

Integrative analyses of conserved WNT clusters and their co-operative behaviour in human breast cancer

Qurrat-ul-Ain¹, Umair Seemab^{1,2}, Sulaman Nawaz^{1,3}, Sajid Rashid^{1*}

¹National Centre for Bioinformatics, Quaid-i-Azam University Islamabad, Pakistan; ²Ludwig-Maximilians-Universität München, Pathologisches Institut der LMU, Thalkirchner Str. 36, D-80337 München Germany; ³Department of BioSciences, COMSATS Institute of Information Technology, Park Road Chak Shahzad Islamabad Pakistan; Sajid Rashid - Email: sajid@qau.edu.pk; Phone: +92-05190644107; *Corresponding author

Received November 05, 2011; Accepted November 16, 2011; Published December 10, 2011

Abstract:

In human, *WNT* gene clusters are highly conserved at specie level and associated with carcinogenesis. Among them, *WNT-10A* and *WNT-6* genes clustered in chromosome 2q35 are homologous to *WNT-10B* and *WNT-1* located in chromosome 12q13, respectively. In an attempt to study co-regulation, the coordinated expression of these genes was monitored in human breast cancer tissues. As compared to normal tissue, both *WNT-10A* and *WNT-10B* genes exhibited lower expression while *WNT-6* and *WNT-1* showed increased expression in breast cancer tissues. The co-expression pattern was elaborated by detailed phylogenetic and syntenic analyses. Moreover, the intergenic and intragenic regions for these gene clusters were analyzed for studying the transcriptional regulation. In this context, adequate conserved binding sites for SOX and TCF family of transcriptional factors were observed. We propose that SOX9 and TCF4 may compete for binding at the promoters of *WNT* family genes thus regulating the disease phenotype.

Keywords: WNT; breast cancer; TCF; SOX

Background:

Breast cancer is one of the major causes of death in human. The incidence of breast cancer is increasing since 1960s with the reduction in survival rate [1]. While most breast cancers arise sporadically, approximately 10% of all cases are associated with hereditary cancer [2]. In recent years, several gene mutations have been identified to be responsible for familial breast cancers [3]. Among them the oncogenic mutations leading to cancer mostly target the genes encoding receptors or downstream signaling components. Many reports suggest that dysfunction of *WNT* signaling pathway is one of the leading cause of cancer in humans [4]. *WNT* was first identified as an oncogene in mouse mammary tumours in early 1980s postulating that *WNT* proteins or components of their signaling pathway could be the underlying cause of breast cancer in humans [5]. *WNT* proteins belong to a large family of secreted proteins that act as extracellular signaling factors. These growth factors are highly conserved and influence many cellular functions such as differentiation, migration, proliferation, apoptosis, and stem cell renewal [6-8]. Altogether, there exist 19 *WNT* genes that encode highly similar proteins. The dynamic expression pattern of

WNT genes may induce multiple signaling pathways depending upon the receptor involved, co-receptor expression and the intracellular signaling proteins [9-11]. However there lies a functional redundancy among the *WNT* family members [12].

Notably, *WNT* genes have rarely been found to be mutated in cancer; however, mutations routinely occur at their downstream targets [13]. For example, 85% of colon cancer cases were due to the loss of function mutations in APC resulting in the elevated β -catenin levels [4]. Similarly, loss of Axin-1 resulted in hepatocellular carcinoma [14]. In breast carcinoma cells, expression of β -catenin is less with poor prognosis rate. Moreover, β -catenin is not localized into the nucleus in breast cancer, indicating that *WNT* signaling may be uncommon in breast tumours. However, other studies reported a positive role of *WNT* signaling in breast cancer [15] and levels of *WNTs* or other components of *WNT* pathway are known to be altered in 50% of breast cancer cases. During the recent years, an increasing number of *WNT* genes have been implicated in the development of mammary gland. These include *WNT-2*,

WNT-4, *WNT-5A*, *WNT-5B*, *WNT-6*, *WNT-7B*, *WNT-10A* and *WNT-10B* [16]. In mouse, several *WNT* genes including *WNT-10A*, *WNT-10B* and *WNT-6* express within mammary line between E11.25 and E11.5 (40-42 somite stage) [17]. More often, *WNT* genes regulate the developmental and morphological aspects of mammary gland, suggesting that malfunctioning of these genes may cause breast cancer. For example, *WNT-2* that normally expresses in stromal compartment was found to be involved in breast carcinoma [18]. Similarly, *WNT-5A* and *WNT-7B* over-expression in mammary tumorigenesis had also been reported [19]. It is quite surprising that in human chromosomes, most of *WNT* genes are located in the form of clusters and are well conserved. For example, *WNT-10A*, *WNT-10B*, *WNT-3A* and *WNT-3* genes that are located in chromosomal regions 2q35, 12q13, 1q42 and 17q21 are homologous to *WNT-6*, *WNT-1*, *WNT-9A* and *WNT-9B*, respectively. These data reveal that ancestrally they were clustered together and duplicated or diverse later on during the course of evolution. In the current study, the expression pattern of two of these conserved clusters, *WNT-1/-10B* and *WNT-6/-10A* was ascertained in human breast carcinoma tissues in parallel to their detailed phylogenetic and syntenic analyses. The combined *in-silico* data enlighten the facts that these genes are not only syntenic but also functionally co-regulate each other. Also, their intergenic regions were analysed for finding the binding sites of regulatory elements. The study was further broadened by increasing the number of species in phylogenetic analysis as well as in the syntenic and intergenic analysis. In order to learn the phylogeny of all 19 *WNT* paralogs, almost all species from every clad were included to examine the duplication and speciation events.

Methodology:

Sequence isolation

From the NCBI database, putative sequences of human *WNT* family members including *WNT-1*, *WNT-2*, *WNT-2B*, *WNT-3*, *WNT-3A*, *WNT-4*, *WNT-5A*, *WNT-5B*, *WNT-6*, *WNT-7A*, *WNT-7B*, *WNT-8A*, *WNT-8B*, *WNT-9A*, *WNT-9B*, *WNT-10A*, *WNT-10B*, *WNT-11* and *WNT-16* were retrieved. Genbank accession numbers of *WNT-1*, *WNT-6*, *WNT-10A* and *WNT-10B* mRNAs are [GenBank:X03072, GenBank:AF079522, GenBank:AB059569 and GenBank:X97057], respectively. Their chromosomal locations and detailed gene annotation studies were performed by using ENSEMBL genome browser. The Gene Ontology (GO) hierarchy annotations were downloaded from the Gene Ontology database.

