COMBINATORIAL GENOMIC DATA REFUTE THE HUMAN CHROMOSOME 2 EVOLUTIONARY FUSION AND BUILD A MODEL OF FUNCTIONAL DESIGN FOR INTERSTITIAL TELOMERIC REPEATS

Jeffrey P. Tomkins, Institute for Creation Research, 1806 Royal Ln Dallas, TX 75010 jtomkins@icr.org

ABSTRACT

Evolutionists allege that human chromosome 2 is the product of an ancient fusion event in an ancient hominid ancestor descended from apes. However, both the alleged site of fusion and the so-called cryptic centromere of human chromosome 2 are situated inside active genes negating the idea of fusion. Not only are the alleged genomic fossils of fusion representative of functional intragenic sequence, but they are also both highly degenerate versions of their supposed evolutionary beginnings, suggesting something other than an evolutionary origin. Given that these data strongly refute an evolutionary fusion scenario, it behooves creationists to propose an alternative model for the functional nature of telomere-like sequences scattered around the internal regions of human chromosomes. Towards this end, new data based on ENOCODE project data sets is provided that further elucidates the regulatory role of interstitial telomeric repeat sequences genome-wide, particularly with respect to their transcription factor binding domain properties and transcription start site associations.

KEY WORDS

Chromosome 2 fusion, DDX11L2 gene, ANKRD30BL transmembrane protein, human evolution, human-chimpanzee, cryptic centromere, interstitial telomere sequence, transcription factor binding, transcription start site

INTRODUCTION

One of the most often used arguments explaining human evolution from a chimpanzee-like ancestor is the alleged fusion of ape chromosomes 2A and 2B in a telomere-to-telomere fashion, resulting in human chromosome 2 (Yunis and Prakash 1982; Ijdo et al. 1991). This scenario attempts to account for the discrepancy in chromosome numbers between humans and great apes. Humans have a diploid chromosome complement of 46 while chimpanzees, orangutans, and gorillas have 48. See Figure 1 for a graphical depiction of the purported fusion event.

The idea of human chromosome 2 fusion is strongly promoted despite the fact that all known chromosome fusion events in extant mammals involve satellite DNA and breaks at or near centromeres (Chaves et al. 2003; Tsipouri et al. 2008; Adega et al. 2009). All genetic data in living mammals up to this point shows that telomere-satelliteDNA or satelliteDNA-satelliteDNA are the hallmark signatures of naturally occurring but rare chromosomal fusion sites in nature, not telomere-telomere fusions (Chaves et al. 2003; Tsipouri et al. 2008; Adega et al. 2009). Some evolutionists may counter this data with the argument that telomere-telomere fusions have been observed in the rearranged aberrant genomes of human cancer cells. However, these genomic aberrations are not

indicative of normal healthy cells, but instead are the products of the failure of mechanisms maintaining genomic integrity in cells - leading to disease and death of the organism (Tanaka et al. 2012; Tanaka et al. 2014; Tu et al. 2015).

An end-to-end fusion of chromosomes as proposed by evolutionists for humans would give a head-to-head telomeric repeat signature of at least 10,000 bases in length due to the fact that human telomeres range in size between 5,000 and 15,000 bases in length (Tomkins and Bergman 2011a, b). However, the alleged the fusion site is exceptionally small in size and only 798 bases in length. Another significant aspect questioning the validity of a telomere-telomere fusion signature is the fact that in evolutionary terms, it is very degenerate given the alleged 3 to 6 million years of divergence from a human-chimpanzee common ancestor (Fan et al. 2002). Given that no major rearrangements within the fusion site appear to have occurred combined with the lack of transposable element insertions, the fusion site should be about 98.5% similar to pristine fused repeats based on standard evolutionary predictions. However, the 798-base fusion signature is only 70% identical to the sequence of a hypothetical pristine fusion of the same size based on a pair-wise global alignment in the Geneious software package.

Some evolutionists may also attempt to claim that the fusion site and the alleged cryptic centromere are positioned where one might expect them to be if a fusion occurred. However, an analysis of the assembled chimpanzee DNA sequences for these chromosomes (panTro4) reveals that not only are they assembled using human chromosome 2 as a scaffold, but they contain many gaps and are full of large numbers of meaningless N's (Tomkins 2017). The letter N is substituted for nucleotides in the areas of DNA that contain unknown sequence instead of the letters A, T, G, or C and the number of N's inserted does not correspond to actual gap sizes, which are unknown. At the time of this research, the new panTro5 version of the chimpanzee genome was released and is currently in
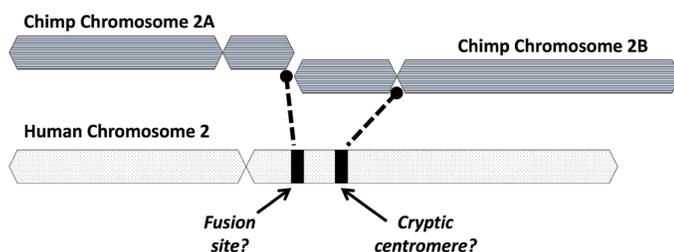

Figure 1. Depiction of the evolutionary model in which chimpanzee-like chromosomes 2A and 2B allegedly fused end-to-end to form human chromosome 2. Chromosomes are drawn to scale according to cytogenetic images in Yunis and Prakash (1982).

the process of being compared by this author to the current hg38 version of the human genome, end-trimmed trace reads, and the previous panTro4 version of chimpanzee. While the new assembly is likely to be greatly improved, it's veracity as an unbiased construction needs to be critically evaluated given the history of human evolutionary bias in previous versions.

