Cellular Pharmacology (1996), 3, 189–194

Identification of muscle-specific calpain and β-sarcoglycan genes in progressive muscular dystrophies

JS Beckmann[1], I Richard[1], O Broux[1], F Fougerousse[1], V Allamand[1], N Chiannilkulchai[1], LE Lim[2], F Duclos[2], N Bourg[1], L Brenguier[1], C Roudaut[1], Y Sunada[2], J Meyer[2], FMS Tomé[3], D Cohen[4], CE Jackson[5], KP Campbell[2] and M Fardeau[3]

[1]*Généthon/CNRS URA 1922, Evry;* [2]*Howard Hughes Medical Institute, Iowa City;* [3]*INSERM U153, Paris;* [4]*CEPH-Fondation Jean Dausset, Paris; and* [5]*Henry Ford Hospital, Detroit*

Summary

Mutations in the gene encoding the muscle-specific calpain, CANP3, recently identified this gene as responsible for autosomal recessive limb-girdle muscular dystrophy type 2A (LGMD2A, MIM 253600). Six CANP3 mutations were found within patients from the Island of la Réunion. A digenic inheritance model was proposed to account for the unexpected presence of multiple independent mutations in this small inbred population. Altogether, a large series of distinct mutations were uncovered, dispersed throughout the entire length of the CANP3 gene. These data demonstrate the wide geographic and molecular heterogeneity of defects in this rare disease. The observation of patients with two null alleles suggests that the skeletal muscular calpain is altered or missing in patients.

Amish 1gmd families from northern Indiana were shown to carry the same CANP3, while involvement of this locus was excluded in families from southern Indiana, despite multiple consanguineous links and common ancestry. A genetic study localised the morbid gene to the pericentromeric region of chromosome 4. Meanwhile, the human cDNA of β-sarcoglycan, a 43 kDa dystrophin-associated glycoprotein, has been cloned and mapped to the same chromosome 4 region. A β-sarcoglycan missense mutation, present in a homozygous state in all these patients, was identified that leads to dramatically reduced expression of β-sarcoglycan and a concomitant loss of α- and γ-sarcoglycans. Thus, the β-sarcoglycan gene is the fifth locus (LGMD2E) involved in autosomal recessive 1gmd.

Keywords: *positional cloning, limb-girdle muscular dystrophy, calpain, β-sarcoglycan*

Introduction

Limb-girdle muscular dystrophies (LGMD) are a group of inherited neuromuscular diseases presenting clinical heterogeneity. LGMDs are characterized by progressive weakness of the pelvic and shoulder girdle muscles, and their genetic etiology has yet to be elucidated. These disorders may be inherited either as an autosomal dominant or recessive trait, the latter being more common with an estimated prevalence of 10^{-5}.[1] In view of the lack of a consensus on the specific nosological definition of each entity, it was proposed at the 30th and 31st ENMC workshop to temporarily lump these autosomal recessive entities under the common denomination of LGMD2.[2] The fact that this decision is far from having generated a unanimous support among clinicians involved in this area illustrates the complexity of this medical and scientific problem.

The first demonstration of a genetic basis for any form of LGMD was obtained in 1991 using biallelic RFLP markers.[3] This study was performed in a genetic isolate from the Isle of La Réunion, localising by linkage analysis a recessive form (LGMD2A, MIM number 253600) to chromosome 15q. Confirmation of this localization was subsequently reported in norther Indiana Amish[4] and Brazilian[5] pedigrees. The latter and subsequent work demonstrated that the autosomal recessive forms (LGMD2) constitute a genetically heterogeneous group involving five genes. These were shown to map to either chromosomes 2p13-p16 (LGMD2B)[4], 13q12 (LGMD2C,[7–9]), 15q15.1 (LGMD2A)[3], 17q12-q21.33 (LGMD2D)[10,11], and 4q12.[12,13] The LGMD2D gene was recently identified: it encodes for the 50 kDA adhalin glycoprotein,[10,11,14,15], now known as α-sarcoglycan. The LGMD2C locus was also just identified: it encodes for the 35 kDa γ-sarcoglycan[9,16]. Both sarcoglycans are members of the dystrophin glycoprotein complex.

