De novo variants and recombination at 4q35: Hints for preimplantation genetic testing in facioscapulohumeral muscular dystrophy

Sara Pini | Floriana Maria Napoli | Enrico Tagliafico | Antonio La Marca | Emma Bertucci | Valentina Salsi | Rossella Tupler

Abstract
Facioscapulohumeral muscular dystrophy (FSHD) has been associated with the deletion of an integral number of 3.3 kb units of the polymorphic D4Z4 repeat array at 4q35. The prenatal identification of this defect can be carried out on chorionic villi or amniocytes, whereas preimplantation genetic testing for monogenic disorders (PGT-M) requires molecular markers linked to the D4Z4 allele of reduced size. In this context the reliability of this association is crucial. To test the informativeness of the nearby polymorphic markers we investigated recombination at 4q35 using the polymorphic markers D4S1523, D4S163 and D4S139 positioned at 0.55, 0.5 and 0.21 Mb proximal to the D4Z4 array respectively. We determined the probability of recombination events to occur in the D4Z4-D4S1523 interval considering 86 subjects belonging to 12 FSHD families and found a recombination frequency of 14% between D4Z4 and D4S1523. Our study also revealed the occurrence of de novo variants and germline mosaicism. These findings highlight the recombinogenic nature of the 4q subtelomere and indicate that caution should be taken when interpreting PGT-M results. It is advisable that a woman who underwent a PGT-M cycle undertakes a prenatal DNA analysis to confirm the size of the D4Z4 alleles carried by the fetus.

Keywords
facioscapulohumeral muscular dystrophy, genetic counseling, linkage analysis, polymorphic markers, preimplantation genetic testing, prenatal diagnosis, recombination frequency

1 | INTRODUCTION
With an estimated prevalence of 1 in 20 000, facioscapulohumeral muscular dystrophy (FSHD, OMIM#158900) is the third most common heritable myopathy. The majority of FSHD cases carry a deletion of an integral number of tandem 3.3 kb units of the polymorphic D4Z4 repeat array at 4q35. D4Z4 alleles with fewer than 10 repeat units in association with the 4qA polymorphism are considered pathogenic even though studies revealed low penetrance and intra-familial clinical variability.
These conditions make genetic counseling challenging and the decision-making process arduous for couples. At present beside prenatal diagnosis (PND), preimplantation genetic testing for monogenic disorders (PGT-M) is also offered. In the case of FSHD PGT-M is based on the indirect analysis of D4Z4 by PCR amplification of nearby polymorphic markers ([https://www.fshdsociety.org/diagnosis/genetic-testing/](https://www.fshdsociety.org/diagnosis/genetic-testing/)). This is because the methodology routinely performed for FSHD PND requires large amounts of high molecular weight genomic DNA from chorionic villi or amniocytes, instead very low quantities of genomic material are required for PGT-M.

Here, we investigate the presence of recombination between the D4Z4 locus and neighboring polymorphic loci used for PGT-M.

Materials and Methods are reported in Data S1.

2 | RESULTS

2.1 | Recombination between D4Z4 and the 4q35 markers

Within the 4q35 locus, four microsatellite and two VNTR markers have been identified (Figure 1). The four microsatellites (D4S2390, D4S1652, D4S2930 and D4S1523) were previously reported to be suitable for PGT-M protocols.

Considering that the 4q subtelomere is highly recombinogenic, we focused on the most telomeric of these four microsatellite polymorphic markers, D4S1523 which maps 0.55 Mb proximal to the D4Z4 array, since a distance of 1 Mb between a pathogenic variant and a polymorphic marker is commonly considered appropriate in marker-based diagnostic approaches. We also included D4S163 and D4S139, which are VNTRs positioned 500 kb and 210 kb centromeric to D4Z4 respectively. Even though VNTRs are not suitable for PGT-M the use of these markers allowed evaluating the occurrence of recombination in the most telomeric part of 4q.

Our analysis showed for the D4Z4-D4S139 interval a maximum LOD score of 7.43 at θ = 0.1 and an optimized maximum LOD score of 7.49 at θ = 0.080 with 3-lod-unit support interval Z(θ)ϵ [0.008, 0.285] (Table S3A). For the D4Z4-D4S163 interval a maximum LOD score of 7.48 at θ = 0.1 was observed, and an optimized maximum LOD score of 7.50 at θ = 0.089 with 3-lod-unit centromeric to D4Z4 respectively. Even though VNTRs are not suitable for PGT-M the use of these markers allowed evaluating the occurrence of recombination in the most telomeric part of 4q.

