while threshold for intolerance is $\leq$0.05. (de Alencar and Lopes, 2010) found the level of functional influence of a specific amino acid substitution to be inversely proportional to TI score. Deleterious predictions with median score of 3.25 or less were made with low confidence due to lack of sequence diversity for protein alignment.

Prediction of functional consequences of nsSNPs using PolyPhen

We used PolyPhen-2 (Polymorphism Phenotyping version 2) (http://genetics.bwh.harvard.edu) in order to predict detrimental outcome of nsSNPs on structural and functional basis. Protein sequence in FASTA format, position of SNP in protein sequence, wild type amino acid and its substitute for each of the 27 variations in coding region of *c-Myc* were endowed with as an input to the algorithm. PolyPhen-2 uses three structure based and eight sequence based predictive features out of total thirteen structure based and nineteen sequence based candidate features which are automatically selected by iterative greedy algorithm (Adzhubei *et al.*, 2010).

Homologues of the submitted sequence were searched using BLAST+ (Camacho *et al.*, 2009) and

aligned using MAFFT (Katoh *et al.,* 2002). Quality of resulting alignments was made better by use of LEON software (Thompson *et al.,* 2004). Secator algorithm formerly executed in ClusPack software was used to gather reliably aligned sequences (Wicker *et al.,* 2001). Analyzed sequences along with its homologues of compact cluster were taken into further consideration. PolyPhen also searches for 3D protein structures and contact information about amino acids in a number of structure based databases in order to firstly calculate the position-specific independent counts (PSIC) scores for allelic variants of particular SNP and then to evaluate PSIC score difference (Adzhubei *et al.,* 2010). On basis of all the information the effect of nsSNP on protein is predicted. If the score shows high reliability that the nsSNP should influence structure or/and function of protein, then it will be predicted as ‘‘possibly damaging’’, if there is an option of “may or may not”, the nsSNP will be assigned as ‘‘probably damaging’’ while being unlikely to phenotypic alteration in protein can designate a nsSNP as ‘‘benign’’ (Ramensky *et al.,* 2002).

Prediction of molecular phenotypic effects of nsSNPs using SNPeffect

SNPeffect (De Baets *et al.,* 2012) was used as a platform to forecast the molecular phenotypic influence of nsSNPs lying in coding region of *c-Myc* protein. It works beyond the scores got on conservational basis and mainly emphasizes to map the effect of SNPs on the capability of cells to uphold suitable concentration of the properly folded proteins in appropriate cellular region i-e protein homeostasis landscape (Powers *et al.,* 2009). For this evaluation, wild type protein sequence in FASTA format was provided to the SNPeffect server (http://snpeffect.switchlab.org) along with each of its variants to be analyzed. Homology hreshold was adjusted to 90%. SNPeffect uses TANGO (Fernandez-Escamilla *et al.,* 2004) which predicts the regions in given protein sequence that are prone to aggregation and measures TANGO score with wild and variant amino acid. Based on difference of TANGO score (dTANGO) the server evaluates effect of these variants on protein aggregation.

Aggregate morphology is more specifically determined by WALTZ server (Maurer-Stroh *et al.,* 2010) that predicts amyloid-forming regions in given protein sequence with accuracy and specificity on basis of dWALTZ score. Chaperone binding propensity is predicted by LIMBO (Van Durme *et al.,* 2009) for the Hsp70 chaperones and effect of variant is determined by dLIMBO score. SNPeffect also uses high resolution crystal structure of proteins from Protein Data Bank (PDB) (Deshpande *et al.,* 2005) and models the variants using the empirical force field FoldX (version 2.5) (Schymkowitz *et al.,* 2005) to evaluate possible effects on stability and binding properties of given protein.

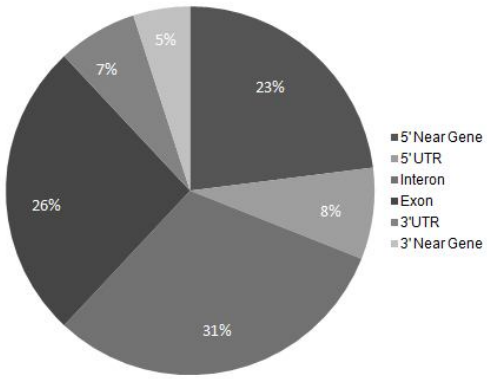

Fig. 2. Distribution of SNPs in human c-Myc1 gene region

RESULTS

Dataset compilation

Our study included retrieval of total 165 human *c-Myc* SNPs from the dbSNP database which were further scrutinized. GeneCards database included four additional pathologically important SNPs of *c-Myc2* proto-oncogene reported in Burkitt lymphoma samples; VAR_063384 (Glu39Asp), VAR_063385 (Pro57Ser), VAR_063386 and VAR_063387 (Asn86Thr). The non-conding region of gene contains 122 SNPs of the total SNPs, out of which 38 (23%) are in the region of 5’ near gene, 13 (8%) are in 5’ UTR, 51 (31%) in intronic region, 11 (7%) in 3’ UTR and 9 (5%) in region of 3’ near gene. Coding region harbors 43 SNPs out of which 29 (17%) are non-synonymous and 14 (9%) are synonymous (Fig. 2). As nsSNPs are supposed to be

