

## Supplementary Information

### Supplementary Table S1

This table presents 47 genes which have been shown to have role in breast cancer onset/progression were fished out from the database along with their known regulator. The binding sequence of each regulator was used as input for USP. The last column shows the unique sequence predicted by USP. E.g. Akt2 gene is regulated by MyoD, CREB, NFkB and AP1 transcription factors. Their binding sequences, along with the starting position of the motif, are given which was used as input for USP. MyoD binding motif in Akt2 gene is 11bp long while the unique sequence predicted for it is 17bp long. The number at the end of the sequence indicates the length of the unique sequence.

Smallest unique sequence in the following list is 13nt long for E2F transcription factor binding site in the promoter of Wnt10B gene.

	Gene	TranscriptionFactor	Chr. No	Starting Position	Binding Sequence	Length	Unique Sequence
1	AKT2	MyoD	chr19	40791913	ATGCCACCTGC	11	TATTAATGCCACCTGC
	AKT2	CREB	chr19	40796835	TGGACTGCGGT	11	CGCAGGCTGGACTGCGGT
	AKT2	NF-kappaB1	chr19	40790600	CAGGGCTCCC	10	CAGGGCTCCCCCAACG
	AKT2	AP-1	chr19	40788299	GCTGTGTCAC	10	GTAGCTGTGTCACCCC
	AKT2	AP-1	chr19	40788585	GGGGCTCAGTG	11	CTAGGGGCTCAGTGACC

2	APC	c-Myc	chr05	112072476	TGGCACGTGCCT	12	ACTTAGCTGGGCATGATGGCACGTGCCTGTAGTCCCAGCTACTT
	APC	HNF-4alpha1	chr05	112071326	TGAAATTTGGCCC	13	ATGCTGAAATTTGGCCC
	APC	C/EBPbeta	chr05	112057062	GATTGCGCCACT	12	GATTGCGCCACTACACTCTAGCCTT
	APC	C/EBPbeta	chr05	112081881	AGATGAAGAAATC	13	CAAGATGGAGAAAGCAT
3	ATM	Sp1	chr11	108083718	TGGGCGGGGC	10	AACCTCCCTGGGCGGGGC
	ATM	CREB	chr11	108079770	GCTGAGGCAGGC	12	GCTGAGGCAGGCAGATCACCTGAGGTCAAGAGTTTGAGG
	ATM	p53	chr11	108139192	AGACATGCTC	10	AGACATGCTCAAGTTCT
	ATM	AP-1	chr11	108138314	GGTGGCTCAC	10	TAAAACCAGGCCAGGCATGGTGGCTCAC
	ATM	c-Jun	chr11	108091071	GGTGACACAG	10	GGTGACACAGAACATAAG
4	BCL2	NF-kappaB1	chr18	60980154	GGGAAGTCCC	10	TTAACAAGGGAAGTCCC
	BCL2	p53	chr18	60992888	ATCTGTACAG	10	CGCATCTGTACAGACC
	BCL2	CREB	chr18	60985408	CCGGTTGACGCTC	13	CCGGTTGACGCTCTC
	BCL2	AP-2alphaA	chr18	60987120	CCGCCTCCGGGCT	13	CCGCCTCCGGGCTGCG
	BCL2	C/EBPalpha	chr18	60988190	TTTTAAGCAATG	12	CACGTTTTAAGCAATG
5	BRCA1	Sp1	chr17	41269521	GCCGAGGCGGGC	12	TTGTAATCCCAGCACTTGGGAGGCCGAGGCGGGC
	BRCA1	Bach1	chr17	41269056	ATGGTACTGTGTA	13	CTGGCAAGAACTTTACAGATGGTACTGTGTA
	BRCA1	p53	chr17	41275932	TAGACATGTC	10	ATCTTACTAGACATGTC
	BRCA1	c-Jun	chr17	41271815	CACTGACTCAAT	12	CACTGACTCAATTCTGAAAC
	BRCA1	AP-1	chr17	41274786	GGACGCAGTG	10	GAATGGCTGGACGCAGTG
6	BRCA2	p53	chr13	32875134	AGGGATGCCC	10	AGGGATGCCCTACCCCT
	BRCA2	NF-kappaB1	chr13	32883640	TGTGCAATTCCT	13	TGTGCAATTCCTGTGC
	BRCA2	CREB	chr13	32870704	TGTGATGTAAGC	12	TTTTGTGATGTAAGCTTA
	BRCA2	CREB	chr13	32889189	GTGAAGTCATC	11	GTGAAGTCATCCACAA
	BRCA2	CREB	chr13	32890087	ATGACGCCATC	11	AGATGACGCCATCTG
7	CCND1	ATF-2	chr11	69453024	AGGTGACCCCACC	13	GCAGGTGACCCCACCTG
	CCND1	CRE-BP1	chr11	69455864	CAGTAACGTCACA	13	CAGTAACGTCACACG
	CCND1	NF-kappaB1	chr11	69462837	TGAGGAGCCCC	11	TGAGGAGCCCCAACAAC
	CCND1	AP-1	chr11	69436584	AGTGGCTCAC	10	CCACAGTGGCCAGGCACAGTGGCTCAC
	CCND1	STAT5A	chr11	69455443	CGTTCTTGAAAAT	13	GGCGTTCTTGAAAAT
8	COL18A1	p53	chr21	46884268	AGGCAGGCCC	10	ATAAGGCAGGCCCTCG

