

TP53 Intron Derived Concentrations Implicate p53

Kevin Bermeister, Jonathan Dyne, Xinghao Yu: Contributing research: Liran Carmel, Khens Kermesh, Adam Bermeister, Mark Kon, Daniel Shnier

31 March 2017 - Codoncex LLC research for participatory review

Introduction

TP53 is a highly studied gene due to its important regulatory role in DNA, RNA (Marcel, Catez, and Diaz 2015), disease and the progression of cancer. TP53 functions as a tumor suppressor and has been described as the “Guardian of the Genome”. p53, the protein coded by TP53 predominantly binds regulatory response elements (“RE”) in introns of multiple target genes. We compared identical subsequences in introns of multiple gene/transcripts including TP53, BRCA1, men1, PELP1, SET, HIF1A, ULBP1/2 and IRF3. Some identical subsequences also contain the core RE’s of known p53 binding sites or TP53’s intronic, autoregulatory response elements (“ARE”) and strongly correlate with known miRNA’s. We describe the mechanism by which p53 concentrations change as stress free cells become adversely affected. We describe how TP53’s intronic ARE’s contribute their preference to p53 binding and propose a computational method to identify targets.

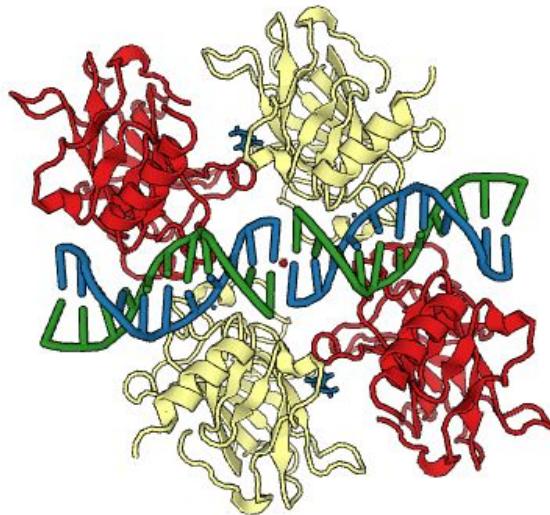
Background

For intronic subsequences that are part of negative feedback loops (Marcel et al. 2010) steady state transcription is, in part 'self-satisfied' by p53 binding ARE’s. Mutant or p53 isoforms may exclusively recognize RE’s or ARE’s by iterating along DNA until identifying their uniquely compatible sequence combinations to which they bind. As soon as a stress condition affects the cell, specific p53 isoforms may be induced to relocate and bind compatible ARE’s and RE’s. More aberrant TP53 transcription may change the concentration of p53 isoforms, which would change ARE and/or RE binding patterns. Scarcity of p53 binding RE’s would also expose those genes to further unregulated or misregulated transcription events that normally include p53 and may bias transcription at other genes.

“p53 can be attenuated directly, by mutation or deletion (Herrero et al. 2016), or indirectly through alterations in methylation, miRNAs, isoform expression and p53 regulators. Six TP53 hotspot mutations and regions potentially affected by methylation are indicated. p53 isoforms arise from the use of two alternative transcription start sites.”

“The existence in the intron of multiple identical subsequences in multiple gene/transcripts that include some or all of p53’s prerequisite [CA-><-AG] and/or [CT-><-TG] and/or [CA/CT-><-TG] (McLure 1998) core binding quarter-sites may indicate a gene/transcripts ARE or RE (Weinberg et al. 2005) state of readiness to bind p53 more completely. It should be noted the ARE or RE quarter sites and variations (Bian and Sun 1997) are each capable

of binding at least one p53 monomer and stronger binding as dimers or tetramers to half and full sites has a tendency to bend DNA and increase affinity.”



[P53 tetramer bound to DNA](#)

Mechanisms of p53 Regulation in Disease

At first we loosely searched [identical subsequences](#) of [multiple gene transcripts](#) that included [CA and/or AG|CT and TG] and/or [CA and CT|TG] where each element [nucleotide pair] was separated by at least one nucleotide. We discovered the intersection of TP53 introns with other genes introns in our database. The lengths of these identical subsequences varied from 8 to more than 55 nucleotides. It is known that the regulatory regions of the most-highly regulated miRNA species are enriched in conserved binding sites for p53 (Hattori et al. 2014). Evidently the discovery, by querying the more frequently recurring quarter and half sites of ARE's and RE's in introns with identical subsequences confirmed that among those longer than 23 nucleotides the significant majority were known miRNAs.

Further, p53 induced (Loayza-Puch et al. 2013) transcription at combinatorial ARE-RE identical subsequences may increase cellular concentrations of these miRNA sequences more rapidly. We consider that increased concentration of ARE-RE subsequences that are capable of maturing as miRNA's suggests they can be present in larger concentrations in regions of the cytoplasm. We found evidence in support that a fraction of the proportion of a given miRNA from an intron is capable of adjusting concentration dynamics such that it affects protein assembly and/or function:

The first relates to a paper on Cytoplasmic Retained Introns (CIRT's) proposing that a small fraction of the localized concentration can tip the balance. “Another important aspect of this CIRT-related phenotypic change is the number of KCNMA1 transcripts retaining i16 that can be detected in cells. Approximately 90% of detectable KCNMA transcripts in the cytoplasm are fully spliced, and 10% of the total KCNMA1 transcript population is comprised of KCNMA1i16 CIRTs. This raises the question of the

physiological relevance of a relatively small number of KCNMA1 transcripts.”
(Buckley et al. 2014)

The second relates to the mechanism of genes in Amyotrophic lateral sclerosis (ALS) that causes concentrations to promote phase separation. “Phase transition by RNA-binding proteins (Taylor, Brown, and Cleveland 2016) that contain low-complexity domains is exquisitely dependent on concentration, and it is probable that the increased accumulation of FUS and hnRNP A1 in the cytoplasm as a consequence of mutations that affect the nuclear localization signals of these proteins is sufficient to drive excess phase separation, as shown by the hyper-assembly of stress granules in cells derived from people with relevant mutations.”