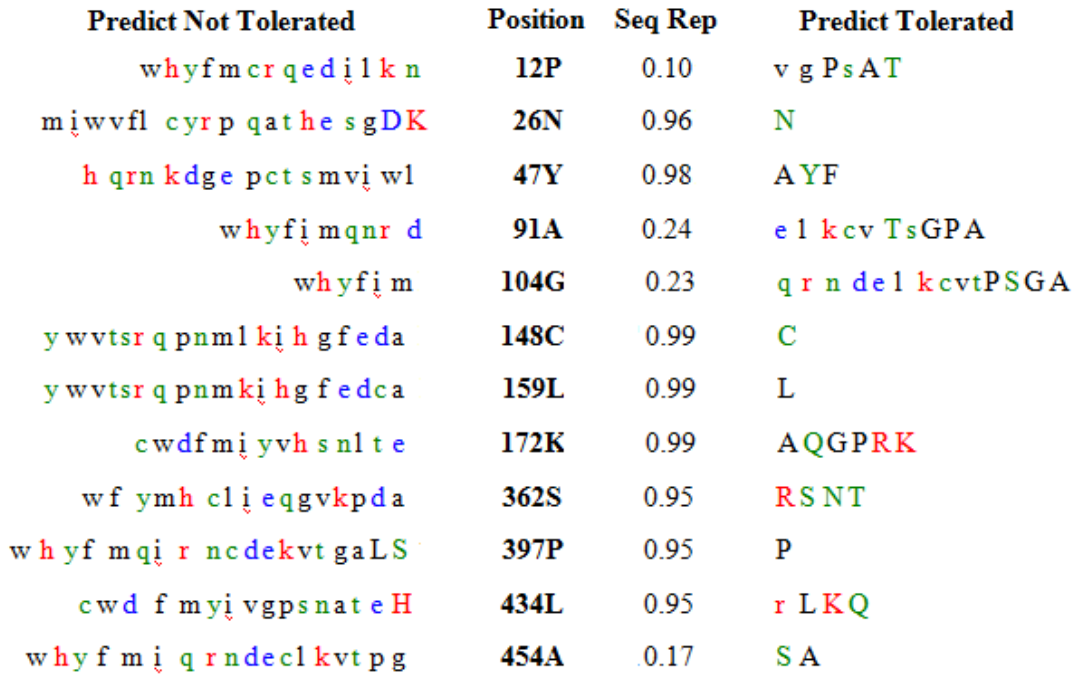

Fig. 3. Prediction of possible substitutions of amino acids in human c-Myc1 protein using SIFT algorithm. Amino acids are shown in color code: basic (red), acidic (blue) non polar (black); uncharged polar (green). Capital letters signify amino acids which come into view during the alignment while lower case letters are the consequence of prediction. 'Seq-Rep' is actually the fraction of sequences that have one of the fundamental amino acids. Low fraction shows that position is unalienable or is severely gapped providing little information which results in poor prediction.

involved in structural and functional alteration of proteins, thus those can result in disease development. As two SSNPs in coding region of *c-Myc* protein are frame shift mutations thus we investigated the probable damage caused by 27 nsSNPs.

Prediction of tolerated and deleterious SNPs and functional consequences of nsSNPs using SIFT

The SIFT program predicts the tolerant or deleterious effect of amino acid change on protein function. The analysis mainly considers the alignment of homologous sequences, physical properties of amino acids and the effect of natural nsSNPs on phenotypic alterations by aligning paralogous and orthologous protein sequences. The prediction of possible substitutions for positions 1-454 amino acids was made. Possible substitutions for those nsSNPs are shown in Figure 3 which were predicted to have some damaging effect by SIFT, PolyPhen or SNPeffect algorithms. 'Seq-Rep' is actually the fraction of sequences that has one of the fundamental amino acids. Low fraction shows that the position is unalienable or is severely gapped providing little information which results in poor prediction. Amino acids are shown in color codes *i.e.*, basic (red), acidic (blue) non polar (black); uncharged polar (green). Capital letters signifies the amino acids which come into view during the alignment while lower case letters are the consequence of prediction.

Sequence homology-based tool of SIFT was used to predict tolerant and deleterious effect of 27 nsSNPs present in coding region of *c-Myc* protein. As a result 8 nsSNPs were found to be deleterious in total, out of which 2 had low confidence due to median score >3.25. (Table I, Fig. 4). Out of these eight nsSNPs, rs139294902 (Pro12Arg), rs4645959 (Asn26Ser), rs114570780 (Tyr47His), rs150308400 (Cys148Tyr) and rs137906262 (Leu159Ile) are located within N-terminal trans-activation domain while rs145561065 (Leu434Phe) and rs143501729

Table I.- SIFT analysis of nsSNPs in coding region of human c-Myc1 protein.

dbSNP id	Chr. Position	mRNA Pos.	dbSNP Allele	Codon Pos.	Amino acid variant	Amino acid Pos.	SIFT prediction	Tolerance Index
rs139294902	128750498	560	C/G	2	Pro [P]/Arg [R]	12	Intolerant §	0.00
rs146505192	128750527	589	T/C	1	Phe [F]/Leu [L]	22	Tolerant	1.00
rs148915481	128750534	596	A/G	2	Asn [N]/Ser [S]	24	Tolerant	0.99
rs4645959	128750540	602	A/ G	2	Asn [N]/Ser [S]	26	Intolerant	0.01
rs114570780	128750602	664	T/C	1	Tyr [Y]/His [H]	47	Intolerant	0.00
rs148228388	128750735	797	C/A	2	Ala [A]/Glu [E]	91	Tolerant	0.06
rs199561469	128750773	835	G/C	1	Gly [G]/Arg [R]	104	Tolerant	0.36
rs150308400	128750906	968	G/A	2	Cys [C]/Tyr [Y]	148	Intolerant	0.00
rs137906262	128750938	1000	C/A	1	Leu [L]/Ile [I]	159	Intolerant	0.02
rs61755060	128750967	1029	G/T	3	Gln [Q]/His [H]	168	Tolerant	0.45
rs147329312	128750977	1039	A/C	1	Lys [K]/Gln [Q]	172	Tolerant	0.11
rs4645960	128750986	1048	G/T	1	Gly [G]/Cys [C]	175	Tolerant	0.12
rs4645960	128750986	1048	G/C	1	Gly [G]/Arg [R]	175	Tolerant	0.13
rs4645961	128751016	1078	G/A	1	Val [V]/Ile [I]	185	Tolerant	0.26
rs112602073	128751121	1183	C/T	1	Pro [P]/Ser [S]	220	Tolerant	0.93
rs147506213	128751160	1222	C/A	1	Pro [P]/Thr [T]	233	Tolerant	0.64
rs186663828	128751185	1247	C/T	2	Ser [S]/Leu [L]	241	Tolerant	0.19
rs148544254	128751200	1262	C/T	2	Pro [P]/Leu [L]	246	Tolerant	0.26
rs150629172	128752729	1415	C/A	2	Pro [P]/His [H]	297	Tolerant	0.15
rs139697494	128752745	1431	C/A	3	His [H]/Gln [Q]	302	Tolerant	0.15
rs146971340	128752799	1485	T/A	3	His [H]/Gln [Q]	320	Tolerant	0.74
rs4645968	128752849	1535	C/T	2	Ala [A]/Val [V]	337	Tolerant	1.00
rs200431478	128752924	1610	C/T	2	Ser [S]/Phe [F]	362	Intolerant	0.01
rs141095253	128753029	1715	C/T	2	Pro [P]/Leu [L]	397	Tolerant	0.05
rs145561065	128753141	1827	G/C	3	Leu [L]/Phe [F]	434	Intolerant	0.01
rs148863193	128753155	1841	G/T	2	Arg [R]/Leu [L]	439	Tolerant	0.12
rs201337668	128753196	1882	T/G	1	Cys [C]/Gly [G]	453	Tolerant	0.64
rs143501729	128753200	1886	C/A	2	Ala [A]/Glu [E]	454	Intolerant §	0.00

Variants with tolerance index ≤0.05 score are considered as deleterious while others are taken to be tolerant. § Predictions are made with low confidence because MSCS > 3.25, thus the interpretations should be made cautiously.

