

Electronic Supplementary Information

***In silico* identification of potential quadruplex forming sequences in lncRNAs of cervical cancer**

Deepshikha Singh^{1, ‡}, Nakshi Desai^{1, ‡}, Viraj Shah¹ and Bhaskar Datta^{*, 1, 2}

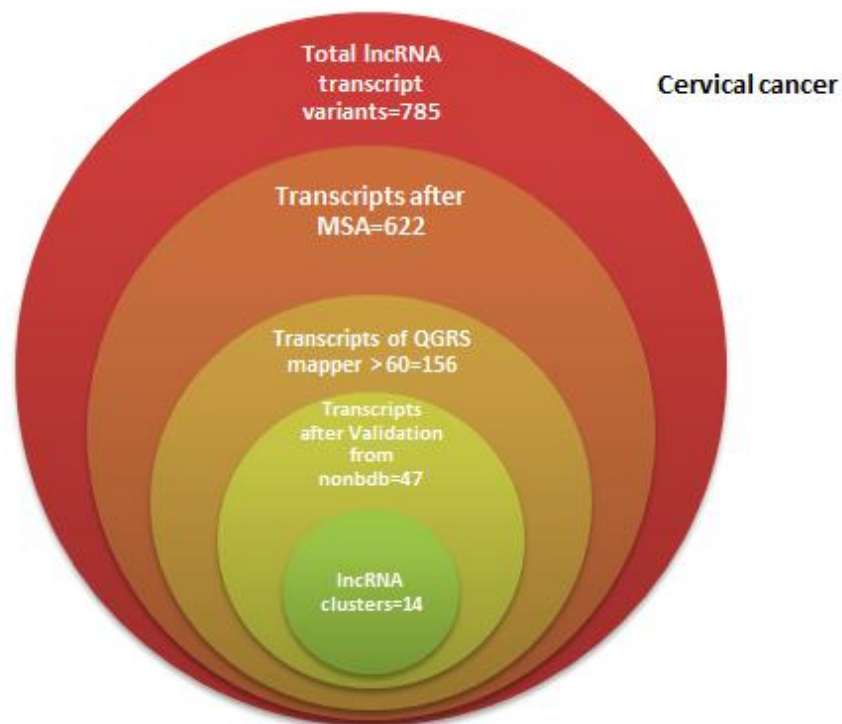
¹Department of Biological Engineering, ²Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat, India

*** Corresponding author**

* Bhaskar Datta, Department of Biological Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar 382355 India.

[‡] These authors contributed equally to the work.

Telephone: 079-2395-2427; Fax: 079-2397-2622; E-mail: bdatta@iitgn.ac.in



Supplementary Figure S1: Distribution of lncRNAs for cervical cancer at different stages of *in silico* analysis.

Supplementary Table S1: DNA oligonucleotides used for in vitro transcription

lncRNA name	Sequence
TERRA	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCCTAACCTAACCTAACCTAATTT CGCTATAGTGAGTCGTATTA
SNGH20	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCAGGCCCAAGCCAAACCCTTTCG CTATAGTGAGTCGTATTA
MEG3	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCCTTGGCCAGGTCCCCCTCCTGAG AATTCCCTTTCGCTATAGTGAGTCGTATTA
LINP1	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCCTTACAGAGTGCCCTGGTCCCCAT ACCCTCTCCTACCCCTTTCGCTATAGTGAGTCGTATTA
CRNDE-R1	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCGAGGCCAGGCCCTAGCCCTTTC GCTATAGTGAGTCGTATTA
CRMDE-R1	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTGCCCGCCCCGAACCCCGACACCCTTTC GCTATAGTGAGTCGTATTA
T7 promotor sequence	TGTGTTCTTGTTCTCTGTCGTTGTTTTGTG

Supplementary Table S2: LncRNA isoforms in clusters and the QGRS analysis of the transcript sequence. QGRS analysis parameters used were as follows- Max Length: 45; Min G group: 3; Loop size: 1-14.

lncRNA H19	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_002196.2	2	14	26	GGGGGGTGGGGGGTGGGGGTGGGGGG	144
			879	36	GGGCAGTGCAGGGTGTACCAGAAGGGGAG TGGGGG	67
Variant 2	NR_131223	2	14	26	GGGGGGTGGGGGGTGGGGGTGGGGGG	144
			879	36	GGGCAGTGCAGGGTGTACCAGAAGGGGAG TGGGGG	67
		2	14	26	GGGGGGTGGGGGGTGGGGGTGGGGGG	144
			879	36	GGGCAGTGCAGGGTGTACCAGAAGGGGAG TGGGGG	67

lncRNA MEG3	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_002766.2	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 2	NR_003530.2	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 3	NR_003531.3	1	537	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 4	NR_033358.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 5	NR_033359.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 6	NR_033360.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 7	NR_046464.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 8	NR_046465.2	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 9	NR_046466.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 10	NR_046467.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 12	NR_046469.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 13	NR_046470.2	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64