Relative Concentration Dynamics - Codondex Computation

Concentration influences (Yu et al. 2016) DNA-Protein-RNA-Protein interactions in both the nucleus and cytoplasm, which can be significant. Further, it is evident that extremely small changes, often in a small fraction of concentrations that results from dosage sensitive genes (Babu et al. 2011) can be immediately significant.

Analogous to oil droplets in water, subcellular concentrations (Smith et al. 2016) are modified by dynamic elements including DNA, RNA, protein, metabolites, molecules and water. This concentrated environment is determinative for binding of particular elements as is the capacity for ligands to prevail in the pressure of condensed space. The effects on intrinsic protein disorder (Babu et al. 2011), including p53 is also concentration dependent and dosage sensitive. Double stranded DNA association and dissociation inherencies, as influenced by epigenetic factors and induced by proteins (Pontius and Berg 1992) may play a role in organizing (Elbarbary, Lucas, and Maquat 2016) DNA subsequence concentrations to be more or less favorable to binding protein, transcription efficiency and translation.

Intra chromosome, subsequences of the DNA molecule can be thought of as concentrations. Each gene is then a particular pattern of subsequence concentrations that relatively and proximally define DNA of the chromosome. Therefore, our interest in specifics of DNA’s relative concentrations led us to review every adjacent combination of any of the four letters A,G,C or T within a gene of interest. Our DNA kmer algorithm computes and maintains a database of every possible subsequence >7 letters of a given DNA sequence. It captures the smallest single nucleotide change in any subsequence >7 for all possible subsequences of a transcript.

By example, from the transcribed DNA sequence GAGCTTCGAG₆ there is potential to momentarily exist, including as derivative RNA subsequences or sub-concentrations >7 |GAGCTTCG₃| |AGCTTCGA₄| GAGCTTCGA₄| GCTTCGAG₃| AGCTTCGAG₄|. Subscript represents the repeat frequency each subsequence in any >7 derivations of GAGCTTCGAG. In other words if two subsequences have the same repeat frequency they will have the same relative concentration and will likely be subjected to similar dissociation dynamics resulting from transcription events. Since DNA nucleotide order is tightly

maintained this model also represents the complete weighted set of RNA subsequences that can be potentially transcribed into the nucleus and cytoplasm. Because various concentrations of repeatedly transcribed DNA will exist as RNA in the cell at various times, we devised an exhaustive model in which a single cell-genes' 'any-time' potential state is computed. Measuring and sequencing a genes' cytoplasmic RNA concentrations and comparing them to the any-time potential state in same cell transcripts may provide a useful comparative. Using this as a diagnostic tool may also provide a sensitive holistic measure to associate non-coding DNA-RNA states with diseases.

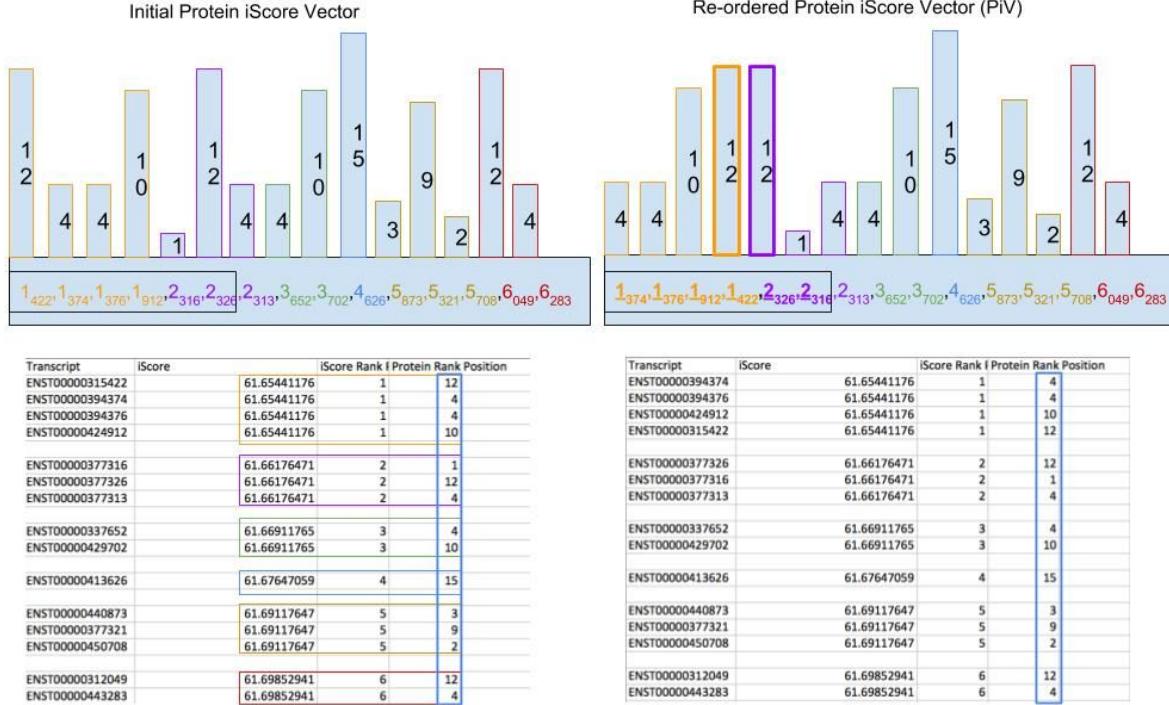
As discussed in this paper, RNA in various forms may bind protein and in some cases cause protein to relocate to regions where it interacts. These types of reactions are abundant in concentrations, which are sometimes determinative to binding and thereafter to downstream reactions. We looked at the intersection of TP53 and IRF3 supported by their known interactions; The novel findings presented by Baresova and colleagues (Baresova et al. 2013) are that "vIRF-3 can significantly decrease the stability of p53 protein, its phosphorylation, tetramer formation, and DNA-binding ability. This results in reduced transcription of the p53-regulated genes, the p21 and Bax genes, which may lead to impaired cell growth or apoptosis. These data indicate that Kaposi's sarcoma-associated herpesvirus evolved mechanisms to attenuate the p53 growth-regulatory pathway and thus contributes to development of virus-associated malignancies."