(Ala454Glu) harbor the C-terminal domain of *c-Myc1* protein (Fig. 1).

Prediction of functional consequences of nsSNPs using PolyPhen

PolyPhen-2 server was used to predict structural and functional effects of 27 nsSNPs reported in the coding region of *c-Myc* proto oncogene protein. Using structure and sequence based predictive features PolyPhen searched for homologous protein sequences for multiple alignment and 3D protein structures to find probability score and predicted possible influence of variants on the protein. Out of 27 mutations, 7 were predicted to be probably damaging with probabilistic scores above 0.85, one variant was forecasted to be possibly damaging with probabilistic score above 0.15 while all others were considered to be benign (Table II, Fig. 4). Out of these eight nsSNPs, rs114570780 (Tyr47His), rs148228388 (Ala91Glu), rs199561469 (Gly104Arg), rs150308400 (Cys148Tyr), rs137906262 (Leu159Ile) are located within N-terminal trans-activation domain while rs200431478 (Ser362Phe), rs141095253 (Pro397Leu) harbor the C-terminal domain of *c-Myc1* protein (Fig. 1).

Prediction of molecular phenotypic effects of nsSNPs using SNPeffect

The SNP effect server was used to predict the effect of molecular phenotype of nsSNPs present in coding region of *c-Myc* protein. dTANGO, dWALTZ and dLIMBO scores were calculated to find effect of 27 nsSNPs on aggregation tendency, amyloid propensity and chaperone binding tendency respectively. Threshold for some significant effect

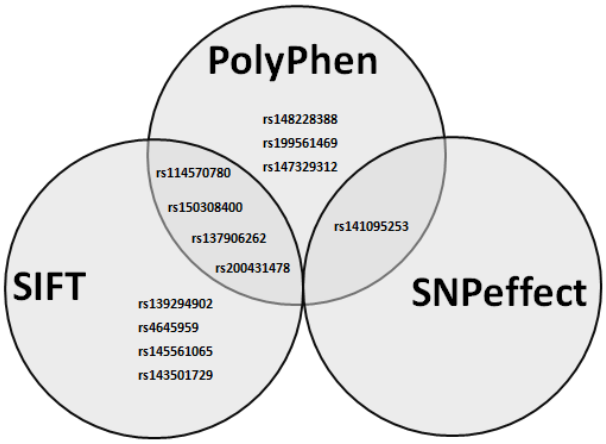

Fig. 4. nsSNPs of human c-Myc1 protein predicted by SIFT, PolyPhen and SNPeffect algorithms to have some biological importance.

Fig. 5. Molecular visualization of the wild type; Pro (left) and variant; Leu (right) residues colored in black at position 397 in human c-Myc1 protein

was ranging from -50 to 50. The dLIMBO score for Pro/Leu variation at position 397 (rs141095253) equals -277.70, which means that the mutation decreases the chaperone binding tendency of our protein. The scores for rest of the variants were not observed to be more than 50 or less than -50. The result shows that none of these nsSNPs other than rs141095253 is able to bring significant change in the aggregation tendency, amyloid propensity and chaperone binding tendency of *c-Myc* protein (Table III). The difference in free energy of the mutation was calculated using empirical protein design force field (FoldX). The submitted sequence had 98.80 percent homology with 1NKP. Using FoldX a homology model was built starting from this PDB. The difference in free energy due to Pro/Leu variation at position 397 was of 2.68 Kcal/mol. This entails that the mutation reduces the protein stability (Fig.5). For rest of the variations SNPeffect could not predict possible changes in the stability of protein which could have been brought by variants due to the unavailability of adequate structural information of *c-Myc* in the form of PDB model. 1NKP presents the X-ray structure of the *c-Myc* protein in complex with the Max protein and DNA. Both proteins form a fork-like structure and bind the major groove of the DNA. It includes 353 to 434 residues of *c-Myc2*. Three SNPs; rs141095253 (Pro397Leu), rs145561065 (Leu434Phe) and rs148863193 (Arg439Leu) of longest isoform *i.e. c-Myc1* are illustrated in Figure 6.

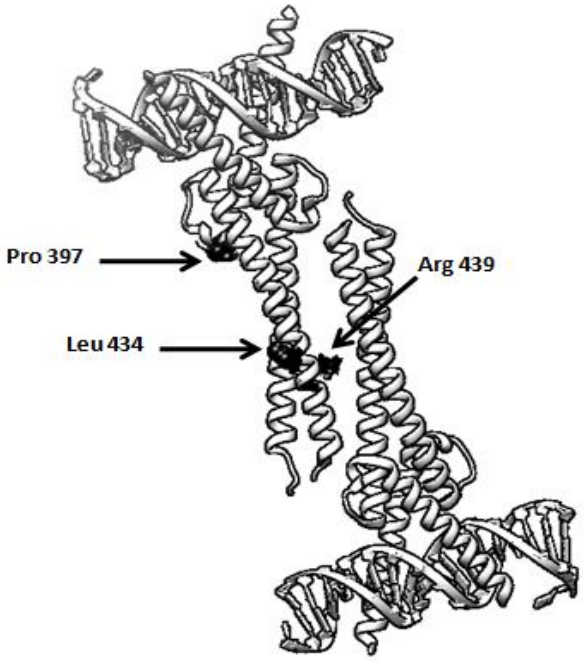

Fig. 6. Illustration of rs141095253 (Pro397Leu), rs145561065 (Leu434Phe) and rs148863193 (Arg439Leu) in 1NKP.