Variant 14	NR_046471.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 15	NR_046472.1	1	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
Variant 16	NR_046473.1	9	561	34	GGGAAATTCTCAGGAGGGGGACCTGGGCCA AGGG	64
			3077	37	GGGTCCTCCTGATGAGGGTGTGGGGTCCCC CCTGGG	65
			3693	31	GGGAGGGAGGAGTGGGGCCACAGAGAGAG GG	62
			3878	38	GGGCATTGTAGACAGGGATGAGGGTCATTG ATCCTGGG	65
			6706	37	GGGGGCAGCAGAGGGTCAGGGAGCAATAG GCCCTGGG	62
			7453	33	GGGAGCACGTGGGAGGCAGGGGAGTGGGG TGGG	70
			7539	21	GGGTGGGATGGGGAGTGCGGG	68
			7810	43	GGGCCACATGAGCTGGGCCTCGTGGGCCTG ATGTGGTGCTGGG	64
			8001	20	GGGGCGGGAGGGGTTTCAGGG	70

lncRNA NEAT 1	NCBI ID	PQS found	Position	Length	PQS	G- score
NEAT1 Variant Menbeta	NR_131012.1	11	14	33	GGGAGGGATGCGCGCCTGGGTGTAGTTGTG GGG	63
			425	30	GGGTGCAGGGCTGCAGGAGGGCCGGGAGG G	68
			2070	28	GGGAGTTGTGGGGACGGGGAGAAGGGG	71
			2171	27	GGGATGAGGGTGAAGAAGGGGAGAGGG	69
			2250	38	GGGAATTCTCTGGGGCTTTGGGGAATTTAGT GCGTGGG	66
			3226	34	GGGAAGGGGATGGGGATTGTGGGGAGAGA GTGGG	71
			4559	34	GGGGAGATAAGGGCTGAACTGGGACTAGCC AGGG	71
			19560	33	GGGGTGGGGACCTTTGGGCTCTGACCCGAA GGG	63
			21342	35	GGGATGATGGGGCTTGAGCAAAGTGGGGGA GGGGG	65
			21866	40	GGGTGGCCTGGGGCAAGTACCGGCTGAGGG TCCAGGTGGG	65

			22619	15	GGGAGGGAGGGAGGG	72
NEAT1 Variant MENEPSIL ON	NR_028272.1	6	14	32	GGGAGGGATGCGCGCCTGGGTGTAGTTGTG GG	63
			425	30	GGGTGCAGGGCTGCAGGAGGGCCGGGAGG G	68
			2070	28	GGGAGTTGTGGGGGACGGGGAGAAGGGG	71
			2171	27	GGGATGAGGGTGAAGAAGGGGAGAGGG	69
			2250	38	GGGAATTCTCTGGGGCTTTGGGGAATTAGT GCGTGGG	66
			3226	34	GGGAAGGGGATGGGGATTGTGGGGAGAGA GTGGG	71

lncRNA SNHG20	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_027058.1	1	1143	21	GGGTTTGGGCTGGGGCCTGGG	72

lncRNA GHET1	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_130107.1	1	1171	43	<u>GGGAAGGGGGGTGGTGCCATTCTGGGAGA</u> <u>GGAGGTTTCTGGG</u>	64

lncRNA CRNDE	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_034105.4	0	-	-	-	-
Variant 2	NR_034106.3	2	230	22	GGGCTAGGGCCTGGGCCTCGGG	71
			279	22	GGGTGTCGGGGTTTCGGGGCGGG	70
Variant 3	NR_110453.2	0	-	-	-	-
Variant 4	NR_110454.2	2	230	22	GGGCTAGGGCCTGGGCCTCGGG	71
			279	22	GGGTGTCGGGGTTTCGGGGCGGG	70
Variant 5	NR_170995.1	2	230	22	GGGCTAGGGCCTGGGCCTCGGG	71
			279	22	GGGTGTCGGGGTTTCGGGGCGGG	70

lncRNA HAGLR	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_110458.1	1	927	39	GGGCCAGTGGGCTGGTACAGACTAGGGAAG AGAAAAGGG	64