Method

For any intron, we partition DNA transcript data by its computed, potential subsequences (potential kmers "pkmers") associated with a signature of mRNA or protein encoded by the transcript. For each of multiple same gene transcripts, pkmers are computed by iterating single nucleotide steps to establish every potential subsequence in our database. Once complete for the entire intron, we count repeats (within the intron computation) for each pkmer and compute rankings. We assimilate ranked results from each step of each transcript-intron to a highly structured multi-transcript, multi-dimensional vector.

For any given pkmer (which we may refer to by offset#) we generate an intra transcript-intron sequence ranking. We compute each pkmers' iScore, derived by summing trailing (3') repeats found in it, which we divide by the pkmer length. The protein signature of every pkmer is constant for the transcript-intron. Any step (next offset#) of the transcript can be compared using the vector comprising iScore:Protein signature as assimilated for all transcripts.

The illustration below is for [offset#48333](#) of 15 transcripts of men1. Each Transcript ID, iScore and Protein Rank Position is recorded in a vector (the illustration subscripts last 3 digits of transcript ID). Since protein signature is constant for every pkmer of a given transcript-intron for a gene, order by the protein signature is static ("PhV"). Rules permit shuffling transcripts offset#'s that share the same iScore, as per transcripts 422 and 326 to maximize transcript assimilation in the vector. Finally we assign vector positions 1-15 to transcripts according to their order, this constitutes Position in Vector ("PiV").



By example, a computation of 50 nucleotides of each transcript-intron subsequence amplifies to ± 4000 rows (steps). Each row contains one of all possible pkmers for the total length 50 intron sequence. We sort rows by pkmer aggregate repeats and identify pkmers that contain the most changes in the vector order at the next row. The vector order changes because of a transcripts iScore:Protein signature assimilation to the vector for each offset#.

Membership to the vector is bounded by nucleotide order of each transcript-intron DNA and transcripts' protein signature. Therefore, a change in a transcript PiV from step (row) to step (row) must be understood differently to the row with the most changes in all transcript PiV orderings. The latter being the biggest PiV change, by repeats for all transcript-intron subsequences for the set of pkmer:Protein assimilations.

The result is a selection method, for any given sequence length, from all pkmers of a set of transcripts. The identification is the pkmers that most strongly modified PiV in a successive step when sorted by aggregate pkmer repeats.

TP53 intronic half-sites implicate p53

We demonstrate that TP53 quarter and half sites implicate p53 according to Codondex iScore-PiV computations. Introns of TP53 reveal known ARE half-sites 5'-PuPuPuC(A/T)|(A/T)GPyPyPy-3' for the p53 dimer or tetramer that infer its four nucleotide core as a strong signal for binding preference. The core combinations CAAG, CATG, CTAG or CTTG in the 10 letter ARE or RE context are sufficient to ensure high degrees of half-site binding specificity. Further, p53 dimers can (McLure 1998) bind reversed, adjacent quarter

sites as a reverse half site core 5'-G(A/T)|(A/T)C-3' - GAAC, GATC, GTAC or GTTC, which interestingly we discovered are strongly [implicated in men1](#).

Analogous to setting a combination in a lock then unlocking-locking using the combination: We investigated whether TP53 intron-1 ARE's alter p53 protein through transcription-translation (Gabunilas and Chanfreau 2016) or post-translational modification (Gu and Roeder 1997) in a way that would transmit to binding domains of p53 monomers (Tafvizi et al. 2011) a specific ARE (or RE) preference. One way may be through DNA to by the mechanism of p53 successfully binding intronic and exonic (Yang and Wu 2004) RE's of genes to form tetramers and induce transcription. Further we considered how this may implicate RE's in other genes p53 is known to bind. For RE's we reviewed identical pkmers (>7 letters) of TP53 transcripts with other gene pkmers that also contain the core binding domain. We queried all identical pkmers with any six (6) letter (1-4-1) half-site cores. We initially used "GCAAGC" and discovered several genes that shared identical pkmers with this 6 letter core including HIF1A with TP53. We queried the identical subsequence cores on multiple gene intron subsequences and discovered their intersections with half-site cores in RE's or ARE's.

