Int J Legal Med (1999) 112 : 355–359

ORIGINAL ARTICLE

C. Seidl · S. Müller · O. Jäger · E. Seifried

Sequence analysis and population data of short tandem repeat polymorphisms at loci D8S639 and D11S488

Received: 18 September 1997 / Received in revised form: 23 March 1998

Abstract Short tandem repeat loci are ideal markers for forensic and paternity case work. A high degree of polymorphism, as determined by gross length measurement, is very often due to complex underlying sequence variation. In the present study, we have studied the sequence structure and population genetics of two short tandem repeat polymorphisms at loci D8S639 and D11S488 in German Caucasians from the region of Hesse. Sequence data revealed a considerable polymorphism at both loci. Locus D8S639 is characterized by a tetranucleotide (AGAT)n repeat pattern with (GAT) and (AGGT) repeats dispersed throughout several alleles. These microvariations lead to alleles differing by one base pair or alleles of identical size. At locus D8S639 we observed 17 allelic lengths comprising 25 different alleles. Alleles at locus D11S488 possess a compound repeat region consisting of (AAAG)n and (GAAG)n repeats. At locus D11S488 we observed 15 allelic lengths with a total of 24 alleles. Allelic lengths increased in size by 4bp increments corresponding to the addition of one tetranucleotide repeat unit. Population data of loci D8S639 and D11S488 revealed a high polymorphism with heterozygosity rates of 0.85 (D8S639) and 0.91 (D11S488).

Key words Short tandem repeat polymorphism · D8S639 · D11S488 · Sequence analysis · Population genetics

Introduction

The repeat structure of alleles at short tandem repeat (STR) loci varies due to the length of the individual repeat units, the number of repeat units and the repeat unit pattern of the individual alleles (Puers et al. 1993; Urquhart et al. 1993; Möller and Brinkmann 1994). Simple repeats contain units of identical length and sequence, compound repeats comprise two or more adjacent simple repeats, complex repeats may contain several repeat blocks of variable length, along with variable intervening sequences (Urquhart et al. 1994). Very often a high polymorphism is associated with considerable sequence variations (Adams et al. 1993; Möller and Brinkmann 1994; Brinkmann et al. 1996; Lareu et al. 1996). Although most STR loci are characterized by a low mutation rate (Weber and Wong 1993), some loci exhibit hypermutability (Talbot et al. 1995) and microsatellite instability is found in a variety of neoplasmas (Perucho 1996; Boland 1996) or has been associated with some human genetic disorders (Mahadevan et al. 1992).

Due to the structural variations of particular STR loci, sequence analysis of individual alleles is essential for the use of these markers for human identification. In the present study, we have therefore analysed the sequence structure and population genetics of STR loci on chromosome 8, locus D8S639 at 8p21-p11 (Nelson et al. 1994) and on chromosome 11, locus D11S488 at 11q24.1-q25 (Browne et al. 1992). Alleles at both loci were originally reported to vary in size by multiples of 4 bp.

Materials and methods

Population sample

Peripheral blood samples were collected in EDTA from healthy unrelated blood donors (locus D8S639 n = 306 and locus D11S488 n = 186) from the region of Hesse (Mid-Germany) and genomic DNA was isolated following the salting-out procedure (Miller et al. 1988).

PCR conditions

Primer sequences for locus D8S639 (8p21-p11) (Nelson et al. 1994) and locus D11S488 (11q24.1.25) (Browne et al. 1992) were used as previously described. All oligonucleotides were commer-

The data presented here are part of the theses of S. Müller and O. Jäger

C. Seidl · S. Müller · O. Jäger · E. Seifried (✉)
Institut für Transfusionsmedizin und Immunhämatologie,
Blutspendedienst Hessen des DRK, Sandhofstraße 1,
D-60528 Frankfurt am Main, Germany
Tel. +49-69-6782 201; Fax +49-69-6782 217

