Am. J. Hum. Genet. 59:912–917, 1996

Expansion of 50 CAG/CTG Repeats Excluded in Schizophrenia by Application of a Highly Efficient Approach Using Repeat Expansion Detection and a PCR Screening Set

Timothy Bowen,[1,2] Carol Guy,[1] Graham Speight,[1] Lisa Jones,[1] Alastair Cardno,[1] Kieran Murphy,[1] Peter McGuffin,[1] Michael J. Owen,[1,2] and Michael C. O'Donovan[1]

[1]Department of Psychological Medicine and [2]Institute of Medical Genetics, University of Wales College of Medicine, Cardiff

Summary

Studies of the transmission of schizophrenia in families with affected members in several generations have suggested that an expanded trinucleotide repeat mechanism may contribute to the genetic inheritance of this disorder. Using repeat expansion detection (RED), we and others have previously found that the distribution of CAG/CTG repeat size is larger in patients with schizophrenia than in controls. In an attempt to identify the specific expanded CAG/CTG locus or loci associated with schizophrenia, we have now used an approach based on a CAG/CTG PCR screening set combined with RED data. This has allowed us to minimize genotyping while excluding 43 polymorphic autosomal loci and 7 X-chromosomal loci from the screening set as candidates for expansion in schizophrenia with a very high degree of confidence.

Introduction

Although there is convincing evidence that genetic factors are the major contributors to susceptibility to schizophrenia, the exact mode of inheritance of this disorder remains unknown (McGuffin et al. 1995). However, several studies have found that, in multiplex families transmitting schizophrenia, the age at onset becomes younger in succeeding generations, an observation consistent with genetic anticipation and the hypothesis of transmission of schizophrenia at least in part through an expanded trinucleotide repeat mechanism (O'Donovan and Owen 1996). However, it should be stressed that the published studies have all been conducted in families originally ascertained for genetic linkage analysis of schizophrenia. As has been discussed elsewhere, such families are almost inevitably biased toward earlier ages at onset in younger generations (Penrose [1945] 1991; O'Donovan and Owen 1996). Since this bias can simulate anticipation, it is as yet premature to conclude that true genetic anticipation occurs in schizophrenia. Nevertheless, the hypothesis of the involvement of an expanded trinucleotide repeat mechanism in schizophrenia has recently received strong support from two independent groups who found that the distribution of maximum CAG/CTG repeat size is greater in schizophrenics than in controls (Morris et al. 1995; O'Donovan et al. 1995) by using the method of repeat expansion detection (RED) (Schalling et al. 1993). These findings have now been further replicated in a large multicenter European study (O'Donovan et al., in press) and in a smaller sample by a third independent group (Vincent et al, 1996).

Unfortunately, the task of identifying the particular expanded locus or loci associated with schizophrenia is not simple for several reasons. First, because RED is a multilocus assay, and because the repeat sequences associated with schizophrenia are only modestly expanded and overlap with the size of repeats found in unaffected individuals, it is not possible to map the loci directly with linkage by using RED as originally suggested (Schalling et al. 1993). Second, although a method of directly mapping expanded trinucleotide repeat sequences by FISH has been developed (Haaf et al. 1996), this is unsuitable for repeats in the size range (60–90 repeats) implicated in schizophrenia by RED (O'Donovan et al. 1995). A third difficulty is that traditional cloning methods result in deletion of the repeat sequences (Gastier et al. 1996). Finally, schizophrenia is likely to be heterogeneous, and therefore, even if direct cloning was straightforward, it would require the construction and screening of multiple libraries from patients.

Because there is at present no easy way to proceed, some groups have approached the problem by identifying cDNAs that contain CAG repeats and then seeking expansions in these candidates in case control studies

Received June 11, 1996; accepted for publication July 26, 1996.

Address for correspondence and reprints: Dr. Michael O'Donovan, Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, Wales. E-mail: wpcmod@cardiff.ac.uk

0002-9297/96/5904-0023$02.00

(e.g., Margolis et al. 1996). However, while this approach has been successful in at least one Mendelian disorder showing anticipation (Kawaguchi et al. 1994), it is laborious and each gene has a low prior probability of being involved in the disease. Recently, a CAG/CTG repeat genomic PCR screening set has been developed (Gastier et al. 1996). This set is based on 338 STSs obtained from 1,138 genomic DNA clones containing five or more CAG/CTG repeats. One hundred forty-one of these repeats contain seven or more repeats and, of these, 45 were found to be polymorphic (Gastier et al. 1996). Although the prior probability of any single repeat being involved in a disease is low, this is offset by the fact that a large number of candidate loci can be readily screened.