In summary, monomer-dimer-tetramer binding RE-ARE's suggests their innate sensitivities are reflected in their determinative structures. Binding events are ultimately governed by the relationship between ligand and DNA sequence, which are influenced by complex factors that diminish the chance a tetramer will form and through which transcription will be promoted. This presents a very wide range of statistical outcomes with dependencies in ligand, RE-ARE sequences, upstream and downstream DNA or RNA concentrations and various epigenetic factors.

[The image](#) below from [Excel file](#) was ordered by repeats (column O) and pkmer length. We searched all pkmers containing AGGCAAGCCC - one of four specific half-site binding sequences of interest and used example [TP53 ENST00000420246 @offset:#512580 \[Start 1010,End 1019\]](#). We highlighted (light green) every pkmer (column G) that contained the full pkmer of offset #512580 and in [continuous light green](#) all offsets# that start at nucleotide #1010 of the intron - column H. We observed changes in the vector order (column E) as implicated by repeats and reverse complement ("coffset" -column R).

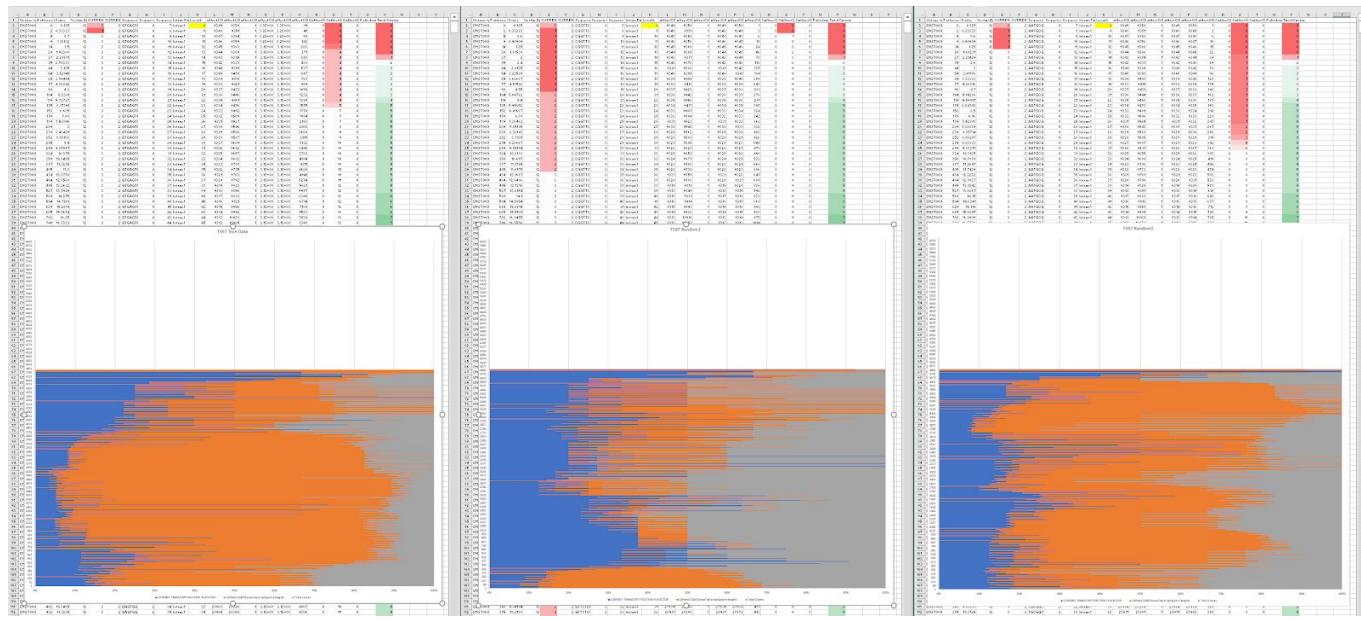
	A	B	C	E	G	H	I	J	R
1	Unique-name	Patterned-seq	Score	CLIPRE	Sequence Text	Sequence	Sequence	Length	CoffseC0
42927	ENST0000420246	539270	12.7027	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1045	37	11
42928	ENST0000420246	522767	5.5	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1029	22	3
42929	ENST0000420246	539231	12.2222	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1044	36	9
42930	ENST0000420246	521744	5.04752	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1029	21	3
42931	ENST0000420246	553920	11.26	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1059	50	22
42932	ENST0000420246	537193	11.7429	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1043	35	8
42933	ENST0000420246	520722	4.6	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1027	20	3
42934	ENST0000420246	553967	10.7765	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1059	49	22
42935	ENST0000420246	536156	11.2647	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1042	34	8
42936	ENST0000420246	519701	4.7579	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1025	19	3
42937	ENST0000420246	551815	10.2917	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1057	48	22
42938	ENST0000420246	535120	10.7873	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1041	33	8
42939	ENST0000420246	518681	3.72222	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1025	18	3
42940	ENST0000420246	550764	17.0805	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1056	47	22
42941	ENST0000420246	534085	10.3125	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1049	32	8
42942	ENST0000420246	517662	3.29412	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1024	17	3
42943	ENST0000420246	543714	17.3261	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1055	46	20
42944	ENST0000420246	533091	8.3871	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1039	31	8
42945	ENST0000420246	516644	2.8125	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1023	16	3
42946	ENST0000420246	548665	16.8444	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1054	45	17
42947	ENST0000420246	532070	9.36867	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1038	30	8
42948	ENST0000420246	515627	2.4	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1022	15	3
42949	ENST0000420246	547617	16.3636	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1053	44	15
42950	ENST0000420246	530986	8.8555	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1037	29	5
42951	ENST0000420246	514611	2	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1021	14	3
42952	ENST0000420246	546570	15.3865	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1052	43	14
42953	ENST0000420246	529955	8.42857	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1036	28	5
42954	ENST0000420246	513596	16.1538	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1029	13	3
42955	ENST0000420246	545524	15.381	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1051	42	12
42956	ENST0000420246	520925	7.96296	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1035	27	5
42957	ENST0000420246	512822	1.25	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1019	12	3
42958	ENST0000420246	544479	14.9324	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1050	41	11
42959	ENST0000420246	527896	7.47614	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1034	26	5
42960	ENST0000420246	515659	0.90909	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1018	11	3
42961	ENST0000420246	543436	14.3	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1049	40	11
42962	ENST0000420246	526668	7	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1033	25	5