Our efforts focussed mainly on the LGMD2A and LGMD2E loci. These genes were identified following either a classical positional cloning (for LGMD2A) or a combined positional and functional cloning (for LGMD2E) strategies. The various steps involved in this quest are reviewed here.

Mapping of a Chromosome 15 Region Involved in LGMD2A

Four years ago, few highly informative genetic markers were available. Fortunately, the Généthon program had started and had already produced a score of new markers.[17,18] In parallel, we also set to develop in a targetted fashion markers from the LGMD2A region. After genotyping the newly available chromosome 15 microsatellite markers, two were found to flank the disease locus within an interval that was assessed, based on segregation studies on the CEPH reference families, as spanning 7 centiMorgans. The

Correspondence: JS Beckmann
Received February 20, 1996

screening of the CEPH YAC libraries with the corresponding probes allowed the isolation of YACs which were used in fluorescence *in situ* hybridization to define the LGMD2A cytogenetic interval as 15q15.1–15q1.1.[19] Different approaches were pursued for the establishment of the physical map of this area. This allowed eventually the assembly of an uninterrupted YAC contig spanning an estimated 10–12 megabases (Figure 1), with a current average STS resolution of 50 kb, and for the 36 polymorphic microsatellites on this map, of 300 kb. Twelve genes and 25 genetic markers were initially positioned in this contig, which is constituted of a minimum of 10 clones.

Preferential Localization of LGMD2A in the Proximal Part of a 1 cM 15q15.1-q15.3 Interval

New microsatellite markers developed using the YACs from the 10–12 Mb LGMD2A YAC contig were genotyped on large, consanguineous LGMD2A pedigrees from different origins.[21] The identification of recombination events in these families allowed the restriction of the LGMD2A region to an estimated 1 cM interval, equivalent to approximately 3–4 Mb (Figure 1), bracketed by D15S514 and D15S222. Linkage disequilibrium data on genetic isolates from the Isle of La Réunion and from the Amish community suggested a preferential location of the LGMD2A gene in the proximal part of this region, an area covered by YAC774G4.

Furthermore, analysis of the interrelated pedigrees from the Isle of La Réunion revealed the existence of at least six different carrier haplotypes. This allelic heterogeneity is incompatible with the *a priori* presumed existence of a founder effect, and suggests that multiple LGMD2A mutations may segregate in this population.

An STS Map of the LGMD2A Region

As YAC clone 774G4 spans a region showing significant linkage disequilibrium with LGMD2A, this 1.6 Mb long YAC was tentatively organized into a contig of overlapping cosmid clones, mainly by STSs screening and inter-Alu PCR hybridization. A total of 70 STSs, including 13 transcription units and 7 microsatellites covers this 1.6 Mb region with an average of 3 cosmids per STS and one STS every 25 kb.[22]

A primary expression map of chromosome 15q15 region containing the recessive LGMD2A gene

In order to progress toward the identification of the gene involved in LGMD2A, a primary transcription map of the LGMD2A genomic region was generated.[23] The direct cDNA selection strategy was used with 3 YACs covering the candidate region and two different muscle cDNA libraries. Seventeen different transcription units were identified among 171 cDNA fragments analysed. Five sequences corresponded to known genes, and twelve to new ones. They were characterized for their sequences, physical positions within the YAC and cosmid contigs, and expression patterns. Among those specifically transcribed in muscle, the calpain gene—whose cDNA sequence was known since 1989[24]—appeared to be a good positional candidate for LGMD2A, although it was not considered, *a priori*, as a functional candidate.

Mutations in the proteolytic enzyme, calpain 3, cause LGMD2A

The CANP3 gene, encoding the large subunit of the muscle-specific calcium-activated neutral protease 3, belongs to the family of intracellular calpains, requiring calcium for their catalytic activities (for a review see Saido[25]). Its genomic organisation was determined, by sequencing cosmids containing this gene, which is composed of 24 exons spread over more than 40 kb[26].