	COL18A1	ATF	chr21	46830693	GGGTGACGTGAGCT	14	GGGTGACGTGAGCTGA
	COL18A1	ATF	chr21	46876206	CCCTGACGTCCGCC	14	CCCTGACGTCCGCCTG
	COL18A1	NF-kappaB	chr21	46826029	TGGGGACCTTCCT	13	CCCGTGGGGACCTTCCT
	COL18A1	NF-kappaB1	chr21	46808663	AGGGAGGCCC	10	AGGGAGGCCCCAGCAGGA
9	CTSD	STAT1	chr11	1777997	ACCCTTTTCCCAGA	14	GAGAACCCTTTTCCCAGA
	CTSD	STAT1beta	chr11	1780035	GGCCCATTTCCAGACC	14	GGCCCATTTCCAGACC
	CTSD	Egr-1	chr11	1779947	CCCTCCCCTC	11	CCCTCCCCTCATTGTG
	CTSD	Egr-1	chr11	1799690	ACCTCCCACCC	11	CGCCTACCTCCCACCC
	CTSD	c-Jun	chr11	1779543	GCTGTGTCAC	10	GCTGTGTCACCATCCCG
10	DCC	p53	chr18	49847510	GAGCCTTCCTTGGCATTTC	19	GAGCCTTCCTTGGCATTTC
	DCC	p53	chr18	49854093	AGACATGTCT	10	ATGCCAAGACATGTCT
	DCC	p53	chr18	49875230	GAGCAAGTCCTGCCATGTT	19	GAGCAAGTCCTGCCATGTT
	DCC	Nkx6-1	chr18	49849114	AGTTTAAATTGTAT	14	TCAAAAGTTTTAATTGTAT
	DCC	Nkx6-1	chr18	49867987	TTTCTAATTAAAAA	14	TTTCTAATTAAAAATGGC
11	EGF	c-Fos	chr04	110814460	TGTGACTCAG	10	TGTGACTCAGCTAGAAGC
	EGF	CREB	chr04	110816874	TGAAGTCA	8	TGAAGTCAGGTAGGGTGATGCCTCCAGCTTTGTTCTTTTGGCTTAGGAT
	EGF	CREB	chr04	110820615	GGTGACGTA	9	GGTGACGTACAGATGGGTTTTTGGTGTGGATGGT
	EGF	CREB	chr04	110822464	GGTGACGCA	9	GGTGACGCAAAGACA
	EGF	CREB	chr04	110825186	GGTGATGTC	9	GGTGATGTCAGAGAGAA
	EGF	CREB	chr04	110834176	GTGAAGTCAGC	11	GTGAAGTCAGCCAGAGCA
	EGF	AP-1	chr04	110814202	ATTAGTCAG	9	CAAATTTAAAAAATTAGTCAG
	EGF	AP-1	chr04	110814460	TGTGACTCAG	10	TGTGACTCAGCTAGAAGC
	EGF	AP-1	chr04	110827406	AATGACTCA	9	GAAAACTCTAATGACTCA
	EGF	AP-1	chr04	110832909	CCTGACTTA	9	CCTGACTTAAGCTGTTT
	EGF	AP-1	chr04	110832774	TCTGACTCA	9	TCTGACTCAGTCCAGA
	EGF	AP-1	chr04	110836255	GGTGGCTCAG	10	CCACCTCTGGTGGCTCAG
	EGF	NF-kappaB1	chr04	110820018	TGGGGAAGTTCTC	13	TGGGGAAGTTCTCTGGAAAATATCCTGCAGAGTGTTTTT
12	EGFR	AP-2alphaA	chr07	55072263	GCCCCGCGC	9	CTGTGAGCCCCGCGC
	EGFR	c-Myc	chr07	55072632	GTGCCAGGTGCTG	13	GCTTCTGTGCCAGGTGCTG
	EGFR	p53	chr07	55083419	AGGCATGCTC	10	AGGCATGCTCCTGGTA