42963	ENST0000420246	519557	0.5	5	GAAGGCAAGCCGGAGGTTTTCAAG	101	1017	10	3
42964	ENST0000420246	542382	13.7436	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1043	39	11
42965	ENST0000420246	525941	5.54167	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1032	24	3
42966	ENST0000420246	541350	13.2105	6	AAGGCAAGCCGGAGGTTTTCAAG	101	1047	38	11
42967	ENST0000420246	509946	0.33333	5	GAAGGCAAG	101	1016	9	1
42968	ENST0000420246	524815	5	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1031	23	2
42969	ENST0000420246	540309	12.7027	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1045	37	11
42970	ENST0000420246	509536	0.125	5	GAAGGCAAA	101	1030	22	2
42971	ENST0000420246	523790	5.5	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1027	19	5
42972	ENST0000420246	539269	12.2222	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1045	36	11
42973	ENST0000420246	527266	5.04762	5	AAGGCAAGCCGGAGGTTTTCAAG	101	1029	21	2
42974	ENST0000420246	554973	19.26	6	GGCAAGCCGGAGGTTTTCAAGA	101	1060	50	20
42975	ENST0000420246	538230	11.7423	5	GGCAAGCCGGAGGTTTTCAAGA	101	1044	35	9
42976	ENST0000420246	527143	4.5	5	GGCAAGCCGGAGGTTTTCAAGA	101	1028	20	2
42977	ENST0000420246	553919	18.7755	6	GGCAAGCCGGAGGTTTTCAAGA	101	1059	49	20
42978	ENST0000420246	537192	11.2647	5	GGCAAGCCGGAGGTTTTCAAG	101	1043	34	8
42979	ENST0000420246	520721	4.15739	5	GGCAAGCCGGAGGTTTTCAAG	101	1027	19	2
42980	ENST0000420246	552866	18.2917	6	GGCAAGCCGGAGGTTTTCAAG	101	1058	48	20
42981	ENST0000420246	536155	10.7873	5	GGCAAGCCGGAGGTTTTCAAG	101	1042	33	8
42982	ENST0000420246	519700	3.72222	5	GGCAAGCCGGAGGTTTTCAAG	101	1026	18	2
42983	ENST0000420246	551014	17.8005	6	GGCAAGCCGGAGGTTTTCAAGA	101	1057	47	20
42984	ENST0000420246	529954	7.96296	5	GGCAAGCCGGAGGTTTTCAAG	101	1041	32	8
42985	ENST0000420246	535119	10.3125	6	GGCAAGCCGGAGGTTTTCAAG	101	1025	21	2
42986	ENST0000420246	518680	3.29412	5	GGCAAGCCGGAGGTTTTCAAG	101	1036	20	2
42987	ENST0000420246	550763	17.3061	7	GGCAAGCCGGAGGTTTTCAAGA	101	1056	46	20
42988	ENST0000420246	524084	9.88707	5	GGCAAGCCGGAGGTTTTCAAG	101	1049	31	8
42989	ENST0000420246	517661	3.275	5	GGCAAGCCGGAGGTTTTCAAG	101	1024	16	2
42990	ENST0000420246	549713	16.8444	7	GGCAAGCCGGAGGTTTTCAAG	101	1037	28	6
42991	ENST0000420246	533050	9.38667	5	GGCAAGCCGGAGGTTTTCAAG	101	1021	13	2
42992	ENST0000420246	516643	2.4	5	GGCAAGCCGGAGGTTTTCAAG	101	1023	15	2
42993	ENST0000420246	540064	16.3535	7	GGCAAGCCGGAGGTTTTCAAG	101	1054	44	15
42994	ENST0000420246	532017	8.69556	5	GGCAAGCCGGAGGTTTTCAAG	101	1038	29	8
42995	ENST0000420246	516766	2.0062	5	GGCAAGCCGGAGGTTTTCAAG	101	1022	14	2
42996	ENST0000420246	547516	15.8837	7	GGCAAGCCGGAGGTTTTCAAG	101	1053	43	13
42997	ENST0000420246	530985	8.42357	5	GGCAAGCCGGAGGTTTTCAAG	101	1055	40	9
42998	ENST0000420246	514510	16.1539	5	GGCAAGCCGGAGGTTTTCAAG	101	1021	13	2
42999	ENST0000420246	546569	15.381	6	GGCAAGCCGGAGGTTTTCAAGA	101	1052	42	12
43000	ENST0000420246	535195	1.25	5	GGCAAGCCGGAGGTTTTCAAGA	101	1036	27	5
43001	ENST0000420246	545523	14.9024	6	GGCAAGCCGGAGGTTTTCAAGA	101	1051	41	10
43002	ENST0000420246	528924	7.5	5	GGCAAGCCGGAGGTTTTCAAG	101	1035	26	6
43003	ENST0000420246	512580	0.09090	5	GGCAAGCCGGAGGTTTTCAAG	101	1019	11	2
43004	ENST0000420246	544478	14.4245	7	GGCAAGCCGGAGGTTTTCAAGA	101	1045	35	9
43005	ENST0000420246	527895	7	5	GGCAAGCCGGAGGTTTTCAAG	101	1034	25	6
43006	ENST0000420246	543434	13.8205	6	GGCAAGCCGGAGGTTTTCAAG	101	1049	39	9
43007	ENST0000420246	519568	0.5	5	GGCAAGCCGGAGGTTTTCAAG	101	1018	10	2
43008	ENST0000420246	528867	5.54167	5	GGCAAGCCGGAGGTTTTCAAG	101	1033	24	6
43009	ENST0000420246	520391	12.2222	6	GGCAAGCCGGAGGTTTTCAAG	101	1048	38	9
43010	ENST0000420246	510566	0.33333	5	GGCAAGCCGGAGGTTTTCAAG	101	1017	9	2
43011	ENST0000420246	525840	6.09656	5	GGCAAGCCGGAGGTTTTCAAG	101	1032	23	3
43012	ENST0000420246	514349	12.7277	6	GGCAAGCCGGAGGTTTTCAAG	101	1047	37	9
43013	ENST0000420246	509545	0.125	5	GGCAAGCCGGAGGTTTTCAAG	101	1015	8	0
43014	ENST0000420246	524914	5.54545	5	GGCAAGCCGGAGGTTTTCAAG	101	1031	22	2
43015	ENST0000420246	540306	12.2222	6	GGCAAGCCGGAGGTTTTCAAG	101	1046	36	9
43016	ENST0000420246	523789	5.04762	5	GGCAAGCCGGAGGTTTTCAAG	101	1030	21	2
43017	ENST0000420246	558627	19.26	6	GGCAAGCCGGAGGTTTTCAAGA	101	1051	50	20
43018	ENST0000420246	539268	11.7433	5	GGCAAGCCGGAGGTTTTCAAGA	101	1045	35	9
43019	ENST0000420246	522765	4.8	5	GGCAAGCCGGAGGTTTTCAAG	101	1039	20	2
43020	ENST0000420246	554972	18.7755	6	GGCAAGCCGGAGGTTTTCAAGA	101	1060	49	20
43021	ENST0000420246	538279	11.2647	5	GGCAAGCCGGAGGTTTTCAAGA	101	1044	34	7
43022	ENST0000420246	527472	4.7579	5	GGCAAGCCGGAGGTTTTCAAG	101	1028	19	2
43023	ENST0000420246	553918	10.2917	6	GGCAAGCCGGAGGTTTTCAAGA	101	1059	48	20
43024	ENST0000420246	537191	10.7873	5	GGCAAGCCGGAGGTTTTCAAGA	101	1043	33	6
43025	ENST0000420246	520720	3.72222	5	GGCAAGCCGGAGGTTTTCAAG	101	1027	18	2
43026	ENST0000420246	558655	17.0805</						