While all of this information is important to consider when examining the plausibility of the fusion model, the most compelling data refuting it came when the actual fusion signature was analyzed in 2013 showing that it's DNA sequence when read in the minus strand orientation is a functional transcription factor binding domain inside the first intron of the DDX11L2 noncoding RNA helicase, where it acts as a second promoter (Tomkins 2013; Figure 2). This data was further verified in a follow-up research report which revealed that the alleged fusion site binds to 11 different transcription factors, including RNA polymerase II, the primary enzyme that transcribes genes (Tomkins 2017). See Figure 3 showing ENCODE-related data from the UCSC genome browser. Additional data presented in the Tomkins 2017 paper showed that along with RNA polymerase binding, is the fact that transcription initiates inside the fusion-like sequence in a classic promoter-like expression pattern (Figure 4). As expected, these data implicating promotor activity also intersect with transcriptionally active histone marks and active chromatin profiles that are key features of gene promoters. As a whole, these combinatorial results strongly indicate that the alleged fusion sequence is a gene promoter, not a random accident of chromosomal fusion.

When the products of the DDX11L2 gene were analyzed, it was found that it encoded RNA transcripts expressed in at least 255 different cell and/or tissue types (Tomkins 2013). The gene produces RNAs of two different lengths—short variants (~1,700 bases long) and long variants (~2,200 bases long). The alleged fusion site functions as a promoter for the shorter variants (Figure 2). Annotation of the transcripts revealed that they contained the capacity for complex post-transcriptional regulation through a variety of microRNA binding sites (Tomkins 2013). A number of the microRNA binding sites were shared with DDX11 protein coding gene transcripts (Tomkins 2013). Both the DDX11L2 and DDX11 genes are significantly co-expressed together in the same tissues (Tomkins 2013). Shared microRNA binding sites and co-expression suggest co-regulation between a protein coding gene and its noncoding RNA pseudogene counterpart, as revealed in the well-documented example of the PTEN protein coding gene and it's PTEN pseudogene counterpart (Johnsson et al. 2013).

If two chromosomes actually fused, then there would be two centromeres present and one of them would have to be deactivated to maintain chromosome stability. Centromeres are specific regions of chromosomes that play an important role in the assembly of the kinetochore—a complex structure that plays a key function in the separation of chromosomes during cell division. Evolutionists propose that an inactivated centromere in a post-fusion scenario would degrade over time and become a cryptic genomic fossil, such as that which is alleged to be present on human chromosome 2.

A recent research report was published seemingly bolstering the evidence of a cryptic centromere in human chromosome 2 (Miga 2016). The author argues this point based on gene synteny (gene order) between human and chimpanzee. Of course, the problem with this premise is based on the artificially contrived assembly of chimpanzee chromosomes 2A and B which are bloated with gaps and assembled based on the human genome. Using an argument based on synteny is fallacious because the conclusion is assumed in the premise. Actual synteny between human and chimpanzee remains to be resolved until an unbiased assembly of the chimpanzee genome is produced.

A problem with the alleged cryptic centromere is that its human alphoid repeat DNA sequence does not closely match chimpanzee centromeres and chromosomes (Archidiacono et al. 1995; Haaf and Willard 1997; Tomkins 2017). In addition to the problem of discontinuity with ape sequence, the alleged cryptic centromere is exceptionally small compared to a real centromere. It is only 41,608 bases in length, but this length includes non-centromeric
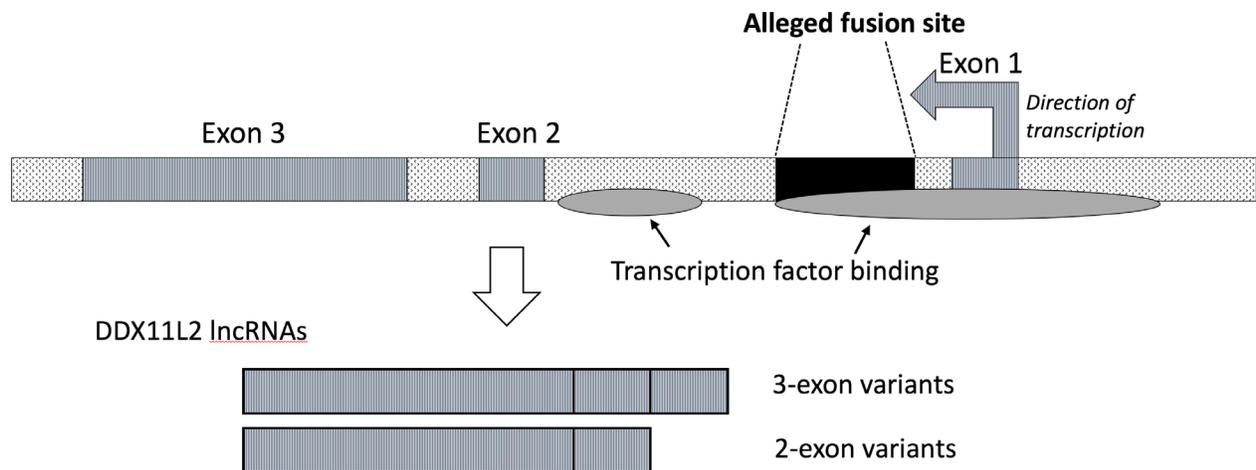