	EGFR	Egr-1	chr07	55087262	GGCGGAGGAGG	11	GCGGGCGGAGGAGGAGA
	EGFR	c-Jun	chr07	55095863	ATGACGTCA	9	ATGACGTCATGATTG
13	ERBB2	PPAR-gamma1	chr17	37825107	AGAGGGGCAGAGGAGA	16	GAGAGGGGCAGAGGAGAG
	ERBB2	p300	chr17	37859823	CTGAGGGAGTAAGAG	15	CTGAGGGAGTAAGAGCT
	ERBB2	p53	chr17	37851371	AGCCATGCCT	10	AGCCATGCCTGCGCA
	ERBB2	c-Myc	chr17	37862772	TTGGCATGTGGGT	13	AGAGGTTGGCATGTGGGT
	ERBB2	N-Myc	chr17	37827240	GCCACGTGCCC	11	CCCCATGCCACGTGCCC
14	ERBB3	AP-1	chr12	56473839	CGTGACTCA	9	CGCGGCCGTGACTCA
	ERBB3	AP-1	chr12	56476154	AATGACTAAC	10	AATGACTAACCTTCCT
	ERBB3	c-Jun	chr12	56480110	GGTGACTGAG	10	CTGGACAGGTGACTGAG
	ERBB3	AP-2alpha	chr12	56474037	GGCCCCCGGGCC	12	GGCCCCCGGGCCGGA
	ERBB3	Hand1	chr12	56477585	CAATACCAGACACTGTA	17	CAATACCAGACACTGTACA
15	ESR1	FOXO3	chr06	152122328	TCAAGTAAACAACAG	15	CTCAAGTAAACAACAG
	ESR1	p53	chr06	152116354	GGTCATGCCT	10	AAGAGCTGGGTGTGGTGGGTCATGCCT
	ESR1	AhR	chr06	152128883	CCGCAGGCTCCCGGGGCA	18	CCCGCAGGCTCCCGGGGCA
	ESR1	AP-1	chr06	152121909	AATGAGTCAT	10	AATGAGTCATTCACTGGC
	ESR1	GATA-3	chr06	152128357	GCCCTATCTC	10	GCCCTATCTCGGTTA
16	FGF3	GATA-1	chr11	69626980	TCCCCATCTG	10	TCCCCATCTGGGGTGTGG
	FGF3	GATA-1	chr11	69630599	GCGCCATCCC	10	TGGGCGCCATCCCCCT
	FGF3	GATA-1	chr11	69637299	CAGCCCTATCT	11	CCCCAGCCCTATCTGGT
	FGF3	HOXA3	chr11	69634226	ACCAATTATG	10	ACCAATTATGCCTCCG
	FGF3	AP-2alphaA	chr11	69632489	CGCCCGGGG	9	CGCTTCCTGCGCCCGGGG
	FGF3	Sp1	chr11	69590303	GGACCCCGGCCC	12	TGGAGGGACCCCGGCCC
	FGF3	Sp1	chr11	69590327	CCGCCCCGGCCCA	12	CCGCCCCGGCCACGGAC
	FGF3	Sp1	chr11	69590348	TCGTCCCGCCCC	12	TCGTCCCGCCCCCTG
	FGF3	MyoD	chr11	69588922	GGGCAGGTGA	10	GCGGGGCAGGTGATCC
	FGF3	Tal-1	chr11	69586801	CTGTAACAGCTGTCTGC	17	CTGTAACAGCTGTCTGC
17	FGFR1	Sp1	chr08	38324406	ACACCCACCCC	12	ACACCCACCCCCCCCCC
	FGFR1	Sp1	chr08	38339815	GTCCCTGCCT	10	AGATCCCCTGTCCCTGCCT
	FGFR1	CREB	chr08	38318566	GGGGCACATCTCAG	14	AGGGGGCACATCTCAG

	FGFR1	CREB	chr08	38324839	TTTGGTGACGACA	13	CCATTTGGTGACGACA
	FGFR1	CREB	chr08	38339242	TGGAGTGCAGTG	12	GTCACCAGGATGGAGTGCAGTGGTGCGATCTT
18	FGFR2	E2F-2	chr10	123356334	GGCGCTAA	9	GGCGCTAAAGAGCTC
	FGFR2	Sox9	chr10	123361710	TGGGTAACAAGA	12	CGCCTGGGTAACAAGAGAGA
	FGFR2	Sox9	chr10	123368265	GCCCTTTGTTCT	12	GCCCTTTGTTCTCCGG
	FGFR2	Sox5	chr10	123349798	GTCAACAATGA	11	CATTAATGTCAACAATGA
	FGFR2	Sox5	chr10	123351358	TCCAACAATAC	11	CTGCTTCCAACAATAC
19	FGFR4	Ik-3	chr05	176517990	CTGCGGGGAACACC	14	CTGCGGGGAACACCGT
	FGFR4	COUP-TF1	chr05	176524543	CTGACCTTCGGACCC	15	CTGACCTTCGGACCCTA
	FGFR4	HNF-1A	chr05	176514211	CGGTAAATCAGTAACC	16	CGGTAAATCAGTAACC
	FGFR4		chr05	176514286	TGGTTACTCATTGCCC	16	TGGTTACTCATTGCCC
	FGFR4	POU3F2	chr05	176516184	AATCAATTCATTTAT	15	GCAATCAATTCATTTAT
	FGFR4	ZID	chr05	176517610	GAGGCTCCATGATC	14	GAGGCTCCATGATCG
	FGFR4	HNF-4alpha2	chr05	176524543	CTGACCTTCGGACCC	15	CTGACCTTCGGACCCTA
	FGFR4	SRY	chr05	176498800	AAATACAATAGG	12	GTGTAAAATACAATAGG
	FGFR4	SRY	chr05	176504259	GTAAACAAATGA	12	GTAAACAAATGAGTACT
20	GH1	SREBP-1a	chr17	61987242	AGTCAGGTGG	10	AGTCAGGTGGCTGCCG
	GH1	SREBP-1a	chr17	61999953	TGCCACATGA	10	CTCGTTGGACCCTTCAGTTTGCCACATGA
	GH1	SREBP-1a	chr17	62007907	AGCCACGTGA	10	GAAGCAGAGCCACGTGA
	GH1	STAT5B	chr17	62011891	CTTCCCAGAA	10	CTTCCCAGAACAGGAGG
	GH1	STAT5A	chr17	62004754	ATAAAACAAGAAC	13	ATAAAACAAGAACTTCA
21	HRAS	STAT3	chr11	555028	GCGGAAGC	8	GCGGAAGCCGAATCT
	HRAS	c-Myc	chr11	542288	GTGCCACTTGGT	12	GCTGAGTGCCACTTGGT
	HRAS	c-Myc	chr11	550537	AAGCCACGTGGA	12	AAGCCACGTGGAAACA
	HRAS	NF-Y	chr11	535617	GCCTACCATTGGCTGCG	17	GCCTACCATTGGCTGCG
	HRAS	Max	chr11	544933	CCTGACCCACGTGACCC	18	CCTGACCCACGTGACCCT
22	IGF1R	p53	chr15	99190658	GGACACGCCC	10	GGACACGCCCCCGA
	IGF1R	STAT1	chr15	99177941	TTCTTGGAAGTGGC	14	CTTCTTGGAAGTGGCC
	IGF1R	STAT1	chr15	99201939	GATTTCCC	8	GTTTTTGGGATTTCCC
	IGF1R	N-Myc	chr15	99190720	CGCCACGTGGTG	13	CGCCACGTGGTGAG

