Hum Genet
DOI 10.1007/s00439-006-0308-9

ORIGINAL INVESTIGATION

Identification of 54 large deletions/duplications in *TSC1* and *TSC2* using MLPA, and genotype-phenotype correlations

Piotr Kozlowski · Penelope Roberts · Sandra Dabora · David Franz · John Bissler · Hope Northrup · Kit Sing Au · Ross Lazarus · Dorota Domanska-Pakiela · Katarzyna Kotulska · Sergiusz Jozwiak · David J. Kwiatkowski

Received: 23 October 2006 / Accepted: 23 November 2006
© Springer-Verlag 2006

Abstract Tuberous sclerosis (TSC) is an autosomal dominant disorder caused by mutations in either of two genes, *TSC1* and *TSC2*. Point mutations and small indels account for most *TSC1* and *TSC2* mutations. We examined 261 TSC DNA samples (209 small-mutation-negative and 52 unscreened) for large deletion/duplication mutations using multiplex ligation-dependent probe amplification (MLPA) probe sets designed to permit interrogation of all *TSC1/2* exons, as well as 15–50 kb of flanking sequence. Large deletion/duplication mutations in *TSC1* and *TSC2* were identified in 54 patients, of which 50 were in *TSC2*, and 4 were in *TSC1*. All but two mutations were deletions. Only 13 deletions were intragenic in *TSC2*, and one in *TSC1*, so that 39 (73%) deletions extended beyond the 5′, 3′ or both ends of *TSC1* or *TSC2*. Mutations were identified in 24% of small-mutation-negative and 8% of unscreened samples. Eight of 54 (15%) mutations were mosaic, affecting 34–62% of cells. All intragenic mutations were confirmed by LR-PCR. Genotype/phenotype analysis showed that all (21 of 21) patients with *TSC2* deletions extending 3′ into the *PKD1* gene had kidney cysts. Breakpoints of intragenic deletions were randomly distributed along the *TSC2* sequence, and did not preferentially involve repeat sequence elements. Our own 20-plex probe sets gave more robust performance than the 40-plex probe sets from MRC-Holland. We conclude that large deletions in *TSC1* and *TSC2* account for about 0.5 and 6% of mutations seen in TSC patients, respectively, and MLPA is a highly sensitive and accurate detection method, including for mosaicism.

Keywords *TSC1* · *TSC2* · *PKD1* · Large deletions · MLPA · Mosaic mutations

Electronic supplementary material The online version of this article (doi:10.1007/s00439-006-0308-9) contains supplementary material, which is available to authorized users.

P. Kozlowski · P. Roberts · S. Dabora · D. J. Kwiatkowski (✉)
Genetics Laboratory, Division of Translational Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, One Blackfan Circle, 6th flr, Rm 216, Boston, MA 02115, USA
e-mail: dk@rics.bwh.harvard.edu

D. Franz · J. Bissler
Deparment of Pediatrics and Neurology, Children's Hospital Medical Center, Cincinnati, OH, USA

H. Northrup · K. S. Au
Division of Medical Genetics, Department of Pediatrics, UT Medical School at Houston, Houston, TX, USA

R. Lazarus
Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

D. Domanska-Pakiela · K. Kotulska · S. Jozwiak
The Children's Memorial Health Institute, Warsaw, Poland

Introduction

Tuberous sclerosis (TSC, MIM#191100) is a relatively common (birth incidence 1 in 6,000; Osborne et al. 1991) autosomal dominant genetic disorder in which hamartomas and hamartias are found in multiple organ systems, including brain, skin, heart, lungs, and kidneys (Curatolo 2003; Gomez et al. 1999; Kwiatkowski 2004). It is due to inactivating mutations in either of *TSC1* (MIM#605284) or *TSC2* (MIM#191092), and follows a

two hit tumor suppressor molecular pathogenesis in most hamartomas.

Two-thirds of TSC patients present as sporadic cases with no family history, and are due to new mutations occurring in either of *TSC1* or *TSC2* (Consortium 1993; van Slegtenhorst et al. 1997). *TSC1* consists of 23 exons, with total genomic extent 53 kb, and coding region of 3.5 kb; while *TSC2* consists of 41 exons with total genomic extent 40 kb, and coding region of 5.4 kb. Mutations in *TSC2* are four to fivefold more common than in *TSC1* among TSC patients, particularly in sporadic cases, while *TSC1* mutations are roughly equally common as *TSC2* mutations in large TSC families (Dabora et al. 2001; Jones et al. 1999; Sancak et al. 2005). This is consistent with the observation that *TSC1* patients have milder overall severity than *TSC2* patients (Dabora et al. 2001; Jones et al. 1999; Sancak et al. 2005).

Multiple TSC patients with large genomic deletions and rearrangements in *TSC2* have been identified (Au et al. 1997; Brook-Carter et al. 1994; Dabora et al. 2000, 2001; Jones et al. 1999; Langkau et al. 2002; Longa et al. 2001), and indeed one such patient was instrumental in the identification of *TSC2* (Consortium 1993). On the other hand, genomic deletions and duplications in *TSC1* are much less common, although several have been identified in recent reports (Longa et al. 2001; Nellist et al. 2005). In addition, the observed frequency of *TSC2* large deletions/duplications may have been biased upward by the association of some *TSC2* deletions, those involving the adjacent *PKD1* locus, with early onset and progressive polycystic kidney disease, leading to enhanced reporting of such individuals and mutations (Brook-Carter et al. 1994; Martignoni et al. 2002; Sampson et al. 1997).

Recently, multiplex ligation-dependent probe amplification (MLPA) has been developed as a novel method for the detection of exonic deletions in multiple human disease genes (Schouten et al. 2002). This method utilizes an oligonucleotide ligation assay with embedded universal primer sequences to permit relatively uniform amplification of multiple (up to 40) regions in the genome, which enables accurate copy number determination of those genomic regions. Here we apply both commercially available *TSC1* and *TSC2* MLPA probe sets as well as our own homemade probe sets to enable comprehensive analysis of all exons within *TSC1* and *TSC2* for deletions and duplications. These assays were applied to a large collection of TSC patients, the majority of which had previously undergone thorough analysis for small mutations within *TSC1* and *TSC2*. All intragenic deletions and duplications were confirmed by long range PCR (LR-PCR) studies. This permits a comprehensive view of the frequency and extent of genomic deletions and duplications within *TSC1* and *TSC2*, and their association with various clinical features of the disease.

Materials and methods

TSC patients

The 261 TSC patient DNA samples we studied were derived from four sources: two TSC clinics in Cincinnati and Warsaw ($n = 48$ and 82, respectively); a TSC mutational diagnostic referral center in Houston, TX, USA ($n = 112$); and mail-in requests for TSC mutation detection directed to DJK ($n = 19$). All TSC samples analyzed were derived from patients who met clinical criteria for a diagnosis of TSC. DNA was isolated from blood samples. Local informed consent was obtained from all patients, and coded numbers were used to label DNA samples. All studies were approved by the IRBs of the relevant institutions.