Figure 2. Simplified illustration of the alleged fusion site inside the second intron of the DDX11L2 noncoding RNA gene. The graphic also shows two versions of short and long transcript variants produced along with areas of transcription factor binding. Arrow in first exon depicts direction of transcription.

insertions of two retroelements: a LPA3/LINE repeat (5,957 bases) and a SVA-E element (2,571 bases) (Figure 5). When we subtract the insertions of these non-centromeric elements, it gives a length of only 33,080 bases which is a fraction of the size of normal human centromeres that range in length between 250,000 and 5,000,000 bases (Aldrup-Macdonald and Sullivan 2014).

However, the most serious problem with the concept of a cryptic centromere is that it is entirely situated inside the protein coding gene ANKRD30BL [Ankyrin Repeat Domain 30B Like] (Tomkins 2017). Interestingly, the alleged cryptic centromere sequence extends across intron and exon regions of the gene with part of it

actually coding for amino acids in the resulting protein (Figure 5). This type of ankyrin repeat protein is associated with the plasma membrane in eukaryotic cells and is implicated in the interaction of the cytoskeleton with integral membrane proteins. (Voronin and Kiseleva 2008). The fact that the alleged cryptic centromere is a functional region inside a protein coding gene is at odds with the idea that it is a defunct centromere.

Not only are both the alleged fusion and cryptic centromere sites questionable in their sequence as to being evolutionarily derived from their hypothesized precursors, they both represent functional intragenic sequence. Given this fusion-negating scenario, it
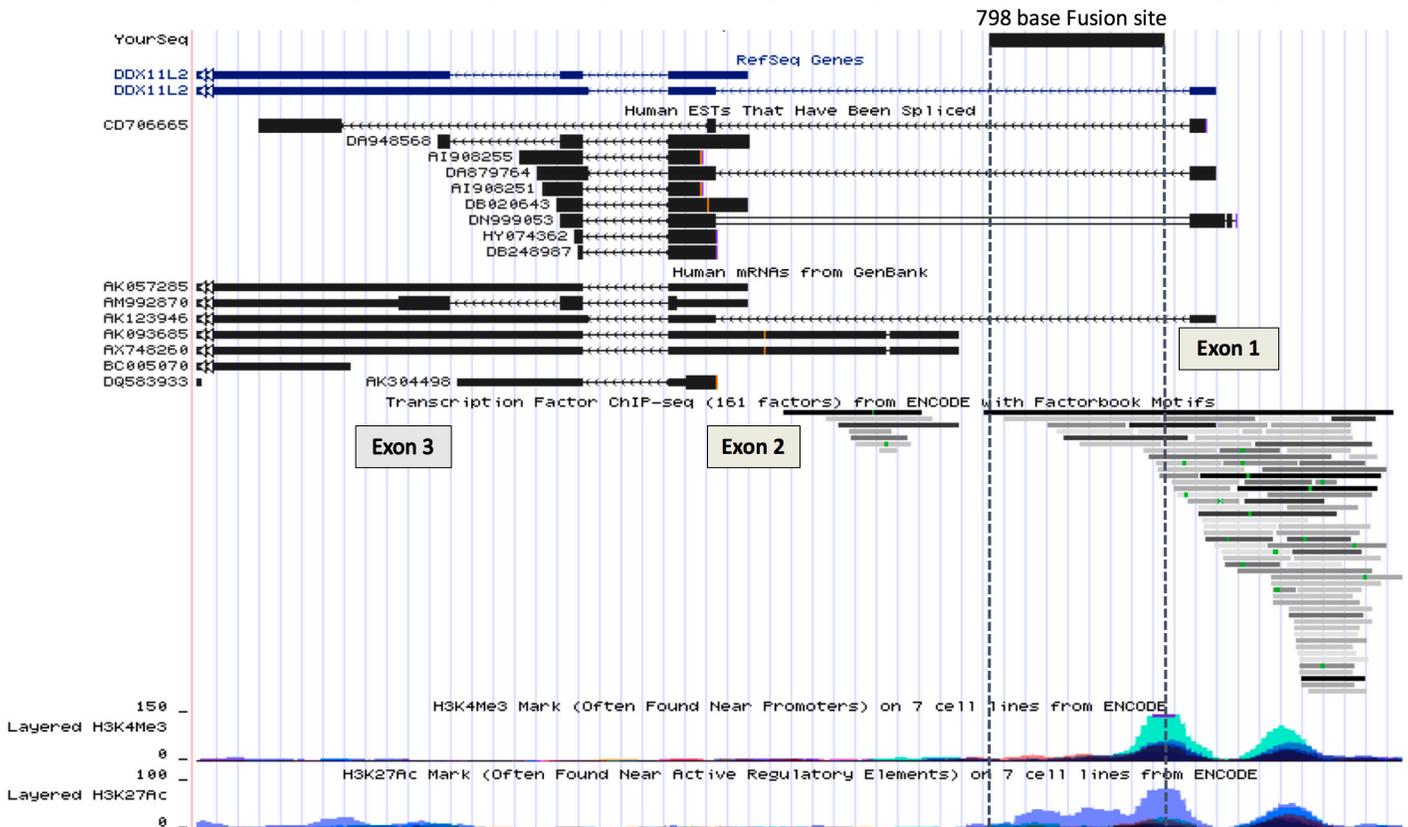