The suggested relative concentration score of each pkmer is represented by repeats (column O). The changes in pkmers PiV at each progressive step suggest the TP53 transcript-intron transmits modification to its associated p53 monomer (Dornan et al. 2003). This may be sufficient to identify subsequences that infer modification to p53 (and ligand) to signal specific transcription outcomes to specific binding combinations of ARE's or RE's that promote transcription of RNA derivatives.

ID Gene-Transcript-Intron-Protein Subsequences

We searched all transcript-intron computations to discover the shortest length pkmer that also included the highest number of length 8 pkmers associated with the biggest change in PiV at the next step (row). We considered for any given pkmer length, that the highest number of PiV changes (biggest total vector shifts) would be of interest. The leading candidate was ENST00000359597.

To [verify PiV variations](#) we reviewed (image below) ENST00000359597 (left) against two randomized versions using a transcript wide random letter reordering (middle) and separately a random pick A.C.T.G of each nucleotide (right). The orange bulge confirmed best and most consistent, per offset# correlation of PiV, coffset and PiV shifts occurred in the ENST00000359597 iScore report (left).



We then identified the frequency that length 8 pkmers with biggest PiV shifts were incorporated in their length 50 pkmers that shared the same sequence start nucleotide in the intron;

Offset	Length	SeqStart	SeqEnd	SeqTxt	PivShift
21	8	6	13	GGATCCAT	10
28	8	7	14	GATCCATT	12
36	8	8	15	ATCCATTG	10
109	12	10	21	CCATTGGAAGGG	12
120	8	15	22	GGAAGGGC	10
528	8	32	39	ACCCCCAC	10
666	8	36	43	CCACCCCA	11
990	8	44	51	ACCCCAGC	11
1128	8	47	54	CCAGCCCC	11
2085	13	59	71	CAGAGACCTGTG	11
2145	8	65	72	CCTGTGGG	11
2278	8	67	74	TGTGGGAA	11
2485	8	70	77	GGGAAGCG	11
4851	8	98	105	TTCTGCTC	11
5151	8	101	108	TGCTCTTG	11

As seen above, these sequences are grouped according to their intron DNA sequence start (SeqStart) positions and are overlapping. We were able to immediately identify using offset#21 that its length 50 counterpart at offset#1218 (below) contained 9 other offsets from the above group (_{subscript} below identifies SeqEnd for each pkmer in the table above).