Phylogenetic analysis

By exploiting the knowledge of 19 paralogs of *WNT* gene family, we studied the evolutionary relationship among these paralogs. For each paralog, the closest putative orthologous sequence was collected using ENSEMBL BLASTP from various vertebrates representing Primates, Rodents, Mammals, and Fish. Orthologous sequences from tunicates (closest relative of vertebrate) and deuterostomes were also assembled to perform a more robust and thorough phylogenetic analysis. Species included in this study are Human (*Homo sapien*), Macaque (*Macaca mullata*), Mouse (*Mus musculus*), Rat (*Rattus norvegicus*), cattle (*Bos Taurus*), Zebra fish (*Danio rerio*), Puffer fish (*Tetradon nigroviridis*), Fruitfly (*Drosophila melanogaster*) and Nematode (*Caenorhabditis elegans*). Multiple sequence alignment (MSA) was performed by ClustalW with default settings and manually

refined the alignment by removing the common gaps of the sequences. MEGA 5.0 was used to construct the trees by using distance based method, Neighbour joining Algorithm.

Tissue Samples collection

Freshly excised breast tumour tissues and adjacent normal breast tissue samples were obtained by surgical operation from 80 volunteers (30-50 years) at the affiliated hospital (Pakistan Institute of Medical Sciences, Islamabad). These tissues were frozen in liquid nitrogen as soon as mastectomy was performed.

RNA Extraction, cDNA Synthesis and RT PCR analysis

Total RNA was extracted from normal as well as cancerous breast tissues by using the RNA mini kit according to the manufacturer's instructions (Invitrogen). Quantification of extracted RNA was performed by nanodrop (ND-1000) and readjusted by DEPC water. The integrity of RNA samples was tested by formaldehyde gel electrophoresis. Subsequently, cDNA synthesis was carried out from 1 µg of total RNA using the first strand cDNA synthesis system (Fermentas) and priming with random hexamer primers as per their standard protocol. Human β-actin primers 5'-AGCGAGCATCCCCAAAGTT-3' (forward primer) and 5'-GGGCACGAAGGCTCATCATT-3' (reverse primer) were used as positive control. The sequences for the primers used are as follows: for *WNT-1*, 5'-GGTTTCIGTACGCTGCTG-3' (forward primer) and 5'-TCTGTAAGCAGGTTCTGGGA-3' (reverse primer); for *WNT-6*, 5'-GGTTATGGACCCTACCAGCA-3' (forward primer) and 5'-GGAAGCTGGAAGCTGGCACTCT-3' (reverse primer); for *WNT-10A*, 5'-GGGAGCCTCACAGAGACATC-3' (forward primer) and 5'-CCAGCACGTCTGAGCTG-3' (reverse primer) and for *WNT-10B*, 5'-GGCTTGCTAGCTCTCTCGAT-3' (forward primer) and 5'-CTGGTTCGTGCCCCCTTGT-3' (reverse primer). Standard PCR reactions were performed by using 2µl of cDNA template and 1µl of each forward and reverse primer in a 25 µl reaction volume. Cycling conditions were 5 min at 95°C, 1 min at 59 - 61°C and 1 min at 72°C. After 35 cycles, the products were resolved by agarose gel electrophoresis and stained with ethidium bromide.

Elucidation of super secondary structures

Sequence conservation analysis of the *WNT-6/-10A* and *WNT-1/-10B* clusters were carried out carefully by visualizing the MSA in GeneDoc tool (K.B. Nicholas and H.B.J. Nicholas) and conserved regions were inspected for presence of domains or motifs. To detect the conserved domains within orthologs, PFAM, PRINTS, BLOCKS, PROSITE, PRODOM and InterProScan databases were used.

In silico Intergenic and Intragenic Study

In order to study the intergenic regions for *WNT-6/-10A* and *WNT-1/-10B* clusters, their exon-intron structures were refined in detail. Both *WNT-10A* consist of four exons with three introns, *WNT-10B* has five exons while *WNT-1* and *WNT-6* comprise four exons and three introns, respectively. Transcriptional orientation of conserved clusters was analysed by using ENSEMBL genome browser. In order to validate the co-regulation of these clusters in various species, Co-express db and GO db were used. Almost 7KB upstream and downstream regulatory sequences of all four homologs *WNT-1*, *WNT-6*,

WNT-10A and *WNT-10B* were extracted from ENSEMBL database and were analysed for the presence of transcriptional factor binding sites by using TrFAST (Transcription factor search and analysis tool) tool (unpublished data) to discover the new unified motifs in these intergenic regions. Furthermore, in order to find the methylation sites near the transcription factor binding sites, Methylator tool was used.

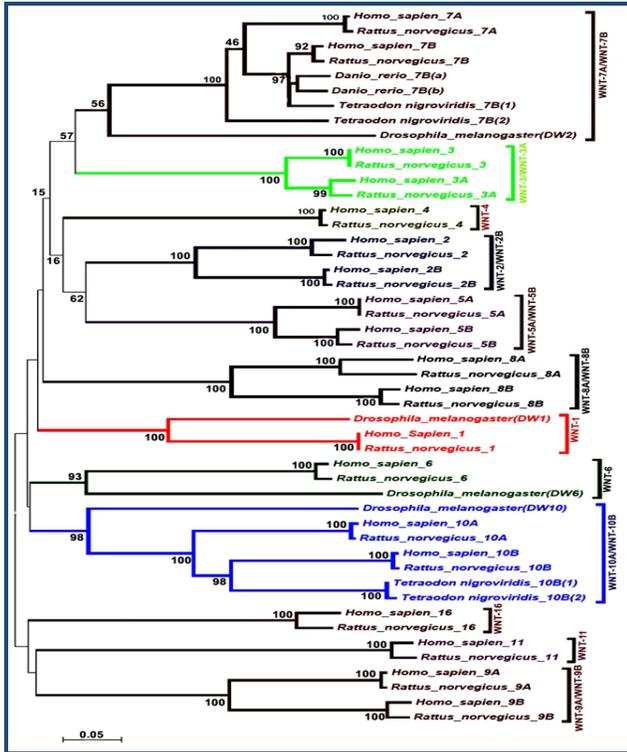


Figure 1: A comprehensive phylogenetic analysis of WNT family, displaying all 19 paralogs in Primates (*Homo sapiens*), Rodents (*Rattus norvegicus*), Fish (*Tetraodon nigroviridis*) and Deuterostome (*Drosophila melanogaster*). An un-rooted tree is constructed using MEGA 5.0 and Neighbour-joining algorithm, boot strap values are shown at each cluster validating the clustering of genes and species. P-distance is used as a measure of evolutionary distance, which also includes the correction for hidden changes.