Eighty-seven of these samples were derived from a previous collection of TSC samples in which comprehensive small mutation detection was performed using DHPLC (Dabora et al. 2001). All samples for which adequate DNA remained from this previous analysis, and which were small-mutation-negative were included in the current analysis. Ten of those samples were thought to have deletions based upon previous analysis by LR-PCR or quantitative PCR with 3–4 primer sets (Dabora et al. 2000, 2001). Similarly, 94 of the samples from Houston were derived from a set of 347 samples screened for small mutations by comprehensive resequencing, and found to be negative. Two of those samples were thought to have deletions based upon previous Southern blot analysis (Au et al. 1997). Twenty-eight samples had been screened for small mutations by Athena Diagnostics Inc., Worcester, MA, USA. Fifty-two samples were studied in which other forms of mutation screening had not been performed.

Information on clinical phenotype of these patients was collected with the clinical investigator blinded as to deletion extent and gene involved. We used a TSC phenotype data instrument previously designed for this purpose (Dabora et al. 2001), that included information on epilepsy, developmental delay/autism, mental retardation, hypomelanotic macules, facial angiofibroma, Shagreen patch, ungual fibroma, forehead plaque, subependymal nodules and giant cell astrocytomas, tubers, kidney cysts and angiomyolipoma, liver angiomyolipoma, retinal lesions, cardiac rhabdomyoma, and lymphangioleiomyomatosis. Renal cystic

disease was graded by ultrasound examination as: grade 1: 1–2 small cysts, <2 cm diameter; 2: >2 small cysts, <2 cm diameter; 3: >2 cysts, at least 1 >2 cm; and 4: classic PKD.

Multiplex ligation-dependent probe amplification (MLPA)

Multiplex ligation-dependent probe amplification was performed following the directions provided by MRC-Holland, Amsterdam, The Netherlands (www.mlpa.com), using one probe set for *TSC2* (SALSA MLPA kit P046 vs. 04) covering 31 of the 41 exons, and one probe set for *TSC1* (SALSA MLPA kit P124 vs. 01) covering 16 of the 23 exons (Schouten et al. 2002). Briefly, 5 µl genomic DNA (20 ng/µl) was incubated at 98°C for 5 min and then after cooling to room temperature mixed with 1.5 µl of probes mix (containing 1.5 fmol of each probe) and 1.5 µl of SALSA hybridization buffer, denatured at 95°C for 2 min and hybridized at 60°C for 16 h. Hybridized probes were ligated at 54°C for 15 min by addition of 32 µl ligation mixture. Following heat inactivation, 10 µl of ligation reaction was mixed with 30 µl of PCR buffer, heated to 60°C, mixed with 10 µl PCR mixture (Salsa polymerase, dNTPs, and universal primers, one of which was labeled with fluorescein) and subjected to PCR amplification for 30 cycles. All reagents except probe mixes for homemade MLPA sets were from MRC-Holland.

Amplification products were 10× diluted in HiDi formamide (ABI, Foster city, CA, USA) containing 1/16 volume of ROX500 size standard (ABI) and then were separated by size on an ABI 3100 Genetic Analyzer (ABI) (Supplemental Fig. 1). Electropherograms were analyzed by GeneMapper v3.5 (ABI), and peak height data were exported to an Excel table (pilot analysis showed equivalent results using peak area data). Excel programs were generated (available upon request) to transform the peak height data to normalized values, such that control samples gave a value of 1 after normalization. Briefly, peak heights for each probe were divided by the average signal from three or more control probes (located on different chromosomes), and then that value was divided by a similar value calculated from reference samples. We used the average values from four to eight reference samples without deletion in *TSC1*/*TSC2* processed concurrently for each analysis.

Multiplex ligation-dependent probe amplification results for which the SD of normalized signal for all probes was >0.15 were considered uninterpretable. When this occurred, the MLPA analysis was repeated, and when seen repeatedly, such DNA samples were not considered further. Preliminary results showed that the quality of DNA strongly influenced MLPA results so that all samples that showed signs of degradation were removed before screening. Each sample was successfully analyzed for at least one MLPA set for

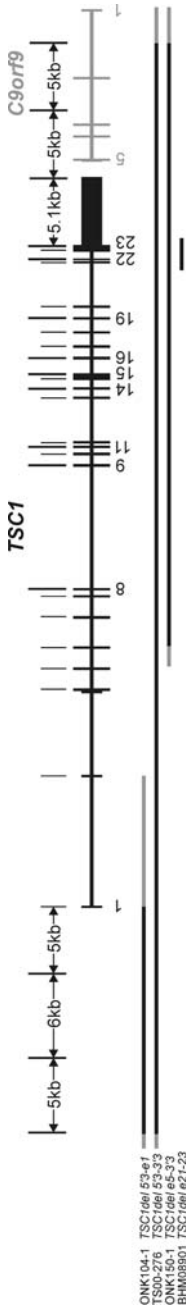

Fig. 1 *TSC1* deletion mutations. Map of *TSC1* and flanking region. Exons are indicated by *vertical lines* with proportional spacing, and numbering of selected exons. *Long thick* and *short thin black lines* indicate the position of homemade and MRC-Holland probes, respectively. Each deletion is represented by a *black horizontal line*, indicating the minimum deleted region. *Gray lines* indicate possible extent to the next undeleted probe for those deletions not confirmed by breakpoint sequencing

each of *TSC1* and *TSC2*; 93% were analyzed successfully for all four probe sets. 2.3, 0, 6.1, and 0.4% of samples were not interpretable for the *TSC1* MRC-Holland, *TSC1* homemade, *TSC2* MRC-Holland and *TSC2* homemade probe sets, respectively.

MLPA assay design

Multiplex ligation-dependent probe amplification probe sets were designed for each of *TSC1* and *TSC2*, to expand the coverage beyond that contained in the MRC-Holland probe sets. The homemade *TSC1* probe set consisted of 18 genomic regions: 7 *TSC1* exons not included in the MRC-Holland set; *TSC1* exons 15 and 23 (duplicate coverage for these relatively large exons); 3 genomic regions extending 16 kb 5′ of *TSC1* exon 1; 3 genomic regions extending 15 kb 3′ of the stop codon of *TSC1*; and 3 genomic regions on different chromosomes to serve as controls (Supplemental Table 1). The homemade *TSC2* probe set consisted of 20 genomic regions: 10 *TSC2* exons not included in the MRC-Holland set; *TSC2* exon 33 (duplicate coverage for a large exon); 3 genomic regions extending 16 kb 5′ of *TSC2* exon 1; 3 genomic regions extending 50 kb 3′ of the stop codon; and 3 control probes (the same as for *TSC1*) (Supplemental Table 2). These homemade probe sets were designed to generate amplification products in the range of 90–156 bp, with 18 and 20 probes for *TSC1* and *TSC2*, respectively. Linker regions in each of these probe sets were designed to generate products differing in size by 3 bp for those of size 90–120 bp, and by 4 bp for those of size >120 bp. One *TSC2* probe (a 3′ flanking probe located 14 kb from *TSC2*) gave variable results, possibly due to genetic variation, and was not considered during analysis.