cially synthesized and the 5′ primers were labeled with fluorescence dye markers (Perkin Elmer, Applied Biosystem Division, Weiterstadt, Germany), locus D8S639 with hexachloro-6-carboxyfluorescein (HEX) and locus D11S488 with 6-carboxylfluorescein (FAM). PCR amplification was performed in a 9600 DNA thermal cycler system (Perkin Elmer, Forster City, Calif., USA) using 50 ng of genomic DNA, 25 pmol of each primer, 200nmol dNTPs, 1 × PCR buffer and 0.5 U Taq polymerase (Appligene Oncor, Heidelberg, Germany) in a final volume of 50 µl. PCR cycling conditions were: (D8S639) 28 cycles at 95° C for 30 s, 60° C for 40 s, 72° C for 30 s; (D11S488) 28 cycles at 95° C for 30 s, 54° C for 40 s, 72° C for 30 s, with a final extension step at 72° C for 5 min. PCR products were analysed using an allelic ladder on a 6% polyacrylamide denaturing sequencing gel containing 8.3 M urea, 6% acrylamide/bis-solution and 1 × tris/borate buffer. Electrophoresis was performed at 30 W for 8 h on an 373A automated DNA sequencer (Applied Biosystem Division, Perkin Elmer, Foster City, Calif., USA) and fragments were analysed using the Genescan 672 software (Applied Biosystem Division, Perkin Elmer).

Solid phase DNA sequence analysis

Sequence analysis was performed in both directions using the PCR primers as sequencing primers. PCR products from homozygous individuals were routinely sequenced. For heterozygous individuals, the PCR products were separated by electrophoresis on 3% agarose gels (Agarose NA, Pharmacia Biotech, Uppsala, Schweden) and purified using Qiaex II gel extraction kit (Qiagen Inc, Chatsworth, Calif., USA). Sequence reactions were set up according to standard protocols (Prism Sequenase DyeDeoxy Terminator Sequencing Kit, Perkin Elmer). Sequence analysis was performed on an 373A automated DNA sequencer using the SeqEditor and Sequence Navigator software (Applied Biosystems Division, Perkin Elmer). Alleles were assigned according to the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (Bär et al. 1997).

Statistical analysis

The polymorphic information content (PIC) was calculated using the formula of Botstein et al. (1980). The matching probability (pM) was calculated by the method of Jones (1972). The expected frequency of heterozygotes was calculated according to Nei and Roychoudhury (1974). Hardy-Weinberg equilibrium was tested using the χ^2-goodness of fit test (Kirby and Puterman 1992).

Results

Sequence structure and variations of D8S639 and D11S488 alleles

At locus D8S639 a total of 66 fragments were sequenced and 17 allelic lengths corresponding to 25 different allelic sequences were found ranging from 316 bp to 371 bp (Table 1). In contrast to the published Genebank sequence of D8S639 (Nelson et al. 1994; Accession number L24797) most of the alleles, with the exception of allele 233, had two (AGAT) repeats between the (GAT) and (AGGT) repeat. The repeat region at locus D8S639 is characterized by a considerable microheterogeneity which resembles complex STR systems. Small alleles from allele 20 to allele 27 exhibit variations in the number of (AGAT) repeats at both ends of the repeat region. The complexity of

Table 1 Sequence variabilities at locus D8S639

Allele		Length (bp)	Repeat region	No. of sequenced alleles
20		316	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)6	1
21		320	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)6	1
22		324	(AGAT)11 GAT (AGAT)2 AGGT (AGTA)9	4
23	23_1	328	(AGAT)11 GAT (AGAT)2 AGGT (AGAT)10	5
	23_2	328	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)9	3
	23_3	328	(AGAT)11 GAT (AGAT)1 AGGT (AGAT)11	4
24	24_1	332	(AGAT)11 GAT (AGAT)2 AGGT (AGAT)11	4
	24_2	332	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)10	3
25	25_1	336	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)10	5
	25_2	336	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)11	4
26	26_1	340	(AGAT)11 GAT (AGAT)2 AGGT (AGAT)13	3
	26_2	340	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)12	2
	26_3	340	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)11	3
27	27_1	344	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)12	3
	27_2	344	(AGAT)14 GAT (AGAT)2 AGGT (AGAT)11	2
27.3		347	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)9	1
28		348	(AGAT)14 GAT (AGAT)2 AGGT (AGAT)12	3
28.3	28.3	351	(AGAT)11 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)11	1
	28.3_2	351	(AGAT)12 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)10	2
29		352	(AGAT)14 GAT (AGAT)2 AGGT (AGAT)13	3
29.3		355	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)10	2
30.3		359	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)11	3
31.2		362	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)5 GAT (AGAT)6	1
31.3		363	(AGAT)13 GAT (AGAT)2 AGGT (AGAT)4 GAT (AGAT)12	2
33.3		371	(AGAT)16 GAT (AGTA)2 AGGT (AGAT)4 GAT (AGAT)11	1