	IGF1R	N-Myc	chr15	99190722	CCCACGTGGTG	11	AGCCGCCCACGTGGTG
23	KIT	Sp1	chr04	55519363	GGAGGCGGAGCT	12	GAGGAATGGCGTGAACCCGGGAGGCGGAGCT
	KIT	AP-2gamma	chr04	55523921	CTCCTCCCGGCG	12	CTCCTCCCGGCGGGCA
	KIT	p53	chr04	55521107	AGACATGGCC	10	AGACATGGCCAATCAGC
	KIT	GATA-2	chr04	55515663	TACCCATCTCC	11	TACCCATCTCCAGAGTC
	KIT	GATA-2	chr04	55525335	GGATAAGCCC	10	GGAGATGGCGTGGAAC
24	MCC	Pax-2	chr05	112622925	ACACAAACCC	10	ACACAAACCCCTCTCTC
	MCC	Pbx1a	chr05	112620680	AACCATCAATCAAGGG	16	GAACCATCAATCAAGGG
	MCC	Pbx1a	chr05	112626525	TGATTTGCCTGATGTT	16	TGATTTGCCTGATGTTGA
	MCC	Pbx1a	chr05	112626076	TAAGTTGATTGCTTTG	16	TAAGTTGATTGCTTTGTAC
	MCC	POU2F1	chr05	112631707	AACTATTTGCATAT	14	AACTATTTGCATATAACG
	MCC	POU2F1	chr05	112630437	CCCCCATTACCAG	14	CCCCCATTACCAGGA
	MCC	Oct-B1	chr05	112629316	ACACATTTGCATAG	14	GCTCACACATTTGCATAG
	MCC	c-Myb	chr05	112629827	AGCCAGTTCAG	11	TCAAGCCAGTTCAGGTC
25	MDM2	p53	chr12	69202750	CTGACTTGTCT	11	TAAGTCCTGACTTGTCT
	MDM2	STAT1	chr12	69188515	TTCTGAGAAATGTG	14	TTCTGAGAAATGTGTGCTTAGGTGA
	MDM2	STAT1	chr12	69202950	GATTTC	8	GATTTCAGTTTTTCATCG
	MDM2	p300	chr12	69188832	TCAGTGAGTGAGTG	14	TCAGTGAGTGAGTGGTGAATC
	MDM2	GATA-3	chr12	69188062	GAGATAGAG	9	GAGATAGAGGGGTTAGGAC
26	MET	p53	chr07	116320299	AGACATGCCT	10	TTTGTTTAGACATGCCT
	MET	GR	chr07	116293524	AGGAACTCTGTGTTATGT	18	AGGAACTCTGTGTTATGTT
	MET	GR	chr07	116316547	CATGACACTCTGTTCTAT	18	CATGACACTCTGTTCTAT
	MET	Sp1	chr07	116312089	TGGGCGGGGC	10	TGTCGTGGGCGGGGC
	MET	NF-1	chr07	116297072	CATGGCAGGAAGCCAGA	17	CATGGCAGGAAGCCAGAT
27	MYC	ITF-2	chr08	128742193	TGGGGACATATGTTGTC	17	TGGGGACATATGTTGTC
	MYC	NF-kappaB	chr08	128735610	TGGAATTCCCC	11	ATCCTTGGAATTCCCC
	MYC	Pax-4a	chr08	128751480	TGTCTCCACCCC	13	CGTTTGTCTCCACCCC
	MYC	NF-AT	chr08	128739805	AGTTTTTCCAATG	13	ACAGTTTTTCCAATGCC
	MYC	PPAR-gamma2	chr08	128746310	CTGACCCCCGCCCTCGTT	18	CTGACCCCCGCCCTCGTT
28	NF2	p53	chr22	29984474	AGGCATGCGC	10	AGGCATGCGCCATCCAT