The main theoretical drawback of applying a genomic CAG/CTG screening set is that the proportion of repeats that are exonic (as are all known pathogenic CAG/CTG repeats) is as yet unknown. However, there is a relative exclusion of CAG repeats from intronic DNA (Stallings 1994), and our own analysis using the gene recognition and analysis internet link (GRAIL) (Xu et al. 1995) suggests that ≥40% of the CAG/CTG repeats in the screening set are likely to lie within exons (authors' unpublished data).

In this study, we have analyzed the repeat sizes at all 45 CAG/CTG loci that were described as polymorphic by Gastier et al. (1996) in patients with schizophrenia. In addition, because our original RED data (O'Donovan et al. 1995) suggested an interaction between repeat size and sex, we also selected primers to amplify all repeats that mapped to the X chromosome, although recent data from our larger study suggests that the original sex effect was a chance finding (O'Donovan et al., in press).

In order to minimize genotyping and increase throughput, patients with enlarged repeats were first identified by RED. A further threefold increase in throughput was achieved by the use of fluorescently labeled PCR products and an ABI 373 DNA fragment analyzer. Our RED data excluded the presence of very large expanded CAG/CTG alleles, which might be impossible to amplify. Nevertheless, because of differential amplification of small and large alleles, in parallel to this high throughput approach, we also employed a highly sensitive radioisotopic detection system to confirm the fluorescent data and safeguard against failure to detect moderately expanded alleles.

Subjects and Methods

Subjects

Ethical approval for molecular genetic studies of schizophrenia was obtained from the South Glamorgan Local Research Ethics Committee, and all subjects provided informed written consent. Probands (n = 70; 48 male, 22 female;) were Caucasians recruited from clinics in Wales. All met DSM-IV (American Psychiatric Association 1994) criteria for schizophrenia. Diagnoses were made with OPCRIT version 3.31 (McGuffin et al. 1991) following a semistructured interview SCAN (Wing et al. 1990) and examination of case notes. The mean age at onset was 24.2 years (SD 8.1), and the mean age at interview and venepuncture was 45.6 years (SD 13.6). Most (82.9%) had a family history of schizophrenia. RED scores obtained from previous studies (O'Donovan et al. 1995 and unpublished data) were used to select 11 patients (5 female, 6 male; mean age at onset 22.2 years, SD 5.6) with maximum RED product sizes of between 60 and 90 repeats.

Amplification

We obtained primers for all the polymorphic autosomal CAG/CTG repeats (n = 44) from the CAG/CTG screening set (Gastier et al. 1996). In addition, we obtained primers for all the CAG/CTG repeats that map to the X-chromosome (n = 7) including one further polymorphic locus. Primer pairs were obtained from Research Genetics with the forward primer of each pair labeled with FAM, TET, or HEX fluorescent dyes.

Amplification was carried out using high-molecular-weight DNA, which was isolated from lymphocytes using phenol-chloroform extraction. Reactions were carried out in a total volume of 12.5 µl containing 1 × PCR buffer (Amersham), 30 ng of DNA, primers at 1 µM, dNTPs at 0.4 mM, MgCl at 1.5 mM, and 0.5 U of *Taq* polymerase (Amersham). Each reaction began with a single denaturation step of 94°C for 5 min. This was followed by 35 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C.

Fluorescent Detection of PCR Products

PCRs were performed individually. The PCR products from each individual subject were then separately arranged in 17 pools, each consisting of the product of three primer pairs with different fluorescine labels. After pooling, the volume of each sample was increased 12-fold with sterile water, and then 1.8 µl were removed for electrophoresis on a standard denaturing 6% polyacrylamide gel. The predicted sizes of products in each pool were similar so that weak signal corresponding to an expanded allele would not be masked by "cross talk." Electrophoretic analysis was carried out using a ABI 373 DNA sequencer (ABI) with TAMRA-labeled molecular weight markers run in each lane. The sequencer was preset for a 10-h run at 2,500 V, 21 mA, and 25 W. The GENESCAN and GENOTYPER software packages (ABI) were used for collection and analysis of data.