Variant 2	NR_110459.1	2	102	31	GGGCTGCCTGGGCAAGGGCGGGGCGAGAG GG	71
			878	39	GGGCCAGTGGGCTGGTACAGACTAGGGAAG AGAAAAGGG	64
Variant 3	NR_110460.1	2	102	31	GGGCTGCCTGGGCAAGGGCGGGGCGAGAG GG	71
			878	39	GGGCCAGTGGGCTGGTACAGACTAGGGAAG AGAAAAGGG	64
Variant 4	NR_110461.1	1	783	39	GGGCCAGTGGGCTGGTACAGACTAGGGAAG AGAAAAGGG	64

lncRNA NORAD	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_027451.1	5	79	30	<u>GGGCCGGCGGGTGAAGTGGGGGGCCCCGGG</u>	70
			346	27	<u>GGGCGGCGCTGGGCGTGGTGGGCCGGG</u>	67
			3482	28	<u>GGGTGGTAGAGGTGGGAAAGGGAAGGG</u>	63
			4098	23	<u>GGGGTGGGGGAAGGGGGGGTGGG</u>	71
			4377	42	<u>GGGATGGTGGGAAGAGGGAGAAGAGGAAGG</u> <u>GTGAAGGGAAGGG</u>	70

lncRNA CASC2	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_026939.1	2	114	34	GGGAACGTGAAGGGGCGGGGCCTCGAGGTC AGGG	64
			193	40	GGGCGCAATGGCGGGGACCCGGGGCTTGAT GCTGCTTGGG	65
Variant 2	NR_026940.1	2	114	34	GGGAACGTGAAGGGGCGGGGCCTCGAGGTC AGGG	64
			193	40	GGGCGCAATGGCGGGGACCCGGGGCTTGAT GCTGCTTGGG	65
Variant 3	NR_026941.1	2	114	34	GGGAACGTGAAGGGGCGGGGCCTCGAGGTC AGGG	64
			193	40	GGGCGCAATGGCGGGGACCCGGGGCTTGAT GCTGCTTGGG	65

lncRNA SLC16A1- AS1	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_103743.1	1	549	29	GGGTGGGGAGACTGGCCAGTGGGTTGGG	61

lncRNA MALAT1	NCBI ID	PQS found	Position	Length	PQS	G- score
Variant 1	NR_002819.4	6	590	31	GGGCCGTGGGGGGCTGGCGGCAACTGGGGG G	60
			3790	35	GGGATGGTCTTAACAGGGAAGAGAGAGGGT GGGGG	63
			4851	27	GGGTGGGCTTTTGTGATGAGGGAGGG	60
			6348	24	GGGAAGGGAGGGGGTGCCTGTGGG	67
			6535	22	GGGATGGGAGGAGGGGGTGGGG	70
			7726	24	GGGGAGGGAAAGGGGGAAAGCGGG	69
Variant 2	NR_144567.1	6	590	31	GGGCCGTGGGGGGCTGGCGGCAACTGGGGG G	60
			3556	35	GGGATGGTCTTAACAGGGAAGAGAGAGGGT GGGGG	63
			4617	27	GGGTGGGCTTTTGTGATGAGGGAGGG	60
			6114	24	GGGAAGGGAGGGGGTGCCTGTGGG	67
			6301	22	GGGATGGGAGGAGGGGGTGGGG	70
			7492	24	GGGGAGGGAAAGGGGGAAAGCGGG	69
Variant 3		5	590	28	GGGCCGTGGGGGGCTGGCGGCAACTGGG	60
			3556	35	GGGATGGTCTTAACAGGGAAGAGAGAGGGT GGGGG	63
			5871	24	GGGAAGGGAGGGGGTGCCTGTGGG	67
			6058	21	GGGATGGGAGGAGGGGGTGGG	70
			7249	24	GGGGAGGGAAAGGGGGAAAGCGGG	69

lncRNA FEZF1- AS1	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_036484.1	1	<u>442</u>	34	<u>GGGAGCCGAGGGCACCGCCGCGGGCGCCGC</u> <u>CGGG</u>	63

lncRNA EWSAT1	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_026949.1	1	1557	22	GGGAGCGGGCTGGCGGGGTGGG	69

lncRNA LINP1	NCBI ID	PQS found	Position	Length	PQS	G- score
Transcript 1	NR_138480.1	1	614	41	<u>GGGGTAGGAGAGGGTATGGGGACCAGGGC</u> <u>ACTCTGTAAGGG</u>	69