	NF2	STAT3	chr22	29984227	GGCTTCCC	8	CACATGGCGGCTGGCTTCCC
	NF2	c-Fos	chr22	29979932	GGTGACGCAC	10	ATGGCATGGTGACGCAC
	NF2	c-Jun	chr22	29979652	TGAGGTCA	8	GGAATCACCTGAGGTCAGGGGTTCGA
	NF2	AP-1	chr22	29992559	GGTGACTCAT	10	GAGCTGGGTGACTCAT
29	NME1	p53	chr17	49222669	AGACTGGGCTGGGCATGGT	19	AGACTGGGCTGGGCATGGTGA
	NME1	p53	chr17	49225595	AGCCATGCCT	10	ATGAATACAGCCATGCCT
	NME1	NF-kappaB	chr17	49226667	TGGGCAATTTCCC	13	TGGGCAATTTCCCCTCTC
	NME1	AP-1	chr17	49212823	AGTGGCTCAC	10	TTTTCAGCCAGGCTCAGTGGCTCAC
	NME1	c-Jun	chr17	49211358	TGACCTCA	8	TGACCTCAGGTGATCTGCCAGTCTCGGCCTCG
30	NRAS	p53	chr01	115262392	AGGGCATGCC	10	AGGGCATGCCACACCT
	NRAS	C/EBPalpha	chr01	115251157	TACATCACCACAC	13	GTTACATCACCACACAT
	NRAS	c-Fos	chr01	115276659	TCTAAGTCAC	10	TCTAAGTCACCTTAAAT
	NRAS	AP-1	chr01	115261942	ATGAATCAG	9	CGGCATTATGAATCAG
	NRAS	MyoD	chr01	115254085	GGTGGTGGACG	11	AAAAAATTAGTCAGGCATGGTGGTGGACG
	NRAS	NF-1	chr01	115250501	TCTGGTCAGACAGCCAA	17	TCTGGTCAGACAGCCAA
31	PGR	Sp1	chr11	100999042	CGGCTCCTCCTC	12	AGCCGCGGCTCCTCCTC
	PGR	p53	chr11	101009115	GAGGCATTTC	10	GAGGCATTTCCTTCTATA
	PGR	c-Myb	chr11	100997784	GCCCCAACTCGCCGCGC	18	GCCCCAACTCGCCGCGC
	PGR	TBP	chr11	100991516	TTTTTTAAAG	10	ACATTTTCATTTTTTAAAG
	PGR	CHOP-10	chr11	100993321	GGAATGCAATGGGC	14	GGAATGCAATGGGCTAT
32	PHB	STAT3	chr17	47502290	ATCGATCTCGGTCTTAAAC	19	ATCGATCTCGGTCTTAAAC
	PHB	Brachyury	chr17	47489173	AAAGATGACAGCTCTG	16	AAAGATGACAGCTCTGTGCCCAGCATCCACT
	PHB	STAT5A	chr17	47489502	TATTTCCAGGAAT	13	TATTTCCAGGAATCATCT
	PHB	STAT5A	chr17	47494783	ACGAAGGATGTTT	13	TATCACGAAGGATGTTT
	PHB	p53	chr17	47509587	TGGGGATGCC	10	CGCTAATGGGGATGCC
33	PLG	NF-kappaB	chr06	161127476	GGGGGCTTCAC	11	ACCCAGGGGGCTTCAC
	PLG	Sp1	chr06	161110973	GGAGGTGGGGTG	12	CACACGGGAGGTGGGGTG
	PLG	Egr-1	chr06	161104377	TGTGTGGGAGT	11	TGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCTTTATAA
	PLG	AP-1	chr06	161107158	GCTGATTCA	9	GCTGATTCAATGCCTAT

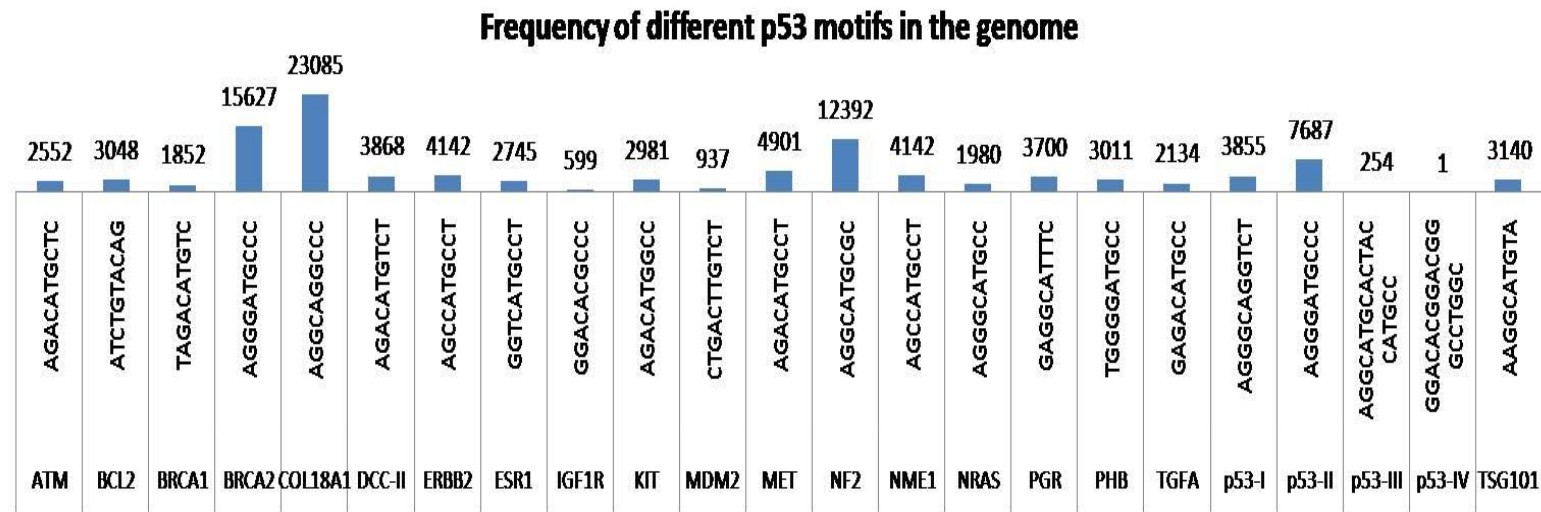
	PLG	c-Jun	chr06	161104403	GGTCACTCAG	10	GGTCACTCAGGACTTGCTTTATAAATCTGGGTGCTT
34	PLAT	AP-1	chr08	42072895	ATTTTAGTCACC	12	TCTGATTTTAGTCACC
	PLAT	NF-kappaB	chr08	42064525	GAAGTAAGTCCCC	13	GGAAGTAAGTCCCC
	PLAT	GR	chr08	42056746	ACAGAACTGTTTATAA	16	ACAGAACTGTTTATAAC
	PLAT	IRF-2	chr08	42059466	GGGTTTCTCTTTT	13	CAACGGGTTTCTCTTTT
	PLAT	CREB	chr08	42058131	GCTGCAGTG	9	GGCCTGAGCCTGGGAGTTTGAGGCTGCAGTG
35	PRL	C/EBPalpha	chr06	22293381	CTATACCTTACCAAGTT	17	CTATACCTTACCAAGTT
	PRL	AP-1	chr06	22290964	TCTGACTCA	9	TCTGACTCATGAGAAATAG
	PRL	c-Jun	chr06	22298075	AACTGCAG	8	AACTGCAGTCACTTGCG
	PRL	FOXO1	chr06	22293119	TGTTGTTTAC	10	AATTCCTTGTTGTTTAC
	PRL	STAT5B	chr06	22293464	CAATTTTAGGAAA	13	GGCACACAATTTTAGGAAA
36	PTH	E4BP4	chr11	13514310	TACTTACTTAACA	13	TACTTACTTAACAGATT
	PTH	c-Fos	chr11	13521454	AGACGCAGTG	10	AATAGACGCAGTGGCT
	PTH	IRF-1	chr11	13507612	GTGAAAGGTAAAA	13	AAAGTGAAAGGTAAAACTC
	PTH	C/EBPbeta	chr11	13511158	CTCAACACTTTG	12	CTCAACACTTTGGGAGGGG
	PTH	aMEF-2	chr11	13510354	CATCTTAAAATAAGTTT	17	CATCTTAAAATAAGTTTGG
37	PTPRF	CREB	chr01	43986943	GGTGATGTC	9	GGTGATGTCTGCTGCCT
	PTPRF	Max1	chr01	43980196	GACCACATGTCC	12	GACCACATGTCCCTTG
	PTPRF	GR	chr01	43983481	TGAGCCACTGTGTCCGTC	18	TGAGCCACTGTGTCCGTCT
	PTPRF	GR-beta	chr01	44005443	TGGAGCCCTCTGTCCTCT	18	TGGAGCCCTCTGTCCTCTC
	PTPRF	HNF-4alpha1	chr01	43989065	GAAGGGGAAAGGCCAAG	17	AGAAGGGGAAAGGCCAAG
38	RB1	E2F	chr13	48878047	TCATGCCGCCCAA	13	TCATGCCGCCCAAAA
	RB1	MyoD	chr13	48878395	TTGACAGGTGTCTG	13	ACTTGACAGGTGTCTG
	RB1	E2F-4	chr13	48877861	TTTTCCCGC	9	TGACGTTTTCCCGC
	RB1	YY1	chr13	48869988	TGGCTGCCATCCTGGACCC	19	TGGCTGCCATCCTGGACCC
	RB1	ATF-2	chr13	48877835	ACGTGACGCCGCG	13	ACGTGACGCCGCGG
39	SNCG	AP-1	chr10	88719062	CCAATGACTCAGCT	14	CCAATGACTCAGCTCTG
	SNCG	Sp1	chr10	88717632	CGAGGTGGGGCC	12	CGAGGTGGGGCCACAG
	SNCG	Sp1	chr10	88718095	TGGGCAGGGC	10	TGGGCAGGGCAGTAAAT
	SNCG	Sp1	chr10	88718679	GGGGGCGAGGCC	12	GGGGGCGAGGCCCTGG