Validation of a candidate gene requires the identification of the pathogenic mutations. A total of 16 nonsense, splice site, frameshift or missense mutations in this muscle-specific gene were initially found to consegrate with the disease in LGMD2A families, six of which were found within Réunion island LGMD2A patients.[26] A digenic inheritance model was proposed to account for the unexpected presence of multiple independent mutations in this small inbred population. One of the mutations encountered in the isle of La Réunion, was found to be a synonymous missense mutation leading to a predicted Gly-to-Gly substitution. In this manner, we could demonstrate that this apparently neutral mutation creates a new splice site, and is thus effectively equivalent to a null mutant.[28] As a matter of fact, to reveal this mutation, we had to screen lymphocyte illegitimate transcription products,[27] since the CANP3 locus shows a tissue-specific pattern of expression and for lack of muscle tissues available from this like from most patients. Last but not least, the finding that calpain mutations cause LGMD2A is the first demonstration of an enzymatic rather than a structural protein defect causing a muscular dystrophy, a defect that may have regulatory consequences, perhaps in signal transduction.[26]

Having identified the role of the calpain gene in LGMD2A, we next analysed the segregation of markers flanking the LGMD2A locus and carried out a search for CANP3 mutations in LGMD2 pedigrees from various geographic origins. Altogether, 42 distinct mutations were uncovered (23 missense, 11 frameshift, 2 nonsense, 4 splice site, and 2 small in-frame deletions), dispersed throughout the entire length of the CANP3 gene (unpublished data). These data indicate that muscular dystrophy caused by mutations in the muscle calpain gene are found in patients from all countries so far examined, with no predominant mutation, and further support the wide heterogeneity of molecular defects in this rare disease.

Characterisation of the murine canp3 gene

cDNA clones encoding mouse muscle-specific calpain were selected by hybridization and sequenced.[29] The mouse gene encodes for an mRNA of similar size to the human CANP3 mRNA, directing the synthesis of an 821 amino acids long protein. A partial characterization of the genomic organisation of the canp3 gene suggests that it follows the same overall structure as its human homologue. The results