Parallel Screening by Hybridization

Individual PCR products were subjected to denaturing electrophoresis on a 6% polyacrylamide gel, after which the DNA was transferred to HYBOND N^+ membrane (Amersham) by capillary transfer overnight. DNA was fixed to the membrane by UV cross-linking. Membranes were prehybridized for 15 min at 55°C, and hybridization was then performed in Rapid Hyb Buffer (Amersham) for 1 h at 55°C with a $(CAG)_{10}$ oligonucleotide probe end labeled with γ32P dATP. Membranes were washed twice at room temperature for 20 min in 1 × SSC/0.1% SDS followed by 30 min at 55°C in 0.1 × SSC/0.1% SDS. Autoradiography was performed with one enhancing screen at −70°C for 1 or 2 nights. Autoradiographs were deliberately overexposed to facilitate visualization of larger products. Expanded Huntingtons disease alleles are easily visualized by this procedure (data not shown).

Results

The results of the fluorescent analysis are summarized in table 1. This lists the marker names, the chromosomes on which they are mapped, the predicted product sizes, the corresponding repeat sizes (from Gastier et al. 1996), the size of the largest PCR product, and the corresponding *maximum* size of repeat. The maximum size of repeat was calculated under the assumption that any size difference observed was due wholly to expansion of the CAG repeat and therefore, in some cases, this is likely to be an *overestimate* of the actual CAG repeat size. In a minority of cases, there was a difference between predicted allele size in bases and the actual size measured by the ABI 373 in mobility units by ~1 bp. In such cases, the inferred size of repeat was rounded to the nearest whole number.

Nine of the markers that were expected to be polymorphic (Gastier et al. 1996) were not polymorphic in this sample (nonpolymorphic markers are indicated with footnote in table 1). However, in another sample of individuals, very uncommon polymorphisms were seen at a further four of these loci (data not shown). One set of primers in the set (GCT1A10) failed to amplify.

The largest number of repeats in an allele was found using marker GCT5E11, in which one allele of 30 repeats (maximum size) was observed. This represents an increase in size of only 7 repeats from the predicted product size (Gastier et al. 1996) but of 18 from the most common allele in our sample. No extra alleles were detected by hybridization that had not already been detected by the ABI system. However, as anticipated, differential amplification resulted in weaker signal sizes for the larger alleles in the fluorescent analysis. This phenomenon was not apparent in the hybridization experiments, suggesting that the reduction in amount of product was compensated for by an increase in the number of labeled probe molecules annealing to it.

The extra specificity of the hybridization assay also allowed us to confirm that all the products which we considered alleles did contain CAG/CTG repeats. This is important, since some of the primer sets gave a number of nonspecific products (as described in the Cooperative Human Linkage Center internet data), which were largely eliminated by hybridization at the stringency described here.

Discussion

A genomic CAG/CTG PCR screening set has recently been developed to facilitate the identification of pathogenic expanded CAG/CTG repeats (Gastier et al. 1996). However, while the application of this methodology to genetically homogeneous disorders is simple and requires the study of only one or two affected individuals, in the case of complex diseases such as schizophrenia, which is likely to be genetically heterogeneous, more subjects must be examined. It follows that, in a study of this nature, power is a critical consideration.

However, because the proportion of schizophrenics who carry expanded CAG repeats is not known, it is difficult to make accurate calculations of power in an unselected sample. For example, if we assume that an expanded repeat is present in 10% of patients, it is necessary to genotype 28 patients to obtain 95% power, but for an expanded repeat that is present in only 5% of patients, the corresponding figure is 58 subjects. For a few markers, this is a trivial undertaking, but in order to screen the large number of loci in the genome with CAG repeats of five or more this becomes very labor intensive.

We have therefore developed a protocol that allows us to minimize genotyping by using a two-stage procedure. Subjects with RED product sizes of 60–90 repeats were chosen, because this size range is more common in schizophrenic patients than controls (O'Donovan et al. 1995) yet is small enough to allow PCR amplification. From our data (O'Donovan et al. 1995), which have been subsequently replicated (authors' unpublished data), there is an excess of ~33% of patients over controls carrying repeat sizes in this putative pathogenic range. On the assumption that this figure is approximately correct, only eight individuals with RED scores within this range need to be genotyped to achieve 95% power. This represents a major improvement in genotyping efficiency. We have sought to increase further the throughput by pooling DNA samples prior to PCR. However, so far, this approach has been unsuccessful, since we have found that larger alleles are effectively competed against.