Figure 3. UCSC genome browser (https://genome.ucsc.edu) view of the DDX11L2 gene, its aligned mRNAs, transcription factor binding, and transcriptionally active histone tracks in relation to the alleged fusion site.
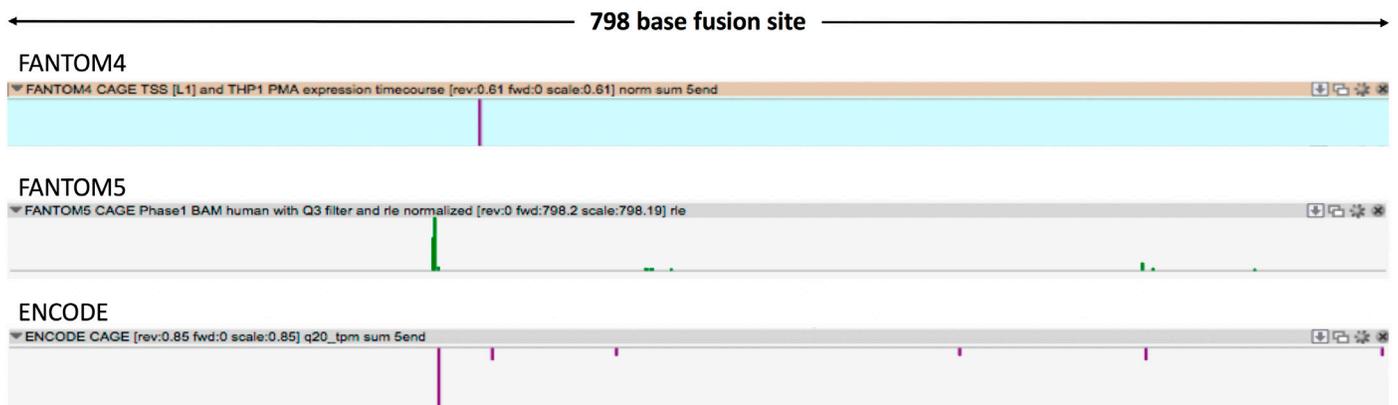

Figure 4. Transcription start site data taken from the FANTOM4, FANTOM5, and ENCODE cap analysis of gene expression (CAGE) databases for the genomic coordinates of the alleged 798 base fusion site (http://fantom.gsc.riken.jp).

behooves both creationist and secular researchers to investigate alternative functions for the types of sequence motifs involved, particularly the alleged fusion site sequence, which is the chief hallmark of the whole fusion-based evolutionary story.

The characterization of interstitial telomeric-like repeats across the human genome has been documented in previous reports using both cytogenetic and bioinformatic techniques (Azzalin et al. 2001; Nergadze et al. 2007; Lin and Yan 2008; Ruiz-Herrera et al. 2008; Bolzan 2017). Based on the evolutionary assumption that the genome is littered with the non-functional remnants of stochastic evolutionary processes, very little research has been done to investigate whether interstitial telomeric-like sites, like those at the alleged fusion site, might actually serve some useful purpose. Most research has focused on the evolutionary origins of interstitial telomeric repeats being accidentally and randomly transferred from chromosome ends along with the possibility that the sites of these alleged events might be associated with chromosome instability and disease (Nergadze et al. 2007; Lin and Yan 2008; Ruiz-Herrera et al. 2008; Bolzan 2017). However, such research has been inconclusive as to the definitive association of these types of sites with human disease (Bolzan 2017). In light of over a decade of inconclusive research related to attempts at associating interstitial telomeric-like repeats with chromosome instability and disease, one evolutionary researcher stated, "future research should focus on the possible biological significance of ITSs by investigating the role of these telomeric-like sequences in the regulation of gene expression" (Bolzan 2017). In light of the research I have uncovered regarding the role of the alleged fusion site in the regulation of gene expression, this statement could not be more timely or profound.