Each probe was composed of two 5′ and 3′ half-probes, each containing unique target specific sequence, stuffer sequence and universal primer sequences on their 5′ and 3′ ends, respectively (Schouten et al. 2002) (Supplemental Tables 1, 2). All probes were synthesized at 100 nM scale and purified by PAGE (IDT, Skokie, IL, USA). To allow ligation 3′ half-probes were synthesized with 5′ phosphate. Target-specific sequences were designed in areas free of repeat elements (RepeatMasker Bedell et al. 2000) and SNPs (dbSNP and our previous observations Dabora et al. 2001) and their melting temperature was determined by RaW-Probe v. 0.15 b using the Go-Oli-Go algorithm (http://www.mrc-holland.com/). Target sequences were designed to consist of at least 21 nt and have a melting temperature of 71 ± 2°C, when possible (Supplementary Tables 1, 2). The unique nature of each target specific sequence was verified by BLAST comparison (BLASTN algorithm with Expect = 1 and without filtering) with reference to human genome sequence. Stuffer sequences were included in both half-probes such that all half-probes had length <78 nt. Portions of the same M13 sequence (AC# V00604) were used to generate each of the stuffer sequences. Universal primer sequences were identical to those in the commercial MLPA kits (MRC-Holland) (Schouten et al. 2002).

All DNA samples showing evidence for a deletion were confirmed in a second MLPA run using the same primers. In addition, all intragenic and several other deletions were confirmed by LR-PCR and sequencing of the breakpoint junction.

LR-PCR methods and sequence analysis of breakpoints

Selected apparent deletions detected by MLPA were confirmed by LR-PCR using LA PCR 2.1 kit (TaKaRa, Shiga, Japan). PCR reactions used a combination of primers located outside the deleted region under conditions recommended by the manufacturer. PCR products representing deleted alleles were either directly sequenced or gel-purified and then sequenced. For two apparent duplications, primers from the 5′ and 3′ extents of the duplications were similarly used for LR-PCR confirmation and identification of the breakpoint.

Statistical analyses

For apparent mosaic deletions, the relative signal intensities for deleted versus non-deleted probes were compared using the one sided unpaired *t*-test (Statistica StatSoft, Tulsa, OK, USA). $P < 0.001$ was considered significant.

To assess whether the observed distribution of breakpoints within each inter-probe segment across the *TSC2* gene departed significantly from what would be expected if breakpoints had been randomly scattered at uniform density, we used a simulation method implemented in Python (http://www.python.org). Under the null hypothesis of uniform distribution, we simulated the total observed number of breakpoints, scattering them across the entire physical length of the region using a uniform distribution from a pseudo-random number generator (Matsumoto and Nishimura 1998). In each simulation run, we counted the number of simulated breakpoints that fell into each inter-probe region. We performed 100,000 simulations to obtain the empirical distribution of breakpoint counts within each of the inter-probe regions under the null hypothesis.

Using these distributions, we determined the probability of the observed number of breakpoints within each inter-probe region. Since we tested 41 inter-probe intervals, we used the false discovery rate (FDR) method, to control family wise error over these 41 tests, keeping the FDR at or below 0.05 (Benjamini and Hochberg 1995).

To assess potential correlations between clinical features and the extent and/or mosaic nature of the deletion mutations we identified, we used Fisher's exact test for 2×2 contingency table analyses, the Mann–Whitney test for categorical variables, and the Student *t*-test for quantitative variables (such as age of onset and tuber number), using GraphPad Prism v. 4 (GraphPad Software, San Diego, CA, USA).

Results

Deletion screening

We screened 261 TSC patient DNA samples for large genomic deletions and duplications, using two sets of MLPA assays for each of *TSC1* and *TSC2*. One assay for each gene consisted of the commercially-available MRC-Holland probe set. The second assay for each gene was a novel set of MLPA probes (oligonucleotides) that we generated (see Methods for details, Supplemental Tables 1, 2). Together the two assays for each gene provided coverage of all of the exons, and included 6 amplification products derived from adjacent genomic regions, 3 on the 5′ side and 3 on the 3′ side of each gene (Figs. 1, 2).

The MLPA assay products, which were fluorescently labeled, were size-separated on an ABI 3,100 capillary sequencer. Fluorescent output was captured and quantified according to peak size, and this raw data was normalized by comparison to a set of reference samples (Supplemental Figs. 1, 3). Samples with deletions typically had normalized signals that were about 0.50 for several probes, corresponding to the exons present in the deletion. To compare the performance of these 4 MLPA sets, 23 DNA samples were examined using all 4 in parallel. The average signal for non-deleted and deleted probes were very similar for all 4 assays, and was very close to 1 and 0.5, respectively, as expected. However, the SD was higher for the MRC-Holland probe sets (0.09 for all probes and 0.1 for control probes only) than for our homemade probe sets (0.05 for all and control probes only). Considering probe signals consistent with non-mosaic deletion (see below), SDs were again higher for the MRC-Holland probe sets than for our homemade probe sets (0.07 and 0.06 vs. 0.04 and 0.03, respectively). In addition, 4 and 2 MLPA analyses for the MRC-Holland *TSC1* and *TSC2* probe sets, respectively, were not interpretable due to a SD > 0.15; while only 1 homemade probe set analysis was not interpretable. These quantitative results are consistent with our general observation that our homemade probe sets were more robust and easier to interpret than the MRC-Holland probe sets.

All MLPA assays that gave results suggestive of a deletion were repeated at least once for confirmation. Since we used two different probe sets for each gene, this provided additional confirmation since multi-exon deletions were the rule (see further below).

Deletion/duplication results

Fifty-four of 261 (21%) TSC DNA samples screened by MLPA were found to have deletions/duplications in either *TSC1* or *TSC2* (Figs. 1, 2). Only two of the samples showed evidence for a genomic duplication mutation (both in *TSC2*), while all others were deletions. Considering only those samples previously screened for small mutations, 50 of 209 (24%) were found to have deletions or duplications. Fifty (93%) of the mutations were found in *TSC2*; while 4 (7%) deletions were found in *TSC1*.

Considering the 48 *TSC2* deletions, 13 (27%) were intragenic in extent, 13 (27%) had size less than 10 kb, and 4 (8%) were deletions of a just a single exon (1, 10, 12, and 15) (Fig. 2). Twenty-one (44%) extended into the genomic region 5′ to *TSC2*, while 24 (50%) extended into the *PKD1* gene 3′ to *TSC2*. Ten (21%) of deletions extended into both the 5′ and 3′ regions outside of *TSC2*; of which 3 (6%) extended beyond the entire length of the MLPA probes used here, over 104 kb.

Considering the four *TSC1* deletions, only one was intragenic in extent, measuring 2.3 kb (Fig. 1). One extended into the region 3′ to *TSC1*; one extended into the 5′ region of *TSC1*; and one extended beyond the gene in both directions, over 80 kb.