Our analysis showed for the D4Z4-D4S139 interval a maximum LOD score of 7.43 at θ = 0.1 and an optimized maximum LOD score of 7.49 at θ = 0.080 with 3-lod-unit support interval Z(θ)ϵ [0.008, 0.285] (Table S3A). For the D4Z4-D4S163 interval a maximum LOD score of 7.48 at θ = 0.1 was observed, and an optimized maximum LOD score of 7.50 at θ = 0.089 with 3-lod-unit support interval Z(θ)ϵ [0.012, 0.293] (Table S3B). For the D4Z4-D4S1523 interval a maximum LOD score of 7.88 at θ = 0.1, and an optimized maximum LOD score of 8.01 at θ = 0.140 with 3-lod-unit support interval Z(θ)ϵ [0.019, 0.498] were detected (Table S3C).
We also evaluated the recombination fraction for intermediate intervals obtaining a maximum LOD score of 9.99 at $\theta = 0.1$ and an optimized maximum LOD score of 10.91 at $\theta = 0.032$ for D4S139-D4S163 interval with 3-lod-unit support interval $Z(\theta) \in [0.001, 0.192]$ (Table S3D). D4S163-D4S1523 interval showed a maximum LOD score of 5.09 at $\theta = 0.1$ and an optimized maximum LOD score of 5.44 at $\theta = 0.046$ with 3-lod-unit support interval $Z(\theta) \in [0.001, 0.290]$ (Table S3E).

**FIGURE 2** Haplotypes of families with de novo alleles. A) Family 5: de novo alleles (light gray) for D4S1523 were detected in subjects II.2 and II.3. The haplotype of subject I.2 is incomplete because of the limited amount of genetic material; B) Family 12: de novo alleles (light gray) for D4S1523 were detected in subjects II.2 and III.1 and III.2. Because of the low allelic heterogeneity of D4S1523 it was not possible to clearly determine the origin of the de novo allele in subject II.2. A D4S139 de novo allele (dark gray) was detected in subject III.1. Non-informative alleles have no colored background. [Colour figure can be viewed at wileyonlinelibrary.com]
Recombination events occurring in the second generations were identified because the haplotypes detected in siblings were only partially overlapping. Since the phase of parental haplotypes was unknown it was not possible to identify the recombinant sibling in the second generation (see in family 12 subjects II.1 and II.2 Figure 2B; family 10 subjects II.1, II.3 and II.5 Figure S1B).

Overall, a 14% chance of recombination was assessed within the D4S1523 microsatellite-D4Z4 genomic interval which is consistent with the evidence that among 30 full informative meiosis four definite recombinants were observed (subject III.1 in family 3, subjects III.3 and III.4 in family 5, subject III.1 in family 10) (Figure 2, Figure S1, Table S4).

### 2.2 Presence of de novo alleles

Analysis of D4S1523 microsatellite segregation in families also revealed the occurrence of de novo alleles in two of the tested families. In family 5, siblings II.2 and II.3 both carry a 125 bp de novo allele at D4S1523. As shown in Figure 2A and in Figure S2A, for the D4S1523 marker the paternal alleles are 121 bp and 125 bp (subject I.1) and maternal alleles are 121 bp and 129 bp (subject I.2) whereas subjects II.2 and II.3 both present two 125 bp alleles, one of which was inherited from the father, and one is apparently de novo (Figure 2A). The presence of the same “de novo” allele of maternal origin suggests the presence of germline mosaicism in the mother, I.2. In family 12, de novo alleles were detected in two different generations. As shown in Figure 2B subject II.2 shows a 121 bp/133 bp genotype: the 133 bp allele was inherited from the mother I.2, who carries a 121 bp/133 bp genotype, whereas no paternal allele was inherited from subject I.1 who carries a 129 bp/129 bp genotype. The fact that II.2 carries the 121 bp/133 bp genotype which is identical to the mother, hinders the possibility of establishing which is the de novo allele between the two and the occurrence of a recombination event. In this family also individuals III.1 and III.2, respectively son and daughter of subject II.2, show de novo alleles. Parental genotypes are 121 bp/133 bp (subject II.2) and 121 bp/121 bp (subject II.3) while III.1 presents a 121 bp/125 bp genotype and III.2 a 121 bp/129 bp genotype. In this case, both siblings inherited a 121 bp allele either maternal or paternal whereas the 125 bp and 129 bp alleles are de novo. In this family we also detected a 9.2 kb de novo allele occurring at D4S139 VNTR in subject III.1 (Figure 2B and Figure S2B).

### 3 DISCUSSION

Nowadays, an increasing number of couples in families with FSHD request prenatal testing to evaluate the risk of having children with the disease. Besides, there is a demand for PGT-M, which might represent a valid alternative to PND.