Towards identifying the possibility of interstitial telomeric-like repeats having purpose and function within a creation model, like those found at the alleged fusion site, I have completed a research project which involves identifying both intact and degenerate telomeric-like repeats within the internal regions of human chromosomes and then intersecting their genomic coordinates with a wide variety of ENCODE project data sets. These results will help build a creationist model of designed and engineered functionality for interstitial telomeric repeats.

METHODS

The GRCh38 version of the human genome was downloaded at ftp://ftp.ensembl.org/pub/release-90/fasta/homo_sapiens/dna/. Each of the chromosome FASTA files were manually end-trimmed in a text editor to remove telomeric sequence from both chromosome ends. Perfect interstitial telomeric motifs (TTAGGG and CCCTAA) of two tandem repeats or more were identified and binned by chromosome. The common degenerate telomeric forms of the motif (TTNGGG and CCCNAA) were also identified in tandem repeats of two or more and binned by chromosome. Data was outputted in bed file format containing data columns of chromosome ID, genome coordinates, feature ID (eg. forward telomere, reverse telomere). Based on genomic coordinates, each identified telomeric repeat was then intersected with a variety of ENCODE data sets: gencode v. 22 gene annotation (www.gencodegenes.org), human lincRNAs (long intergenic noncoding RNA) genes from hg38 (https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38), lincRNAs from a publication (Hangauer et al. 2013), robust enhancers, permissive enhancers, ubiquitous enhancers, transcription start site association enhancers (http://enhancer.binf.ku.dk/presets/), remap transcription factor binding (http://tagc.univ-mrs.fr/remap/index.php?page=download), and the trusight inherited disease loci (https://support.illumina.com/downloads/trusight_inherited_disease_product_files.html). ENCODE-related BED files based on previous versions of the human genome (e.g. hg16 and hg18) were converted to GRCh38 coordinates using the Python program CrossMap and the appropriate liftover chain conversion files (http://crossmap.sourceforge.net). Multiple hits to the same ENCODE-related target bed file were reduced to a single unique output hit to remove redundancy produced by multiple overlapping telomeric repeats at a single locus. Python scripts and modules written by author Tomkins for locating and intersecting interstitial telomeric sequence can be viewed and/or downloaded at https://github.com/jt-icr/interstitial_telomeric_repeats.

RESULTS

In this current study, I have queried the entire interstitial space of the latest version of the human genome (hg38) using a Python module I wrote for the identification of perfect and degenerate interstitial telomeric sequence (ITS) sites comprising signatures of 2 repeats or larger. These ITS sites were then intersected using their genome coordinates with a wide variety of ENCODE-related data
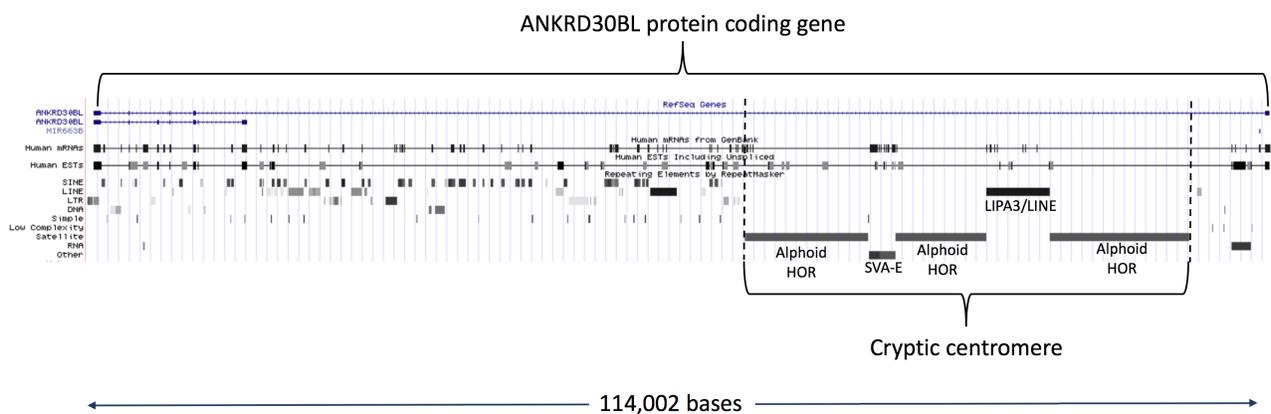

Figure 5. The 41,608 base cryptic centromere region on chromosome 2 that is positioned within the ANKRD30BL protein coding gene.

sets described as follows.

