HHS Public Access

Author manuscript

Psychiatr Genet. Author manuscript; available in PMC 2016 May 04.

Published in final edited form as:

Psychiatr Genet. 2009 April ; 19(2): 64–71. doi:10.1097/YPG.0b013e3283207ff6.

Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations

Heather A. Bruce[1], **Nancy A. Sachs**[1], **Dobrila D. Rudnicki**[1], **Stephanie G. Lin**[1], **Virginia L. Willour**[1], **John K. Cowell**[2], **Jeffrey Conroy**[2], **Devin E. McQuaid**[2], **Michael Rossi**[2], **Daniel P Gaile**[2], **Norma J. Nowak**[2], **Susan E. Holmes**[1], **Pamela Sklar**[7], **Christopher A. Ross**[1,4,5,6], **Lynn E. DeLisi**[3], and **Russell L. Margolis**[1,4,6]

[1]The Johns Hopkins Univ. School of Medicine, Dept. of Psychiatry, Baltimore, MD

[2]Roswell Park Cancer Institute Dept. of Cancer Genetics, Buffalo, NY

[3]New York University Dept of Psychiatry, New York, NY

[4]The Johns Hopkins Univ. School of Medicine, Neurology, Baltimore, MD

[5]The Johns Hopkins Univ. School of Medicine, Neuroscience, Baltimore, MD

[6]Program in Cellular and Molecular Medicine, Baltimore, MD

[7]Psychiatric Disease Initiative, Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA.

Abstract

Objectives—Genomic copy number variations (CNVs) are a major form of variation in the human genome and play an etiologic role in several neuropsychiatric diseases. Tandem repeats, particularly with long (> 50bp) repeat units, are a relatively common yet underexplored type of CNV that may significantly contribute to human genomic variation and disease risk. We therefore performed a pilot experiment to explore the potential role of long tandem repeats as risk factors in psychiatric disorders.

Methods—A bacterial artificial chromosome (BAC)-based array comparative genomic hybridization (aCGH) platform was used to examine CNVs in genomic DNA from 34 probands with schizophrenia or schizoaffective disorder.

Results—The aCGH screen detected an apparent deletion on 5p15.1 in two probands, caused by the presence in each proband of two low copy number (short) alleles of a tandem repeat that ranges in length from < 10 to > 50 3.4 kb units in the population examined. Short alleles partially segregate with schizophrenia in a small number of families, though linkage was not significant. An association study showed no significant difference in repeat length between 406 schizophrenia cases and 392 controls.

Corresponding Author: Russell L. Margolis, M.D. CMSC 8-121, 600 N. Wolfe Street, Baltimore, MD 21287 rmargoli@jhmi.edu, telephone: 410-614-4262 fax: 410-614-0013..

Conclusion—Though we did not demonstrate a relationship between the 5p15.1 repeat and schizophrenia, our results illustrate that long tandem repeats represent an intriguing type of genetic variation that have not been previously studied in connection with psychiatric illness. aCGH can detect a small subset of these repeats, but systematic investigation will require the development of specific arrays and improved analytic methods.

Keywords

schizophrenia; megasatellite; copy number variant; repeat; array comparative genomic hybridization; psychosis; polymorphism; deletion; duplication

INTRODUCTION

CNVs have emerged as an important source of genetic variation (Iafrate et al. 2004; Sebat *et al.,* 2007). CNV typically refers to a segment of DNA > 1000 bp in length that varies among genomes as a result of duplication or deletion (Scherer *et al.,* 2007). CNVs have been implicated in both neurodegenerative and neurodevelopmental disorders, including familial Parkinson's disease (Singleton *et al.,* 2003; Farrer *et al.,* 2004)), autism (Sebat *et al.,* 2007; Weiss *et al.,* 2008), and syndromic mental retardation (Sharp *et al.,* 2006). The well-established association between 22q11 deletion and schizophrenia (Pulver *et al.,* 1994; Bassett *et al.,* 1998; Murphy *et al.,* 1999), and the association between CNVs and other disorders of the central nervous system, suggest that additional CNVs may have an etiological role in schizophrenia.

Compelling evidence associating schizophrenia with rare recurrent CNVs has recently emerged. Two small studies (Moon *et al.,* 2006; Wilson *et al.,* 2006) initially yielded results of uncertain significance (Sutrala *et al.,* 2007). Analysis of a set of 891 candidate genes examined by oligonucleotide array did not detect CNVs associated with schizophrenia (Sutrala *et al.,* 2008). However, a genome wide examination of 51 sporadic and 42 familial cases of schizophrenia, using a high density BAC array, yielded 13 CNVs of potential interest, including a 1.4 Mb duplication on chromosome 15q13.1 in a sporadic case absent in 372 controls, and a .25 Mb deletion on 2p16.3 in an affected sibpair and their unaffected mother (Kirov *et al.,* 2008). An oligonucleotide microarray analysis found substantially higher rates of novel copy number variations in 150 people with schizophrenia compared to 268 matched controls; the rate of CNVs was particularly high in younger-onset cases (Walsh *et al.,* 2008). A 15% higher rate of CNVs was also detected in cases (N = 3391) compared to controls (N = 3181) in a large-scale microarray analysis by the International Schizophrenia Consortium; in addition to the 22q11 region, deletions on 15q13.2 and 1q21.1 were strongly associated with schizophrenia (The International Schizophrenia Consortium *et al.,* 2008). A multistage analysis of a large subset of the the Icelandic population by deCODE genetics, using a different oligonucleotide platform, found strong evidence for rare CNVs associating with schizophrenia at 1q21.1, 15q11.2 and 15q13.3 (Stefansson *et al.,* 2008).