Eight of the DNA samples analyzed showed evidence for mosaicism (seven deletions and one duplication), with average normalized probe signals of 0.69–0.83 for the deletions, and 1.19 for the duplication, for multiple adjacent exons (Fig. 3), in contrast to expected values of 0.5 and 1.5. These findings were repeated at least twice for confirmation and were consistent. For all of these mosaic cases, the difference in signal between probes showing a deletion and those not showing a deletion was very significant, using the *t*-test, with $P < 10^{-3}$. In addition, 6 of 8 mosaic mutations were seen using both homemade and MRC-Holland probe sets, due to the extent of the deletion, and

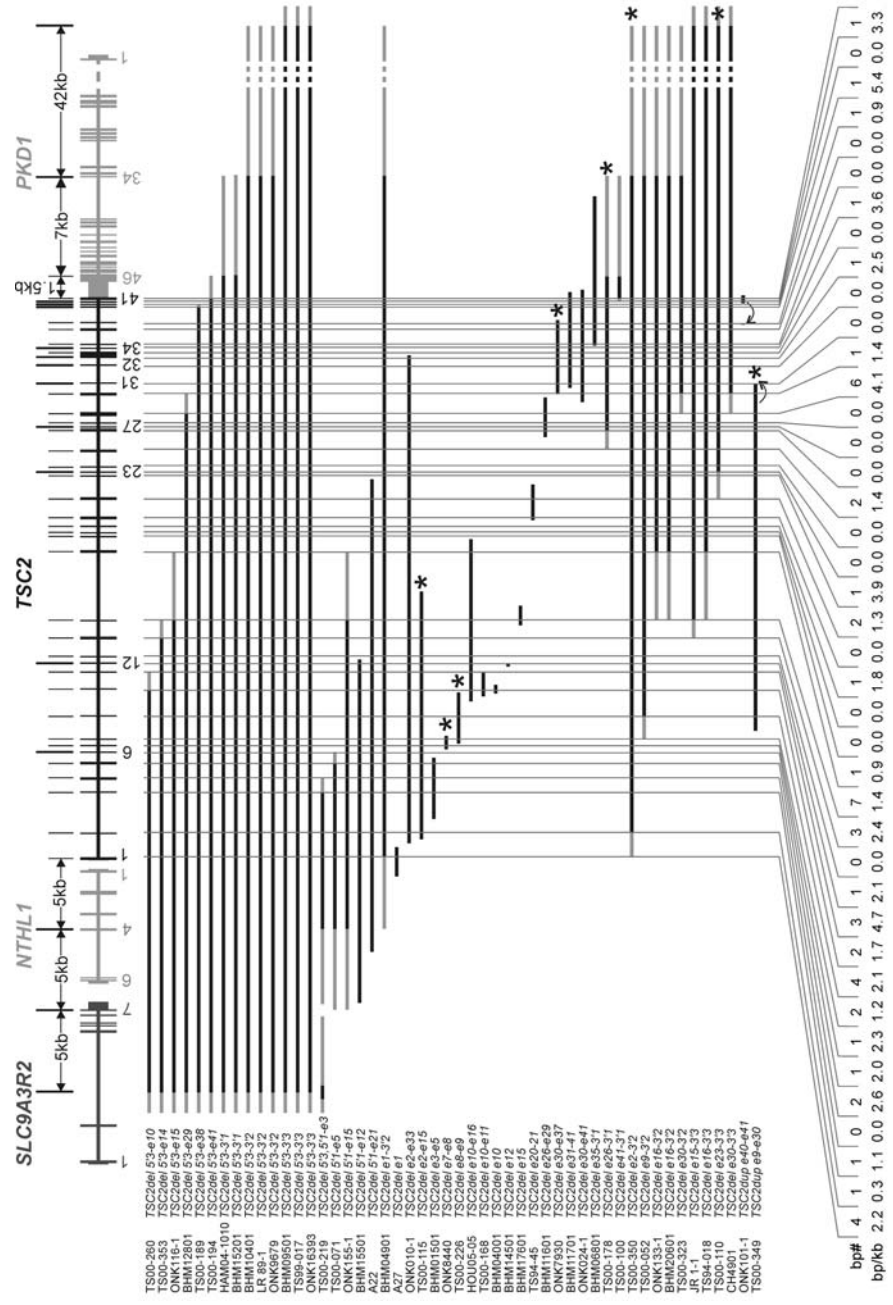

Fig. 2 *TSC2* deletion/duplication mutations. Top, map of *TSC2* and flanking region. Exons are indicated by *vertical lines* with proportional spacing, and numbering of selected exons. *Long thick* and *short thin black lines* indicate the position of homemade and MRC-Holland probes, respectively. Each deletion is represented by a *black horizontal line*, indicating the minimum deleted region. *Gray lines* indicate possible extent to the next undeleted probe for those deletions not confirmed by breakpoint sequencing. Two duplications are shown at the bottom with *black lines* indicating the size of the duplicated region and an *arrow* showing the position of insertion (both duplications were confirmed by LR-PCR/sequencing). Mosaic mutations are indicated by *asterisks*. Bottom, the number of internal *TSC2* deletion breakpoints and their normalized frequency (/kb) in each intron

comparison using the combined data was highly significant, with $P < 10^{-7}$. These values were also significantly different from the expected value of 0.5 (1.5 for duplications), as well as the observed values in heterozygous deletions/duplication ($P < 0.0001$ for both comparisons, by *t*-test). The probe signals seen correspond to mosaicism at 34–62%.

Five DNA samples appeared to have a *TSC2* single exon deletion (two in exon 10, three in exon 34) based upon MLPA analysis with the MRC Holland probe set.