The full Gencode22 data set used in this study contains all comprehensive protein coding gene annotations, all comprehensive lncRNA gene annotations, all polyA features (polyA_signal, polyA_site, pseudo_polyA), 2-way consensus (retrotransposed) pseudogenes, and tRNA genes (Derrien et al. 2012). In total, there are 195,178 genic features and their corresponding coordinates in the BED file used for intersecting ITS sites (Table 1). Approximately 2.7% of these genic features contained at least one ITS site of 2 repeats or more. Over 5,000 ITS sites of two repeats or larger intersected with these various Gencode22 annotations.

Long intergenic noncoding RNAs (lincRNA) are long noncoding RNA genes located in the intergenic protein space in the genome. Like lncRNA genes, lincRNA genes have complex promoters, are alternatively spliced and transcribed and tend to be highly cell-type specific in their expression patterns (Ulitsky and Bartel 2013). They comprise a hotly pursued area of biomedical research due to their association with human health and cellular development (Guttman et al. 2011; Ulitsky et al. 2011; Batista and Chang 2013). Two different lincRNA datasets were queried: the UCSC lincRNAs and a much larger lincRNA dataset from a publication by Hangauer et al. (2013) which produced 730 and 300 ITS site intersections, respectively (Table 1).

Enhancer regions in the genome regulate the proper temporal and cell type specific activation of gene expression in higher eukaryotes (Dickel et al. 2013; Andersson et al. 2014). Both transcription factor binding and transcription start sites are hallmarks of enhancers. Two data sets of enhancers were used in this study. Robust enhancers are transcribed at a significant expression level in at least one primary cell or tissue sample while all known transcribed enhancers comprise the permissive set, producing numbers of ITS intersections of 63 and 64, respectively (Table 1).

Transcription start site associations (TSS) are defined as TSSs that correlate with transcriptional, epigenetic, and transcription factor binding within 500 kb of the TSS (Andersson et al. 2014). The goal of such research is to link enhancers to their target genes. Therefore, a dataset of 64,621 enhancer TSS associations was queried with the ITS sites in which 5,002 intersections were found (Table 1). A surprisingly large 8% of these TSS associated regions intersected with ITS sites.

Transcription factor binding sites are determined via the biochemical association (binding) of transcription factors to genomic DNA sequence (Furey 2012; Mundade et al. 2014). A comprehensive data set of transcription factor binding sites comprising 8.8 million genomic locations across the human genome (Griffon et al. 2015) was queried with ITS site resulting in 4,489 intersections.

Given that much of the evolutionary speculation surrounding the implications of ITS sites as being chromosomal aberrations and playing a role in chromosome breakage and human disease, I also decided to determine if they could be associated with known heritable disease. Therefore, a dataset of 8,801 inherited disease loci developed by the Illumina Corporation and used in the screening of human disease (Kingsmore 2012; Saunders et al. 2012) was queried with the ITS sites. The database contains 550 genes, including coding exons, intron-exon boundaries, and regions harboring pathogenic mutations. Only 5 ITS sites could be intersected with disease related loci, and they were all degenerate ITS of 12 bases in length. This does not implicate them as being a part of the pathology of the locus, but that they were found in these genomic segments. Interestingly, all five were located in exons of protein coding genes. One ITS site was located in the last exon of the peroxisomal biogenesis factor 10 (PEX10) gene on chromosome 1. A second was located in the last exon of the alkylglycerone phosphate synthase (AGPS) gene on chromosome 2. A third was located in the last exon of the desmoplakin (DSP) gene on chromosome 6. A fourth was located in the last exon of the tripeptidyl peptidase 1 (TPP1) gene on chromosome 11. The fifth

Table 1. Results from the intersection of genome-wide ITS sites two repeats or larger with various ENCODE-related data sets.

Data set	Number data set entries	Number perfect TTAGGG intersections	Number Degenerate TTNGGG intersections	Number perfect CCCTAA intersections	Number Degenerate CCCNAA intersections	Total ITS intersections
Gencode22	195,178	258	2,347	249	2,343	5,197
UCSC lincRNAs	21,131	59	299	85	287	730
lincRNAs from Hangauer et al. (2013)	59,177	6	138	26	130	300
Permissive enhancers	42,888	3	24	3	34	64
Robust enhancers	38,443	3	23	3	34	63
Enhancer transcription start site associated regions	64,621	204	2,242	258	2,298	5,002
Remap transcription factor binding	8,822,477	498	1,740	445	1,806	4,489
Tru-sight inherited disease	8,801	0	4	0	1	5

was located in a central exon of the huge 25-exon NPC Intracellular Cholesterol Transporter 1 (NPC1) gene on chromosome 18.

DISCUSSION

Since my original 2013 publication evaluating the genomic evidence for fusion, the data refuting the chromosome 2 fusion hypothesis have become even more compelling. The alleged fusion site is clearly a functional second promoter in the second intron of the DDX11L2 gene (Tomkins 2013) and the alleged cryptic centromere site is a key sequence within the protein coding gene ANKRD30BL covering both intronic and exonic sequence (Tomkins 2017).