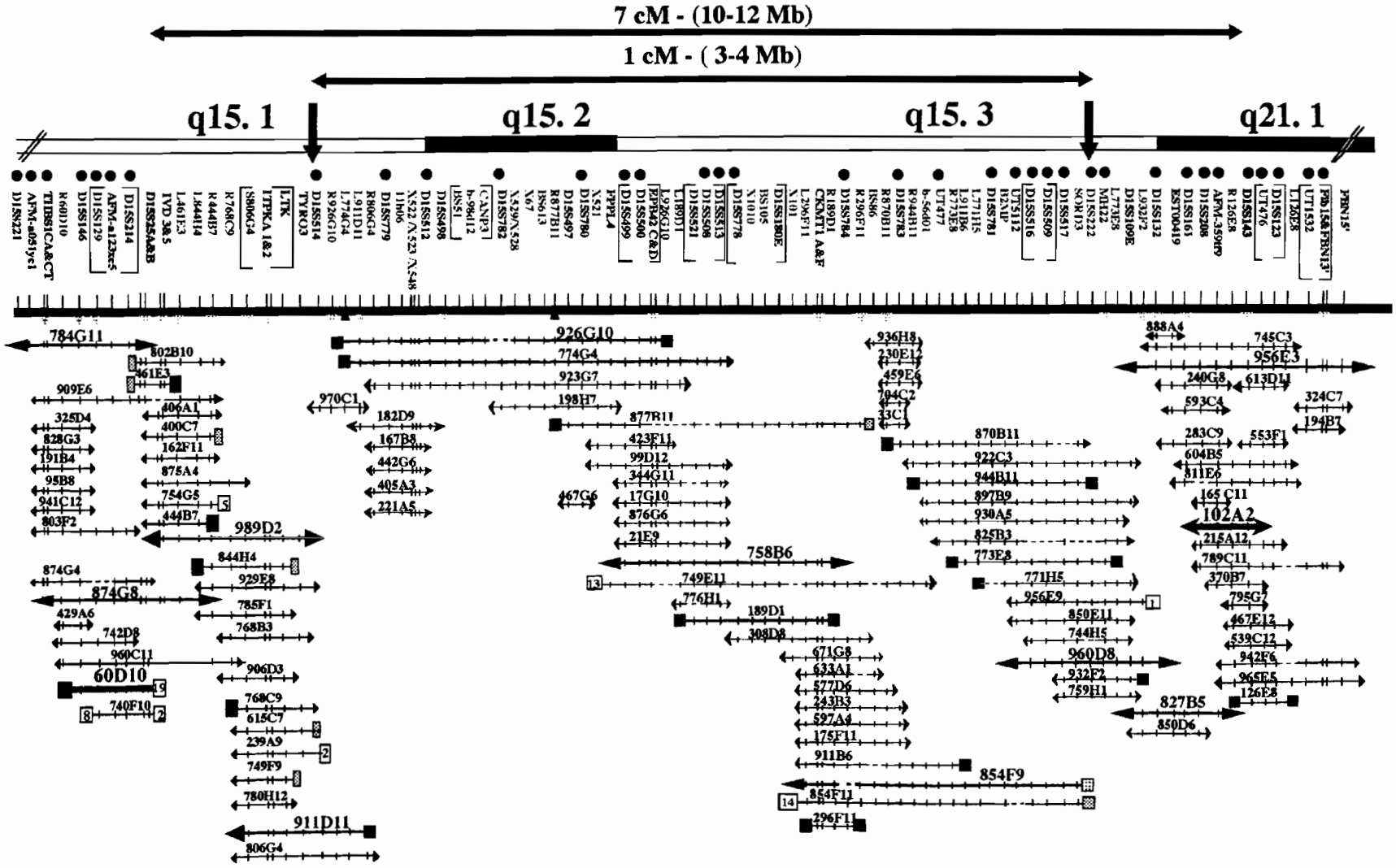

Figure 1 The map shows the contiguous array of YAC clones which are presented as lines (Only clones with more than one characterized STS are drawn). The names of the STSs are given above the line representation of the contig. The position of polymorphic microsatellite markers is indicated by the presence of a dot above the marker name. The STSs are drawn equidistant except when they are issued from the same gene (and hence are arbitrarily ordered), but for the 5′ end of the fibrillin gene. Adjacent markers which cannot be oriented with respect to this map are placed within brackets. Inverted triangles: STSs having supplied YACs upon screening of the CEPH libraries. Upright triangles: STSs which failed to yield YACs in this screening. Bars on full lines indicate positive STSs. Dotted lines mean that a clone was not found to be positive for a given internal STS. Arrows at the end of the clone indicate that a clone was found to be negative for the adjacent STSs. The red lines represent the minimal tiling path consisting of 10 YAC clones spanning the contig. The chromosome 15 YAC end-STSs (R or L) are shown in full squares, and YAC end constituted of repeated sequences as dotted squares. The ends are arbitrarily drawn whenever both ends failed to map within the contig. One extremity of clone 911B6, a double insert and mitotically unstable YAC, is shown as a hatched square, in the middle of contig, as this clone was shown to be positive for external STSs. A drawing showing approximate G-banding pattern of the corresponding region of chromosome 15 is placed above the contig presentation. The smallest recombinant interval defining the LGMD2A region is shown by the vertical arrows on the regional chromosome map, and the distance spanning the recombinant intervals is given above.

Table 1 Percent homology of mouse canp3 with other muscle-specific calpain cDNAs or proteins

Sequence compared		*Rat*	*Human*	*Chicken*
DNA	cDNA coding sequence	95.9	90.1	74.9
	5′ UT of cDNA	93.8	75.0	n.d.
Amino acid	Deduced amino acid sequence	98.7	93.5	77.3
	Allowing for conservative substitutions[a]	100	99.5	94.6

[a]Conservative substitutions using the Fasta PAM250

obtained on somatic cell hybrid panel indicates either a localization on mouse chromosome 2 or on chromosome 4. Information of synteny conservation with human support the chromosome 2 localization. Interspecies comparison of the canp3 DNA sequences shows a high level of identity both at the DNA and amino acid levels. (Table 1). The degree of similarity is indicative of the fundamental biological role of this protein. Together with the other known animal sequences, the mouse canp3 sequence would be useful to assess missense mutations in LGMD2A patients. This sequence information can be used to study the spacio-temporal and tissue-specific expression as well as to develop knock-out mouse models. It represents thus a useful tool for gene-regulation, physiopathological studies of the disease as well as for therapeutic investigations.