DISCUSSION

Myc proto-oncogene protein family comprises of very crucial and extensively studied transcription factors characterized by containing basic helix-loop-helix leucine zipper (bHLH LZ) motif (Vennstrom *et al.,* 1982). Over expression and deregulation of *c-Myc*, N-Myc and L-Myc can become basic cause of tumor development (Oster *et al.,* 2002) and polymorphisms especially SNPs may play an important role in it. SNPs may influence the

Table II.- PolyPhen analysis of nsSNPs in coding region of human c-Myc1 protein

dbSNP id	Amino acid variant	Amino acid position	PolyPhen prediction	PolyPhen score
rs139294902	Pro [P]/Arg [R]	12	Benign	0.027
rs146505192	Phe [F]/Leu [L]	22	Benign	0.008
rs148915481	Asn [N]/Ser [S]	24	Benign	0.007
rs4645959	Asn [N]/Ser [S]	26	Benign	0.160
rs114570780	Tyr [Y]/His [H]	47	Probably Damaging	0.955
rs148228388	Ala [A]/Glu [E]	91	Possibly damaging	0.452
rs199561469	Gly [G]/Arg [R]	104	Probably Damaging	0.915
rs150308400	Cys [C]/Tyr [Y]	148	Probably Damaging	0.999
rs137906262	Leu [L]/Ile [I]	159	Probably Damaging	0.997
rs61755060	Gln [Q]/His [H]	168	Benign	0.155
rs147329312	Lys [K]/Gln [Q]	172	Probably Damaging	0.971
rs4645960	Gly [G]/Cys [C]	175	Benign	0.048
rs4645960	Gly [G]/Arg [R]	175	Benign	0.269
rs4645961	Val [V]/Ile [I]	185	Benign	0.248
rs112602073	Pro [P]/Ser [S]	220	Benign	0.135
rs147506213	Pro [P]/Thr [T]	233	Benign	0.002
rs186663828	Ser [S]/Leu [L]	241	Benign	0.270
rs148544254	Pro [P]/Leu [L]	246	Benign	0.273
rs150629172	Pro [P]/His [H]	297	Benign	0.001
rs139697494	His [H]/Gln [Q]	302	Benign	0.041
rs146971340	His [H]/Gln [Q]	320	Benign	0.327
rs4645968	Ala [A]/Val [V]	337	Benign	0.009
rs200431478	Ser [S]/Phe [F]	362	Probably Damaging	0.914
rs141095253	Pro [P]/Leu [L]	397	Probably Damaging	0.999
rs145561065	Leu [L]/Phe [F]	434	Benign	0.100
rs148863193	Arg [R]/Leu [L]	439	Benign	0.193
rs201337668	Cys [C]/Gly [G]	453	Benign	0.00
rs143501729	Ala [A]/Glu [E]	454	Benign	0.014

Variations with probabilistic score above 0.85 and 0.15 are considered to be “Probably damaging” and “possibly damaging” respectively while all the resting are categorized to be “Benign”.

disease susceptibility, disease progression as well as the treatment responses (De Baets *et al.,* 2012; Noreen *et al.,* 2015). Molecular association studies emphasize on nsSNPs involving substitution of amino acids. These nsSNPs of coding region draw more attention of researchers as these can result in structural and functional alterations of proteins. Detailed information in this concern can be obtained using sequence and structure based tools of bioinformatics to select crucial and functional nsSNPs of *c-Myc* to get into deeply detailed research.

In Burkitt lymphoma samples the *c-Myc2* proto-oncogene contained a cluster of four discussable SNPs in 2nd and 3rd exon (Rabbitts *et al.,* 1983; Bhatia *et al.,* 1993). VAR_063384 (Glu39Asp) involves substitution of Glutamate (E) to Aspartate (D) at position 39. Both amino acids are medium sized and acidic with similar physico-chemical properties. VAR_063385 (Pro57Ser) involves substitution of medium sized and hydrophobic Proline (P) to small sized and polar Serine (S) at position 57. VAR_063386 (Pro59Ala) involves substitution of medium sized and hydrophobic Proline (P) to small sized and hydrophobic Alanine (A) at 59th amino acid. VAR_063387 (Asn86Thr) involves substitution Asparagine (N) to Threonine (T) at position 86. Both amino acid residues are of medium size and polar with similar physico-chemical properties. VAR_063385 (Pro57Ser) and VAR_063386 (Pro59Ala) at position 57 and 59 respectively are of more importance because these have an adjacent threonine residue at position 58. Phosphorylation of serine at position 62 by extracellular signal regulated kinase (ERK) or CDK kinases increases

Table III.- SNPeffect analysis of nsSNPs in coding region of human c-Myc1 protein

dbSNP id	dTANGO	Aggregation tendency	dWALTZ	Amyloid propensity	dLIMBO	Chaperone binding tendency
rs139294902	0.38	No effect	-0.05	No effect	0.00	No effect
rs146505192	-1.21	No effect	-0.93	No effect	0.00	No effect
rs148915481	-0.01	No effect	-0.36	No effect	0.00	No effect
rs4645959	0.00	No effect	-0.06	No effect	0.00	No effect
rs114570780	0.04	No effect	-10.16	No effect	0.00	No effect
rs148228388	-14.97	No effect	0.00	No effect	0.00	No effect
rs199561469	0.41	No effect	-0.05	No effect	0.00	No effect
rs150308400	9.73	No effect	0.24	No effect	0.00	No effect
rs137906262	0.00	No effect	8.63	No effect	0.00	No effect
rs61755060	0.04	No effect	-0.28	No effect	0.00	No effect
rs147329312	-0.41	No effect	0.05	No effect	0.00	No effect
rs4645960	0.00	No effect	0.00	No effect	0.00	No effect
rs4645960	0.41	No effect	-0.05	No effect	0.00	No effect
rs4645961	0.00	No effect	-0.00	No effect	0.00	No effect
rs112602073	0.00	No effect	0.00	No effect	0.00	No effect
rs147506213	0.00	No effect	0.07	No effect	0.00	No effect
rs186663828	0.00	No effect	0.04	No effect	0.00	No effect
rs148544254	0.00	No effect	0.00	No effect	0.00	No effect
rs150629172	0.04	No effect	-0.00	No effect	0.00	No effect
rs139697494	-0.05	No effect	0.01	No effect	0.00	No effect
rs146971340	-0.05	No effect	1.21	No effect	0.00	No effect
rs4645968	0.00	No effect	0.00	No effect	0.00	No effect
rs200431478	0.00	No effect	0.00	No effect	0.00	No effect
rs141095253	1.17	No effect	0.06	No effect	-277.70	Decrease
rs145561065	0.00	No effect	0.16	No effect	-0.05	No effect
rs148863193	-0.68	No effect	-0.02	No effect	0.02	No effect
rs201337668	0.00	No effect	-0.00	No effect	0.00	No effect
rs143501729	0.00	No effect	0.00	No effect	0.00	No effect

Variations with dTANGO, dWALTZ and dLIMBO between -50 and 50 are supposed to have no effect on aggregation tendency, amyloid propensity and chaperone binding tendency, respectively.