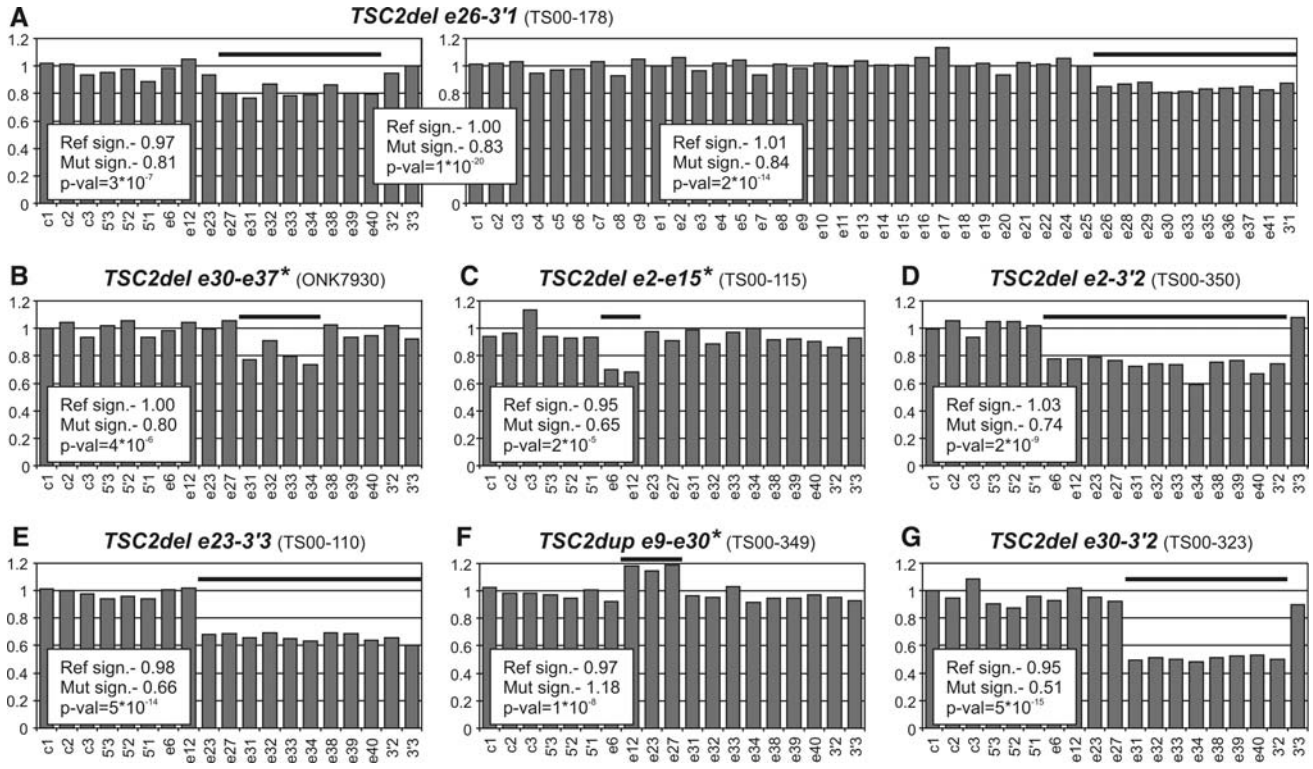

Fig. 3 Mosaic mutations in *TSC2*. Normalized peak height graphs are shown for 7 TSC DNA samples (**a**–**g**). Each *bar* represents the normalized peak height for the probe indicated on the *X*-axis (c1–c9 are controls, e1–e41 are *TSC2* exons, 5′1–5′3 and 3′1–3′3 are 5′ and 3′ flanking probes, respectively). The heavy *black lines* indicate probes with reduced or increased signal. **a** Graphs for both homemade and MRC-Holland *TSC2* probe set results. **b**–**g** Graphs for homemade *TSC2* probe set results. **a**–**f** Graphs show mosaic deletions, while **g** is a non-mosaic deletion. Insets in each graph show the average normalized peak height for non-deleted (*Ref* sign.) and deleted/inserted (*Mut* sign.) probes with *P*-value (*t*-test) for the comparison. For A, statistics for the combined analysis is also shown. *Asterisks* indicate mutations confirmed by LR-PCR/sequencing

Upon sequence analysis of these exons, the samples were found to have small mutations either overlapping the ligation site (a 4 bp deletion, seen twice) or located −2 or −11 nucleotides from the ligation site (two single base substitutions, seen once or twice).

Analysis of breakpoints

To provide definitive confirmation of MLPA results, we identified junction breakpoints for all intragenic deletions of each gene, as well as several additional deletions that did not extend beyond the MLPA probe sets used here (Supplemental Table 3). This was done by long range PCR analysis using primers external to the range of deletion as defined by MLPA. Twenty of 52 (including four mosaic) DNA samples with apparent deletions by MLPA analysis, as well as both duplications (one mosaic), were confirmed by this means.

Breakpoint sequencing demonstrated that there was no apparent pattern or consistency to the sites and sequences involved in these genomic rearrangements. Most strikingly, 3 of 20 deletions were not simple deletions, but included transposition of DNA from other genomic regions or inversion of short DNA sequences nearby the deletion (Fig. 4, Supplemental Table 3). Similar complex deletions have been reported recently in the *CFTR* gene (Ferec et al. 2006 and references therein). In 14 of 20 deletions, the breakpoints occurred at short repeats of 2–7 nucleotides that appeared to have mediated the deletion event. In two other deletions occurring in Alu sequences, exact repeats of 9 and 21 nt were seen. In addition, one duplication also had an overlapping 8 nt element on both ends of the duplicated sequence. Although the *TSC2* gene consists of 24% repetitive sequence, of which 17% is Alu sequence, only 3 of 19 (16%) cloned *TSC2* deletion breakpoints were found to consist of Alu elements, and in one of these the Alu sequences had opposite orientations. These observations indicate that Alu-mediated unequal homologous recombination events represent a minor contribution to *TSC2* deletions.

Analysis of the distribution of the breakpoints, whether cloned or inferred by the precise exon limits from MLPA, was also performed. To determine whether there were regions in *TSC2* that were hotspots