Results and Discussion:

Phylogenetic Analysis

Our findings including orthologous sequences search, conservation pattern analysis in parallel to their genomic annotations studies, extracted newer details about WNT genes in corroboration with the previously documented data [20]. In order to resolve the phylogenetic relationship among WNT family proteins, an un-rooted tree was constructed for the 19 WNTs by neighbour joining method (Figure 1). Some of the WNTs including WNT-7A/-7B, WNT-2/-2B, WNT-3/-3A, WNT-5A/-5B, WNT-9A/-9B and WNT-8A/-8B are observed in the form of clusters showing the ancestral duplication before vertebrate-invertebrate split. It seems from the tree that one of the homolog of these clusters was present in ancestral genome and subsequently diverged. However, the other copy has evolved either through simultaneous or by whole genome duplications. The most recent cluster found in the tree consists

of WNT-7A/-7B, with *Drosophila* DW-2 making an out-group to it. Clearly, two copies of WNT-7B are distantly related to each other in *Tetraodon norvegicus* while closely related in case of *Danio rerio* (zebra fish). WNT-3/-3A being most greatly conserved with WNT-7A/-7B lie next to it, followed by WNT-4 and another cluster of WNT-2/-2B and WNT-5A/-5B. A conserved cluster of WNT-8A/-8B is found aligned with above mentioned clusters. Going deeper into tree, a cluster including WNT-10A/-10B received a spectacular bootstrap support (100%) with WNT-1 and WNT-6 falling outside. Also, the two relevant branches of WNT-1 and WNT-6 contain a very significant bootstrap values (i.e. 100% and 93% respectively). *Drosophila* contains independent copies of both WNT-6 and WNT-10, showing the exact/true relationship to the given cluster. Notably, WNT-10B is independently duplicated in case of *Tetraodon norvegicus* and WNT-16 is the closest relative of this gene cluster, obeying the true specie tree pattern. The last cluster under observation in the phylogenetic tree comprises WNT-11, WNT-9A and -9B, falls outside the whole tree (Figure 1).

The tree topology of WNT protein family is an extended form of ((AB) C) D) type topology. Phylogeny of vertebrate WNT proteins suggests that most of the duplication events occurred after vertebrate and invertebrate split and prior to fish-tetrapoda divergence. However, this observation is not valid for all 19 WNT paralogs. The duplication events which gave rise to WNT-7A/-7B, WNT-1, WNT-6 and WNT-10A/-10B seem to be occurred prior to urochordate-vertebrate split (Figure 2), thus concluding that these three genes (WNT-1, WNT-6, WNT-10) were conserved in ancestor and remained conserved till *Homo sapiens* together with the orientation of these clusters. Altogether, we hypothesize that WNT-1, WNT-6, WNT-10A and WNT-10B have been originated due to ancestral duplication events and are involved in various basic developmental processes or signaling pathways.

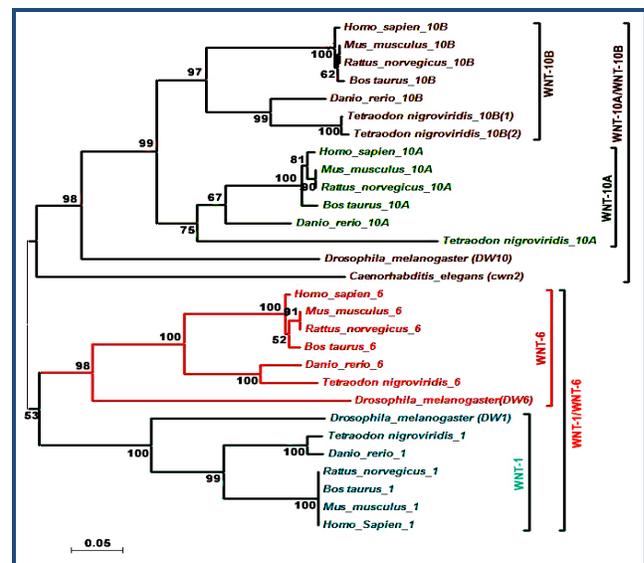


Figure 2: Elaborated phylogenetic history of WNT-6/-1 and WNT-10A/-10B clusters closely related to each other from *Homo sapiens* to *Caenorhabditis elegans*. Independent duplication is observed in case of *Tetraodon norvegicus* WNT-10B. Both clusters illustrate the true reconciliation with species tree. Moreover, the

tree clearly indicates the fact that there is an independent copy of *WNT-1*, *WNT-10* and *WNT-6* in *Drosophila melanogaster*, which depicts the ultra-conservation of this cluster in invertebrates along with its chromosomal orientation. Un-rooted tree is constructed using MEGA 5.0 and one of distance based method (Neighbour Joining Algorithm). Bootstrap values are shown for individual branches to confirm the validity of clustering.

Synthetic Analysis

Both clusters *WNT-10A/6* and *WNT-1/10B* lie at the same loci on human chromosomes 2q35 and 12q13, respectively. In order to explore the synteny of these four genes (*WNT-1*, *WNT-6*, *WNT-10A*, *WNT-10B*), ~1MB locus of human chromosome 2 and 12 was studied and compared with macaque and mouse, which revealed that all the neighbouring genes were almost similar in organization and at sequence level. By performing MSA of *WNT* orthologs, several conserved elements had been isolated that were found to be necessary for multiple co-regulatory functions. Higher syntenic conservation was achieved in case of *WNT-6/10A* cluster as compared to *WNT-1/10B* in vertebrates that was consisted of about 105 MB syntenic region at chromosome 2q35, while the other cluster *WNT-1/10B* exhibited approximately 17 MB region conservation with its ortholog (Figure 3A and B).