With current copy number detection technology, an increase or decrease in signal is typically interpreted as a simple deletion or duplication of an entire region with uncertain endpoints. However, the human genome includes polymorphic tandem repeats of various lengths, with

several thousand containing repeat units of greater than 50 bp (Benson 1999), so that what appears to be a deletion or duplication may actually reflect change in the number of repetitive units of a repeat. This is of potential importance in relating CNVs to human disease. Repeat length variations below the threshold for detection by currently available arrays, including expansions of tri-(Orr and Zoghbi 2007), tetra-(Liquori *et al.,* 2001), and pentanucleotide (Matsuura *et al.,* 2000) repeats, contraction of a dodecamer repeat (Lalioti *et al.,* 1997), and copy number variation of a 24mer repeat (Mead 2006) directly cause a variety of neurodegenerative diseases. However, at least one disease, facioscapulohumeral muscular dystrophy (FSHD) (van der Maarel *et al.,* 2007), is caused by a contraction of a repeat (termed D4Z4) consisting of 3.3 kb units. We therefore hypothesize that variations in the length of long tandem repeats may contribute to the genetic risk of psychiatric disorders, including schizophrenia. To preliminarily investigate this hypothesis, we characterized the structure of a tandem repeat initially detected as a deletion in two individuals with schizophrenia by aCGH screen, and determined the extent of length polymorphism in normal and patient populations.

METHODS

Human subjects

DNA for aCGH was obtained from lymphoblastoid cell lines of 28 unrelated probands with schizophrenia and 6 unrelated probands with schizoaffective disorder who were ascertained as part of a large sib-pair study of schizophrenia in US pedigrees (DeLisi *et al.,* 2002). The sample size was chosen to provide a >80% likelihood of detecting a CNV present in 5% of cases of familial schizophrenia. All individuals gave written informed consent for participation under The State University of New York at Stony Brook IRB supervision. In each case, consensus DSMIII-R diagnoses were made by at least two research psychiatrists using information derived from the Diagnostic Interview for Genetic Studies (DIGS), version 3.0 (Nurnberger et al. 1994). Diagnoses were unchanged when reviewed using DSM-IV criteria; schizoaffective probands were all of the depressed subtype. Each proband had a sibling with either schizophrenia (26) or schizoaffective disorder (8). Additionally, 28 of the probands had at least one other first-degree family member with schizophrenia, schizoaffective disorder, bipolar disorder, schizotypal personality disorder or psychotic disorder not otherwise specified, while the remaining six had an affected second-degree relative. Ethnicity, as determined by grandparents, was predominantly European (27 probands, including the two with evidence of CNV at 15p15.1), with one African proband, four probands with at least one grandparent claiming at least partial Native American, Caribbean, or Indian background, and one proband of unknown ethnicity.

Case and control DNA samples for the association analysis were obtained from the NIMH Genetics Initiative (http://nimhgenetics.org/). Schizophrenia pedigrees were ascertained by three extramural sites (Columbia University, Harvard University, and Washington University). DSM-III-R and DSM-IV diagnoses were made using information obtained from a DIGS interview, external informants, and medical records. Controls, selected to represent the U.S. population, were ascertained through a contract with Knowledge Networks (San Jose). Each subject was administered a validated self-assessment inventory of lifetime major

depression, anxiety disorders, substance use disorders, bipolar disorder, or psychotic disorders. We assayed a subset of the collection consisting of 406 probands with schizophrenia and 392 controls. Controls were limited to individuals 51 years or older, to minimize inclusion of individuals still at risk of developing schizophrenia, and to individuals who did not answer yes to screening questions asking if they had been diagnosed, or received treatment for, schizophrenia, schizoaffective disorder, auditory hallucinations, or delusions. Controls were matched to cases on race and sex; both samples were 60% male and 31 % African American. The remainder were predominantly European American, with less than 1% of predominately Hispanic or Native American ethnicity. A full breakdown of ethnicity, based on grandparent origin, is available from the NIH Genetics Initiative. An additional subsample of individuals with schizophrenia (N = 34) and controls (N = 43) ascertained in the Azores were examined (Pato *et al.,* 2004). Only DNA samples that had not been subjected to whole genome amplification were used in this analysis.

Array Comparative Genomic Hybridization

aCGH was performed using the human 6K RPCI-11 BAC array developed at Roswell Park Cancer Institute (Cowell *et al.,* 2004). In brief, PCR representations of 6,116 BACs spanning the human genome at 0.5-1 MB intervals were arrayed in triplicate. Experimental genomic DNA was labeled with Cy3 and co-hybridized with pooled Cy5-labelled DNA from 10 cytogenetically normal individuals (sex-matched to the experimental case). Signal intensity was normalized such that any element of the array with a signal-to-background ratio of less than 6 was excluded, and log2 ratios of case to control were obtained for each BAC representation.

qPCR

Each genomic DNA sample was assayed for copy number in triplicate by quantitative PCR using standard methods on an ABI 7900HT. ABI Assays by Design™ were used to generate the custom primers and probes. Results were normalized to RNaseP.

RT-PCR

Analysis was performed with cDNA obtained from whole brains of three healthy individuals, frontal cortex of two individuals with HDL2, and hippocampus of one healthy individual. Primers were designed with Primer3 (http://primer3.sourceforge.net/). PCR products were subjected to electrophoresis through a 1.5% agarose gel and stained with ethidium bromide.

Pulsed-Field Gel Electrophoresis (PFGE)

Genomic DNA was isolated from lymphoblastoid cells (CHEF mammalian DNA plug kit, Bio-Rad). A single plug was digested overnight with HindIII and half of the digested gDNA plug was separated on a 1% agarose gel (CHEF system, Biorad). The gel was blotted and hybridized at 50°C (ExpressHyb, Clontech) with a PCR-generated, [α-32P]dCTP-labelled 1024 bp probe, specific for the 5p15.1 repeat.

Genetic Analysis

Parametric and non-parametric linkage analyses was performed using GENEHUNTER v2.1_r5 beta (Kruglyak *et al.,* 1996; Nyholt 2002).