Table IV.- Combined score of SIFT, PolyPhen and SNPeffect prediction

dbSNP id	Amino acid position	SIFT prediction	PolyPhen prediction	SNPeffect prediction	Combined score
rs139294902	12	1	0	0	1
rs4645959	26	1	0	0	1
rs114570780	47	1	1	0	2
rs148228388	91	0	1	0	1
rs199561469	104	0	1	0	1
rs150308400	148	1	1	0	2
rs137906262	159	1	1	0	2
rs147329312	172	0	1	0	1
rs200431478	362	1	1	0	2
rs141095253	397	0	1	1	2
rs145561065	434	1	0	0	1
rs143501729	454	1	0	0	1

dbSNPid, amino acid position and combined score of nsSNPs in coding region of human c-Myc1 protein predicted to have some biological role using SIFT, PolyPhen and SNPeffect.

Myc cellular stability. Then phosphorylation of threonine at position 58 is brought about by GSK3-β which triggers dephosphorylation of S62 by protein phosphatase 2A (PP2A) and subsequently ubiquitin-mediated degradation is brought about through SCF-Fbwx7 (ubiquitin ligases) (Hann, 2006). Thus any variation in this domain can gain some biological importance affecting this series of phosphorylation, dephosphorylation and ubiquitinations.

Out of total 165 SNPs related to longest isoform of human *c-Myc*, 27 were nsSNPs found in coding region along with two frame shift mutations; rs35631115 and rs67856294. Our main emphasis was to evaluate the possible role of nsSNP in the coding region. One SNP dbSNPid: rs4645960 found at position 175 has Gly [G] amino acid in wild type protein, which can be substituted by either Cys [C] or by Arg [R]. In our study we considered both variants separately to find if they have different effects.

In order to find tolerant or deleterious effect of amino acid substitution on protein function, we used the SIFT algorithm (Ng and Henikoff, 2001) which mainly aligns homologous sequences, considers physical properties of amino acids and possible effects of naturally occurring non synonymous polymorphisms on protein. Total 8 nsSNPs were found to be deleterious (Fig. 4), out of which 2 had low confidence due to scarcity of available sequences for alignment (Table I). PolyPhen-2 server (Adzhubei *et al.*, 2010) was used to forecast structural and functional effects of the SNPs of interest. PolyPhen searched for homologous protein sequences and 3D protein structures to predict the possible potential influence of variants on protein. Among total 27 variations, 7 were predicted to be probably damaging, one variant as possibly damaging while all others were considered to be benign (Table II, Fig. 4).

Effect of four nsSNPs in coding region; rs114570780 (Tyr47His), rs150308400 (Cys148Tyr), rs137906262 (Leu159Ile) and rs200431478 (Ser362Phe) were predicted by SIFT algorithm to be intolerant and by PolyPhen-2 server to be probably damaging (Table IV). Four nsSNPs; rs4645959 (Asn26Ser), rs114570780 (Tyr47His), rs150308400 (Cys148Tyr) and rs137906262 (Leu159Ile) are located within N-terminal trans-activation domain of *c-Myc1* containing 2 highly conserved domains called MBI and MBII. In *c-Myc1* the MBI will range approximately 61-78 amino acids and MBII will span region of 144-158 amino acids (Fig. 1). This region harboring MBI and MBII is essential for transactivation of target genes involved in process of apoptosis, cell proliferation and transformation (Wirtenberger *et al.*, 2005). In MBI domain a series of potential phosphorylations by MAPK, CDKs or GSK3β take place in cell cycle dependent manner.

These phosphorylation events regulate transcription and transformation. MBII encodes the hydrophobic part of the *c-Myc* protein. One transcription activation domains (TAD) is located upstream of MBI and the other is upstream and overlapping region of MBII. A repression element is identified between the MBI and MBII domains while other overlaps MBII itself (Sakamuro and Prendergast, 1999). N-terminal trans-activation domain is also reported to interact with TBP, α-Tubulin, p107 as well as adaptor proteins; TRRAP, AMY-1, Bin1, Pam, and MM-1 (Sakamuro and Prendergast, 1999; Beijersbergen *et al.*, 1994; Gu *et al*,. 1994; Hateboer *et al.*, 1993). Thus, interaction of these proteins and presence of sequences related to activation or repression activity in N-terminal trans-activation domain suggests its key role in apoptosis, chromatin modeling, cell cycle and transcriptional regulation (Sakamuro and Prendergast, 1999). Presence of genetic variants in this region can effect this regulation significantly.

The rs4645959 involves substitution of medium sized Asparagine (N) to small sized Serine (S) residue. According to SIFT Prediction this variant is intolerant (Tolerance index = 0.01). This nsSNP was reported to be putatively functional being capable of altering the secondary structure of protein. Heterozygous carriers of this variant among Polish and German individuals were observed to have an increased risk of familial breast cancer (Wirtenberger *et al.*, 2005). Two nsSNPs; rs141095253 (Pro397Leu) and rs145561065 (Leu434Phe) are found within while rs200431478 (Ser362Phe) is located near the C-terminal domain which spans 375-452 amino acid region of *c-Myc1*

(Fig. 1). This C-terminal DNA binding segment contains BR-HLH-LZ domain. Another BR-HLH-LZ domain containing protein; Max is identified as a closely related, obligate and physiological heterodimerization partener of *c-Myc* (Blackwood and Eisenman, 1991). The HLH-LZ domains of both proteins form strong heterodimers and recognize the E-Box containing promoters to regulate transcriptional activation, cellular transformation, cell cycle progression and apoptosis (Amati *et al.*, 1992, 1993). Other than Max some proteins like Nmi, TFII-I, YY-1, AP-2, BRCA1 and Miz-1 are also reported to interact with C-terminal region of *c-Myc*. Based on these interactions, C-terminal region is supposed to control the access of the N-terminal trans-activation domain to specific genetic loci (Sakamuro and Prendergast, 1999). Presence of nsSNPs in C-terminal region can alter these interactions and consequently can lead to an impaired functioning of *c-Myc* protein in cell.