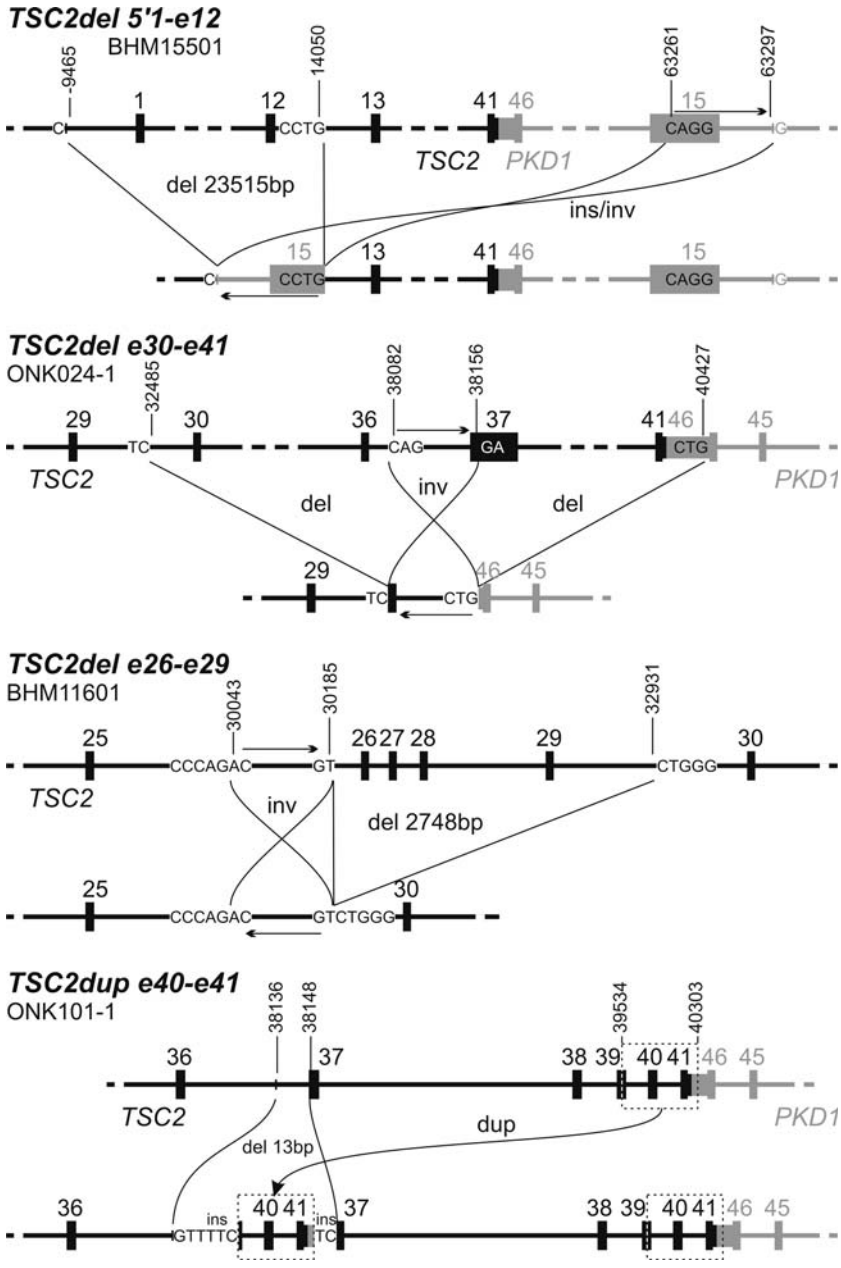

Fig. 4 Complex mutations in *TSC2*. Diagrams of four complex mutations in *TSC2* are shown. Exons and introns are indicated, and nucleotides are numbered from the 1st nucleotide of exon 1. The top sample has a 23,515 bp deletion (exons 1–12 of *TSC2)* with insertion of a 37 bp sequence from *PKD1*. The second sample has a 7,942 bp deletion (exons 30–41 of *TSC2*) with retention of an inverted 74 bp region from the middle of the deleted region. The third has a 2,748 bp deletion (exons 26–29 of *TSC2*), with inversion of the adjacent 143 bp. The fourth is an insertion of 770 bp, exons 39–41 of *TSC2*, into intron 36 of *TSC2*. Sequence elements having a potential role in these complex rearrangements are shown

for deletion events, we compared the distribution of breakpoints among the 41 inter-probe intervals with a random distribution of breakpoints obtained by simulation (Fig. 2). Five inter-probe regions appeared somewhat enriched in breakpoints with nominal $P < 0.05$. However, none of these remained significant after correction for multiple testing by the FDR method (Benjamini and Hochberg 1995). The most significant enrichment for breakpoints was seen in the inter-probe interval containing intron 29, which contained six breakpoints ($P = 0.002$, prior to correction). This intron has length 1.4 kb and contains one AluY and one MLT1B repeat element. We also observed that there was a good correlation between number of breakpoints and inter-probe interval size ($R = 6.6$, $P = 0.000003$). Together these results indicate there is a random distribution of breakpoints in *TSC2*, without a marked hotspot.

This formal analysis could not be performed for *TSC1*, as they were only four intragenic breakpoints, each of which was found in a different inter-probe interval (introns 1, 4, 20, and 3′UTR). Only one of these occurred in *TSC1* intron 20, previously reported as a hotspot for *TSC1* genomic rearrangements (Longa et al. 2001).

Genotype-phenotype correlations

We also examined the potential association between various aspects of clinical phenotype (Dabora et al.

2001) and extent and type of deletion in the *TSC2* gene. First, we compared clinical manifestations (see Methods for details) between patients whose deletion extended into the 5′ region of the *TSC2* gene ($n = 21$) with all others ($n = 33$). There were no significant differences in clinical features between these two groups, while their ages were similar (median age 9 for each subset; $P = 0.89$ by *t*-test). Second, we compared clinical manifestations between patients whose deletion extended into the 3′ region of *TSC2*, involving *PKD1* ($n = 24$), with all others ($n = 30$). Similar to previous observations (Sampson et al. 1997), we found that renal cystic disease was significantly more common in subjects whose deletions extended into the adjacent *PKD1* gene. All 21 *PKD1* deletion patients had kidney cysts, while only 8 of 25 of all other patients had kidney cysts ($P < 10^{-4}$). Furthermore, the average grade of renal cysts in the *PKD1* deletion group was 2.9, in comparison to 1 (indicating minimal involvement, see Methods) for the non-*PKD1* deletion group ($P = 0.0008$); and renal function was below normal in 4 of 14 (29%) of the former compared with 0 of 15 in the latter subgroup ($P = 0.04$). In addition, the *PKD1* deletion subgroup of patients were somewhat younger than the non-*PKD1* subgroup (median ages 9 vs. 11, respectively; $P = 0.10$ by *t*-test).

The eight patients who had mosaic deletions/duplications consisted of seven women and one man ($P = 0.07$), and had somewhat milder TSC manifestations than the other subjects. Four of 7 (57%) mosaic subjects had epilepsy compared to 41 of 45 (91%) non-mosaic subjects ($P = 0.04$). Other comparisons were not statistically significantly different, at least partially due to the small number of mosaic cases available. To compensate for this lack of power, we used the paired *t*-test to compare each of the clinical features for these two sets of subjects. In this analysis, mosaic patients had milder overall clinical severity than non-mosaic patients at $P = 0.007$. In addition, the eight mosaic patients were somewhat older than non-mosaic patients (average age 15.1 vs. 10.2; median ages 9 vs. 9; for mosaic versus non-mosaic subgroups respectively, $P = 0.28$ by *t*-test).

Discussion

We have studied 261 patients, the majority of which (209) had been previously comprehensively screened for small mutations within *TSC1* and *TSC2*. Because small mutations are detected in ~75% of TSC patients (Dabora et al. 2001; Jones et al. 1999; Sancak et al. 2005), we have screened the equivalent of 888 (=209 × 4 + 52) TSC patients for deletions and duplications using MLPA. We detected 54 genomic mutations in these patients, indicating that mutations detectable by MLPA account for 6.1% of all mutations seen in TSC patients, and that 5.6% of these genomic mutations occur in *TSC2* and 0.5% occur in *TSC1*. This current collection of 54 genomic mutations in *TSC1/TSC2* is by far the largest number of such mutations ever reported.

Similar to multiple previous reports on other human disease genes (Schouten et al. 2002; Wimmer et al. 2006), we found that MLPA is an efficient and simple technique for analysis of genomic deletions and duplications. Included in this set of patient DNA samples analyzed were six samples in which genomic disruptions had been identified previously, either by Southern blotting (Au et al. 1997) or LR-PCR (Dabora et al. 2000). All six of these mutations were detected in a blinded fashion by MLPA, confirming that it is a highly sensitive technique. In addition, using our combination of MLPA probe sets, the precise extent of deletions/duplications were delineated, enabling relatively facile LR-PCR analysis for definitive confirmation through sequencing of breakpoints. The precision afforded by this technique is illustrated by the fact that a TSC sample previously thought to consist of a deletion (identified by LR-PCR) was characterized as a partial duplication of *TSC2* in which the duplicated truncated portion was inserted such that an apparent deletion was detected by LR-PCR (Fig. 4).