#1218 GGATCCAT₁₃T₁₄G₁₅GAAGGG₂₁C₂₂AGGCCCACCACCCCCAC₃₉CCCA₄₃ACCCCAGC₅₁CCC₅₄C

We then confirmed (below) that the identical sequence data, referred to earlier in this paper contains ENST00000359597, offset#21 and although not a classical p53 binding site, it conforms with lesser binding site rules.

Gene	Transcript ID	Offset#	Identical Sequences	StartSeq	EndSeq	Length
TP53	ENST00000508793	250986	GGATCCAT	708	715	8
TP53	ENST00000413465	21	GGATCCAT	6	13	8
TP53	ENST00000359597	21	GGATCCAT	6	13	8
BRCA1	ENST00000357654	201295	GGATCCAT	634	641	8
BRCA1	ENST00000489037	148785	GGATCCAT	545	552	8
BRCA1	ENST00000491747	205120	GGATCCAT	640	647	8
BRCA1	ENST00000492859	201295	GGATCCAT	634	641	8
BRCA1	ENST00000468300	205120	GGATCCAT	640	647	8
BRCA1	ENST00000591534	201295	GGATCCAT	634	641	8
BRCA1	ENST00000461798	205120	GGATCCAT	640	647	8
BRCA1	ENST00000493919	205120	GGATCCAT	640	647	8
BRCA1	ENST00000352993	201295	GGATCCAT	634	641	8
BRCA1	ENST00000461221	148785	GGATCCAT	545	552	8
BRCA1	ENST00000591849	201295	GGATCCAT	634	641	8
BRCA1	ENST00000346315	205120	GGATCCAT	640	647	8
BRCA1	ENST00000471181	201295	GGATCCAT	634	641	8
BRCA1	ENST00000470026	148785	GGATCCAT	545	552	8
BRCA1	ENST00000477152	201295	GGATCCAT	634	641	8
BRCA1	ENST00000586385	148785	GGATCCAT	545	552	8
BRCA1	ENST00000493795	205120	GGATCCAT	640	647	8
BRCA1	ENST00000494123	205120	GGATCCAT	640	647	8
BRCA1	ENST00000354071	201295	GGATCCAT	634	641	8
BRCA1	ENST00000497488	201295	GGATCCAT	634	641	8
BRCA1	ENST00000478531	205120	GGATCCAT	640	647	8
BRCA1	ENST00000476777	201295	GGATCCAT	634	641	8
BRCA1	ENST00000309486	205120	GGATCCAT	640	647	8

We further confirmed that the ENST00000420246 sequence at offset #21 (below) is also an identical sequence together with several other TP53 transcripts.

Gene	Transcript ID	Offset#	Identical Sequences	StartSeq	EndSeq	Length
HIF1A	ENST00000394997	73623045	TCCTGACT	12134	12141	8
HIF1A	ENST00000394997	154238266	TCCTGACT	17563	17570	8
HIF1A	ENST00000323441	73623045	TCCTGACT	12134	12141	8
HIF1A	ENST00000323441	154238266	TCCTGACT	17563	17570	8
HIF1A	ENST00000557538	50195190	TCCTGACT	10019	10026	8
HIF1A	ENST00000557538	119328076	TCCTGACT	15448	15455	8
HIF1A	ENST00000539097	50195190	TCCTGACT	10019	10026	8
HIF1A	ENST00000539097	119328076	TCCTGACT	15448	15455	8
HIF1A	ENST00000337138	73623045	TCCTGACT	12134	12141	8
HIF1A	ENST00000337138	154238266	TCCTGACT	17563	17570	8
TP53	ENST00000514944	21	TCCTGACT	6	13	8
TP53	ENST00000503591	21	TCCTGACT	6	13	8
TP53	ENST00000445888	21	TCCTGACT	6	13	8
TP53	ENST00000420246	21	TCCTGACT	6	13	8
TP53	ENST00000269305	21	TCCTGACT	6	13	8
TP53	ENST00000455263	21	TCCTGACT	6	13	8
TP53	ENST00000604348	21	TCCTGACT	6	13	8
TP53	ENST00000509690	21	TCCTGACT	6	13	8
ULBP2	ENST00000367351	8555316	TCCTGACT	4136	4143	8

We confirmed in P53 that the length 8 offset# associated with the set of highest scoring transcript-wide PiV position shifts will have a SeqStart where its Length(8+n) will also contain the most other length 8 offset subsequences of the set.

The discovery of gene-wide, multi-transcript, shortest length offsets containing the most number of biggest PiV shifts is a novel way to discover relevant intronic subsequences. Further testing of these subsequences in biology will confirm whether their inherent, relative concentration makes them capable of affecting p53 or more specifically binding performance of the p53 monomer as suggested in this paper.

Conclusion

Codondex is a novel multi-transcript, single cell analysis of inherent and relative concentration classifications in DNA. It is assembled by DNA/intron pkmers to identify heritable effects on protein by comparing modifications to the order of transcripts in the PiV, protein vector. It is proposed that concentration change induces different combinations of p53, according to their inherited traits to translocate where they bind highly specific ARE's or RE's, which are subsequently transcribed. Further work is required to identify TP53 ARE's and determine their capacity to modify p53 ligand or binding conformation through specific concentration changes the transcribed TP53 feature induced.

Babu, M. Madan, Robin van der Lee, Natalia Sanchez de Groot, and Jörg Gsponer. 2011. "Intrinsically Disordered Proteins: Regulation and Disease." *Current Opinion in Structural Biology* 21 (3): 432–40.