However, when nsSNPs were subjected to SNPeffect server to evaluate their possible effect on molecular phenotype of protein, most of the variations were neither predicted to effect aggregation tendency and amyloid propensity nor the chaperone binding tendency of *c-Myc* protein. It seems as if these variations are really capable of bringing some damaging alterations to *c-Myc*, then these are certainly other than the aggregation tendency, amyloid propensity and the chaperone binding tendency of the protein as per prediction of SNPeffect. Only the Pro/Leu variation at position 397 (rs141095253) was shown to decreases the chaperone binding tendency and protein stability of *c-Myc* (Table III). 1NKP presents the X-ray structure of the *c-Myc* protein in complex with the Max protein and DNA including 353 to 434 residues of *c-Myc2*. Three SNPs; rs141095253 (Pro397Leu), rs145561065 (Leu434Phe) and rs148863193 (Arg439Leu) of longest isoform *i.e. c-Myc1* are shown (Fig. 6). Further analysis can be performed using some other computational tools and availability of PDB structure can help in making more precise and reliable evaluations.

Our results can be helpful and interesting especially to those epidemiologists who are interested in large-scale population based studies. Moreover, on basis of our predictions, the real effect of nsSNPs particularly rs114570780 (Tyr47His), rs150308400 (Cys148Tyr), rs137906262 (Leu159Ile) and rs200431478 (Ser362Phe) on *c-Myc* can be assured by conducting well defined *in vitro* and *in vivo* studies.

ACKNOWLEDGEMENT

We would like to acknowledge National University of Sciences and Technology (NUST) and Higher Education Commission (HEC) Islamabad Pakistan for providing assistance.

Conflict of interest

Authors do not declare any potential conflict of interest.

REFERENCES

AMATI, B., DALTON, S., BROOKS, M.W., LITTLEWOOD, T.D., EVAN, G.I. AND LAND, H., 1992. Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature*, **359**: 423-426.

AMATI, B., LITTLEWOOD, T.D., EVAN, G.I. AND LAND, H., 1993. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO J.*, **12**: 5083-5087.

ADZHUBEI, I.A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V.E., GERASIMOVA, A., BORK, P., KONDRASHOV, A.S. AND SUNYAEV, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat Methods.*, **7**: 248-249.

BEIJERSBERGEN, R.L., HIJMANS, E.M., ZHU, L. and BERNARDS, R., 1994. Interaction of c-Myc with the pRb-related protein p107 results in inhibition of c-Myc-mediated transactivation. *EMBO J.*, **13**: 4080-4086.

BERTUCCI, F., LAGARDE, A., FERRARI, A., FINETTI, P., CHARAFE-JAUFFRET, E., VAN LAERE, S., ADELAIDE, J., VIENS, P., THOMAS, G., BIRNBAUM, D. AND OLSCHWANG, S., 2012. 8q24 Cancer risk allele associated with major metastatic risk in inflammatory breast cancer. *PLoS One*, **7**: e37943.

BHATIA, K., HUPPI, K., SPANGLER, G., SIWARSKI, D., IYER, R. AND MAGRATH, I., 1993. Point mutations in the c-Myc transactivation domain are common in Burkitt's lymphoma and mouse plasmacytomas. *Nat Genet,*. **5**: 56-61.

BLACKWOOD, E.M. AND EISENMAN, R.N., 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science*, **251**: 1211-1217.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. AND MADDEN, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinform.,* **10**: 421.

COWLING, V.H., CHANDRIANI, S., WHITFIELD, M.L. AND COLE, M.D., 2006. A conserved Myc protein domain, MBIV, regulates DNA binding, apoptosis, transformation, and G2 arrest. *Mol. Cell Biol.,* **26**: 4226-4239.

DE ALENCAR, S.A. AND LOPES, J.C., 2010. A comprehensive in silico analysis of the functional and structural impact of SNPs in the IGF1R gene. *J. Biomed. Biotechnol.,* **2010**: 715139.

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. *Nucl. Acids Res.,* **40**: D935-939.

DESHPANDE, N., ADDESS, K.J., BLUHM, W.F., MERINO-OTT, J.C., TOWNSEND-MERINO, W., ZHANG, Q., KNEZEVICH, C., XIE, L., CHEN, L., FENG, Z., GREEN, R.K., FLIPPEN-ANDERSON, J.L., WESTBROOK, J., BERMAN, H.M. AND BOURNE, P.E., 2005. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucl. Acids Res.,* **33**: D233-237.

EASTON, D.F., POOLEY, K.A., DUNNING, A.M., PHAROAH, P.D., THOMPSON, D., BALLINGER, D.G., STRUEWING, J.P., MORRISON, J., FIELD, H., LUBEN, R., WAREHAM, N., AHMED, S., HEALEY, C.S., BOWMAN, R., MEYER, K.B., HAIMAN, C.A., KOLONEL, L.K., HENDERSON, B.E., LE MARCHAND, L., BRENNAN, P., SANGRAJRANG, S., GABORIEAU, V., ODEFREY, F., SHEN, C.Y., WU, P.E., WANG, H.C., ECCLES, D., EVANS, D.G., PETO, J., FLETCHER, O., JOHNSON, N., SEAL, S., STRATTON, M.R., RAHMAN, N., CHENEVIX-TRENCH, G., BOJESEN, S.E., NORDESTGAARD, B.G., AXELSSON, C.K., GARCIA-CLOSAS, M., BRINTON, L., CHANOCK, S., LISSOWSKA, J., PEPLONSKA, B., NEVANLINNA, H., FAGERHOLM, R., EEROLA, H., KANG, D., YOO, K.Y., NOH, D.Y., AHN, S.H., HUNTER, D.J., HANKINSON, S.E., COX, D.G., HALL, P., WEDREN, S., LIU, J., LOW, Y.L., BOGDANOVA, N., SCHURMANN, P., DORK, T., TOLLENAAR, R.A., JACOBI, C.E., DEVILEE, P., KLIJN, J.G., SIGURDSON, A.J., DOODY, M.M., ALEXANDER, B.H., ZHANG, J., COX, A., BROCK, I.W., MACPHERSON, G., REED, M.W., COUCH, F.J., GOODE, E.L., OLSON, J.E., MEIJERS-HEIJBOER, H., VAN DEN OUWELAND, A., UITTERLINDEN, A., RIVADENEIRA, F., MILNE, R.L., RIBAS, G., GONZALEZ-NEIRA, A., BENITEZ, J., HOPPER, J.L., MCCREDIE, M., SOUTHEY, M., GILES, G.G., SCHROEN, C., JUSTENHOVEN, C., BRAUCH, H., HAMANN, U., KO, Y.D., SPURDLE, A.B., BEESLEY, J., CHEN, X., MANNERMAA, A., KOSMA, V.M., KATAJA, V., HARTIKAINEN, J., DAY, N.E., COX, D.R. AND PONDER, B.A., 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* **447**: 1087-1093.