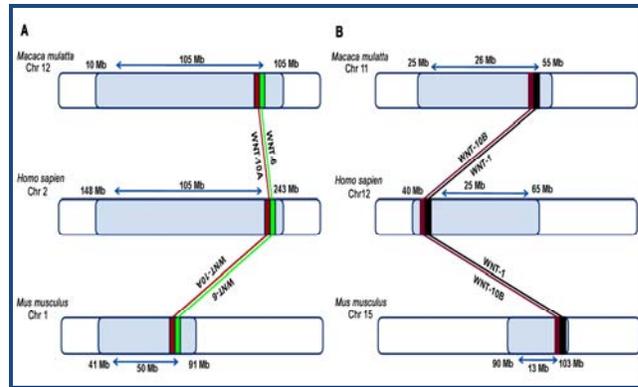


Figure 3: Chromosomal location of two clusters of *WNT* (*WNT-1/6* and *WNT-10A/10B*) and their surrounding syntenic regions in Human, Macaque and Mouse. (A) 105 MB region of chromosome 2 in *Homo sapiens* is conserved with *Macaca mulatta* (chromosome 12). 50 MB region of *Mus musculus* chromosome 1 has similar gene order syntenic with this 105 MB region. This conserved region in *Homo sapiens*, *Mucaca mulatta* and *Mus musculus* contains a cluster of *WNT-10A* and *WNT-6* (B) Chromosomal location of *WNT-1* and *WNT-10B* genes in *Homo sapiens*, *Macaca mulatta* and *Mus musculus*. The conserved region, approximately 17MB (in blue) contains the other two *WNT* genes under study in form of a cluster.

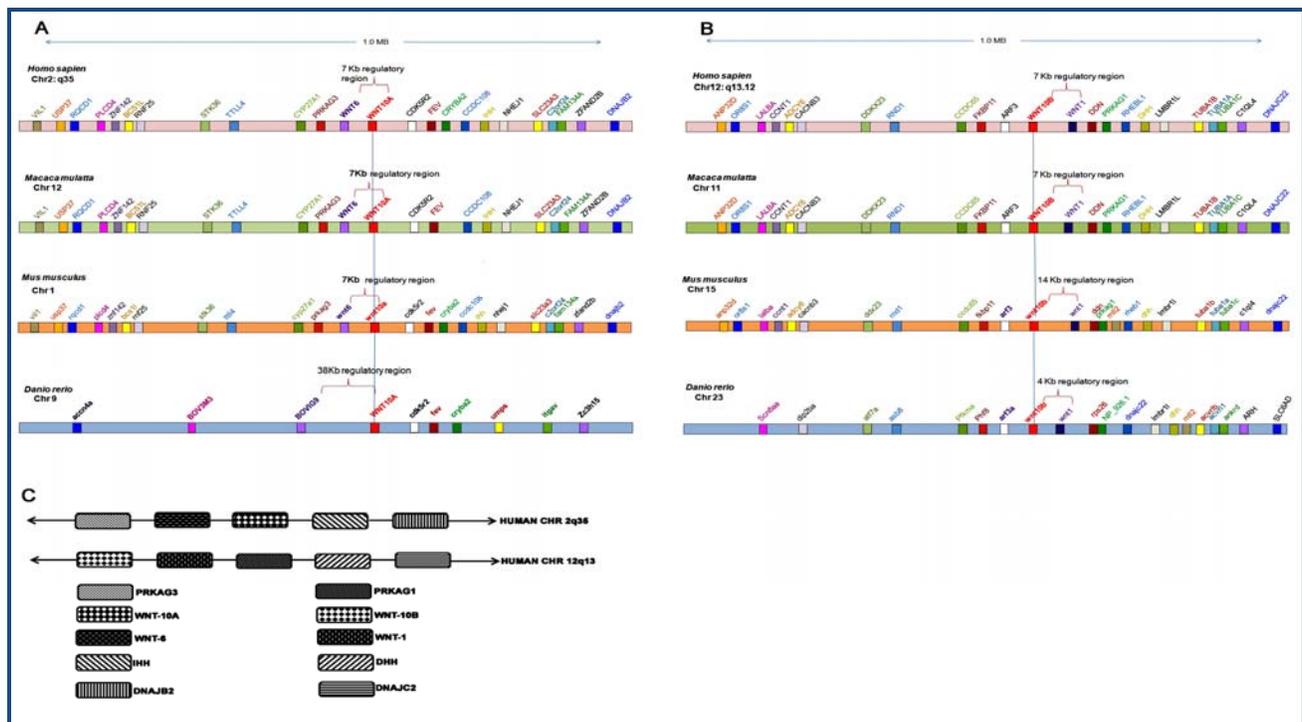


Figure 4: Synteny of a conserved locus (1 MB in size) in *Homo sapiens*, *Macaca mulatta*, *Mus musculus* and *Danio rerio*. (A) *WNT-10A* and *WNT-6*, seems to remain conserved in all species. (B) Shows the other homologous cluster of *WNT-1* and *WNT-10B*. Interesting information here is the conserved regulatory region between *WNT-10A* and *WNT-6* and *WNT-10B* and *WNT-1*. This region is almost 7KB in Human, Mouse and Macaque. However in case of Zebra fish it is of 38KB, which means that independent duplications have occurred in this particular intergenic sequence, which is strange and under study. (C) Chromosomal co-linearity of homologous *WNT* clusters in *Homo sapiens*. *WNT10-A* is homologous to *WNT-10B*, while *WNT-6* is homologous to *WNT-1* at human chromosomes 2 and 12, respectively. Along with these, there are also some other genes whose genetic paralogs exist on these locus including *PRKAG3*, *IHH* and *DNAJB2*, despite of the fact that their exact orientation is not same.

Interesting fact about these syntenic regions is that they also comprise similar pattern of organization for the neighbouring genes and their respective genetic paralogs including *PRKAG3*, *DHH*, *IHH*, *DNAJB2*, *DNAJC2* and *LMBR1*. Surprisingly, the locations of these *WNT* clusters differ between lineages. For example, *WNT-6* is located upstream of *WNT-10A*; while its homolog *WNT-1* is located downstream to *WNT-10B* (Figure 4A, B). Similar is the case with *PRKAG3* and its genetic paralog *PRKAG1*. *IHH*, paralog of *DHH* is exhibiting the similar organization but it lacks *LMBR1* gene, which is present downstream to *DHH* (Figure 4C). Three genetic paralogs of *TUBA* gene (*TUBA1A*, *TUBA1B* and *TUBA1C*) are present at *WNT-1* gene locus while absent in case of *WNT-6* locus.