Baresova, P., J. Musilova, P. M. Pitha, and B. Lubyova. 2013. "p53 Tumor Suppressor Protein Stability and Transcriptional Activity Are Targeted by Kaposi's Sarcoma-Associated Herpesvirus-Encoded Viral Interferon Regulatory Factor 3." *Molecular and Cellular Biology* 34 (3): 386–99.

Bian, J., and Y. Sun. 1997. "p53CP, a Putative p53 Competing Protein That Specifically Binds to the Consensus p53 DNA Binding Sites: A Third Member of the p53 Family?" *Proceedings of the National Academy of Sciences of the United States of America* 94 (26): 14753–58.

Buckley, Peter T., Mugdha Khaladkar, Junhyong Kim, and James Eberwine. 2014. "Cytoplasmic Intron Retention, Function, Splicing, and the Sentinel RNA Hypothesis." *Wiley Interdisciplinary Reviews. RNA* 5 (2): 223–30.

Dornan, David, Harumi Shimizu, Neil D. Perkins, and Ted R. Hupp. 2003. "DNA-Dependent Acetylation of p53 by the Transcription Coactivator p300." *The Journal of Biological Chemistry* 278 (15): 13431–41.

Elbarbary, Reyad A., Bronwyn A. Lucas, and Lynne E. Maquat. 2016. "Retrotransposons as Regulators of Gene Expression." *Science* 351 (6274): aac7247.

Gabunilas, Jason, and Guillaume Chanfreau. 2016. "Splicing-Mediated Autoregulation Modulates Rpl22p Expression in *Saccharomyces Cerevisiae*." *PLoS Genetics* 12 (4): e1005999.

Gu, W., and R. G. Roeder. 1997. "Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain." *Cell* 90 (4): 595–606.

Hattori, Hiroyoshi, Rekin 's Janky, Wilfried Nielfeld, Stein Aerts, M. Madan Babu, and Ashok R. Venkitaraman. 2014. "p53 Shapes Genome-Wide and Cell Type-Specific Changes in microRNA Expression during the Human DNA Damage Response." *Cell Cycle* 13 (16): 2572–86.

Herrero, Ana B., Elizabeth A. Rojas, Irena Misiewicz-Krzeminska, Patryk Krzeminski, and Norma C. Gutiérrez. 2016. "Molecular Mechanisms of p53 Deregulation in Cancer: An Overview in Multiple Myeloma." *International Journal of Molecular Sciences* 17 (12). doi:10.3390/ijms17122003.

Loayza-Puch, Fabricio, Jarno Drost, Koos Rooijers, Rui Lopes, Ran Elkon, and Reuven Agami. 2013. "p53 Induces Transcriptional and Translational Programs to Suppress Cell Proliferation and Growth." *Genome Biology* 14 (4): R32.

Marcel, V., F. Catez, and J-J Diaz. 2015. "p53, a Translational Regulator: Contribution to Its Tumour-Suppressor Activity." *Oncogene* 34 (44): 5513–23.

Marcel, V., V. Vijayakumar, L. Fernández-Cuesta, H. Hafsi, C. Sagne, A. Hautefeuille, M. Olivier, and P. Hainaut. 2010. "p53 Regulates the Transcription of Its Δ133p53 Isoform through Specific Response Elements Contained within the TP53 P2 Internal Promoter." *Oncogene* 29 (18): 2691–2700.

McLure, K. G. 1998. "How p53 Binds DNA as a Tetramer." *The EMBO Journal* 17 (12): 3342–50.

Pontius, B. W., and P. Berg. 1992. "Rapid Assembly and Disassembly of Complementary DNA Strands through an Equilibrium Intermediate State Mediated by A1 hnRNP Protein." *The Journal of Biological Chemistry* 267 (20): 13815–18.

Smith, Jarrett, Deepika Calidas, Helen Schmidt, Tu Lu, Dominique Rasoloson, Geraldine Seydoux, and Julie Ahringer. 2016. "Spatial Patterning of P Granules by RNA-Induced Phase Separation of the Intrinsically-Disordered Protein MEG-3." *eLife* 5 (December). eLife Sciences Publications Limited: e21337.

Tafvizi, Anahita, Fang Huang, Alan R. Fersht, Leonid A. Mirny, and Antoine M. van Oijen. 2011. "A Single-Molecule Characterization of p53 Search on DNA." *Proceedings of the National Academy of Sciences of the United States of America* 108 (2): 563–68.

Taylor, J. Paul, Robert H. Brown Jr, and Don W. Cleveland. 2016. "Decoding ALS: From Genes to Mechanism." *Nature* 539 (7628): 197–206.

Weinberg, Richard L., Dmitry B. Veprintsev, Mark Bycroft, and Alan R. Fersht. 2005. "Comparative Binding of p53 to Its Promoter and DNA Recognition Elements." *Journal of Molecular Biology* 348 (3): 589–96.

Yang, Huanjie, and Gen Sheng Wu. 2004. "p53 Transactivates the Phosphatase MKP1 through Both Intronic and Exonic p53 Responsive Elements." *Cancer Biology & Therapy* 3 (12): 1277–82.

Yu, Isseki, Takaharu Mori, Tadashi Ando, Ryuhei Harada, Jaewoon Jung, Yuji Sugita, and Michael Feig. 2016. "Biomolecular Interactions Modulate Macromolecular Structure and Dynamics in Atomistic Model of a Bacterial Cytoplasm." *eLife* 5 (November). doi:10.7554/eLife.19274.