Table 1

CAG/CTG Repeat Sizes

Primer Pair	Chromosome	Predicted Size (bp)	Repeat Type	Largest Fragment (bp/[CAG]n)
GCT16B08	6	314	(CTA)11-(CTG)8	321/10
GCT8E11	3	291	(CA)16(CAG)7	291/7
GCT5E11	3	254	(GCT)23	275/30
ACT3F12	13	251	(TAG)21-(CGG)4(CAG)6	254/7
GCT3B11	16	241	(GCT)8(GT)16	263/15
GCT14E02	4	238	(CTG)7	244/9
GCT8C05	3	237	(CAG)11	276/24
GCT13H07	7	236	(CTG)7	242/9
GCT8B09	2	231	(CAG)10	234/11
GCT7C09	15	230	(GCT)3GTT(GCT)9	246/14
GCT6EO7	12	224	(CTG)12	227/13
GCT4B10	3	210	(AGC)16	216/18
GCT5C07	2	197	(CAG)7[a]	197/7
GCT10C10	20	194	(CAG)14?	206/18 (?)
GCT4C10	X	192	(ATC)8-(ATC)4(AGC)7	192/7
GCT1A11	12	186	(CAG)6	189/7
GCT3B12	3	182	(GCT)7[a]	182/7
GCT5E07	6	179	(CAG)11GAG(CAG)4	182/12
GCT3E06	18	171	(GCT)7	173/8
GCT1C8	15	167	(TGC)8[a]	170/9
GCT14E12	X	165	(AGC)5[a]	168/6
GCT6F03	4	170	(GCT)12	175/14
GCT14C12	11	161	(CTG)6[a]	164/7
GCT17B07	3	158	(CTG)7	161/8
GCT1A10	3	168	(AGC)15	Not amplified
GCT10D04	17	155	(GCT)13	158/14
GCT10F04	5	151	(GCT)10	160/13
GCT6G02	6	161	(AGC)8	175/12
ACT3E01	10	151	(ACT)5(GCT)11	154/12
GCT10B12	1	150	(GCT)10	153/11
GCT14H07	15	150	(GCT)7[a]	150/7
GCT10G11	Unknown	150	(GCT)12	156/14
GCT16D06	6	149	(AGC)7[a]	153/8
GCT10G08	1	148	(GCA)11	151/12
GCT11B12	2	147	(GCT)7	147/7
GCT17B09	9	145	(AGC)11	178/22
GCT16B04	4	142	(CTG)9(ATG)8	148/11
GCT4B05	6	140	(AGC)11	142/12
GCT12D12	X	139	(TGA)3(TGC)6(TGA)5[a]	139/6
GCT5E05	5	134	(GCT)8	137/9
GCT11G03	20	133	(GCT)8[a]	133/8
GCT13F07	8	125	(GCT)9[a]	125/9
GCT8B03	3	123	(GCT)7	123/7
GCT7D06	X	122	(CAG)7[a]	119/6
GCT9C05	7	117	(GCT)7[a]	120/8
GCT5D07	18	117	(GCT)8	121/9
GCT5D10	X	117	(GCT)5[a]	117/5
GCT13B01	X	117	(CAG)5CGG(CAG)5[a]	120/6
GCT7G03	4	113	(GCA)7	122/10
GCT6D06	X	110	(AGC)6[a]	108/5
GCT6E11	17	105	(GCT)10	109/11

NOTE.—Predicted product sizes and corresponding repeat sizes are from Gastier et al. (1996).
[a] Marker not polymorphic.

In this study, we have studied 11 individuals, which yields an estimated power of 99%. Despite the high power, we did not observe any CAG or CTG repeats within the size range of interest in schizophrenia (Morris et al. 1995; O'Donovan et al. 1995). The largest repeat allele we have detected (at locus GCT5E11) is only 30 repeats, smaller than the smallest pathogenic trinucleotide repeat yet described. We conclude, therefore, that none of the loci studied remain plausible candidates for expansion in schizophrenia. Furthermore, because we have carefully selected subjects, we have obtained sufficient power to conclude that the exclusion of these loci can be viewed with considerable confidence.

Although we have failed to identify the CAG/CTG repeats that are expanded in patients with schizophrenia, the observation that CAG repeats are disproportionately distributed in exons (Stallings 1994) suggests that we have sampled a large number of (as yet unknown) candidate genes in this study. Therefore, we would anticipate that the work presented here will complement that of other groups who are taking cDNA-based approaches. The repeats we have examined that are genuinely located in exons will have flanking sequence homologous to cDNAs (for example the STS GCT7B03 (CHLC) corresponds to the repeat within CAGR1, the human homologue of the *Caenorhabditis elegans* fate-determining gene (Margolis et al. 1996)). Thus, our data will allow other groups to exclude candidate cDNAs without having to clone the entire cDNA or undertake case control studies.