Supplementary Chart S2: Sequences of RNAs selected for PQS identification

SNHG20

AAGUUGCUGACGGAGCUACUUCCGCCCAGGGGAGUGGGAAGUAAGUGGAGACACGUGCU
UUGGCCUGUUGGAGGGGAAACCCGCUCUCGCCUCCUGGUGGUCGCCGACUCGCAGUCCGC
AGGAUGACUCAGGGCAGCCUGACCACAGUUCCGCCGCCAUCGCCCCUAGCCUGCGGAU
UGGGCUCGCCCCGGGACGAUAACAGAGCUCUGCCGGGGGCUGGAGGCACUGACCGGGUG
ACCAGAGACCCAGAGACCAGACCCCUCCACGGCGCCCGGGAUUUCGGGGACGGCUUCUC
CCAUCGCAAGUUUCAACAGAAAAUGAGAAAUAUCCCCCGACGAUUGGCUGAAAAACAUGC
AGCAACCACUAUUUUCUUCUGCCCCUCGUUGAUGAGAGCAUUCGAAGUGACCUCAGCA
GGGCAUCCAGGUCAGUUUCUGGAAGACUUGUGUGUGUGAUGAAUGAAUAUCUGGUUUU
GUCUCUGCUGGGCCUGUGUGCCUGGAAAGGAAUUGUUUUGGCCUAGGAUCAUCCAGGUU
UGUUUGGUUUAGUUCUUGACCACAUAUUUUUUGAAUGGUGACUGCUUAAGACCCUGUUG
UGUAUGGCUAUAAAUAGAUACACGCCAAGGUGACCACAUAUCUCUGUGGUUCCUGUGCCC
GCCGGCCAUGUGUCUGAGUGUGUAGCCUCUCAUCAUCUAAAGGAACUUUGGCUGCAGAG
GGGAGGCCUGGUCCCAUGGGAAGUUUUGGGAGCGCAGCAGCAGGUGGGUCAAACGACC
CAGCAAGUGCCUCCUAGCACUAGAGGCGGAGGCCACCACGUUGUCCACGUGGGGUUU
CUGAUCCCAGCUCCCCACCAGCCUGCUGGACCUCGGGCAGGUCCCUCUUGUUGUACCU
CCAUUACUUCUUCAGUAAGAUGGGGACACUGAAGAUGACCAUGCCUCCCACCAGAUUGG
UGCUUUUUUUGUUUUAUAAGGGACAGAGUCACUAUGUUGCCUAUAGGCUGGUCUCA
AGCUCUUGGCCUCAGUUUCCUGCCUCAGCCUCAAUAUGCUGGGGUGACAGGCAUGA
GCCACUGCACUGGGCCAGAUUGGUACAUAUUGAGGGUUAAAUGAGCAAUCCUAUAAAGC
ACUCAGGAGAGGGCUGUGGGGUUUGGGCUGGGGCCUGGGAGCACUUUAGACGUGGUAGC
UAUGUUGUUGUCACCUUUCUCCUCUUAAGCAUUUUUUGUUUGUUUUGUUUUGUUU
UUGAGACAGAGUUUCGCUCUUGUGGCCCAGGCUGGAGUGCGGUGGUGUAGCCUCGGCUC
ACUGCAUCCUCCGCCUCCCGGGUUAAGUGAUGGGAUUACAGGUGCCUGCCACAACGCCC
AGCUAAUUGUUUGUAUUUUUAGUAGAGACGAGGUUUCACCAUGUUGGCCAGGCUGGUCU
CGAACUCCUGACCUCAGGUGAUCCACCUGCCUUGGCCUCCCAAAGUGCUGGGAUUACAGG
UAUGAGCUACCGCGCCCGGCACAAAGCAUUUUCAGCUGUGAAAUCCCUGGGGUGCUGC
AGUAGAGGAAGGUUUGGGGUCUUGACGGGUGCUAUUUGCCACGGAAAGAUGCCUCCUG
CUCCCAGGGCAGGAGAGCCGAGGUAAGACUUACUGUAGGCUGUCGUUUUUUUUGUUUGU
UUUUUGUCUUUGCGAUGGAGUCUCACUCUGUCGCCAGGCUGGAGUGCAGUGGCAUGAUC
UUGGCUCACUGCAGAACCUCACCUCUCCAGGUUCAAGCGAUUCUCCUGCCUCAGCCUCC
AAGUAACUGGGAUUACAGGCACAUGCCCCACAACCAGCUAAUUUUUUUAUUUUUAGUAG
AGACAGGGUUUCACAUUGUGCCAGGCUGGUCUUGAACUCCUGACCUCAGGUGAUCCGC
CCGCCUCGGCCUCCCAAAGUGCUGGGAUUACAGACAUGAGCCACUGCGCCCAGCCAGGCU
GUUGUUUUUUUACCUCUUGUUUGCACAUAUUGGGCCACUCACAAGAGUGUAUACCCUG
UGAUAAACAGUUACCUACAUCUCCUCUGCAUGCUUGUCUUUAGAGGAAGGAAAUGUAU
UAAUUGCCCAAAGUAAUAUAUUGUGUUAAGAUGUGAUUAUAUACUGGGGAAAAAAAAG
UGUAUAUUGACAUUUCUGGAAUAAACCACUUUGAUUCCCAAAAAAAAAAAAAA