While the sequence of the alleged cryptic centromere is a unique coding and noncoding part of a large protein coding gene, the alleged fusion site represents a regulatory sequence that could have implications for the genomic mystery of ITS sites genome-wide. This question also has key importance for building the creation model of purpose and design within the human genome.

The presence and preponderance of interstitial telomeric-like repeats in the human genome has been well established by a variety of cytogenetic and bioinformatics techniques (Azzalin et al. 2001; Nergadze et al. 2007; Lin and Yan 2008; Ruiz-Herrera et al. 2008; Bolzan 2017). However, little is known of their possible function and most research has primarily focused on their evolutionary origins and the possibility that they might be associated with chromosome instability and disease. Because of these past research efforts based on the evolutionary presupposition that the genome is littered with the remnants of purposeless random evolution, no research has been done to investigate whether ITS sites might actually serve some functional purpose. Interestingly, a recent review by Bolzan (2017) from a naturalistic perspective on ITS sites stated, "future research should focus on the possible biological significance of ITSs by investigating the role of these telomeric-like sequences in the regulation of gene expression".

Thus, the purpose of this paper was not to prove that ITS sites are abundant in the human genome, a fact already well established, but to ascertain what their role might be based on the premise that the genome was created with function and purpose, although subject to degeneration over time, a process commonly referred to as genetic entropy (Sanford 2010).

The most important finding of this current research was the confirmation of the possible role of ITS sites in gene regulation as depicted by the high numbers of ITS sites associated with enhancer transcription start site associated regions and transcription factor binding. In addition, the same approximate number of intersections (~5,000) were also associated with Gencode22 annotations. These data closely follow the results from the comprehensive analysis of the alleged fusion site I conducted in two previous reports (Tomkins 2013; Tomkins 2017) and strongly suggest that many ITS sites are involved in gene regulation, especially in regard to their role in transcription factor binding and transcription start site activity.

Another interesting minor aspect to this study was the investigation of the five intersections of ITS sites with inherited disease loci. Amazingly, all five ITS were located in the exons of protein coding genes. Many exons not only code for proteins, but have also been found to contain gene regulatory sequence in addition to other imbedded codes that regulate both transcription and translation (Tomkins 2015).

CONCLUSION

As a key component of the evolution paradigm, scientists have argued that human chromosome 2 is the product of an ancient fusion event in a hominid ancestor following the divergence of humans and apes based on the alleged genomic remnants of a fusion event. These telomeric-like genomic signatures are thought to represent the actual fusion site and a cryptic centromere. Recent genomic data, however, indicates that both the alleged site of fusion and the cryptic centromere of human chromosome 2 are positioned inside functional genes. Furthermore, the alleged site of fusion has been proven to be a functional promoter in the second intron of a noncoding RNA gene. Given that these data strongly refute evolutionary claims, this current study was undertaken to help develop an alternative creationist model of intelligent design based on the idea that these features are functional characteristics of unique engineering by the Creator. By comprehensively intersecting interstitial telomeric repeats genome-wide with ENCODE-related data sets, this study helps elucidate the regulatory role of interstitial telomeric repeat sequences, particularly with respect to their transcription factor binding domain properties and transcription start site associations.

REFERENCES

Adega, F., H. Guedes-Pinto, and R. Chaves. 2009. Satellite DNA in the karyotype evolution of domestic animals--clinical considerations. *Cytogenet Genome Res* 126 (1-2):12-20.

Aldrup-Macdonald, M. E., and B. A. Sullivan. 2014. The past, present, and future of human centromere genomics. *Genes (Basel)* 5, no.1:33-50.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507, no. 7493:455-61.

Archidiacono, N., R. Antonacci, R. Marzella, P. Finelli, et al. 1995. Comparative mapping of human alphoid sequences in great apes using fluorescence in situ hybridization. *Genomics* 25, no. 2:477-84.

Azzalin, C. M., S. G. Nergadze, and E. Giulotto. 2001. Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma* 110, no. 2:75-82.

Batista, P. J., and H. Y. Chang. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152, no. 6:1298-307.

Bolzan, A. D. 2017. Interstitial telomeric sequences in vertebrate chromosomes: Origin, function, instability and evolution. *Mutat Res* 773:51-65.

Chaves, R., F. Adega, J. Wienberg, H. Guedes-Pinto, et al. 2003. Molecular cytogenetic analysis and centromeric satellite organization of a novel 8;11 translocation in sheep: a possible intermediate in biarmed chromosome evolution. *Mamm Genome* 14, no. 10:706-10.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22, no. 9:1775-89.

Dickel, D. E., A. Visel, and L. A. Pennacchio. 2013. Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci* 368, no. 1620:201

Fan, Y., E. Linardopoulou, C. Friedman, E. Williams, et al. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome*

Res 12, no. 11:1651-62.

Furey, T. S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* 13, no.12:840-52.

Griffon, A., Q. Barbier, J. Dalino, J. van Helden, et al. 2015. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res* 43, no. 4:e27.