RESULTS

34 patients with schizophrenia or schizoaffective disorder from multiplex families were screened for CNVs using the Roswell Park 6K BAC array. A threshold of + or − 0.5 log2 ratio of patient to control signal was used to select possible copy number variants. The strongest array signal was on chromosome 5p15.1 in proband S56 and indicated a deletion. This was also the most robust signal as it was generated by two overlapping BACs (RP11-88l18 and RP11-91d21) with log2 ratios of – 1.0 and −0.8 respectively (Fig. 1). Proband S200 also had diminished log2 ratios (−0.5 and −0.6 respectively) at this locus. A bioinformatic investigation of the locus defined by the 2 BACs in the human reference sequence (hg 18 build 36) revealed a tandem repeat overlapping the BACs (Kent *et al.,* 2002). We therefore chose to investigate this locus in more detail.

The basic repeat unit is 3434 bp in length. In the reference sequence, there are 19 units with 99-100% identity in tandem (16 oriented in the positive direction and 3 in the negative direction) (Fig. 2A). There are 7 units, of lesser identity and with greater separation between units, located in the telomeric and centromeric flanking regions surrounding the central 19 units. A BLAST search indicates that the repeat unit does not exist elsewhere in the human reference sequence.

A section of 405 bp within each 3434 bp repeat unit is highly conserved across species (Fig. 2B). Genscan (Burge and Karlin 1997) predicts a gene of 552 bp that includes the highly conserved region of each repeat unit (Fig. 2C). The predicted product is a protein of 183 amino acids. A stretch of 35 amino acids at the C-terminus of the predicted protein shows 76% identity with a conserved region at the C-terminus of TAF11, a TFIID subunit which binds to TFIIA and is involved in RNA polymerase II transcription initiation (Robinson *et al.,* 2005). No human ESTs have been mapped to the conserved region. However, RT-PCR indicated transcription of the conserved region in human brain; the central 249 bp (Fig. 2D) and the encompassing 438 bp were both amplified. Two unspliced human ESTs map within the repeat unit but outside of the conserved region. U83515 is 205 bp long, 99% identical to the genomic sequence, and from unspecified human tissue. AV730429 is 386 bp long, 99% identical to the genomic sequence, and from hypothalamus. The nearest known gene to the tandem repeat (220 kb telomeric) is *Brain Acid-Soluble protein 1* (*BASP*1); the protein product is involved in axonal growth and guidance.

The presence of a tandem repeat led us to suspect that the diminished log2 ratio detected at the 5p15.1 locus in subjects S56 and S200 was caused by a variation in the number of units of the tandem repeat as opposed to a simple deletion of the region. To examine this possibility, we designed three qPCR reactions (Fig. 2A), one in the center of the tandem repeat and one in either flanking region. qPCR in the center of the tandem repeat unit demonstrated a reduced signal in S56 (normalized quantity mean of 0.2) and S200 (normalized quantity mean of 0.4) consistent with a decrease in copy number, whereas

qPCR at telomeric and centromeric regions flanking the tandem repeat demonstrated no difference in copy number.

To further characterize this region we took advantage of the HindIII sites that flank the repeat and used PFGE to determine the length of each allele. Proband S56 had one allele of approximately 24 kb and one of 36 kb, corresponding to approximately 11 and 14 repeat units respectively. PFGE of 11 other samples showed restriction fragment lengths ranging from 105 kb to 190 kb consistent with tandem repeat lengths ranging from 34 to 59 units (Fig. 3). Overall, including four unrelated individuals shown in Fig. 4, 12 of 15 unrelated individuals were heterozygotes.

qPCR values at the 5p15.1 repeat locus were determined in the original 34 probands examined by BAC array, and an additional 46 probands with schizophrenia or schizoaffective disorder, each with at least one affected first degree relative with schizophrenia or schizoaffective disorder, from the same sib-pair study. Five probands had qPCR values of ≤ 0.5. Lymphoblastoid cell lines were available from family members of probands S56, S1549, S200, and S785 (qPCR values were 0.19, 0.32, 0.39 and 0.47, respectively); all available cell lines from family members were used for PFGE to determine the stability of repeat length in vertical transmission and the extent to which short repeats segregate with disease. Fig. 4A demonstrates segregation of short alleles with disease in the family of proband S56. Proband S1549 had an affected sib; both inherited a short allele (Fig. 4B). In the family of proband S200 (Fig. 4C), short alleles segregate with disease in the brother and first cousin of the proband. In the family of proband S785 (Fig. 4D), all affected individuals have at least one short allele, but the unaffected sister (S726) of the proband also carries a short allele. No evidence of repeat length change was detected during vertical transmission in any family, though small changes in length would potentially fall below the threshold of detection. Overall, a short repeat (defined as 21 or fewer repeat units, as estimated by band size on PFGE) was carried by 12 individuals with schizophrenia, schizoaffective disorder, or psychosis not otherwise specified, and by one individual unaffected at the time of examination. Parametric and nonparametric analyses were performed using GENEHUNTER. For parametric analysis, assigned variables were phenocopy rate = .01, penetrance = .5, disease allele frequency and risk allele = .01, and dominant transmission. Affected status was defined as a diagnosis of schizophrenia, schizoaffective disorder, or psychosis NOS. These parameters yielded a LOD score of 0.686 with all families contributing. Non-parametric analysis yielded an NPL of 1.161 ($p = .126$), with information content = .327.

We next performed a case-control study to investigate the potential association of the 5p15.1 repeat with schizophrenia in a large and heterogeneous population. To validate this approach and establish that qPCR measures the combined allele length, we performed qPCR and PFGE on the same individuals. Fig. 5 demonstrates that the qPCR value is highly correlated with the total length of the two alleles. Like BAC arrays, qPCR does not provide allelic information. Our analysis therefore assumes that disease association with a single short allele would be reflected as shift in mean qPCR across the affected population, or that association with disease is based on the combined length of the two alleles.