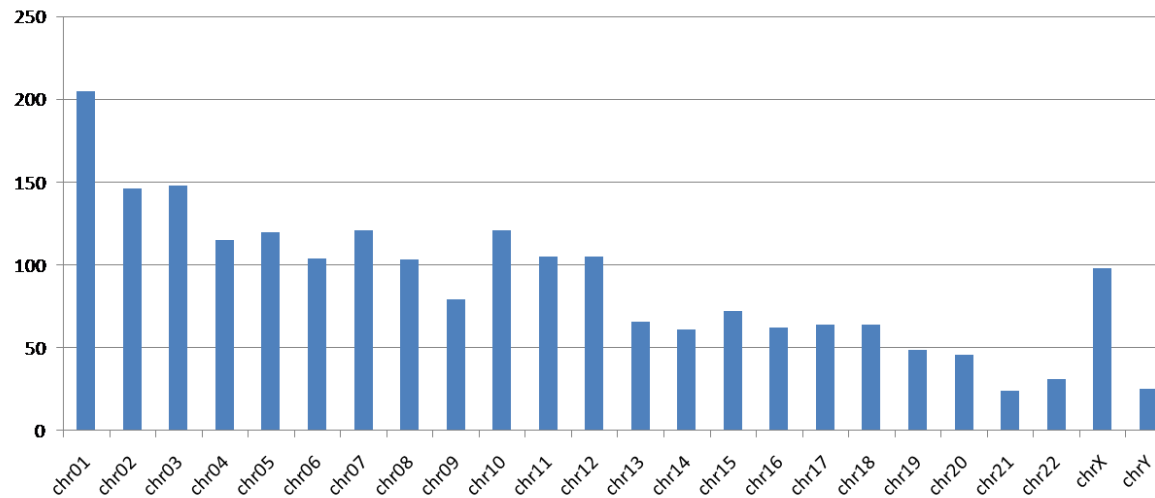
	SNCG	Sp1	chr10	88720307	TGGGCGGGGC	10	TGGGCGGGGCAGCTGCC
40	SRC	STAT3	chr20	35962565	GGCTTCCC	8	GGCTTCCCGACTAGT
	SRC	AP-1	chr20	35956199	TGAATGAATCACCC	14	CTGTGAATGAATCACCC
	SRC	NF-kappaB1 AP-2alpha isoform	chr20	35960516	TGGGGATCCCT	11	TGGGGATCCCTGCCTCCG
	SRC	4	chr20	35975107	GCCCCGGGG	9	AGGGGCCCCGGGGTCAG
	SRC	STAT1	chr20	35953855	TTCTGGGAAGTGGG	14	TTCTGGGAAGTGGGTTG
41	TFAP2C	AP-2alphaA	chr20	55189297	AGCCCCCAGGCC	12	AGCCCCCAGGCCCTAAA
	TFAP2C	AP-2alphaA	chr20	55200469	ACGCCTGGAGGCG	13	ACGCCTGGAGGCGAC
	TFAP2C	AP-2alphaA	chr20	55202326	CGCTCCCAGGCG	12	AGACGCTCCCAGGCG
	TFAP2C	AP-2alphaA	chr20	55205375	TGCCCCGAGGCC	12	TGCCCCGAGGCCCGGGC
	TFAP2C	N-Myc	chr20	55192015	GCCACGTGGCT	11	ACAGCCACGTGGCTCCC
42	TGFA	NF-kappaB1	chr2	70776933	AAGAAAAGTCCCC	13	CTATAAGAAAAGTCCCC
	TGFA		chr2	70780673	CCGGACAGTCCCC	13	GACCGACAGTCCCC
	TGFA		chr2	70782497	AGGGGTCCCC	10	AGGGGTCCCCCACCT
	TGFA		chr2	70785769	GGGAAAGTCCT	11	GCCCCTGGGAAAGTCCT
	TGFA		chr2	70786492	AGGGAAGGACCCC	13	AGGGAAGGACCCCCG
	TGFA		chr2	70794995	AAAGAAAATCCCC	13	AAAGAAAATCCCCTAAACA
43	TGFA	NF-kappaB	chr2	70773652	CTGGAGTGCAGTGG	14	AGGGCCAGGCTGGAGTGCAGTGG
	TGFA	NF-kappaB	chr2	70787960	GAGACATGCC	10	GAGACATGCCACCTTG
	TGFA	p53	chr2	70794296	CACCATGGCAGGGCCTTCC	19	CACCATGGCAGGGCCTTCC
	TGFA	FOXD1	chr2	70792412	TCCCCTGTTTACTTTGG	17	TCCCCTGTTTACTTTGG
	TGFA	FOXF2	chr2	70792586	TCCCTTGTTTACTATGT	17	TCCCTTGTTTACTATGT
	TGFA	RFX1	chr2	70776888	TATGTTGCCACAGTAACAG	19	TATGTTGCCACAGTAACAG
	TGFA	FOXO3b	chr2	70792586	TCCCTTGTTTACTAT	15	TCCCTTGTTTACTATGT
44	TP53	p53	chr17	7578504	AGGGCAGGTCT	11	AGGGCAGGTCTTGCCAGT
	TP53	p53	chr17	7592715	AGGGATGCCC	10	TCTAGGAGAGGGATGCCC
	TP53	p53	chr17	7603194	AGGCATGCACTACCATGCC	19	AGGCATGCACTACCATGCCAGCTAATTTTTTTTTTC
	TP53	p53	chr17	7606362	GGACACGGACGGGCCTGGC	19	GGACACGGACGGGCCTGGC
	TP53	NF-kappaB	chr17	7590848	GGGGAAAACCCCA	13	GGGGAAAACCCCAATCC