Functional Annotation Assay

Protein sequences of *WNT-1*, *WNT-6*, *WNT-10A* and *WNT-10B* collected from Human (*Homo sapiens*), Mouse (*Mus musculus*), Macaque (*Macaca mullata*) and Zebrafish (*Danio rerio*) were subsequently aligned. Based on the multiple sequence alignment (MSA), it is evident that there exists more than 80% conservation at sequence level among the selected species. Figure 5A, B and C represent the MSA of *WNT-1* and *WNT-6* proteins while Figure 6A, B and C represent the MSA of *WNT-10A* and *WNT-10B*. These results reveal that *WNT* proteins exhibit distinctive distribution of conserved motifs and share similar pattern or number of conserved motifs. *WNT* family signature (C-[KR]-C-H-G-[LIVMT]-S-G-x-C) present in *WNT* ligands is involved in *WNT* receptor signaling pathway as well as in calcium modulating pathway and signal transduction activity [21]. They also comprise palmitoylation site (ECKWQFRFRWNC) which is required for accurate processing of *WNTs* [22]. Somatostatin receptor Type-1 signature (SPGTRGRACNSSAPDLGCD) is involved in G-protein coupled receptor activity. HIV TAT domain signature (CLCRFHWCCVVQ) was also found conserved in *WNT* proteins, which is a signature for trans-activating response element of HIV and has proven to be a successful antiviral drug developmental target [23]. A conserved motif of cysteine-knot family signature is found in all *WNTs* at specie level. Almost 22-23 conserved cysteines are present at the C-terminus of *WNTs*, which help in proper folding of *WNTs* due to formation of disulphide bridges. Secondly, these conserved cysteine motifs help *WNT* ligands to bind with Frizzled (FZD) receptors to initiate the *WNT* signaling cascade.

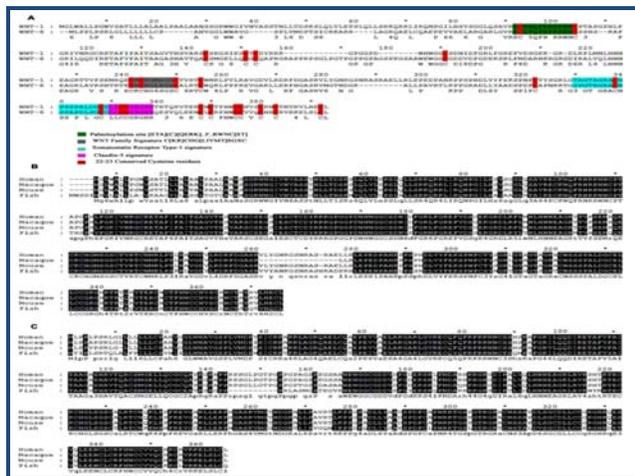


Figure 5: Alignment file of *WNT-1* and *WNT-6* showing the conservation of sequence between four species and also

displaying conserved sequence motifs. (A) Sequence alignments of *WNT-1* and *WNT-6* illustrating the conserved sequence motifs. The conserved motifs including Palmitoylation site (IPR014788), *WNT* family signature (PS00246), Claudin-5 signature, S locus related glycoprotein 1 binding pollen coat and Somatostatin receptor signature (IPB001116A) are represented by colours as indicated. Motif enrichment analysis is performed by using PRINTS, PRODOM, Blocks, PFAM and InterProScan, respectively. (B) MSA of *WNT-1* sequences and (C) MSA of *WNT-6* derived from *Homo sapiens*, *Macaca mullata*, *Mus musculus* and *Danio rerio*. Black colour is indicating the level of conservation at amino acid level.



Figure 6: Multiple sequence alignment of *WNT-10A* and *WNT-10B*. (A) Sequence alignments of *WNT-10A* and *WNT-10B* illustrating the conserved sequence motifs. The conserved motifs including Palmitoylation site (IPR014788), *WNT* family signature (PS00246), Claudin-5/8 signature, Rhodopsin like GPCR family signature (IPR000276), HIV- TAT domain (IPR001831) and 4-disulphide core signature (IPR015874) are represented by colours as indicated. Motif enrichment analysis is performed by using PRINTS, PRODOM, BLOCKS, PFAM and InterProScan, respectively. (B) MSA of *WNT-10A* sequences and (C) MSA of *WNT-10B* derived from *Homo sapiens*, *Macaca mullata*, *Mus musculus* and *Danio rerio*. Black color indicates the level of conservation at amino acid level.

In *WNT-10A/WNT-6* cluster, both genes co-regulate in *Mus musculus*. By genomic annotation of *WNT-10A* and *WNT-6* in mice, we came to know that both genes play important role in G-protein coupled receptor (GPCR) binding, signal transduction activity, pattern and axis formation, female gonad development, uterine bud morphogenesis, odontogenesis, nephron tube formation and in *WNT* receptor signaling pathway [24]. The data is further supported by co-regulation values calculated by co-express database. Although the co-regulation value is less than 0.5 but the Mutual Rank (MR) score, i.e. 9.9, strongly implicate their co-expression in *WNT* signaling pathway. In humans, *WNT-10A* and *WNT-6* also regulate the same basic developmental processes like odontogenesis, cell response to stimuli and *WNT* receptor signaling pathway. Along with this *WNT-10A* is also involved in tongue, skin and hair development [25]. Homologous cluster

of *WNT-1/WNT-10B* is involved in regulation of WNT signaling pathway and Notch pathway and in several transcriptional regulatory activities. *WNT-1* and *WNT-10B* are also involved in brain segmentation, mid brain-hind brain boundary development, axis specification and differentiation and growth of mesoderm and mammary glands. Moreover, they also exhibit GPCR and FZD binding activity. In humans, *WNT-1* and *WNT-10B* are important developmental proteins being probable ligands for CNS development and in various regions of tissues [25].