FERNANDEZ-ESCAMILLA, A.M., ROUSSEAU, F., SCHYMKOWITZ, J. AND SERRANO, L., 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.,* **22**: 1302-1306.

FIGUEIREDO, J.C., KNIGHT, J.A., CHO, S., SAVAS, S., ONAY, U.V., BRIOLLAIS, L., GOODWIN, P.J., MCLAUGHLIN, J.R., ANDRULIS, I.L. AND OZCELIK, H., 2007. Polymorphisms cMyc-N11S and p27-V109G and breast cancer risk and prognosis. *BMC Cancer,* **7**: 99. DOI: 10.1186/1471-2407-7-99

GONG, W.F., ZHONG, J.H., XIANG, B.D., MA, L., YOU, X.M., ZHANG, Q.M. AND LI, L.Q., 2013. Single nucleotide polymorphism 8q24 rs13281615 and risk of breast cancer: meta-analysis of more than 100,000 cases. *PLoS One.,* **8**: e60108.

GU, W., BHATIA, K., MAGRATH, I.T., DANG, C.V. and DALLA-FAVERA, R., 1994. Binding and suppression of the Myc transcriptional activation domain by p107. *Science,* **264**: 251-254.

GUL, A., KHAN, M.A., ZUBAIR, M., AHMAD, AHMAD, I. and KHAN, M.A., 2015. Prioritization of Disease Prone Exons in INPP5E Gene, Associated With Joubert Syndrome, by in silico Analysis of Non-Synonymous SNPs. *Pakistan J. Zool.,* **47**: 601-605.

GUO, Y., FANG, J., LIU, Y., SHENG, H.H., ZHANG, X.Y., CHAI, H.N., JIN, W., ZHANG, K.H., YANG, C.Q. AND GAO, H.J., 2011. Association between polymorphism rs6983267 and gastric cancer risk in Chinese population. *World J. Gastroenterol.,* **17**: 1759-1765.

HAIMAN, C.A., PATTERSON, N., FREEDMAN, M.L., MYERS, S.R., PIKE, M.C., WALISZEWSKA, A., NEUBAUER, J., TANDON, A., SCHIRMER, C., MCDONALD, G.J., GREENWAY, S.C., STRAM, D.O., LE MARCHAND, L., KOLONEL, L.N., FRASCO, M., WONG, D., POOLER, L.C., ARDLIE, K., OAKLEY-GIRVAN, I., WHITTEMORE, A.S., COONEY, K.A., JOHN, E.M., INGLES, S.A., ALTSHULER, D., HENDERSON, B.E. AND REICH, D., 2007. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.,* **39**: 638-644.

HANN, S.R., 1995. Methionine deprivation regulates the translation of functionally-distinct c-Myc proteins. *Adv. exp. med. Biol.,* **375**: 107-116.

HANN, S.R., 2006. Role of post-translational modifications in regulating c-Myc proteolysis, transcriptional activity and biological function. *Semin Cancer Biol.,* **16**: 288-302.

HANN, S.R., KING, M.W., BENTLEY, D.L., ANDERSON,

C.W. AND EISENMAN, R.N., 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell*, **52**: 185-195.

HATEBOER, G., TIMMERS, H.T., RUSTGI, A.K., BILLAUD, M., VAN 'T VEER, L.J. AND BERNARDS, R., 1993. TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. *Proc. natl. Acad. Sci. USA*, **90**: 8489-8493.

HENRIKSSON, M. AND LUSCHER, B., 1996. Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv. Cancer Res.*, **68**: 109-182.

KATOH, K., MISAWA, K., KUMA, K. AND MIYATA, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.*, **30**: 3059-3066.

LEE, L.A. AND DANG, C.V., 2006. Myc target transcriptomes. *Curr. Top. Microbiol. Immunol.*, **302**: 145-167.

LEMAITRE, J.M., BUCKLE, R.S. AND MECHALI, M., 1996. c-Myc in the control of cell proliferation and embryonic development. *Adv. Cancer Res.*, **70**: 95-144.

MAURER-STROH, S., DEBULPAEP, M., KUEMMERER, N., LOPEZ DE LA PAZ, M., MARTINS, I.C., REUMERS, J., MORRIS, K.L., COPLAND, A., SERPELL, L., SERRANO, L., SCHYMKOWITZ, J.W. AND ROUSSEAU, F., 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods*, **7**: 237-242.

MEYER, A., SCHURMANN, P., GHAHREMANI, M., KOCAK, E., BRINKHAUS, M.J., BREMER, M., KARSTENS, J.H., HAGEMANN, J., MACHTENS, S. AND DORK, T., 2009. Association of chromosomal locus 8q24 and risk of prostate cancer: a hospital-based study of German patients treated with brachytherapy. *Urol. Oncol.*, **27**: 373-376.

MEYER, N. AND PENN, L.Z., 2008. Reflecting on 25 years with MYC. *Nat. Rev. Cancer*, **8**: 976-990.

MILLER, D.M., THOMAS, S.D., ISLAM, A., MUENCH, D. AND SEDORIS, K., 2012. c-Myc and cancer metabolism. *Clin. Cancer Res.*, **18**: 5546-5553.

NG, P.C. AND HENIKOFF, S., 2001. Predicting deleterious amino acid substitutions. *Genome Res.*, **11**: 863-874.

NG, P.C. AND HENIKOFF, S., 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucl. Acids Res.*, **31**: 3812-3814.

NOREEN, M., SHAH, M.A., MALL, S.M., CHOUDHARY, S., HUSSAIN, T., AHMED, I., JALIL, S.F. AND RAZA, M.I., 2012. TLR4 polymorphisms and disease susceptibility. *Inflamm Res.*, **61**: 177-188.

NOREEN, M. and ARSHAD, M., 2015. Association of TLR1, TLR2, TLR4, TLR6, and TIRAP polymorphisms with disease susceptibility. *Immunol Res.*, **62**: 234-252.

OS'KINA, N.A., BOIARSKIKH, U.A., LAZAREV, A.F., PETROVA, V.D., GANOV, D.I., TONACHEVA, O.G., LIFSHITS, G.I. AND FILIPENKO, M.L., 2012. [Association of chromosome 8q24 variants with prostate cancer risk in the Siberian region of Russia and meta-analysis]. *Mol. Biol. (Mosk).* **46**: 234-241.