LINP1

CCCUCGUCUCCUUGACUCUGGGUGGGCUGUGUGACUUUCCUUGCUCAAUAGAAUGAG
AAGGAGGUGGCGCACCGGCUACAAGGCUAGACCGGGGGCUCGCAUAUCUCCACUUGCAG
CUGCCACUGCCAUAAGAAGACUUGCUCAGCCAGCUCUCCUGGGCCCCAGGAAAAGGACAA
GUGUCACCCAGAACAGAGCCACCUCAGCUGCACUCAGCUCUCCAGAAGCUGGCCAGCCGG
UCCAGUACACCUUUUAAAUAUUGUCCUCUACGUGCCGGUGUGGAAGUAGCCCGGAUGCA
AUUGAAUGAACACAGACGGUGCUUUCAGGACGGCGCUGUGCUUUCAGGAUGGUGCU
GUGCUUUCAUUCAUUUGGGUAGCUCCUCUGUGAGCCUCCAGCGCCGACUGCAGAGCCCC

CACUCUCCAGCCUGCAAGACCCCCGAAAUUCAAGCCACACAAAGAAAGGAGGAGGGGGCCG
UUGGCAUUUACUGAACCUUAUAAAACUGUCAGCAAAACAGCCCUUAGGCUUGGACUCCC
UGCUAGCCGGGUUUUACGGUGCUGAAGUCAGCAUCUUGAUUCAGCUGCAUAAAUAUUCU
CCUGCAGUCCUGCAAGGCCUGGGGUAGGAGAGGGUAUGGGGACCAGGGCACUCUGUAAG
GGCUGGGAUAGGAACCCCAGGGAAUAAGACAGACCAACUGCGGGACUUCAGACUCCACU
GCAGCCGGGAUCGGGUUGUUGUUAUUUCUUAAGCAAUUUCUAAAUUCUGUAUUGACUC
UCUCAUGCAUGUAACUGAUCCUUAAGAUUUUGUCAGCCAAAUGACUAAAGGAUUUAGAAA
UAAAAAAAAA

MEG3

AGCCCCUAGCGCAGACGGCGGAGAGCAGAGAGGGAGCGCGCCUUGGCUCGCUGGCCUUGG
CGGCGGCUCUCAGGAGAGCUGGGGCGCCACGAGAGGAUCCCUCACCCGGGUCUCUCCU
CAGGGAUGACAUAUCCGUCCACCUCUUGUCUUAAGGACCACCUCUCUCAUGCUGA
GCUGCUGCCAAGGGGCCUGCUGCCCAUCUACACCUCACGAGGGGCACUAGGAGCACGGUUU
CCUGGAUCCCACCAACAUAACAAGCAGCCACUCACUGACCCCCAGGACCAGGAUGGCAAA
GGAUGAAGAGGACCGGAACUGACCAGCCAGCUGUCCCUCUUAACCUAAAGACUUAACCA
AUGCCCUAGUGAGGGGGCAUUGGGCAUUAAGCCUGACCUUUGCUAUGCUCUAUACUUG
ACUCUAUGAGUACUUCUUAUAGUCUUGUCUUGUGUUCACCUGCUAGCAAACUGGAGU
GUUCCCCUCCCCAAGGGGGUGUCAGUCUUGUCGACUCUGUCAUACCCUUAUGA
UGUCCUGAAUGGAAGGAUCCCUUUGGGAAAUUCUCAGGAGGGGGACCUGGGCCAAGGGC
UUGGCCAGCAUCCUGCUGGCAACUCCAAGGCCUUGGGUGGGCUUCUGGAAUGAGCAUGC
UACUGAAUACCAAAGGCACGCCCAGCCUCUCUGAAGAUCUCCUUAUCCUUAUCCUGGGG
GAAUGGGGUCGAUGAGAGCAACCUCCUAGGGUUGUUGUGAGAAUUAUUUGAGAUAAAA
GAGGCCUCAGGCAGGAUCUGGCAUAGAGGAGGUGAUCAGCAAAUGUUUGUUGAAAAGGU
UUGACAGGUCAGUCCCUUCCCACCCUCUUGCUUGUCUUAUUGUCUUAUUUAUUCUCC
AACAGCACUCCAGGCAGCCCUUGUCCACGGGCUCUCCUUGCAUCAGCCAAGCUUCUUGAA
AGGCCUGUCUACACUUGCUGUCUUCUUCUCCUACCUCCAAUUUCCUCUUAACCCACUGC
UUCUGACUCGCUCUACUCCGUGGAAGCACGCUCACAAAGGCACGUGGGGCCGUGGCCCGG
CUGGGUCGGCUGAAGAACUGCGGAUGGAAGCUGCGGAAGAGGCCCUUGAUGGGGCCACC
AUCCCGGACCCAAGUCUUCUUCUGGGCGGGCCUCUCGUCUCCUUCUUGGUUUGGGCGGAA
GCCAUCACCUGGAUGCCUACGUGGGAAGGGACCUCGAAUGUGGGACCCCAGCCCCUCUCC
AGCUCGAAAUCCCUCACAGCCACGGGGACACCCUGCACCUAUUCCCACGGGACAGGCUG
GACCCAGAGACUCUGGACCCGGGGCCUCCCCUUGAGUAGAGACCCGCCUCUGACUGAUG
GACGCCGUCAGCUGGGGUCAGACCCGUGGGCUGGACCCUGCCCACCCCGCAGGAACCC
UGAGGCCUAGGGGAGCUGUUGAGCCUUCAGUGUCUGCAUGUGGGAAGUGGGCUCCUUA
CCUACCUCACAGGGCUGUUGUGAGGGGCGCUGUGAUGCGGUUCCAAAGCACAGGGCUUG
GCGACCCCCACUGUGCUCUCAAUAAAUGUGUUAUCCUGUCUUAACAAAAA