We tested 406 patients and 392 controls, all ascertained in the United States and selected from the NIMH Genetics Initiative repository. The mean qPCR signal showed no significant difference between cases and controls, 0.82 and 0.85 respectively, p=0.1 by two tailed unpaired t-test. The result did not change when individuals with European or African ethnicity were analyzed separately. Distribution of qPCR values among cases and controls is shown in Fig. 6. The difference between cases and controls remained insignificant when qPCR values were distributed into 0.1 unit bins, with individuals with qPCR values > 1.3 or < 0.4 consolidated (df =10, x^2 = 15.17, p = 0.13). Similarly, no significant difference was detected between patients and controls (N = 34 and 42 respectively) in a subpopulation of a sample collected in the Azores.

DISCUSSION

We determined that a diminished signal at chromosome 5p15.1 detected in two probands with schizophrenia by aCGH represented short alleles of a highly polymorphic tandem repeat with a unit length of 3.4 kb. Each repeat unit contains a central highly conserved region that is transcribed in human brain. We found partial segregation of short alleles with schizophrenia in a small sample of selected families but no significant evidence linkage. A case-control analysis did not detect evidence of shorter combined allele length in individuals with schizophrenia compared to controls.

The most important result in this pilot study is demonstration that signal changes detected by aCGH and other platforms may reflect changes in repeat length rather than simple deletions or duplications. Proper interpretation of this type of signal will require in depth bioinformatic and experimental characterization of the region in question. Some, but not all, long tandem repeats are likely to be detected by current commercial platforms used for CNV detection (i.e., Illumina Human 1M Beadchip™, www.illumina.com). BAC based copy number arrays, designed to detect large duplications or deletions, typically do not have the resolution to directly detect a change in tandem repeat length, though long and highly variables repeats such as that described here at 5p15.1 are exceptions. Previous SNP based platforms, optimized for allelic discrimination, avoid repetitive areas; for instance, the Affymetrix SNP 6.0 Structural Variation™ and Illumina Human Hap 650v3™ do not include probes for the 5p15.1, D4Z4 or rs447 (described below) repeats. Previous annotations in the Database of Genomic Variants (http://projects.tcag.ca/variation/) (Iafrate *et al.,* 2004)) have described changes at the 5p15.1 locus as large gains or losses rather than as repeat length variations.

Only two repeats with unit lengths and degrees of polymorphism similar to the 5p15.1 repeat have been described in the literature. Tandem repeat rs447, located at 4p16.1, varies from 17 to 95 units; each unit is 4746 bp in length (Kogi *et al.,* 1997; Gondo *et al.,* 1998) and contains a central highly conserved region encoding a deubiquitinating enzyme (Saitoh *et al.,* 2000). Of greater interest, as noted above, is the tandem repeat D4Z4, located on 4qter (as well as 10qter) with a unit length of 3.3 kb and associated with FSHD (van der Maarel *et al.,* 2007). In normal individuals the repeat varies between 11 and 100 units per allele while over 95% of FSHD patients have an allele with 1-10 units. The relationship between repeat length and disease is complex, as the allelic background on which the shortened repeat

resides influences penetrance (Lemmers *et al.,* 2002; Lemmers *et al.,* 2007). Proposed pathogenic mechanisms of D4Z4 contraction include copy number effect of a gene located within the highly conserved portion of the repeat unit or dysregulation of a neighboring gene (Clapp *et al.,* 2007; van der Maarel *et al.,* 2007). Similar dysregulation effects have been observed in other CNVs (Sutrala *et al.,* 2007). Given the presence of a transcript generated from the highly conserved region in each 5p15.1 repeat unit, and the proximity of *BASP1*, a gene implicated in neurodevelopment, either mechanism provides a plausible pathogenic link between a CNS phenotype and the repeat.

Despite potential relevance, our data provide little support for a genetic relationship between the 5p15.1 repeat and schizophrenia. The small number of families and the small size of the families included in our linkage analysis substantially limit the power of the analysis, and therefore it remains possible that contraction of this repeat could be an uncommon risk factor for schizophrenia. In addition to sample size, our association study is limited by the loss of information through averaging of the two alleles; the qPCR signal represents the sum of the repeat length, so that small alleles could be masked by larger alleles in any given individual. However, our analysis would detect an association of combined repeat length with schizophrenia, or a shift in repeat length distribution towards shorter allele length in schizophrenia compared to controls if short alleles were a relatively common risk factor. Further, it remains possible that, as in FSHD, short alleles only contribute to disease risk when residing on a particular allelic background.

We conclude that long tandem repeats are a relatively unexplored yet critical component of genomic variation, with potential relevance to normal human variation and to complex diseases such as schizophrenia. While commercially available array-based strategies will detect signals from length variations in some tandem repeats, typical analyses have thus far not systematically distinguished repeat length variation from simple deletions and duplications. Consistent and thorough genome-wide detection and analysis of length variation of tandem repeats will likely require further refinement of arrays designed to detect CNVs and systematic analytic methods, specifically targeted to this form of genomic variation. The importance of this effort is highlighted by the increasing evidence that CNVs contribute to the risk for developing schizophrenia.

ACKNOWLEDGEMENTS

The authors thank P Zandi, R Reeves, H Lorenzi, C Callahan, A Bachani, A Seixas, J Hwang, W Chung, K Chambert, J Pevsner, E Roberson, and E Johnson for advice and technical assistance; C Pato, M Pato, A Medeiros, C Carvalho, A Macedo, A Dourado, I Coelho, J Valente, M Soares, C Ferreira, and M Azevedo for Azores sample collection; and J. R. DePaulo for on-going support.

Grant support: This work was supported in part by NIH MH015330, MH082262, and the Stanley Medical Research Institute.

References

Bassett AS, Hodgkinson K, Chow EW, Correia S, Scutt LE, Weksberg R. 22q11 Deletion Syndrome in Adults with Schizophrenia. Am J Med Genet. 1998; 81:328–337. [PubMed: 9674980]

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nuc Acids Res. 1999; 27:573–580.