b*: GAGTGATGGAAGAAAACAAGTAGC TAGCTGAAGATTATTCAGAGAATAAAAAAGCAGAGTATAAAG TCCAGCAGTGCACTTGCAGTATTGTTAACTGAAGTATTCTAATTATCCTAGTGGCTTGATTTGGACAT TGGGATTCTAGATTTTTCCTCCTATCTCTGCTTTTGTCAT**(AGAT)nGAT(AGAT)n(AGGT)1(AGAT)n**A TGATTTGTCATT TTTCTAAGTTCTTGTGTTT CCTGACTTTACATTTTTTGGTTGAG

* Sequence of repeat region (bold) and flanking regions including primer sequences are underlined

Table 2 Sequence variabilities at locus D11S488

Allele		Length (bp)	Repeat region	No. of sequenced alleles
26		242	(AAAG)14 (GAAG)4 GAAAG (GAAG)4 GAAAGAAAGGAAAG (GAAG)4	1
27		246	(AAAG)15 (GAAG)4 GAAAG (GAAG)4 GAAAGAAAGGAAAG (GAAG)4	1
28		250	(AAAG)12 (GAAG)7 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)5	3
29	29_1	254	(AAAG)7 (GAAG)11 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	4
	29_2	254	(AAAG)8 (GAAG)10 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	2
30	30_1	258	(AAAG)7 (GAAG)12 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
	30_2	258	(AAAG)14 (GAAG)7 GAAAG (GAAG)4 GAAAGAAAGGAAAG (GAAG)5	2
31	31_1	262	(AAAG)7 (GAAG)13 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	5
	31_2	262	(AAAG)10 (GAAG)11 GAAAG (GAA)5 GAAAGAAAGGAAAG (GAAG)5	3
32	32_1	266	(AAAG)7 (GAAG)14 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
	32_2	266	(AAAG)11 (GAAG)10 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	2
	32_3	266	(AAAG)10 (GAAG)11 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	2
	32_4	266	(AAAG)12 (GAAG)9 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
	32_5	266	(AAAG)15 (GAAG)7 GAAAG (GAAG)4 GAAAGAAAGGAAAG (GAAG)6	2
	32_6	266	(AAAG)15 (GAAG)6 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	2
33	33_1	270	(AAAG)13 (GAAG)9 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
	33_2	270	(AAAG)12 (GAAG)10 GAAAG (GAA)5 GAAAGAAAGGAAAG (GAAG)6	3
34	34_1	274	(AAAG)14 (GAAG)9 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
	34_2	274	(AAAG)13 (GAAG)10 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
35		278	(AAAG)15 (GAAG)9 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
36		282	(AAAG)15 (GAAG)10 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
38		290	(AAAG)16 (GAAG)11 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	4
39		294	(AAAG)13 (GAAG)15 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	3
41		302	(AAAG)17 (GAAG)13 GAAAG (GAAG)5 GAAAGAAAGGAAAG (GAAG)6	1

b*: AGGCAATAGAGACCCTGTG AAAG AAGA (AAAG)5 GTAAGAA **(AAAG)n (GAAG)n GAAAG (GAAG)n** GAAAG AAAG GAAAG **(GAAG)n** GAAAG (AAAG)2 AGAGAGGGAG (AAAG)2 GAAG GAAACTAA CTCTCAGTAACAATTCATCATC

* Sequence of repeat region (bold) and flanking regions including primer sequences are underlined

the system increases in larger alleles with variations in the number of the (GAT) and (AGAT) repeats. Several of the larger alleles have additional (GAT) repeats giving rise to alleles that differ by only 1 bp in size. In these cases, the (GAT) trinucleotide repeats are included in the allele assignment.