CRNDE

GCAGCCGCAGCCCCUGGCGCUAACGGUCGGUAACGGCCCCGCGCGCGCCGCCCGGGGG
CUCGCGCCAGCCACGAGGGAGCGUCCGCGGCCCGCGCGCCCGCGCGGCGGAGGAGAGGUG
AGCCCCCGCCGGGCCAGGCCUCUGGCCGCGCCGUCGCCCCUCUAGUCGUGUCCCCUCG
UGGGCCGAACGGACGCGGCGGUGCCCCGCGCCGACCAGACGUCCCGUGGGCUAGGGCCU
GGGCCUCGGGCCGCGUCGGCGCCGGUCGAGCCUCUCCGGGUGUCGGGGUUCGGGGCGGGC
GCGCGUGGGCGUGGCUCUUCUGUCCACGCCUGUCCCUUCGUCGCCGCGGCUCUCGUCCG
GGACACGGCUUCCGGAGUAGAGCCCUUGGAGGUGUUAAGUGUGAUGCUUCCAUAAUAC
AUUUGGAUGCUGUCAGCUAAGUUCACUUCUGAACUAAGGGGUUCCUCCAAAUGUUGGCU
GAAAUUCAUCCCAAGGCUGGUCUGCAAAGUCUGCAAUUAUAAUGGAGCUACUGUACUG
GCUAUUGGAAGGAGGAGAUUCUGAAGAUAAAGGAGGUAAAACCUGUUUAGAAAUUAAAA

AUGAGUUACGAUUUAAAGAAAAUUCAGAUGACUCAUUGUGAGUGCUAGUUCUCUUGUA
GGAUGCCACUGGAAAUGUUGAAAUGAAAAUUAUUCAGCCGUUGGUCUUUGAAAUUUCCU
GUGAUGUGUUUCAUUCUAGAUGCAAAGAACAUGGAAAAUCAAGUGCUCGAGUGGUU
UAAAUAUGUUUUGGGUAUUCUGUUUAUAGACUAUAAUACUUUCCAAUUAAAAUCCU
CAGUUGUCACGCAGAAGAAGGUUAAGCUGUAUUUGAUUGCCAGUUUUACUGAAAAUGCU
UAGUAUUUUACAGUAUCACCAAAUUAUUUUUGUUUAGCCAAGGUUAUAGGAAAAUAAA
AUAAAUUGUAUAGGUUGACUUUUUUUCUAAAAUGUCUUUAUUGGAUUGAAUGAAUGUUU
AUACCUGAAAAAAAAGGUUCAAAAAA

Supplementary Chart S2: FASTA Sequences of proteins under study in this work.

>TP53

WVDSTPPPGTRVRAVAIYKQSQHMTEVVRRCPHHERCSDSD

>CDKN2A

KGAAAAELGPGGGENMVRRLVTLRIRACGPPRVFVHHISWFTGEWAAPGAPAAVALVLM
LLRSQRLGQQPLRRPGHDDGQRPSGAAAAPRRGAQLRRPRHSHPTRARRCPGGLPGHAGGAA
PGRGAAGRARCLGPSARGPG