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Molec Biol. 1997; 268:78–94. [PubMed: 9149143]

Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007; 81:264–279. [PubMed: 17668377]

Cowell JK, Wang YD, Head K, Conroy J, McQuaid D, Nowak NJ. Identification and characterisation of constitutional chromosome abnormalities using arrays of bacterial artificial chromosomes. Brit J Cancer. 2004; 90:860–865. [PubMed: 14970865]

DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, et al. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiat. 2002; 159:803–812. [PubMed: 11986135]

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, et al. Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. Ann Neurol. 2004; 55:174–179. [PubMed: 14755720]

Gondo Y, Okada T, Matsuyama N, Saitoh Y, Yanagisawa Y, Ikeda JE. Human megasatellite DNA RS447: copy-number polymorphisms and interspecies conservation. Genomics. 1998; 54:39–49. [PubMed: 9806828]

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36:949–951. [PubMed: 15286789]

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Gen Res. 2002; 12:996–1006.

Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Molec Genet. 2008; 17:458–465. [PubMed: 17989066]

Kogi M, Fukushige S, Lefevre C, Hadano S, Ikeda JE. A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization. Genomics. 1997; 42:278–283. [PubMed: 9192848]

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Amer J Hum Genet. 1996; 58:1347–1363. [PubMed: 8651312]

Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature. 1997; 386:847–851. [PubMed: 9126745]

Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 2002; 32:235–236. [PubMed: 12355084]

Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Amer J Hum Genet. 2007; 81:884–894. [PubMed: 17924332]

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001; 293:864–867. [PubMed: 11486088]

Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet. 2000; 26:191–194. [PubMed: 11017075]

Mead S. Prion disease genetics. Europ J Hum Genet. 2006; 14:273–281. [PubMed: 16391566]

Moon HJ, Yim SV, Lee WK, Jeon YW, Kim YH, Ko YJ, et al. Identification of DNA copy-number aberrations by array-comparative genomic hybridization in patients with schizophrenia. Biochem Biophys Res Comm. 2006; 344:531–539. [PubMed: 16630559]

Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiat. 1999; 56:940–945. [PubMed: 10530637]

Nyholt DR. GENEHUNTER: your ‘one-stop shop’ for statistical genetic analysis? Hum Hered. 2002; 53:2–7. [PubMed: 11901265]

Orr HT, Zoghbi HY. Trinucleotide repeat disorders. Ann Rev Neurosci. 2007; 30:575–621. [PubMed: 17417937]

Pato CN, Pato MT, Kirby A, Petryshen TL, Medeiros H, Carvalho C, et al. Genome-wide scan in Portuguese Island families implicates multiple loci in bipolar disorder: fine mapping adds support on chromosomes 6 and 11. Am J Med Genet: Part B. 2004; 127B:30–34. [PubMed: 15108176]

Pulver AE, Nestadt G, Goldberg R, Shprintzen RJ, Lamacz M, Wolyniec PS, et al. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. J Nerv Ment Dis. 1994; 182:476–478. [PubMed: 8040660]

Robinson MM, Yatherajam G, Ranallo RT, Bric A, Paule MR, Stargell LA. Mapping and functional characterization of the TAF11 interaction with TFIIA. Molec Cell Biol. 2005; 25:945–957. [PubMed: 15657423]

Saitoh Y, Miyamoto N, Okada T, Gondo Y, Showguchi-Miyata J, Hadano S, et al. The RS447 human megasatellite tandem repetitive sequence encodes a novel deubiquitinating enzyme with a functional promoter. Genomics. 2000; 67:291–300. [PubMed: 10936051]

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, et al. Challenges and standards in integrating surveys of structural variation. Nat Genet. 2007; 39:S7–15. [PubMed: 17597783]

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, et al. Strong association of de novo copy number mutations with autism. Science. 2007; 316:445–449. [PubMed: 17363630]

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet. 2006; 38:1038–1042. [PubMed: 16906162]

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. Science. 2003; 302:841. [PubMed: 14593171]

Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008 in press.

Sutrala SR, Goossens D, Williams NM, Heyrman L, Adolfsson R, Norton N, Buckland PR, Del-Favero J. Gene copy number variation in schizophrenia. Schizophr Res. 2007; 96:93–99. [PubMed: 17826036]

Sutrala SR, Norton N, Williams NM, Buckland PR. Gene copy number variation in schizophrenia. Am J Med Genet, Part B. 2008; 147B:606–611. [PubMed: 18163393]

The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008 in press.

van der Maarel SM, Frants RR, Padberg GW. Facioscapulohumeral muscular dystrophy. Biochim Biophys Acta. 2007; 1772:186–194. [PubMed: 16837171]

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008; 320:539–543. [PubMed: 18369103]

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. New Eng J Med. 2008; 358:667–675. [PubMed: 18184952]

Wilson GM, Flibotte S, Chopra V, Melnyk BL, Honer WG, Holt RA. DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. Hum Molec Genet. 2006; 15:743–749. [PubMed: 16434481]