Locus D11S488 is characterized by a compound repeat structure consisting of (AAAG) and (GAAG) repeats (Table 2). This locus has a basic sequence structure of $(AAAG)_{7–17}(GAAG)_{4–15}GAAAG(GAAG)_{4–5}GAAAGAA-AGGAAAG(GAAG)_{4–6}$. Sequence analysis was performed on 60 fragments giving a total of 15 allelic lengths that comprise 24 alleles. The complexity of this system is located mainly in the 5′ region of the repeat array with variations in the number of $(AAAG)_n$–$(GAAG)_n$ repeats. In contrast, the 3′ end of the repeat region is characterized by a tetra/pentameric repeat array and exhibits low variability with 4–6 copies of a (GAAG) repeat.

Population genetics

There was no deviation from Hardy-Weinberg equilibrium (D8S639, $\chi^2 = 8.23$, df = 55, $P > 0.9$; D11S488, $\chi^2 = 6.58$, df = 62, $P > 0.9$). Locus D8S639 exhibits a unimodal distribution of allele frequencies whereas locus D11S488 is characterized by a bimodal distribution of alleles (Table 3). Both STR systems have a considerable polymorphism with PIC values of 0.821 (D8S639) and 0.888 (D11S488) and heterozygosity rates of 0.85 (D8S639) and 0.91 (D11S488).

Discussion

In the present study we observed a considerable sequence variation among alleles at locus D8S639 and locus D11S488. Originally, both STR systems were described as regular tetranucleotide repeat polymorphisms (Browne et al. 1992; Nelson et al. 1994). Sequence analysis revealed that the loci D8S639 and D11S488 are both characterized by alleles with sequence variations in the repeat regions but identical length when analysed by PCR. Allelic lengths at locus D11S488 differ by 4 bp, whereas at locus D8S639 we observed a more complex sequence structure of alleles giving rise to alleles that differ only by 1 bp in size. Population data revealed that both loci are very polymorphic and display a considerable number of alleles. Due to the presence of (AAAG) and (GAAG) repeats, the STR system at locus D11S488 could be classified as a compound STR, such as locus D11S554 (Adams et al. 1993) or locus D12S391 (Lareu et al. 1996). In contrast the STR system at locus D8S639 has intermediate microvariations with similiarities to complex STR sys-

Table 3 Allele frequencies and additional population genetic parameters at STR locus D8S639 and D11S488

D8S639 N = 306 (612)		**D11S488** N = 186 (372)	
Allele (bp[1])	Frequency	Allele (bp[1])	Frequency
20 (316)	0.008	26 (242)	0.003
21 (320)	0.003	27 (246)	0.005
22 (324)	0.023	28 (250)	0.038
23 (328)	0.111	29 (254)	0.099
24 (332)	0.152	30 (258)	0.091
25 (336)	0.274	31 (262)	0.081
26 (340)	0.224	32 (266)	0.156
27 (344)	0.087	33 (270)	0.156
27.3 (347)	0.008	34 (274)	0.129
28 (348)	0.039	35 (278)	0.075
28.3 (351)	0.005	36 (282)	0.040
29 (352)	0.016	37 (286)	0.040
29.3 (355)	0.010	38 (290)	0.048
30.3 (359)	0.028	39 (294)	0.027
31.2 (362)	0.003	41 (302)	0.005
31.3 (363)	0.008		
33.3 (371)	0.002		
observed HR	0.849		0.905
expected HR	0.829		0.836
PIC	0.821		0.888
pM	0.057		0.023
MEC	0.636		0.785

[1] Length in base pairs as determined by sequence analysis
HR = Heterozygosity rate
PIC = Polymorphic information content
pM = Probability of match
MEC = Mean exclusion chance

tems, such as locus HumACTBP2 (Urquhart et al. 1993; Möller and Brinkmann 1994).

Slipped-strand mispairing during DNA replication has been proposed to be a major mechanism for these length alterations (Levinson and Goodman 1987; Schlotterer and Tautz 1991; Strand et al. 1993; Richards and Sutherland 1994). Furthermore, mutations resulting in insertions and/or deletion of single nucleotides or transitions/transversions can cause di- or trinucleotides that are dispersed throughout repeated sequences (Stallings et al. 1991; Urquhart et al. 1993). The (GAT) repeat at locus D8S639 is probably the result of such a single base pair deletion whereas the (AGGT) repeat could have arisen by A > G transition. The (AGAT) expansion in relation to the (GAT) repeat can be followed through several allelic lengths resulting in a potential 'branching effect' in the generation of the length polymorphism at this STR locus. However, in contrast to other STR systems, i.e. locus D11S554 (Adams et al. 1993) or locus D21S11 (Brinkmann et al. 1996), alleles at locus D8S639 do not reveal a clear pattern of sequence families.