Expression of WNT ligands in breast cancer

Conventionally, several WNTs out of 19 manifest in human breast including *WNT-1*, -2, -4, -5A, -5B, -7A, -8B, -9A, -9B, -10B, and -11. However *WNT-1* and *WNT-4* exhibit an up-regulation in tumour [26]. Being putative paralog of *WNT-10A*, presence of human *WNT-10B* in both cancerous as well as in non-cancerous cell lines is obvious. It also shows differential expression in case of benign and malignant tumours [27]. *WNT-3A*, *WNT-4*, *WNT-6*, *WNT-8B* and *WNT-9A* play a positive role in breast cancer malignancy, while in several malignant tumours, the expression of *WNT-5A*, *WNT-9B* and *WNT-16* are switched off [28]. Some of the WNTs like *WNT-3*, *WNT-4* and *WNT-7B* express in breast cancer cell lines. While *WNT-2*, *WNT-3*, *WNT-4* and *WNT-7B* expression has also been monitored in breast cancer tissues (in case of mice, *WNT-1* and *WNT-3* are involved in tumorigenesis). Various WNT ligands, including, *WNT-3A*, *WNT-4*, *WNT-6*, *WNT-8B*, *WNT-9A* and *WNT-10B* ectopically up-regulate in human mammary epithelial cells (HMEC). However, *WNT-5A* and *WNT-5B* exhibit totally opposite pattern of signaling [28].

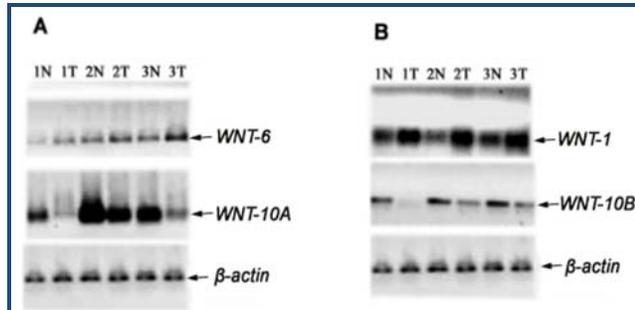


Figure 7: Reverse Transcriptase PCR analysis of *WNT-1*, *WNT-6*, *WNT-10A* and *WNT-10B*. (A) Expression analysis of two homologs *WNT-6* and *WNT-10A*, working in synergistic manner. *WNT-6* showing up-regulation in tumour whereas *WNT-10A* in normal. (B) Expression analysis of *WNT-1* and *WNT-10B*, also acting synergistically, *WNT-1* up regulating in tumour just like its closest homolog *WNT-6* and *WNT-10B* up regulating in normal tissue sample like *WNT-10A*. β -actin is shown as positive control.

Expression investigation of *WNT-1*, *WNT-6*, *WNT-10A* and *WNT-10B* were performed for invasive ductal carcinoma of breast (mammary glands) in human in order to confirm their co-regulation. RNA was extracted from each control and diseased sample and then proceeded to cDNA synthesis and PCR amplification. Our results confirmed that all these WNTs express in human breast. *WNT-1* and *WNT-6* were found to be up-regulated in tumour samples while *WNT-10A* and *WNT-10B* up-regulated in normal tissue (Figure 7A and B).

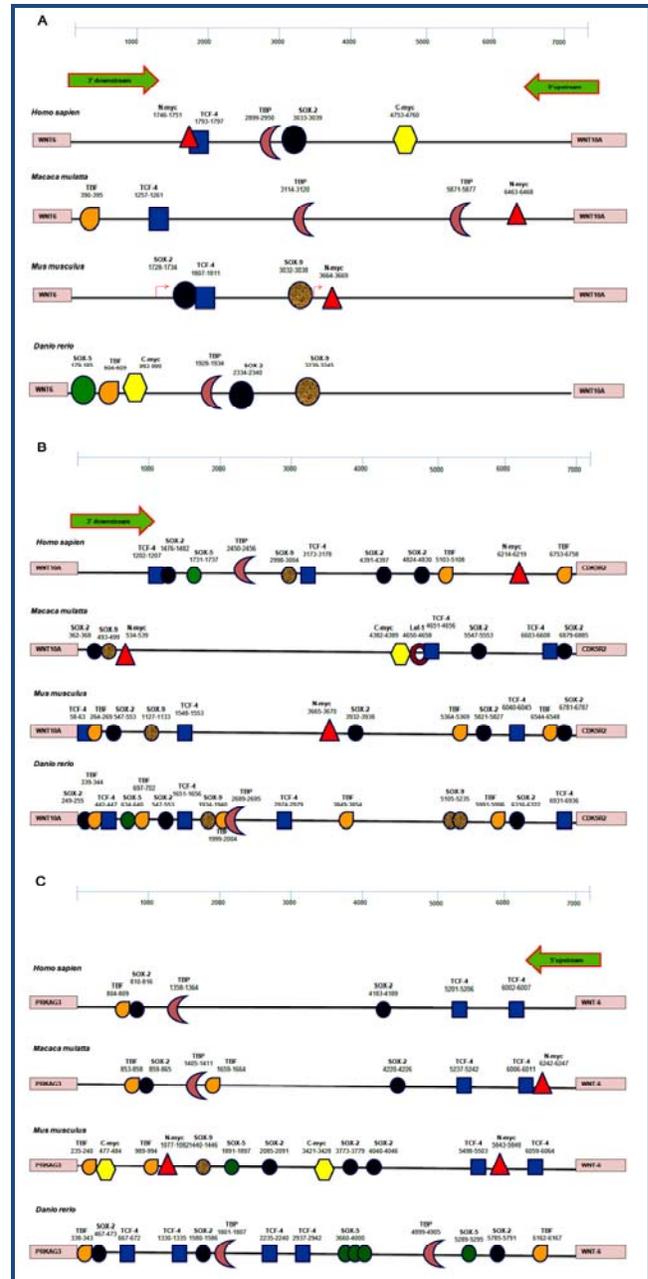


Figure 8: Determination of novel cis-regulatory elements (CREs) at intergenic region of co-linear WNT genes (*WNT-6* and *WNT-10A*) forming a cluster. (A) Intergenic region of 7000bp between *WNT-6* and *WNT-10A* showing conserved motifs of TCF-4, SOX family, myc family, TBF and TBP. (B) 7000bp downstream region of *WNT-10A*. Transcription factor binding sites are present for Sox-9 and TCF-4. In case of mouse, there is also a conserved site of LEF-1 along with TCF-4. (C) 5'upstream regulatory region analysis of *WNT-6*. This analysis is showing the similar results as depicted in A and B. TFBSs of TCF-4 and SOX remained conserved in almost all orthologs, supporting the hypothesis.