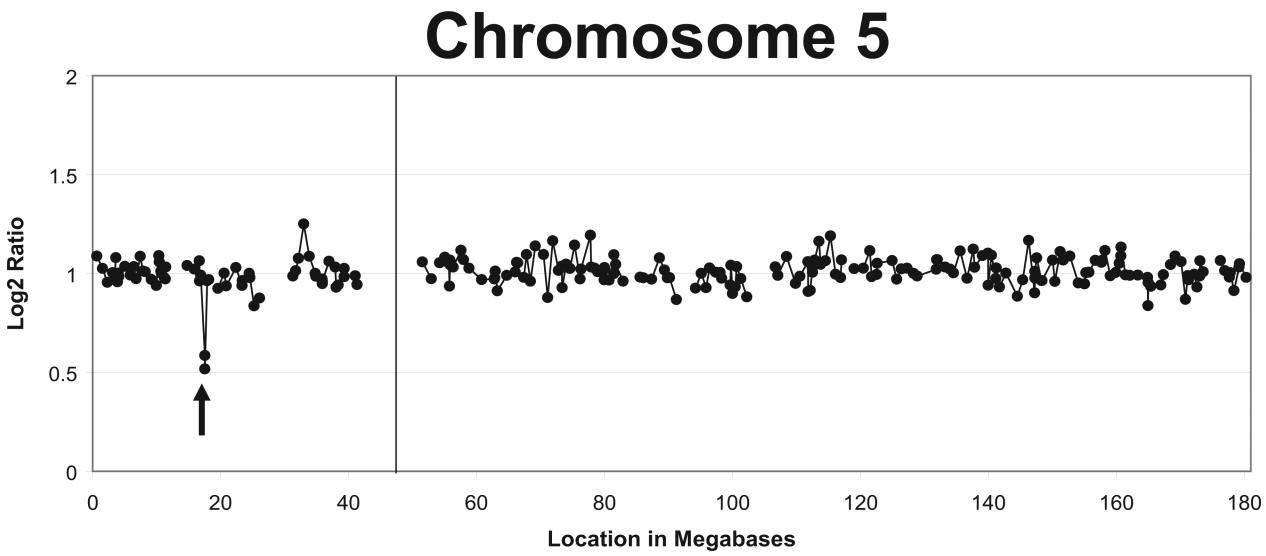

Figure 1.

BAC array results demonstrating diminished log2 signal ratio on chromosome 5p15.1 in DNA from study participant S56. The Y axis depicts the log2 ratio of hybridization signal compared to pooled controls. The X axis depicts location in megabases along chromosome 5. The arrow identifies the loci defined by BACs RP11-88L18 and RP11–91D21.

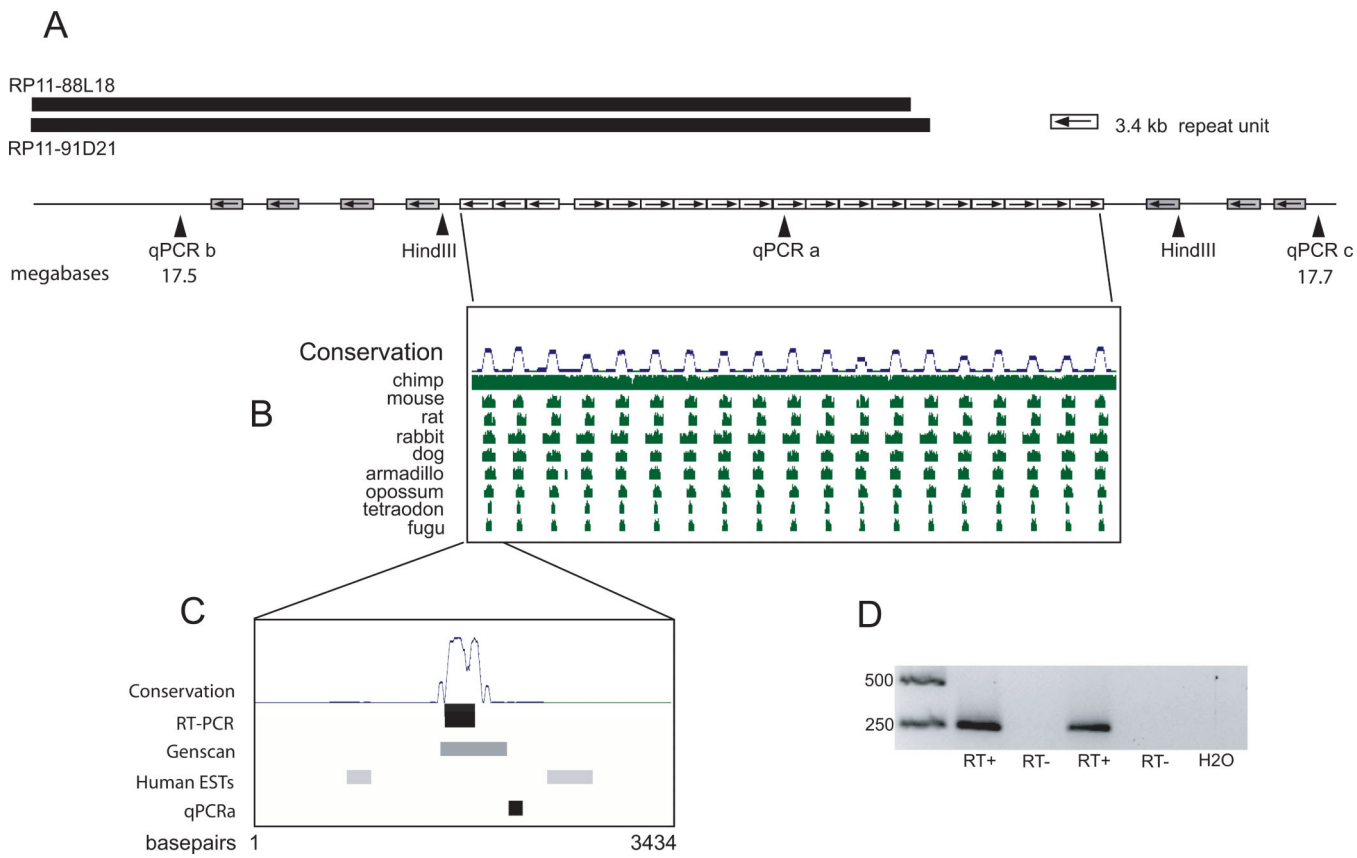

Figure 2.