Genetic heterogeneity of autosomal recessive LGMD in a genetic isolate (Amish) and evidence for a new locus

Cases of autosomal recessive limb-girdle muscular dystrophy among members of the old order of Amish of northern and southern Indiana were already described.[30,31] The families of these communities are interrelated by multiple consanguineous links and common ancestry which can be traced back to the 19th century in the Canton of Bern, Switzerland. It is upon studying these families[4] that the chromosome 15 localisation of the LGMD2A locus was confirmed. All Amish LGMD patients from northern Indiana were subsequently shown to be homozygous for the same Arg769Glu missense CANP3 mutation.[26]

Allamand *et al*[32] reported the exclusion of the LGMD2A locus in six Amish kindreds from southern Indiana that are related by multiple consanguineous links to the same northern Indiana families in which the involvement of the chromosome 15 locus was previously demonstrated. These findings disclosed genetic heterogeneity of LGMD in this Indiana Amish isolate. In view of the high consanguinity and the similar clinical presentation of all Amish LGMD patients, this result was totally unexpected. Furthermore, genetic analyses also ruled out the possible involvement of the other known LGMD2 loci, thus demonstrating that a mutation within at least one additional locus leads to this condition. Several candidate genes putatively involved in neuromuscular disorders were also excluded. A systematic genome-wide genetic study undertaken with six of these families eventually uncovered genetic linkage to the pericentromeric region of chromosome 4,[12] thereby providing legitimacy to the LGMD2E locus.

Identification of the LGMD2E locus

Meanwhile, the human cDNA of β-sarcoglycan, a 43 kDa dystrophin-associated glycoprotein (also known as A3b), was cloned and mapped to the same chromosome 4 region.[12] Several chromosome 4 pericentromeric markers and a β-sarcoglycan intragenic polymorphic CA repeat cosegregated perfectly with the disease in Amish families from Southern Indiana with autosomal recessive limb-girdle muscular dystrophy. In addition, reconstruction of haplotypes allowed identification of chromosomes descending from a common ancestor. Furthermore, a Thr-to-Arg missense mutation, present in a homozygous state in all patients, was identified within the β-sarcoglycan gene that leads to a dramatically reduced expression of β-sarcoglycan in the sarcolemma and a concomitant loss of α- and γ-sarcoglycan, thereby disrupting the integrity of the dystrophin glycoprotein complex.[12] Similar evidence were also obtained concurrently by another lab[13]. These demonstrated that a defect in β-sarcoglycan is responsible for the recessive muscular dystrophy in southern Indiana Amish families. Thus the β-sarcoglycan gene is the fifth locus (LGMD2E) involved in autosomal recessive limb-girdle muscular dystrophy, and the third sarcoglycan involved in a myopathy (Table 2).

Conclusion

The year that just ended seems to have been a good vintage year for geneticists and clinicians interested in autosomal recessive progressive muscular dystrophies. Indeed in a little over a year, four LGMD2 genes have been identified, three of which belong to the dystrophin-associated protein complex, the fourth one being an enzyme (Table 2. These discoveries raise more questions than answers. But this is not a matter of complaint. They are also a new source of hope. This knowledge can, on the one hand, already be used to learn more about the physio-pathological processes involved. On the other hand, careful examination of the

Table 2 Molecular etiology of autosomal recessive progressive myopathies

Locus symbol (synonym)	*MIM*	*Chromosomal localisation*	*Protein (size, kDa)*	*Sub-cellular localisation*	*Identification method*	*Function*
LGMD2A	253600	15q15	calpain (94)	?	pos. cloning	?
LGMD2C (SCARMD1)	253700	13q12	γ-sarcoglycan (35)	membrane of muscle cells	C.G.S.	link between laminin-2 & dystrophin
LGMD2D (SCARMD2)	600119	17q12	α-sarcoglycan (50)		C.G.S.	
LGMD2E	600900	4q12	β-sarcoglycan (43)		pos. + funct. cloning	
CMD[a]	156225	6q22				component of basal lamina
EDMD[b]	310300	Xq28			pos. cloning	vesicular transport?