>NUCLEOLIN

MVKLAKAGKNQGDPPKMAPPPKEVEEDSEDEEMSEDEEDDSSGEEVIPQKKGKAAATSARKV
VVSPTKKVAVATPAKKAATVTPGKKAATPAKKTVTPAKAVTTPGKKGATPGKALVATPGKKGAA
IPAKGAKNGKNAKKEDSDEEEDDDSEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDEDDED
EDEDEDDDDDEEDDSEEEAMETTPAKGKKAQVVPVKAKNVAEDEDEEEDDEDEDDDDDEDDED
DDDEDDEEEEEEEEEEPVKEAPGKRKKEMAKQKAAPAKKQKVEGTEPTTAFNLFVGNLNFNKS
PELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDLEKALELTGLKVFGNEIKLEKPKGKDSK
KERDARTLLAKNLPYKVTQDELKEVFEDAAEIRLVSKDGKSKGIAYIEFKTEADAETFEKQGT
GRSISLYYTGEKGQNQDYRGGKNSTWSGESKTLVLSNLSYSATEETLQEVFEKATFIKVPQNQNGKS
KGYAFIEFASFEDAALNSCNKREIEGRAIRLELQGPRGSPNARSQPSKTLFVKGLSEDTTEETLKE
FDGSVRARIVTDRETSSKGFVDFNSEEDAKAAKEAMEDGEIDGNKVTLDWAKPKGEGGFGR
GGGRGGFGGRGGGRGGFGGRGGFGGRGGFGGRGGFGGRGGGGDHKPQGGKTKFE

>DHX36 isoform 1

MSYDYHQNWGRDGGPRSSGGGYGGGPAGGHGNGRSGGGGGGGGGGRGGRGRHPGHLKGREI
GMWYAKKQKQKNKEAERQERAVVHMDERREEQIVQLLSVQAKNDKESEAQISWFAPEDHGYG
TEVSTKNTPCSENKLDIQEKKLINQEKKMFIRNRSYIDRDSEYLLQENEPDGTLDQKLLEDLQKKK
NDLRYIEMQHFREKLPSYGMQKELVNLIDNHQVTVISGETGCGKTTQVTQFILDNYIERGKGSACRI
VCTQPRRISAISSAERVAERAESCGSGNSTGYQIRLQSRPRKQGSILYCTTGILQWLQSDPYLSSVS
HIVLDEIHERNLQSDVLMTVVKDLLNFRSDLKVLMSATLNAEKFSEYFGNCPMIHIPGFTFPVVEYL
LEDVIEKIRYVPEQKEHRSQFKRGMQGHVNRQEKEEKEAIYKERWPDYVRELRRRYSASTVDVIEM
MEDDKVDLNLIVALIRYIVLEEDGAILVFLPGWDNISTLHDLMSQVMFKSDKFLIPLHSLMPTVN
QTQVFKRTPPGVRKIVATNIAETSITIDDVVYVIDGGKIKETHFDTQNNISTMSAEWVSKANAKQR
KGRAGRVQPGHCYHLYNGLRASLLDDYQLPEILRTPLEELCLQIKILRLGGIAYFLSRLMDPPSNEAV
LLSIRHLMELNALDKQEELTPLGVHLARLPVEPHIGKMILFGALFCCLDPVLTIAASLSFKDPFVIPLG
KEKIADARRKELAKDTRSDHLTVVNAFEGWEEARRRGFRYKDYCWEYFLSSNTLQMLHNMKGQ
FAEHLLGAGFVSSRNPKDPESNINSNEKIIKAVICAGLYPKVAKIRLNLGKKRKMVKVYTKTDGLV
AVHPKSVNVEQTDHFHYNWLIYHLKMRTSSIYLDCTEVSPYCLLFFGGDISIQKDNQETIAVDEWI
VFQSPARIAHLVKELRKELDILLQEKIESPHPVDWNDTKSRDCAVLSAIIIDLIKTQEKATPRNFPPRFQ