Genomic structure of 5p15.1**. A**. Representation of 200 kb of chromosome 5p15.1 from the human reference sequence (hg 18 build 36) illustrating the relative location and orientation of 26 repeat units. The units colored white have > 99% identity; the units colored grey have identity ranging from 88% to 93%. Shown are BACs detecting the copy number change, the qPCR sites used to define the copy number change (telomeric qPCR begins at 17,521,414 bp; centromeric qPCR ends at 17, 718,335 bp and is located in a spliced EST-BM682321) and the HindIII restriction sites (17,571,914 and 17,648,982 respectively) used in pulse field experiments. Small arrows in the repeat units depict their orientation. **B.** Conservation pattern across species is shown for the central 19 tandem units (http://genome.ucsc.edu). **C**. Representation of a single 3434 bp repeat. The figure depicts the relative position of the qPCR product used to measure copy number, the PCR product used to detect transcription, Genscan gene prediction, human ESTs, and the conservation across the repeat unit (http://genome.ucsc.edu). **D.** Transcript expression of the conserved region of 5p15.1 tandem repeat units. Shown is RT-PCR performed with cDNA obtained from frontal cortex of 2 individuals. The same result was obtained using cDNA from whole brain of 3 individuals and hippocampus of 1 individual. Ladder is in lane 1 with units in base pairs. A band of the predicted 249 base pairs located in the central conserved region of the repeat unit is shown in lanes 2 and 4. The region was amplified in the RT+ samples but not in the corresponding RT- samples or the water blank.

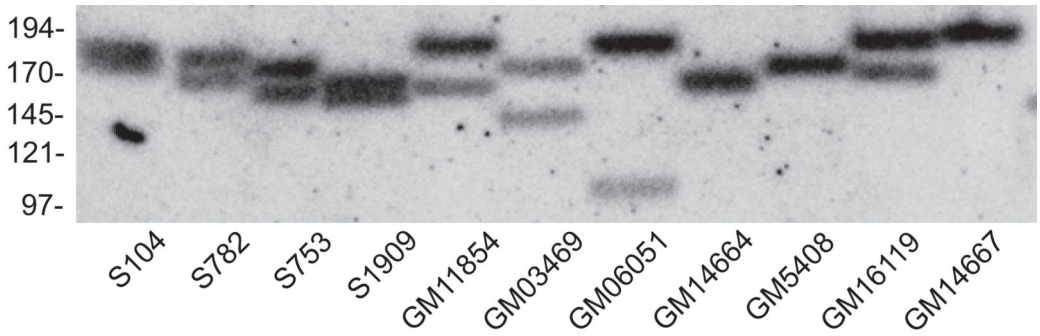

Figure 3.

The 5p15.1 tandem repeat length is variable. PFGE of HindIII restriction fragments containing the repeat from probands and controls, detected with a repeat specific probe. S prefix = study probands, GM prefix = controls. Ladder shown in kilobases.

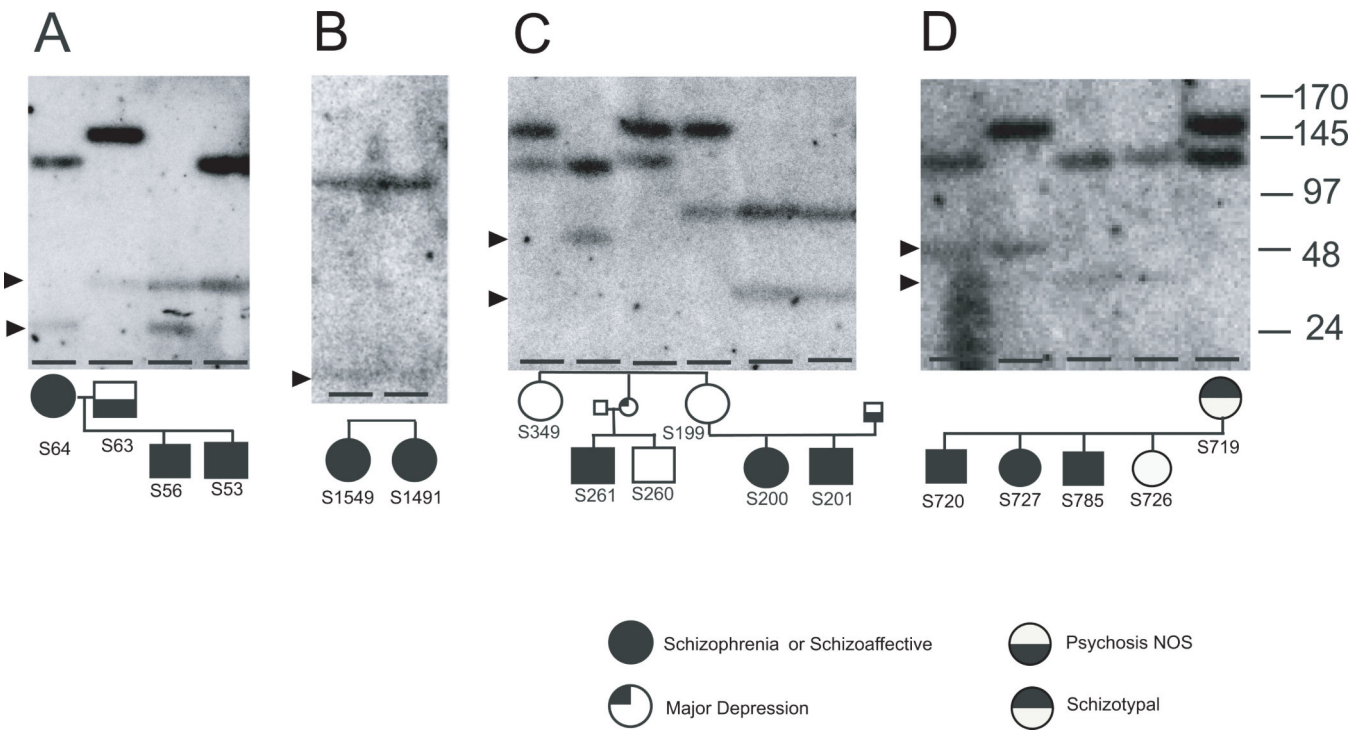

Figure 4.