45	TP53	NF-kappaB	chr17	7607641	AGAGGAATGTCACCC	15	AGAGGAATGTCACCCCTG
	TP53	E2F	chr17	7572967	TTATGGCGGGAGG	13	TTATGGCGGGAGGTA
	TP53	E2F	chr17	7578450	CATGGCGCGGACGCGG	16	CATGGCGCGGACGCGG
	TP53	AhR	chr17	7589736	GTTGCAGGCGACCCGCGG	18	GTTGCAGGCGACCCGCGG
	TP53	AhR	chr17	7606300	CTCTCTCGCAGGACCGGG	18	CTCTCTCGCAGGACCGGG
	TSG101	HOXA9B	chr11	18541113	TTGACAGTTTCACGT	15	TTGACAGTTTCACGTAC
	TSG101	GR-beta	chr11	18561399	AGGTCTTGTATGTCATGT	18	AGGTCTTGTATGTCATGT
	TSG101	GR-alpha	chr11	18539812	GCAGAACTGATTGCCC	16	GCAGAACTGATTGCCC
	TSG101	GR-alpha	chr11	18557485	CATCCTGACCAACATG	16	TGGTCAGGAGATCAAGACCATCCTGACCAACATG
	TSG101	GR-alpha	chr11	18561400	GGTCTTGTATGTCATG	16	GGTCTTGTATGTCATG
46	TSG101	SRY	chr11	18555909	TCTTATTGTTTAC	13	TCTTATTGTTTACAATA
	TSG101	FOXF2	chr11	18555910	CTTATTGTTTACAATAC	17	CTTATTGTTTACAATAC
	TSG101	p53	chr11	18565471	AAGGCATGTA	10	AAGGCATGTATCTAGG
	VIM	c-Myb	chr10	17258492	GCTAACTGCC	10	TTAATTGCTAACTGCC
	VIM	c-Myb	chr10	17272364	GGGGAAGGCGGTGGAGGG	18	GGGGAAGGCGGTGGAGGGA
	VIM	AP-1	chr10	17257390	GTCATGACTCATTG	14	GTCATGACTCATTGCC
	VIM	AP-1	chr10	17258358	CATGAGTCAC	10	CATGAGTCACCAAACT
	VIM	c-Jun	chr10	17268982	ATAGTGAGTCAGTC	14	ATAGTGAGTCAGTCAGCA
	VIM	c-Jun	chr10	17270535	CCGGTGACTAAGCG	14	CCGGTGACTAAGCG
	VIM	MyoD	chr10	17254926	AGCACCTGCT	10	CCAAAGAAGCACCTGCT
47	VIM	MyoD	chr10	17269446	ACCCCCTGCC	10	TGCAGTACCCCCTGCC
	VIM	MyoD	chr10	17271896	CGGCAGGTGGA	11	CGGCAGGTGGACCAGCT
	VIM	MyoD	chr10	17276770	AGACAGGTGCA	11	ACCGGAGACAGGTGCA
	VIM	Nkx2-2	chr10	17273557	CACAAGTATTT	11	CACAAGTATTTGCTAAC
	VIM	Nkx2-2	chr10	17279346	TTCAAGTGCCT	11	TTCAAGTGCCTTTCTGC
	WNT10B	C/EBPalpha	chr12	49364119	CTCATTGCTTAGAGC	15	CTCATTGCTTAGAGCCC
	WNT10B	C/EBPalpha	chr12	49367508	ACACCCATTGCCTAATA	17	ACACCCATTGCCTAATA
	WNT10B	PPAR-gamma1	chr12	49357239	TAGTAGAGACGGGGTT	16	CGCAGGCGCCCGCCACCACGCCTGGCTAATTTTTTTTTGTATTTTTAGTA
	WNT10B	PPAR-gamma1	chr12	49361981	CTCACCCTGCCCCTTC	16	CTCACCCTGCCCCTCCAGC
	WNT10B	PPAR-gamma1	chr12	49371708	CTCCCCCTTTTCCTTC	16	GTCTCCCCCTTTTCCTTC