DGYYS

>DHX36 isoform 2

MSYDYHQNWGRDGGPRSSGGGYGGGPAGGHGGNRRSGGGGGGGGGGGRGGRGRHPGHLKGREI
GMWYAKKQGQKNKEAERQERAVVHMDERREEQIVQLLSVQAKNDKESEAQISWFAPEDHGYG
TEVSTKNTPCSENKLDIQEKKLINQEKKMFRIIRNRSYIDRDSEYLLQENEPDGTLDQKLLDLQKKK
NDLRYIEMQHFREKLPSYGMQKELVNLIDNHQVTVISGETGCGKTTQVTQFILDNYIERGKGSACRI
VCTQPRRISAISSVAERVA AERAESCGSGNSTGYQIRLQSRLPRKQGSILYCTTGIIQLWLQSDPYLSSVS
HIVLDEIHERNLQSDVLMTVVKDLLNFRSDLKVILMSATLNAEFSEYFGNCPMIHIPGFTFPVVEYL
LEDVIEKIRYVPEQKEHRSQFKRGFMQGHVNRQEKEEKEAIYKERWPDYVRELRRRYSASTVDVIEM
MEDDKVDLNLIVALIRYIVLEEEDGAILVFLPGWDNISTLHDLMSQVMFKSVNQTQVFKRTPPGV
RKIVIAATNIAETSITIDDVVYVIDGGKIKETHFDTQNNISTMSAEWVSKANAKQRKGRAGRVQPGHC
YHLYNGLRASLLDDYQLPEILRTPLEELCLQIKILRLGGIAYFLSRLMDPPSNEAVLLSIRHLMELNAL
DKQEELTPLGVHLARLPVEPHIGKMILFGALFCCLDPVLTIAASLSFKDPFVIPLGKEKIADARRKEL
AKDTRSDHLTVVNAFEGWEEARRRGFRYEKDYCWEYFLSSNTLQMLHNMKGQFAEHLLGAGFVS
SRNPKDPESNINSNEKIIKAVICAGLYPKVAKIRLNLGKKRKMVKVYTKTDGLVAVHPKSVNVEQ
TDFHYNWLIYHLKMRTSSIYLYDCTEVSPYCLLFFGGDISIQKDNDQETIAVDEWIVFQSPARIAHLV
KELRKELDILLQEKIESHPVDWNDTKSRDCAVLSAIIDLIKTQEKATPRNFPPRFQDGYYS

>SRSF1 isoform1

MSGGGVIRGPAGNNDCRIYVGNLPPDIRTKDIEDVFYKYGAIRDIDLKNRRGGPPFAFVEFEDPRDA
EDAVYGRDGYDYDGYRLRVEFPSPRSGRGTGRGGGGGGGGGAPRGRYGPSPRRSENRRVVVSGLPSPG
SWQDLKDHMREAGDVCYADVYRDGTGVVEFVRKEDMTYAVRKLDNTKFRSHEGETAYIRVKVD
GPRSPSYGRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSRHSRSRSRT

>SRSF9

MSGWADERGGEGDGRIYVGNLPTDVREKDLEDLFYKYGRIREIELKNRHGLVPFAFVRFEDPRDAE
DAIYGRNGYDYGQCRLRVEFPPTYGGRGGWPRGGRNGPPTRRSDFRVLVSGLPSPGSWQDLKDH
MREAGDVCYADVQKDGVGMVEYLRKEDMEYALRKLDDBTKFRSHEGETSYIRVYPERSTSYGYSRSR
SGSRGRDSPYQSRGSPHYFSPFRPY

>TLS

MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGY
GTQSTPQGYGSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSSSYGQPQSGSYSQQPS
YGGQQQSYGQQQSYNPPQGYGQQNQYNSSSGGGGGGGGGGNYGQDQSSMSSGGGSGGGYGNQ
DQSGGGGSGGYGQQDRGGRGRGGSGGGGGGGGGGYNRSSGGYEPRGRGGGRGGRGGMGSDR
GGFNKFGGPRDQGSRHDSEQDSDNNTIFVQGLGENVTIESVADYFKQIGIKTNKKTGQPMINLY
TDRETGKLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGGGNGRGGGRGGRG
PMGRGGYGGGGSGGGGRGGFPSSGGGGGGGQQRAGDWKCPNPTCENMNFNFSWRNECNQCKAPK
PDGPGGGPGGSHMGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGGDRGGFGPGKMDS
RGEHRQDRRERPY

>Telomeric repeat-binding factor 2

MAAGAGTAGPASGPGVVRDPAASQPRKRPGREGGEGARRSDTMAGGGGSSDGSGRAAGRASRS
SGRARRGRHEPGLGGPAERGAGEARLEEAVNRWVLKFYFHEALRAFRGSRYGDFRQIRDIMQALL
VRPLGKEHTVSRLLRVMQCLSRIIEGENLDCSFDMEAEELTPLESAINVLEMIKTEFTLAEAVVESSRK
LVKEAAVIICIKNKEFEKASKILKKHMSKDPTTQKLRNDLLNIIREKNLAHPVIQNFYSYETFQQKMLR
FLESHLDDAEPYLLTMAKKALKSESAASSTGKEDKQAPAGPVEKPPREPARQLRNPPTTIGMMTLK
AAFKTLGAQDSEAAFAKLDQKDLVLPTQALPASPAKKNKRPRKDENESSAPADGEGGSELQPKN

KRMTISRLVLEEDSQSTEPSAGLNSSQEAASAPPSKPTVLNQPLPGEKNPKVPKGKWNSSNGVEEKE
TWVEEDELQVQAAPDEDSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAAISKNYPFVNRTAV
MIKDRWRTMKRLGMN