Stability of repeat length inheritance and partial segregation of repeat length with disease in four small pedigrees. Bands with fewer than 22 repeat units are indicated by arrows for each pedigree. Size markers are shown in kilobases. Individual S276 carries a short allele but was not affected at last contact.

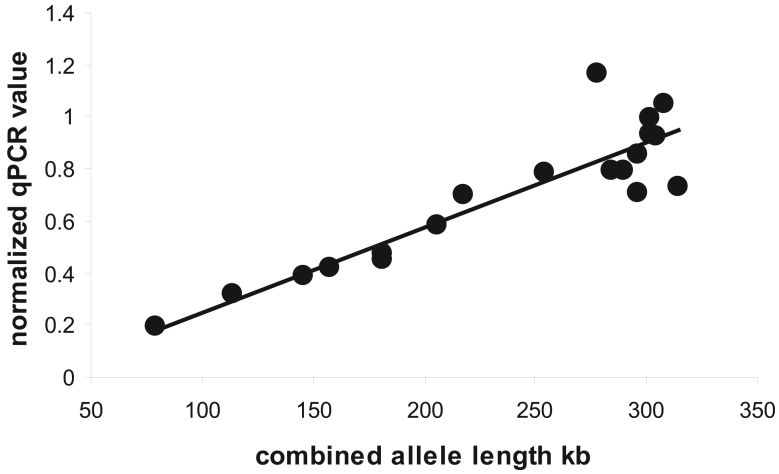

Figure 5.

qPCR is a valid measure of combined allele length. Approximate repeat length as determined by PFGE for each allele is strongly correlated with normalized qPCR results for the combined alleles. $r^2 = 0.82$. The increasing scatter at longer repeat lengths reflects limitations of PFGE resolution.

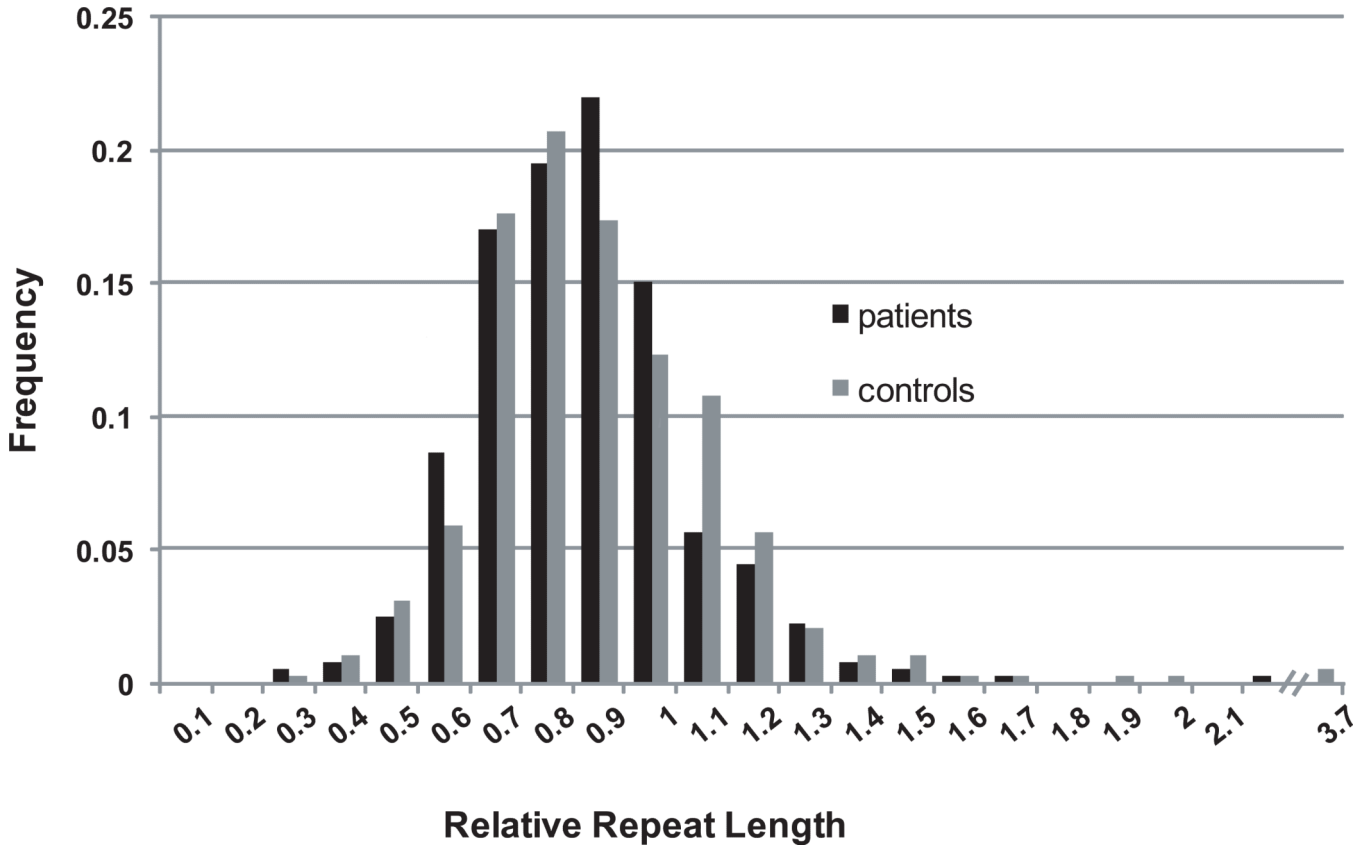

Figure 6.

Distribution of 5p15.1 repeat length in 406 patients and 392 controls. Relative repeat length, as indicated by qPCR, binned as shown on the X-axis. Y axis indicates overall frequency of lengths in each bin.