Guttman, M., J. Donaghey, B. W. Carey, M. Garber, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, no.7364:295-300.

Haaf, T., and H. F. Willard. 1997. Chromosome-specific alpha-satellite DNA from the centromere of chimpanzee chromosome 4. *Chromosoma* 106, no. 4:226-32.

Hangauer, M.J., I.W. Vaughn, and M.T. McManus. 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* 9 no.6:e1003569.

Ijdo, J.W., A. Baldini, D. C. Ward, S. T. Reeders, et al. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci U S A* 88, 20:9051-5.

Johnsson, P., A. Ackley, L. Vidarsdottir, W. O. Lui, et al. 2013. A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol* 20 no.4:440-6.

Kingsmore, S. 2012. Comprehensive carrier screening and molecular diagnostic testing for recessive childhood diseases. *PLoS Curr*:e4f9877ab8ffa9. doi: 10.1371/4f9877ab8ffa9.

Lin, K. W., and J. Yan. 2008. Endings in the middle: current knowledge of interstitial telomeric sequences. *Mutat Res* 658, no.1-2:95-110.

Miga, K. H. 2016. Chromosome-Specific Centromere Sequences Provide an Estimate of the Ancestral Chromosome 2 Fusion Event in Hominin Genomes. *J Hered.* 108, no.1:45-52

Mundade, R., H. G. Ozer, H. Wei, L. Prabhu, et al. 2014. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* 13, no.18:2847-52.

Nergadze, S. G., M. A. Santagostino, A. Salzano, C. Mondello, et al. 2007. Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol* 8, no.12:R260.

Ruiz-Herrera, A., S. G. Nergadze, M. Santagostino, and E. Giulotto. 2008. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res* 122, no.4:219-28.

Sanford, J. 2010. *Genetic Entropy and the Mystery of the Genome.* 3rd ed. Waterloo, NY: FMS Publications.

Saunders, C. J., N. A. Miller, S. E. Soden, D. L. Dinwiddie, et al. 2012. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med* 4, no. 154:135.

Tanaka, H., S. Abe, N. Huda, L. Tu, et al. 2012. Telomere fusions in early human breast carcinoma. *Proc Natl Acad Sci USA* 109 no. 35:14098-103.

Tanaka, H., M. J. Beam, and K. Caruana. 2014. The presence of telomere fusion in sporadic colon cancer independently of disease stage, TP53/KRAS mutation status, mean telomere length, and telomerase activity. *Neoplasia* 16 no.10:814-23.

Tomkins, J. 2013. Alleged human chromosome 2 "Fusion Site" encodes an active DNA binding domain inside a complex and highly expressed gene—negating fusion. *Answers Research Journal* 6, no. 2013:367-375.

Tomkins, J. 2015. Extreme Information: Biocomplexity of Interlocking Genome Languages. *Creation Research Society Quarterly* 51:187-201.

Tomkins, J., and J. Bergman. 2011a. The chromosome 2 fusion model of human evolution—part 2: re-analysis of the genomic data. *Journal of Creation* 25, no. 2:111-117.

Tomkins, J., and J. Bergman. 2011b. Telomeres: implications for aging and evidence for intelligent design. *Journal of Creation* 25, no. 1:86-97.

Tomkins, J.P. 2017. Debunking the Debunkers: A Response to Criticism and Obfuscation Regarding Refutation of the Human Chromosome 2 Fusion. *Answers Research Journal* 10 :45-54.

Tsipouri, V., M. G. Schueler, S. Hu, Nisc Comparative Sequencing Program, et al. 2008. Comparative sequence analyses reveal sites of ancestral chromosomal fusions in the Indian muntjac genome. *Genome Biol* 9, no. 10:R155.

Tu, L., N. Huda, B. R. Grimes, R. B. Slee, et al. 2015. Widespread telomere instability in prostatic lesions. *Mol Carcinog.* 55, no. 5:842

Ulitsky, I., and D. P. Bartel. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* 154 no. 1:26-46.

Ulitsky, I., A. Shkumatava, C. H. Jan, H. Sive, et al. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147 no. 7:1537-50.

Voronin, D. A., and E. V. Kiseleva. 2008. Functional role of proteins containing ankyrin repeats. *Cell and Tissue Biology* 49, no. 12:989-99.

Yunis, J. J., and O. Prakash. 1982. The origin of man: a chromosomal pictorial legacy. *Science* 215 no.4539:1525-30.

THE AUTHOR

Jeffrey Tomkins is the Director of Life Sciences at the Institute for Creation Research (ICR). He has a Ph.D. in Genetics from Clemson University, an M.S. in Plant Science from the University of Idaho, and a B.S. in Agriculture Ed. from Washington State University. He was a faculty member in Genetics and Biochemistry at Clemson University for a decade, publishing 57 secular research papers and seven book chapters. Dr. Tomkins specializes in genomics and has published 27 peer-reviewed creation science journal papers, two books, and a wide variety of semi-technical articles in the ICR magazine *Acts & Facts*.