Nonetheless, it is important to recognize that MLPA has limitations. First, in our experience, the technique is relatively sensitive to DNA quality. We observed in preliminary experiments that DNA degradation enhanced variation in quantitative MLPA results, making them difficult to interpret. The non-random distribution of DNA strand breaks in degraded DNA likely contributes to this observation. Salts or other contaminants may also influence the thermodynamics of DNA melting and/or reanealing in a sequence-sensitive manner (see also http://www.mrc-holland.com/pages/support_faq2pag.html#ANCHOR_Txt6). Second, MLPA will not detect all forms of genomic rearrangements that disrupt the function of (e.g.,) the *TSC2* gene. Inversions, and genomic deletions in which the deleted fragment is retained and inserted elsewhere in the genome, are but two of the genomic rearrangements, which occur in the human genome to cause disease, but will not be detected in general by MLPA. However, the frequency of these more complex genomic events appears to be relatively low in comparison to the more common simple genomic deletions that are detected handily by MLPA (Sampson et al. 1997).

Mutations in *TSC2* are ~four times more common than mutations in *TSC1*, for reasons that are far from clear. The coding region of *TSC2* is about 50% larger than *TSC1* and the number of exons is nearly doubled; the number of nonsense mutations and small indel mutations in the two genes are roughly proportional to this size difference. However, similar to missense mutations and splicing point mutations, large deletions are much less common in *TSC1* than *TSC2*. Our analysis of both the distribution of breakpoints within *TSC2*, and the precise breakpoint sequences, fails to provide a clear reason for the relatively high rate of genomic mutations that occur in *TSC2*. It is clear that, in contrast to *BRCA1* and several other genes (Mazoyer 2005; Shaw and Lupski 2004), Alu-mediated or other repetitive sequence-mediated recombination does not explain the occurrence of the vast majority of deletion mutations in *TSC2*. Very short sequence repeats (2–7 nucleotides) were present at 70% of the junctions of the deleted segments, consistent with a strand slippage mechanism (Albertini et al. 1982) for these deletions (Supplemental Table 3), and arguing against a meiotic exchange mechanism. In addition, the eight mosaic mutations could not occurred by meiotic exchange. It is possible that the polypurine/polypyrimidine tract of *PKD1* intron 21, located 20 kb downstream of the *TSC2* gene, may play a role in these events (Patel et al. 2004), but further study is required. *TSC2* has a much higher GC content than *TSC1* (60% vs. 43%), but *TSC1* contains many more repeat elements than *TSC2* (32% vs. 25% total sequence). *TSC2* also contains a higher frequency of sequence elements thought to be involved in recombination events (Myers et al. 2005): CCTCCCT 0.85 copies/kb vs. 0.24 copies/kb, and CCCCACCCC 0.32 copies/kb vs. 0.04 copies/kb, in *TSC2* vs. *TSC1*, respectively, which may contribute to this difference in deletion frequency.

Similar to previous studies (Brook-Carter et al. 1994; Martignoni et al. 2002; Sampson et al. 1997), we found a strong association between the presence of renal cysts in these TSC patients, with involvement of the *PKD1* gene in the deletion event. Although kidney cysts were also common (32% incidence) in patients whose deletions did not involve *PKD1*, none of those subjects had renal function impairment, and their level of renal cystic disease was minimal with only 1–2 cysts of size < 2 cm. These observations confirm that severe, progressive polycystic kidney disease is associated with combined deletion of *TSC2* and *PKD1*, consistent with an important role of each gene in the pathogenesis of renal cyst development. In contrast, patients with other types of *TSC2* deletion mutations (as well as point mutations in *TSC2* Dabora et al. 2001)] do not have clinically significant polycystic kidney disease.

There were no unusual clinical manifestations or apparent worsening of phenotype associated with extension of *TSC2* deletions into the adjacent *NTHL1* and *SLC9A3R2* genes, suggesting that those genes are haplo-sufficient. This observation contrasts with previous hypothesis (Imai et al. 1998). Although our comparison had limited power ($n = 21$ and 33 for the 5′ deletion and non 5′ deletion groups, respectively), there was no hint of a clinical difference between these two groups.

Somatic mosaicism for deletion/duplication mutations was observed in 8 of 54 (15%) of patients. Their detection was facilitated by the precision afforded by MLPA analysis, and the fact that multi-exon deletions (96% of those identified here) provide the opportunity for assessment of the relative signal intensity for multiple deleted exons. This 15% frequency of mosaicism is higher than that reported in general for mutational analyses of TSC patients, but is somewhat lower than what was seen in a previous analysis of genomic mutations affecting both *TSC2* and *PKD1*, 7 of 27 (26%) (Sampson et al. 1997). However, that study extended the analysis for mosaicism to the first affected individual in each family, increasing their detection rate, an approach not possible with this patient population. It is likely that the frequency of mosaicism is under-recognized for *TSC1* and *TSC2* point mutations and small indels, due to the limited methods available for quantitative detection. Similarly, some of the DNA samples analyzed here showed hints of mosaicism for deletion mutations at a relatively low level (<20%). However, these hints could not be confirmed statistically, and thus are not reported. We also noted that patients with mosaicism had a somewhat milder clinical phenotype than those without, although the small number of mosaic subjects limited the power of this comparison. These observations in aggregate provide further evidence for the hypothesis that somatic mosaicism accounts significantly for the fraction of TSC patients who have no detectible mutation after the usual screening procedures (Dabora et al. 2001).

We found that generation of 18–20-plex MLPA probe sets by direct synthesis of oligonucleotides was relatively simple, following a straightforward algorithm for selection of target sequence. Only one of 32 probes designed in this way failed to work the first time. The spacing of the size of amplification products was also straightforward and effective, without need for modification. This was likely partially due to the inclusion of universal sequence for the stuffer regions, minimizing variation in amplification product migration characteristics. Moreover, we found that the reproducibility of our homemade 18–20-plex probe sets as assessed both qualitatively and quantitatively, as presented above, was superior to that

of the 40-plex MRC-Holland probe sets. Since the size range of the products of our probe sets was much narrower than that of the MRC-Holland probe sets (90–156 bp vs. 130–481 bp, respectively), we suspect that this is due to greater variation in the performance of the PCR reaction when the PCR products have a greater size range. In addition, this oligonucleotide synthesis approach is much more rapid and efficient than generation of probes using M13 (Schouten et al. 2002).