Intergenic Study for the analysis of cis-regulatory elements

WNT signaling pathway alters the expression of target genes in an instructive fashion and hence determines the cell fate.

Normally, the genes regulated directly or indirectly by WNTs are quite often regulatory growth factors, which play a key role in controlling development [29]. Some of the best-studied examples include members of the HMG-box proteins binding to the specific sites at the target gene promoter region. We examined the conserved intergenic region of about 7 KB between *WNT-6* and *WNT-10A* (Figure 8A) and have speculated that TCF-4 and SOX-9 are the responsible regulatory elements, that may act as activator and inhibitors respectively, hence interceding the signaling pathway. The regulatory region used for analysis was further extended from 7 KB downstream of *WNT-6* (Figure 8B) to 7 KB upstream of *WNT-10A* (Figure 8C). The putative transcription factor binding sites (TFBSs) were found by using TrFAST tool created in our lab (unpublished data). The criterion used to select a TFBS was based on two key points. Firstly, the presence of consensus transcription factor sequence at regulator region and secondly, similar conservation pattern in all orthologous species.

The common transcriptional regulatory network model that we obtained from *in-silico* study may regulate WNT signaling in various cellular contexts. 5' upstream region of *WNT-10A* contains the binding sites for myc family (n-myc and c-myc) of transcription factors, TCF-4 and SOX family (SOX-2). Similarly, 5' flanking region of *WNT-6* also contain these TFBS, however, their number almost doubled than that of *WNT-10A*. Association of SOX-2 with cyclinD1 promoter, containing TCF-binding sites raise the β -catenin aroused cyclinD1 transcription [29].

Intergenic region of 7KB between *WNT-1* and *WNT-10B* showing conserved motifs of TCF-4, SOX family, myc family, TBF and TBP. Presence of multiple binding sites for TCF-4 and Sox-9 in human and macaque strongly provide the evidence supporting the hypothesis that these two might be the responsible regulatory elements acting as activator and suppressor. C-myc and n-myc binding sites are also shown to be conserved in vertebrates that might play role in tumorigenesis; (B) Downstream regulatory region of *WNT-1* showing multiple sites for Sox-9 and TCF-4, depicting their strong conservation in case of both genes of *WNT-1/WNT-10B* cluster. Additionally, overlapping TFBSs were also analyzed using TrFAST tool (unpublished data) and are shown. TBP and TBF showing strong conservation in fish are DNA binding motifs; (C) Shows the similar results as elucidated in A and B. TFBSs of TCF-4 and SOX remain conserved in almost all orthologs.

Furthermore, *in-silico* analysis revealed that all these conserved non-coding elements (CNEs) are present in *Homo sapien*, *Macaca mulata*, *Mus musculus* and *Danio rerio* (the species under observation for intergenic analysis), irrespective of the exact orientation on genome. The second homologous WNT cluster of *WNT-1* and *WNT-10B* (Figure 9A, B, and C) exhibits more or less the similar pattern of gene expression due to the plethora of TFBSs located at their upstream and downstream regulatory regions. The binding sites downstream of *WNT-10B* are same as upstream of *WNT-1*, and there are also various sites of TCF-4, SOX-2 and myc family present downstream of *WNT-10B*, due to which it might shows mild expression in tumour breast samples relative to *WNT-1*, which is found to be down regulated. Other studies, however led us to the conclusion that both TCF-4 and SOX-2 can be transactivators in some specific promoter context, thus raising the β -catenin dependent transcription in human breast carcinoma. Similarly, SOX-9 binding sites were localized at the downstream region of *WNT-10A*. Recent studies suggest that SOX-9 antagonizes the WNT signaling either by promoting the β -catenin degradation or by inhibiting β -catenin transcriptional activity [30]. Increased cytoplasmic expression of SOX-9 has been reported to associate with higher grade human breast tumours [31]. These findings support our data that SOX-9 could be a mediator for the inhibition of *WNT-10A* expression in invasive breast cancer.

Conclusion:

Regulation of ancestrally duplicated WNT cluster in breast cancer in comparison to the surrounding tissue is not surprising by considering the structural similarity of TCF and SOX factors, their affinity for binding with DNA sequences and in chromatin remodelling. Genetic studies have shown that SOX-9 inhibits the association between TCF and β -catenin by competing the binding for β -catenin [32]. Similarly, for some target promoters, both SOX and TCF may compete for DNA binding or repress the transcription by direct interaction. Although, it is unknown how SOX represses the WNT signaling, it is likely that either SOX factors inhibit TCF-4 and DNA binding or appear to block the binding of TCF-4 and β -catenin.

The findings reported herein suggest that there exists a coordinated expression pattern of *WNT-1-10B* and *WNT-6-10A* clusters in human breast cancer. Expression studies revealed an up-regulation of *WNT-1* and *WNT-6* in tumour tissues.