Data from our diagnostic center for FSHD indicates that 5.2% of subjects carrying a DRA, either presenting a clinical phenotype or being healthy, with age between 20 and 40 years at some point requested preconception counseling. Since 2018 10 out of 70 couples who requested preconception counseling also requested information about preimplantation diagnosis.

The main benefit of preimplantation diagnosis for high-risk families is the confidence to begin a pregnancy without the physical and emotional trauma of a possible therapeutic termination. Nevertheless, in FSHD this option should be well pondered, as a direct detection of D4Z4 allele size is not applicable to single cell methodologies used for PGT-M. To address these needs a multiplex PCR method, which requires less genetic material, has been developed by Barat-Houari et al. The Authors selected four polymorphic microsatellite markers (D4S2390, D4S1652, D4S2930 and D4S1523) 0.55–1.23 Mb proximal to the D4Z4 array. This study estimated a recombination fraction \( \theta = 0.021 \) in the interval D4S2390-D4S1523 studying 22 pedigrees and a recombination fraction \( \theta = 0.0 \) with LOD score equal to 2.98 between D4S1523 and D4Z4 (Figure S3).

In our study the selected polymorphic markers lay in a region between 0.55 Mb and 0.21 Mb proximal to D4Z4. We found that the recombination chance between D4Z4 and D4S163 and D4S139 VNTRs is >8% and is 14% between D4Z4 and the D4S1523 microsatellite. We also observed a low heterozygosity of the D4S1523 microsatellite in our population, which made the identification of the parental origin of some alleles ambiguous.

Remarkably, we also detected six de novo alleles, which may be a relevant issue for PGT-M based on the inferred association of the D4Z4 allele to a specific haplotype. In the case of the D4S1523 microsatellite the new alleles can be generated by polymerase slippage, a phenomenon occurring at microsatellite stretches. In the case of D4S139 and D4S163 VNTRs the presence of de novo alleles might be related to the recombinogenic nature of VNTRs and other tandemly arrayed repetitive elements. To understand the nature of these phenomena we analyzed the published genomic sequence of 4q35 subtelomere and found various Alu, long interspersed element-1 (LINE-1, L1) and short interspersed elements (SINEs) in the 500 kb region proximal to D4Z4. It is possible that our findings might reflect the germline L1 activity that makes L1, Alu, and SVA integrations commonplace, thus providing novel information about mechanisms at the basis of the frequent identification of structural variants at 4q35.

Our study highlights points that should be considered when undertaking a PGT-M cycle for FSHD. First, the D4S1523 marker allele was uninformative in 14 meioses out of 30 because we were unable to assign the linkage phase between the marker allele and a D4Z4 contraction. In 10 meioses the affected parent was homozygous for D4S1523 marker allele. In four meioses, the segregation analysis was not informative. In particular, it has to be noted that in cases where both parents are heterozygotes for the same marker alleles 50% of meioses will be informative if phase is known. Second, if a PGT-M cycle for FSHD is performed based on the D4S1523 marker, the 14% probability of recombination events and the occurrence of de novo variations should be considered.

Overall, it has to be acknowledged that the use of D4S1523 and other 4q35 markers for PGT-M might give rise to cases of misinterpretation. In fact, if PGT-M was performed on the informative families
analyzed in this work, D4Z4 allele size would have been correctly predicted only in 12 out of 16 (75%) F3 meioses using D4S1523 marker.

Also considering ESHRE PGT-M Consortium good practice recommendations for the detection of monogenic disorders, the D4S1523 microsatellite marker should not be considered suitable for PGT-M application. This risk makes PND necessary to confirm the size of the D4Z4 allele and this information should be given to couples considering PGT-M.

In conclusion, considering the low penetrance and inter/intrafamilial clinical variability described in FSHD families which make genetic counseling challenging and the decision-making process arduous, PGT-M should only be offered in FSHD at present where the patients have already chosen to opt for a CVS. The purpose of PGT-M is then to increase the likelihood that a subsequent CVS will indicate an unaffected fetus.

AUTHOR CONTRIBUTIONS
Sara Pini, Floriana Maria Napoli, Valentina Salsi and Rossella Tupler contributed to study design, molecular analysis, data collection, data analysis and interpretation, literature search, preparing of figures/tabs and manuscript writing. Antonio La Marca and Emma Bertucci contributed in manuscript editing.

ACKNOWLEDGMENT
We are indebted to all patients and their families for participating in this study.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/cge.14250.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT
The Miogen database was approved by the Provincial Ethics Committee of Modena (2712/CE). Informed written consent was obtained from all study participants, in accordance with the ethical standards of the 1964 Declaration of Helsinki.

REFERENCES

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.