The following abbreviations stand for: C.G.S.: candidate gene strategy; pos. cloning: positional cloning; funct. clo.: functional cloning.
[a]See Reference 34.
[b]See Reference 35.

phenotype-genotype relationships[33] will enable a clearer definition of the specific nosological boundaries of each one of these similar, yet different, clinical entities, a knowledge that could be of great immediate benefit to the patients themself. And last but not least, it opens prospects for new therapeutic avenues.

Acknowledgements

We are particularly indebted to all patients, their families and clinicians. We thank Drs M Conneally, C Moomaw, C Slaughter, MR Passos-Bueno, J Tischfield and M Zatz for their participation in parts of this work. We also acknowledge many colleagues for providing clones, chromosome 4 specific cosmid library; microsatellite repeat probes; for performing muscle biopsies; and finally, members from our respective lab for their help. This research was supported by grants from the Muscular Dystrophy Association, The European Commission within the context of the European Gene Mapping (EUROGEM) project, and Association Française contre les Myopathies (AFM), and the Groupement de Recherche Européen sur le Génome. VA is funded by a grant from AFM. KPC is an investigator of the Howard Hughes Medical Institute.

References

1 Emery AEH. Population frequencies of inherited neuromuscular disease — a world survey. *Neuromuscular Disorders* 1991; **1**: 19–29.
2 Bushby KMD and Beckmann JS. Report of the 30th and 31st ENMC International Workshops on the limb-girdle muscular dystrophies — proposal for a new nomenclature. *Neuromuscular disorders* 1995; **5**: 337–343.
3 Beckmann JS *et al*. A Gene for Limb-Girdle Muscular Dystrophy Maps to Chromosome 15 by Linkage. *Cr Acad Sci* 1991; **312 (Série III)**: 141–148.
4 Young K *et al*. Confirmation of linkage of limb-girdle muscular dystrophy to chromosome 15. *Genomics* 1992; **13**: 1370–1371.
5 Passos-Bueno M *et al*. Evidence of genetic heterogeneity for the adul form of limb-girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families. *J Med Genet* 1993; **30**: 385–87.
6 Bashir R *et al*. A gene for autosomal recessive limb-girdle muscular dystrophy maps to chromosome 2p. *Hum Mol Genet* 1994; **3**: 455–457.
7 Ben Othmane, *et al*. Linkage of Tunisian autosomal recessive Duchenne-like muscular dystrophy to the pericentromeric region of chromosome 13q. *Nature Genet* 1992; **2**: 315–317.
8 Azibi K *et al*. Severe childhood autosomal recessive muscular dystrophy with the deficiency of the 50 kDa dystrophin-associated glycoprotein maps to chromosome 13q12. *Hum Mol Genet* 1993; **2**: 1423–28.
9 Noguchi S *et al*. Mutations in the dystrophin-associated glycoprotein γ-sarcoglycan in chromosome 13 muscular dystrophy. *Science* 1995; **270**: 819–822.
10 Roberds S *et al*. Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. *Cell* 1994; **78**: 625–633.
11 McNally EM *et al*. Human adhalin is alternatively spliced and the gene is located on chromosome 17q21. *Proc Natl Acad Sci USA* 1994; **91**: 9690–9694.
12 Lim LE, Duclos F, Broux O *et al*. β-Sarcoglycan: characterization and role in limb-girdle muscular dystrophy linked to 4q12. *Nature Genetics* 1995; **11**: 257–265.
13 Bönnemann CG *et al*. Mutations in the dystrophin-associated glycoprotein β-sarcoglycan (A3b) cause autosomal muscular dystrophy with disintegration of the sarcoglycan complex. *Nature Genet* 1995; **11**: 266–273.
14 Picolo F *et al*. Primary adhalinopathy: a common cause of autosomal recessive muscular dystrophy of variable severity. *Nature Genetics* 1995; **10**: 243–245.
15 Kawai H *et al*. Adhalin gene mutations in patients with autosomal recessive childhood onset muscular dystrophy with adhalin deficiency. *J Clin Invest* 1995; **96(3)**: 12021207.
16 Jung *et al*. Absence of gamma-sarcoglycan (35 DAG) in autosomal recessive muscular dystrophy linked to chromosome *l*3q*l*2. Febs Letters 1996.
17 Weissenbach J *et al*. A second-generation linkage map of the human genome. *Nature* 1992; **359**: 794801.
18 Gyapay G *et al*. The 1993–94 Généthon human genetic linkage map. *Nature Genet* 1994; **7**: 246–329.
19 Fougerousse F *et al*. Mapping of a Chromosome 15 Region Involved in Limb Girdle Muscular Dystrophy. *Human Mol Genet* 1994; **3**: 285–293.
20 Pereira de Souza A *et al*. Targeted development of microsatellite markers from inter-alu amplification of YAC clones. *Genomics* 1994; **19**: 391–393.
21 Allamand V *et al*. Preferential localisation of the limb-girdle muscular dystrophy type 2A gene in the proximal part of a 1 cM 15q15.2–q15.3 interval. *Am J Hum Genet* (1995a); **56**: 1417–1430.
22 Richard I *et al*. A partial cosmid contig of the limb girdle