The approach outlined here is applicable to all complex diseases in which expanded trinucleotide repeats are proposed, for example, bipolar disorder, Parkinson disease, and Crohn disease (Morrison 1993). However, it is very likely that in these diseases only a fraction of cases (if any) will be attributable to expanded repeats, and therefore the improvement in efficiency achieved by our approach will represent a major reduction in workload. On the assumption that there are 1,500 CAG/CTG loci of five or more repeats (Gastier et al. 1996), 58 subjects (see above) would require almost 90,000 genotypings for a complete scan, whereas the 8 subjects required by this two-phase approach will need only 12,000 genotypings. For fluorescent detection, there is a restriction to three markers per lane, since further multiplexing by multiple loadings may obscure products from large alleles. Twelve thousand genotypes represents only 120 gels, and therefore, by adopting the protocol described here, the use of a complete screening set (when it is available) is feasible for a single worker in a few months.

Although it is possible to restrict the sample to subjects with modest-sized repeats, because of preferential amplification of the smaller alleles, for confident exclusion of loci, we would recommend that if the fluorescent phase is negative, then hybridization analysis (which did not show a bias to detecting smaller repeats) should be undertaken. Because only a fraction of each PCR is used for fluorescent detection, hybridization analysis does not require duplication of effort, and, while slower by a factor of three, it remains practical for a single worker to undertake within ~9 mo. This may therefore be a viable alternative either for groups without access to fluorescent detection apparatus or for whom speed of analysis is relatively unimportant. However, the cost of hybridization membrane ensures that this is not an inexpensive alternative.

We would stress that at the time of this study, only about $\frac{1}{3}$ to $\frac{1}{2}$ of the total expected number of CAG/CTG repeats had been identified (Gastier et al. 1996). Furthermore, we have prioritized only those repeats that were thought to be polymorphic and that map to the X-chromosome. In future studies, we will select markers that map to candidate chromosomes or that are predicted by GRAIL to lie in coding sequences.

Acknowledgments

This work was funded by the Medical Research Council and the Welsh Scheme for the Development of Health and Social Research. We also thank Nigel Williams, Jo Daniels, and Mark Rees for advice concerning the ABI system.

References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th ed. American Psychiatric Association, Washington, DC

Gastier JM, Brody T, Pulido JC, Businga T, Sunden S, Hu X, Maitra S, et al (1996) Development of a screening set for new (CAG/CTG)n dynamic mutations. Genomics 32:74–85

Haaf T, Sirugo G, Kidd KK, Ward DC (1996) Chromosomal localization of long trinucleotide repeats in the human genome by fluorescence in situ hybridization. Nat Genet 12: 183–185

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q 32.1. Nat Genet 8:221–228

Margolis RL, Stine OC, McInnis MG, Ranen NG, Rubinsztein DC, Leggo J, Jones Brando LV, et al (1996) cDNA cloning of a human homologue of the Caenorhabditis elegans cell fate-determining gene mab-21: expression, chromosomal localization, and analysis of a highly polymorphic (CAG)n trinucleotide repeat. Hum Mol Genet 5:607–616

McGuffin P, Farmer A, Harvey I (1991) A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the OPCRIT system. Arch Gen Psychiatr 48:764–770

McGuffin P, Owen MJ, Farmer AE (1995) Genetic basis of schizophrenia. Lancet 346:678–682

Morris AG, Gaitonde E, McKenna PJ, Mollon JD, Hunt DM (1995) CAG repeat expansions and schizophrenia: association with disease in females and with early age-at-onset. Hum Mol Genet 4:1957–1961

Morrison PJ (1993) Trinucleotide repeat repeat repeat. Lancet 342:385–86

O'Donovan MC, Guy C, Craddock N, Bowen T, McKeon P, Macedo A, Maier W, et al. Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. Psychol Med (in press)

O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ, et al (1995) Expanded CAG repeats in schizophrenia and bipolar disorder. Nat Genet 10: 380–381

O'Donovan MC, Owen MJ (1996) Dynamic mutations and psychiatric genetics. Psychol Med 26:1–6

Penrose LS [1945] (1991) Survey of cases of familial mental illness. Eur Arch Clin Neurosc 240:314–324

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet 4:135–139

Stallings RL (1994) Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. Genomics 21:116–121

Vincent JB, Klempan T, Parikh SS, Sasaki T, Meltzer HY, Sirugo G, Cola P, et al (1996) Frequency analysis of large CAG/CTG trinucleotide repeats in schizophrenia and bipolar affective disorder. Mol Psychiatr 1:141–148

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, et al (1990) Schedules for clinical assessment in neuropsychiatry. Arch Gen Psychiatr 47:589–593

Xu Y, Mural R, Uberbacher EC (1995) Correcting sequencing errors in DNA coding regions using dynamic programming. Comput Appl Biosci 11:117–124