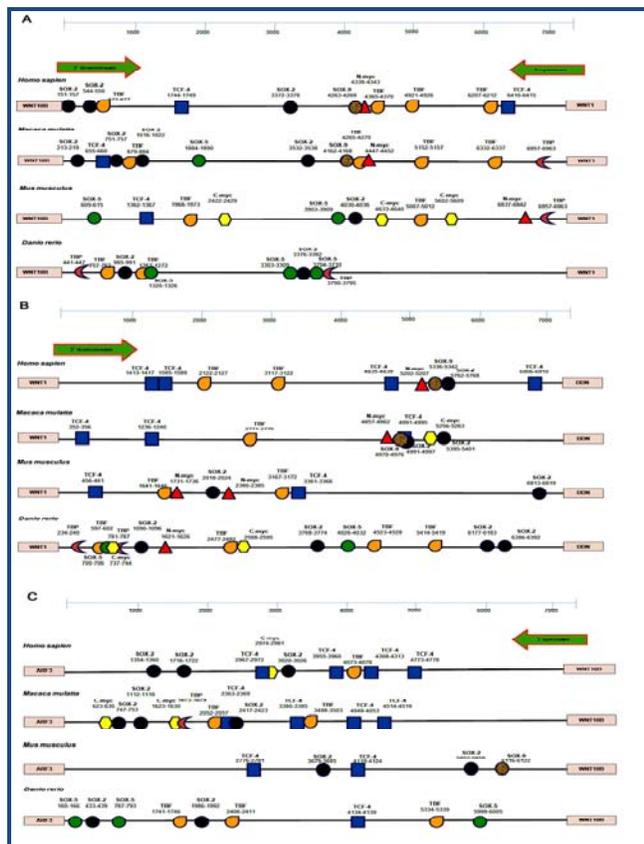


Figure 9: Novel cis-regulatory elements (CREs) at intergenic region of homologous *WNT-1* and *WNT-10B* genes. (A) ISSN 0973-2063 (online) 0973-8894 (print) Bioinformatics 7(7):339-346 (2011)

However, to our surprise, the expressions of their corresponding homologues *WNT-10B* and *WNT-10A* were decreased in the same tissues. These findings led us to study the functional synergy of this co-expression. The existence of conserved binding sites for the TCF-4 and SOX family of TFs at the flanking regions of *WNT-10A/WNT-6* and *WNT-1/WNT-10B* indicated a co-operative regulation of transcriptional activity. Existence of 20 SOX proteins and almost same number of WNT molecules in vertebrates working in a co-operative manner require more attention to delineate the cell-specific transcriptional events and their contribution in the tumourigenesis.

Competing interests:

The authors declare that they have no competing interests.

Authors' contributions:

QA, Figure 1A, 1B, 1C, 2A, 2B, 2C, 3, 4, 5A, 5B, 6a, 6B, 6C, 7A, 7B experimental design, Phylogenetic and syntenic analysis and results compilation; US, Figure 8A, 8B, 8C, 9A, 9B, 9C, Transcription factor binding site prediction and analysis using TrFAST tool; SN, Manuscript writing and editing; SR, research design, manuscript preparation, editing and analysis. All authors read and approve the final manuscript.

Acknowledgement:

We are thankful to Pakistan Institute of Medical Sciences, Islamabad, Pakistan for providing human breast cancer samples used in this study. We also thank Nousheen and Zahida for providing expert technical assistance and for critical reading of the manuscript. This study was supported by HEC, Pakistan.

References:

- [1] Hermon C & Beral V, *Br J Cancer*. 1996 **73**: 955 [PMID: 8611414]
- [2] Easton DF. *Breast Cancer Res* 2002 **4**: 179 [PMID: 12223120]
- [3] Easton DF. *Breast Cancer Res* 1999 **1**: 14 [PMID: 11250676]
- [4] Polakis P. *Genes Dev* 2000 **14**: 1837 [PMID: 10921899]
- [5] Nusse R & Varmus HE, *Cell* 1982 **31**: 99 [PMID: 6297757]
- [6] Almeida M *et al.* *J Biol Chem* 2005 **280**: 41342 [PMID: 16251184]
- [7] Katoh M & Katoh M, *Clin Cancer Res*. 2007 **13**: 4042 [PMID: 17634527]
- [8] Kurayoshi M *et al.* *Cancer Res*. 2006 **66**: 10439 [PMID: 17079465]
- [9] Liu G *et al.* *Mol Cell Biol*. 2005 **25**: 3475 [PMID: 15831454]
- [10] Holmen SL *et al.* *J Biol Chem*. 2002 **277**: 34727 [PMID: 12121999]
- [11] Umbhauer M *et al.* *Embo J*. 2000 **19**: 4944 [PMID: 10990458]
- [12] Kispert A *et al.* *Development* 1998 **125**: 4225 [PMID: 9753677]
- [13] Tetsu O & McCormick F, *Nature*. 1999 **398**: 422 [PMID: 10201372]
- [14] Satoh S *et al.* *Nat Genet*. 2000 **24**: 245 [PMID: 10700176]
- [15] Cowin P *et al.* *Curr Opin Cell Bio*. 2005 **17**: 499 [PMID: 16107313]
- [16] Saitoh T & Katoh M, *Int J Mol Med*. 2002 **10**: 345 [PMID: 12165812]
- [17] Hens JR & Wyszolmerski JJ, *Breast Cancer Res*. 2005 **7**: 220 [PMID: 16168142]
- [18] Gavin BJ & McMahon AP, *Mol Cell Biol*. 1992 **12**: 2418 [PMID: 1373817]
- [19] Katoh M. *Int J Oncol*. 2002 **20**: 373 [PMID: 11788904]
- [20] Miller JR. *Genome Biol*. 2002 **3**: REVIEWS3001 [PMID: 11806834]
- [21] Dale TC. *Biochem J*. 1998 **329**: 209 [PMID: 9425102]
- [22] Tanaka K *et al.* *J Biol Chem*. 2002 **277**: 12816 [PMID: 11821428]
- [23] Chin DJ *et al.* *J Virol*. 1991 **65**: 1758 [PMID: 1900539]
- [24] Blake JA *et al.* *Nucleic Acid res*. 2003 **31**: 193 [PMID: 12519980]
- [25] Rebhan M *et al.* *Bioinformatics*. 1998 **14**: 656 [PMID: 9789091]
- [26] Ayyanan A *et al.* *Proc Natl Acad Sci. USA* 2006 **103**: 3799 [PMID: 16501043]
- [27] Bui TD *et al.* *Oncogene*. 1997 **14**: 1249 [PMID: 9121776]
- [28] Benhaj K *et al.* *Oncol Rep*. 2006 **15**: 701 [PMID: 16465433]
- [29] Wodarz A & Nusse R, *Annu Rev Cell Dev Biol*. 1998 **14**: 59 [PMID: 9891778]
- [30] Topol L *et al.* *J Biol Chem*. 2009 **284**: 3323 [PMID: 19047045].
- [31] Chakravarty G *et al.* *Exp Biol Med*. 2011 **236**: 145 [PMID: 21321311]
- [32] Akiyama H *et al.* *Genes Dev*. 2004 **18**: 1072 [PMID: 15132997]

Edited by P Kanguane

Citation: Ain *et al.* *Bioinformation* 7(7): 339-346 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.