OSTER, S.K., HO, C.S., SOUCIE, E.L. AND PENN, L.Z., 2002. The myc oncogene: MarvelouslY Complex. *Adv. Cancer Res.*, **84**: 81-154.

PEI, Y.L., ZHANG, H.L. AND HAN, H.G., 2013. Polymorphism of 8q24 rsl3281615 and breast cancer risk : a meta-analysis. *Tumour Biol.*, **34**: 421-428.

POMERANTZ, M.M., AHMADIYEH, N., JIA, L., HERMAN, P., VERZI, M.P., DODDAPANENI, H., BECKWITH, C.A., CHAN, J.A., HILLS, A., DAVIS, M., YAO, K., KEHOE, S.M., LENZ, H.J., HAIMAN, C.A., YAN, C., HENDERSON, B.E., FRENKEL, B., BARRETINA, J., BASS, A., TABERNERO, J., BASELGA, J., REGAN, M.M., MANAK, J.R., SHIVDASANI, R., COETZEE, G.A. AND FREEDMAN, M.L., 2009. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.*, **41**: 882-884.

POWERS, E.T., MORIMOTO, R.I., DILLIN, A., KELLY, J.W. AND BALCH, W.E., 2009. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.*, **78**: 959-991.

RABBITTS, T.H., HAMLYN, P.H. AND BAER, R., 1983. Altered nucleotide sequences of a translocated c-myc gene in Burkitt lymphoma. *Nature*, **306**: 760-765.

RAMENSKY, V., BORK, P. AND SUNYAEV, S., 2002. Human non-synonymous SNPs: server and survey. *Nucl. Acids Res.*, **30**: 3894-3900.

SAKAMURO, D. AND PRENDERGAST, G.C., 1999. New Myc-interacting proteins: a second Myc network emerges. *Oncogene*, **18**: 2942-2954.

SAVAS, S., KIM, D.Y., AHMAD, M.F., SHARIFF, M. AND OZCELIK, H., 2004. Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. *Cancer Epidemiol. Biomark. Prev.*, **13**: 801-807.

SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. AND SERRANO, L., 2005. The FoldX web server: an online force field. *Nucl. Acids Res.*, **33**: W382-388.

SOLE, X., HERNANDEZ, P., DE HEREDIA, M.L., ARMENGOL, L., RODRIGUEZ-SANTIAGO, B., GOMEZ, L., MAXWELL, C.A., AGUILO, F., CONDOM, E., ABRIL, J., PEREZ-JURADO, L., ESTIVILL, X., NUNES, V., CAPELLA, G., GRUBER, S.B., MORENO, V. AND PUJANA, M.A., 2008. Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility. *BMC Genomics*, **9**: 12.

SPOTTS, G.D., PATEL, S.V., XIAO, Q. AND HANN, S.R., 1997. Identification of downstream-initiated c-Myc

proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. *Mol. Cell Biol.*, **17**: 1459-1468.

TAKATSUNO, Y., MIMORI, K., YAMAMOTO, K., SATO, T., NIIDA, A., INOUE, H., IMOTO, S., KAWANO, S., YAMAGUCHI, R., TOH, H., IINUMA, H., ISHIMARU, S., ISHII, H., SUZUKI, S., TOKUDOME, S., WATANABE, M., TANAKA, J., KUDO, S.E., MOCHIZUKI, H., KUSUNOKI, M., YAMADA, K., SHIMADA, Y., MORIYA, Y., MIYANO, S., SUGIHARA, K. AND MORI, M., 2013. The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. *Ann. Surg. Oncol.*, **20**: 1395-1402.

THOMPSON, E.B., 1998. The many roles of c-Myc in apoptosis. *Annu. Rev. Physiol.*, **60**: 575-600.

THOMPSON, J.D., PRIGENT, V. AND POCH, O., 2004. LEON: multiple aLignment Evaluation Of Neighbours. *Nucl. Acids Res.*, **32**: 1298-1307.

TOMLINSON, I., WEBB, E., CARVAJAL-CARMONA, L., BRODERICK, P., KEMP, Z., SPAIN, S., PENEGAR, S., CHANDLER, I., GORMAN, M., WOOD, W., BARCLAY, E., LUBBE, S., MARTIN, L., SELLICK, G., JAEGER, E., HUBNER, R., WILD, R., ROWAN, A., FIELDING, S., HOWARTH, K., SILVER, A., ATKIN, W., MUIR, K., LOGAN, R., KERR, D., JOHNSTONE, E., SIEBER, O., GRAY, R., THOMAS, H., PETO, J., CAZIER, J.B. AND HOULSTON, R., 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.*, **39**: 984-988.

VAN DURME, J., MAURER-STROH, S., GALLARDO, R., WILKINSON, H., ROUSSEAU, F. AND SCHYMKOWITZ, J., 2009. Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS. Comput. Biol.*, **5**: e1000475.

VENNSTROM, B., SHEINESS, D., ZABIELSKI, J. AND BISHOP, J.M., 1982. Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29. *J. Virol.*, **42**: 773-779.

WANG, Y.P., ZHANG, J., ZHU, H.Y., QIAN, C.L., LIU, H., JI, F. AND SHEN, Z.Y., 2014. Common variation rs6983267 at 8q24.1 and risk of colorectal adenoma and cancer: evidence based on 31 studies. *Tumour Biol.*, **35**: 4067-4075.

WICKER, N., PERRIN, G.R., THIERRY, J.C. AND POCH, O., 2001. Secator: a program for inferring protein subfamilies from phylogenetic trees. *Mol. Biol. Evol.*, **18**: 1435-1441.

WIRTENBERGER, M., HEMMINKI, K., FORSTI, A., KLAES, R., SCHMUTZLER, R.K., GRZYBOWSKA, E., BERMEJO, J.L., WAPPENSCHMIDT, B., BUGERT, P., BUTKIEWICZ, D., PAMULA, J., PEKALA, W., ZIENTEK, H., BARTRAM, C.R. AND BURWINKEL, B., 2005. c-MYC Asn11Ser is associated with increased risk for familial breast cancer. *Int. J. Cancer*, **117**: 638-642.

ZHU, Y., SPITZ, M.R., AMOS, C.I., LIN, J., SCHABATH, M.B. AND WU, X., 2004. An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. *Cancer Res.*, **64**: 2251-2257.

(Received 25 August 2014, revised 20 May 2015)