WNT10B	E2F	chr12	49366397	GTTAGGCGC	9	ACGGGTTAGGCGC
WNT10B	GATA-3	chr12	49363169	CCCCATCT	9	CCCCATCTCTGGGCTG



**Fig. S2.** Frequency of total occurrences of the 27, p53 binding motifs (from Table 1) present in 20 of its target genes in the human genome. It is evident that each decamer motif is repeated thousands of the time in the genome as predicted by USP programme. This result is predicted by GeneFreq programme of USP.



**Fig. S3**

Here, the distribution of the 4112 ERB-B2 p53 motif, predicted by Frequency Counter (GeneFreq), in the Human genome is shown on each chromosome. Clearly each human chromosome has multiple binding motifs for the p53 motif present in ERB-B2 gene. This highlights the problems in targeting the p53 motif of ERB-B2 gene which may result in the drug binding to all these positions and creating undesired effects. This emphasizes the need for targeting locus specific unique sequence for p53 motif in the ERB-B2 gene.

**Table S4.** A comparison of the melting temperatures ( $T_m$ ) of different p53REs.

Unique motif sequence ( $n$ )	$T_m$ ( $n$ )	$T_m$ ( $n-1$ )	$T_m$ ( $n-2$ )
AGACATGCTCAAGTTCT (17)	61.33	60.04	58.64
ATCTGTACAGACCTTAT (17)	59.17	58.2	58.64
TAGACATGTCTTTTCTTCCC (20)	<b>61.79</b>	59.41	56.82
AGGGATGCCCTACCCC (16)	<b>70.6</b>	68.44	66.05
AGGCAGGCCCTCGGCA (16)	74.28	74.32	71.3
GAGCCTTCCTTGGCATTTC <u>A</u>	66.2	65.6	64.58
AGACATGTCTTTGGCAC (17)	<b>63.06</b>	60.95	60.11
GAGCAAGTCCTGCCATGTT <u>A</u>	66.94	67.54	67.85
AGCCATGCCTGCGCA (15)	70.4	70.25	66.79
GGTCATGCCTGTAATCCCAGCACGTTGGGAGGCTGAG GT (39)	84.14	83.42	82.48
GGACACGCCCCCGA (15)	73.83	73.4	71.88
AGACATGGCCAATCAGC (17)	<b>65.65</b>	62.33	61.09
CTGACTTGTCTCCAGCTG (18)	64.99	64.36	63.25

AGACATGCCTAATTTTTAT (19)	57.09	56	56.14
AGGCATGCGCCATCCAT (17)	69.98	69.68	69.42
AGACTGGGCTGGGCATGGT	73.34	71.93	69.98
AGCCATGCCTTTTCCCCAT (19)	68.31	67.85	67.38
AAAAAGAGAGGGCATGCC	<b>64.58</b>	62.19	58.66
GAGGCATTTCTTCTATA (17)	56.57	56.82	56.82
TGGGATGCCCAGAGT (16)	<b>68.76</b>	66.97	66.05
GAGACATGCCCACCTTG (17)	66.09	65.55	65.99
CACCATGGCAGGGCCTTCC	<b>72.57</b>	70.7	69.98
AGGGCAGGTCTTGGCC (16)	<b>70.6</b>	68.44	65
AGGGATGCCCCAGAGCT (17)	71.27	70.6	67.46
AGGCATGCACTACCATGCCCAGCTAATTTTTTTTTC (36)	74.96	74.38	74.43
GGACACGGACGGGCCTGGC (19)	<b>77.98</b>	75.6	73.87
AAGGCATGTATCTAGG (16)	<b>59.58</b>	56.68	55.03

**Fig. S5.**

Major groove (upper panel) and Minor groove dimensions (lower panel) of different p53 motifs do not reveal any common DNA geometry. p53-DNA complexes, for which crystal structures are available, were analysed by 3DNA programme to extract major and minor groove dimensions.