muscular dystrophy type 2A region. *Mammalian Genome* 1995a; **6**: 754–756.

23 Chiannilkulchaï N *et al.* A primary expression map of the chromosome 15q15 region containing the LGMD2A gene. *Human Mol Genet* 1995; **4**: 717–726.

24 Sorimachi H *et al.* Molecular cloning of a novel mammalian calcium-dependent protease distinct from both m- and mu-type. Specific expression of the mRNA in skeletal muscle. *J Biol Chem* 1989; **264**: 20106–20111.

25 Saido T, Sorimachi H, Suzuki K. Calpain: new perspectives in molecular diversity and physiological–pathological involvement. *FASEB J* 1994; **8**: 814–822.

26 Richard I *et al.* A novel mechanism leading to muscular dystrophy: mutations in calpain 3 cause limb girdle muscular dystrophy type 2A. *Cell* 1995b; **81**: 27–40.

27 Chelly J, Concordet JP, Kaplan JC, and Kahn A. Illegitimate transcription: Transcription of any gene in any cell type. *Proc Natl Acad Sci USA* 1989; **86**: 2617–2621.

28 Richard I, Beckmann JS. How neutral are synonymous codon mutations. *Nature* 1995; **10**: 259.

29 Richard I, Beckmann JS. Molecular cloning of mouse canp3, the gene associated with limb-girdle muscular dystrophy 2A in human. *Mammalian Genome* 1996, **7**: (in press).

30 Jackson CE, Carey JH. Progressive muscular dystrophy: autosomal recessive type. *Pediatrics* 1961; **28**: 77–84.

31 Jackson CE, Strehler DA. Limb-girdle muscular dystrophy: clinical manifestations and detection of preclinical disease. *Pediatrics* 1968; **41**: 495–502.

32 Allamand V *et al.* Genetic heterogeneity of autosomal recessive limb-girdle muscular dystrophy in a genetic isolate (Amish) and evidence for a new locus. *Human Mol Genet* 1995b; **4**: 459–463.

33 Fardeau M *et al.* Juvenile limb-girdle muscular dystrophy. Clinical, histopathologial and genetic data on a small community living in the Réunion Island. *Brain* 1996; **119**: 295–308.

34 Helbling-Leclerc A *et al.* Mutations in the laminin α2-chain gene (LAMA2) cause merosin-deficient congenital muscular dystrophy. *Nature Genetics* 1995; **11**: 216–218.

35 Bione S *et al.* Identification of a novel gene responsible for Emery–Dreyfuss muscular dystrophy. *Nature* 1994; **8**: 323–327.