Locus D11S488 in contrast is characterized by allelic lengths differing by multiples of 4 bp. The polymorphism of this STR system predominantly derives from variations in the number of (AAAG) and (GAAG) repeats. Furthermore, this STR system exhibits a considerable number of alleles with sequence variations of identical fragment sizes (i.e. allele 32 consists of six sequence variations). Repeat slippage during DNA replication could be a potential cause for these sequence variations at this locus.

The presence of 'iso'-alleles or alleles differing only by single base pairs in length makes sizing of these STR systems important. In the present study using denaturing conditions and an allelic ladder, we were able to define the 1 bp differences observed at locus D8S639. The use of non-denaturing gels could also be useful, since 'iso'-allelic sequence variation can be resolved. However, the reproducibility of non-denaturing gel electrophoresis is reduced in comparison to denaturing gel conditions. Thus, using denaturing gels 'iso'-allelic sequence variation have to be treated as allelic groups when used for gene mapping or human identification.

Acknowledgements We wish to thank Prof. Dr. Klaus D. Zang (Institut für Humangenetik, Universität Homburg-Saar, Germany) for helpful discussion and Mrs. Maria Kilp for technical assistance.

References

Adams M, Urquhart A, Kimpton C, Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Hum Mol Genet 2:1373–1376

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations – Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110:175–176

Boland CR (1996) Setting microsatellites free. Nat Med 2:972–974

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:182–190

Brinkmann B, Meyer E, Junge A (1996) Complex mutational events at the HUMD21S11 locus. Hum Genet 98:60–64

Browne D, Gen M, Evans GA, Clark SP, Litt M (1992) Tetranucleotide repeat polymorphism at the D11S488 locus. Hum Mol Genet 2(1):89

Jones DA (1972) Blood samples: probability of discrimination. J Forensic Sci Soc 12:355–359

Kirby LT, Puterman ML. (1992) Probability and statistical analysis. In: Kirby LT (ed) DNA fingerprinting. An introduction. Stockton Press, New York, pp 149–177

Lareu MV, Pestoni C, Schürenkamp M, Rand S, Brinkmann B, Carracedo A (1996) A highly variable STR at the D12S391 locus. Int J Legal Med 109:134–138

Levinson G, Goodman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Early-Macdonald J, de Jong PJ, Wieringa B, Korneluk RG (1992) Myotonic dystrophy mutation: An unstable CTG repeat in the 3′ untranslated region of the gene. Science 255:1253–1255

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33) Sequencing data reveals considerable polymorphism. Int J Legal Med 106:262–267

Nei M, Roychoudhury AK (1974) Sampling variance of heterozygosity and genetic distance. Genetics 76:379–390

Nelson L, Lu J, Petterson M, Riley R, Ward K (1994) Tetranucleotide repeat polymorphism at the D8S639 locus. Hum Mol Genet 3(7):1209

Perucho M (1996) Microsatellite instability: the mutator that mutates the other mutator. Nat Med 2:676–681

Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958

Richards RI, Sutherland GR (1994) Simple repeat DNA is not replicated simply. Nat Genet 6:114–116

Schlotterer C, Tautz D (1991) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyziz RK (1991) Evolution and distribution of (GT)n repetetive sequences in mammalian genomes. Genomics 10:807–815

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting mismatch repair. Nature 365:274–276

Talbot CC Jr, Avramopoulos D, Gerken S, Chakravarti A, Armour JA, Matsunami N, White R, Antonarakis SE (1995) The tetranucleotide repeat polymorphism D21S1245 demonstrates hypermutability in germline and somatic cells. Hum Mol Genet 4:1193–1199

Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Hum Genet 92:637–638

Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences - a survey of twelve microsatellite loci for use as forensic identification markers. Int J Legal Med 107:13–20

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128