One single exon deletion (exon 12) was detected by our homemade probe set for *TSC2* but not the MRC-Holland *TSC2* probe set. We found that all single exon deletions identified by MLPA must be confirmed by other means to identify false positive findings due to variation within the probe binding regions. This is relatively easily done for internal exons, but is more difficult for exons at the 5′ or 3′ end of these genes. The addition of three probes to each end of each gene in our homemade sets provided the ability to both confirm deletions of the first and last exons, and also to delineate the extent of such deletions.

Acknowledgments We thank all of the patients and clinicians who contributed DNA samples and clinical information; Mei Lin and Dawn Ciulla for assistance with capillary electrophoresis and peak height determinations; NIH NS31535 for support.

References

Albertini AM, Hofer M, Calos MP, Miller JH (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29:319–328

Au KS, Rodriguez JA, Rodriguez E Jr, Dobyns WB, Delgado MR, Northrup H (1997) Mutations and polymorphisms in the tuberous sclerosis complex gene on chromosome 16. Hum Mutat 9:23–29

Bedell JA, Korf I, Gish W (2000) MaskerAid: a performance enhancement to RepeatMasker. Bioinformatics 16:1040–1041

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Brook-Carter PT, Peral B, Ward CJ, Thompson P, Hughes J, Maheshwar MM, Nellist M, Gamble V, Harris PC, Sampson JR (1994) Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease- a contiguous gene syndrome. Nat Genet 8:328–332

Consortium TEcts (1993) Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75:1305–1315

Curatolo P (2003) Tuberous sclerosis complex: from basic science to clinical phenotypes. Mac Keith Press, Cambridge

Dabora SL, Jozwiak S, Franz DN, Roberts PS, Nieto A, Chung J, Choy YS, Reeve MP, Thiele E, Egelhoff JC, Kasprzyk-Obara J, Domanska-Pakiela D, Kwiatkowski DJ (2001) Mutational analysis in a cohort of 224 tuberous sclerosis patients indicates increased severity of TSC2, compared with TSC1, disease in multiple organs. Am J Hum Genet 68:64–80

Dabora SL, Nieto AA, Franz D, Jozwiak S, Van Den Ouweland A, Kwiatkowski DJ (2000) Characterisation of six large deletions in TSC2 identified using long range PCR suggests diverse mechanisms including alu mediated recombination. J Med Genet 37:877–883

Ferec C, Casals T, Chuzhanova N, Macek M Jr, Bienvenu T, Holubova A, King C, McDevitt T, Castellani C, Farrell PM, Sheridan M, Pantaleo SJ, Loumi O, Messaoud T, Cuppens H, Torricelli F, Cutting GR, Williamson R, Ramos MJ, Pignatti PF, Raguenes O, Cooper DN, Audrezet MP, Chen JM (2006) Gross genomic rearrangements involving deletions in the CFTR gene: characterization of six new events from a large cohort of hitherto unidentified cystic fibrosis chromosomes and meta-analysis of the underlying mechanisms. Eur J Hum Genet 14:567–576

Gomez M, Sampson J, Whittemore V (1999) The tuberous sclerosis complex, 3rd edn. Oxford University Press, Oxford, UK

Imai K, Sarker AH, Akiyama K, Ikeda S, Yao M, Tsutsui K, Shohmori T, Seki S (1998) Genomic structure and sequence of a human homologue (NTHL1/NTH1) of Escherichia coli endonuclease III with those of the adjacent parts of TSC2 and SLC9A3R2 genes. Gene 222:287–295

Jones AC, Shyamsundar MM, Thomas MW, Maynard J, Idziaszczyk S, Tomkins S, Sampson JR, Cheadle JP (1999) Comprehensive mutation analysis of TSC1 and TSC2-and phenotypic correlations in 150 families with tuberous sclerosis. Am J Hum Genet 64:1305–1315

Kwiatkowski DJ (2004) TSC1 and TSC2 and tuberous sclerosis. In: Epstein CJ, Erickson RP, Wynshaw-Boris A (eds) Inborn errors of development. Oxford University Press, New York, pp 910–919

Langkau N, Martin N, Brandt R, Zugge K, Quast S, Wiegele G, Jauch A, Rehm M, Kuhl A, Mack-Vetter M, Zimmerhackl B, Janssen B (2002) TSC1 and TSC2 mutations in tuberous sclerosis, the associated phenotypes and a model to explain observed TSC1/ TSC2 frequency ratios. Eur J Pediatr 161:393–402

Longa L, Saluto A, Brusco A, Polidoro S, Padovan S, Allavena A, Carbonara C, Grosso E, Migone N (2001) TSC1 and TSC2 deletions differ in size, preference for recombinatorial sequences, and location within the gene. Hum Genet 108:156–166

Martignoni G, Bonetti F, Pea M, Tardanico R, Brunelli M, Eble JN (2002) Renal disease in adults with TSC2/PKD1 contiguous gene syndrome. Am J Surg Pathol 26:198–205

Matsumoto M, Nishimura T (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans Model Comput Simul 8:3–30

Mazoyer S (2005) Genomic rearrangements in the BRCA1 and BRCA2 genes. Hum Mutat 25:415–422

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321–324

Nellist M, Sancak O, Goedbloed MA, van Veghel-Plandsoen M, Maat-Kievit A, Lindhout D, Eussen BH, de Klein A, Halley DJ, van den Ouweland AM (2005) Large deletion at the TSC1 locus in a family with tuberous sclerosis complex. Genet Test 9:226–230

Osborne JP, Fryer A, Webb D (1991) Epidemiology of tuberous sclerosis. Ann N Y Acad Sci 615:125–127

Patel HP, Lu L, Blaszak RT, Bissler JJ (2004) PKD1 intron 21: triplex DNA formation and effect on replication. Nucleic Acids Res 32:1460–1468

Sampson JR, Maheshwar MM, Aspinwall R, Thompson P, Cheadle JP, Ravine D, Roy S, Haan E, Bernstein J, Harris PC (1997) Renal cystic disease in tuberous sclerosis: role of the polycystic kidney disease 1 gene. Am J Hum Genet 61:843–851

Sancak O, Nellist M, Goedbloed M, Elfferich P, Wouters C, Maat-Kievit A, Zonnenberg B, Verhoef S, Halley D, van den Ouweland A (2005) Mutational analysis of the TSC1 and TSC2 genes in a diagnostic setting: genotype—phenotype correlations and comparison of diagnostic DNA techniques in tuberous sclerosis complex. Eur J Hum Genet 13:731–741

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57

Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13(Spec No 1):R57–R64

van Slegtenhorst M, de Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, Lindhout D, van den Ouweland A, Halley D, Young J, Burley M, Jeremiah S, Woodward K, Nahmias J, Fox M, Ekong R, Osborne J, Wolfe J, Povey S, Snell RG, Cheadle JP, Jones AC, Tachataki M, Ravine D, Kwiatkowski DJ (1997) Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. Science 277:805–808

Wimmer K, Yao S, Claes K, Kehrer-Sawatzki H, Tinschert S, De Raedt T, Legius E, Callens T, Beiglbock H, Maertens O, Messiaen L (2006) Spectrum of single and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. Genes Chromosomes Cancer 45:265–276