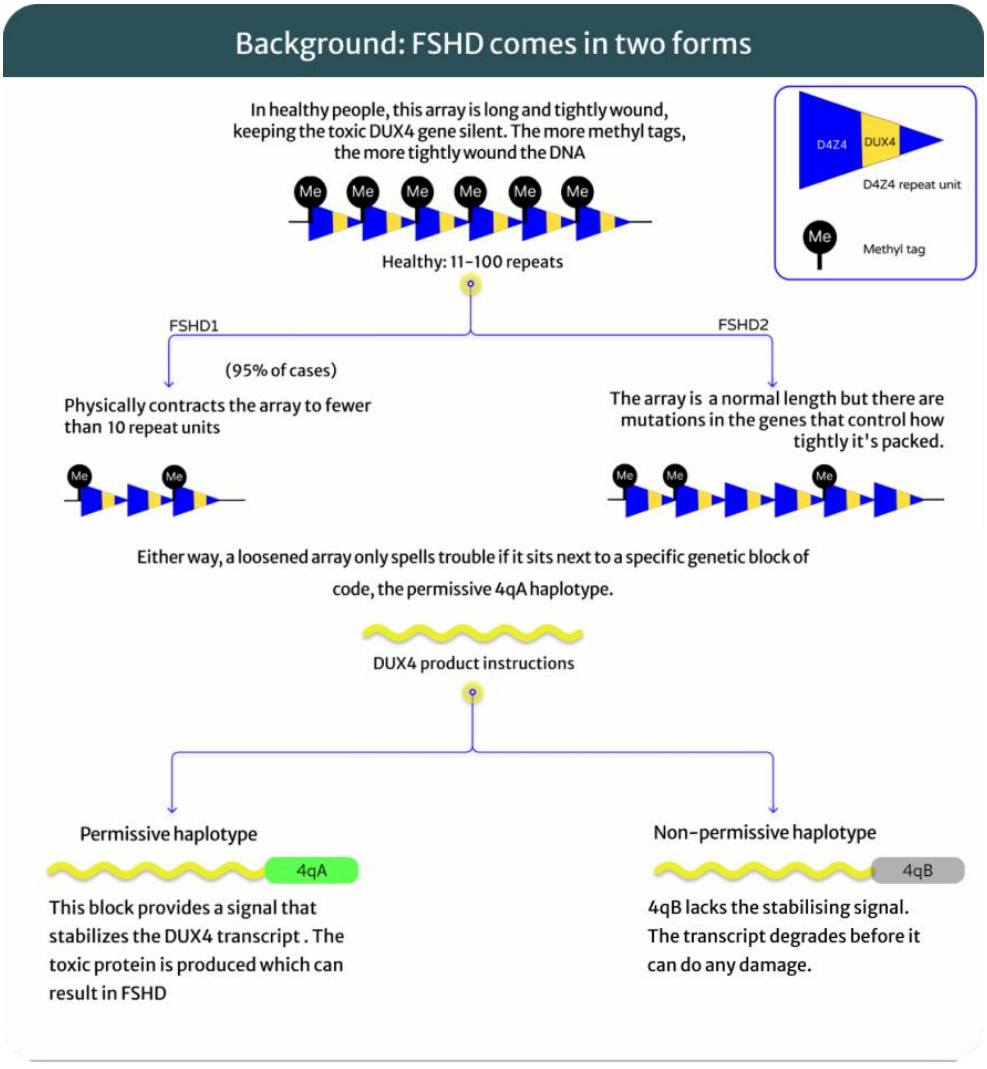
Uncovering D4Z4's Hidden Genomic Relatives: FSHD Diagnostics in the Era of Gap-less References.
By Charmilla N

Medical researchers, like detectives, build a case file of evidence to solve the mystery of a disease. When better-suited technologies emerge, we can re-evaluate and improve on previous knowledge.

Imagine a detective investigating a suspect who has a twin. Over the years, the detective has learned how to tell them apart. What the detective doesn't know is that the neighbourhood is full of distant cousins who also resemble the suspect enough to fool you at a glance.

This is roughly where genetic testing for facioscapulohumeral muscular dystrophy (FSHD) has stood for decades. FSHD is the 3rd most common inherited muscular dystrophy. It traces back to a highly repetitive stretch of DNA called the D4Z4 macrosatellite, at 4q35 on chromosome 4. Each repeat unit is about 3.3 kilobases long and contains the DUX4 gene, which produces a toxic protein when expressed. Its near-identical twin on chromosome 10 at 10q26 (99.5% similar) has long been accounted for in diagnostic tests.



FSHD throws two genetic curveballs. The first is incomplete penetrance: having the right genetics doesn't guarantee symptoms. Roughly 3-4.6% of the general population have the right genetics but never develop FSHD. The second is variable expressivity: in affected people, severity ranges from mild muscle weakness to being wheelchair-bound, even within the same family.

Diagnostic frameworks almost exclusively focused on chromosomes 4 and 10. Potential look-alikes were overlooked because older reference genomes left repetitive regions blank, and they were too difficult to sequence and map with older technologies. Is every "positive ID" the real culprit, or an unassuming doppelgänger, or perhaps an accomplice?

This is the mystery that a team of scientists set out to investigate. They utilized the new, gap-less Telomere-to-Telomere (T2T-CHM13) human genome reference, along with long-read sequencing, to resolve these regions.

Using the chromosome 4 D4Z4 repeat as a search query, they scanned the entire T2T reference and found D4Z4-like repeats on at least ten other chromosomes: 1, 3, 9, 13, 14, 15, 18, 21, 22, and Y. To validate the significance of these findings, they checked primers used for FSHD diagnosis against the doppelgängers, using special single-chromosome hybrid cells. The primers accidentally amplified sequences on the doppelgängers failing to be fully locus specific. This suggests a fraction of the "DUX4 signal" reported in the clinical literature may have been background noise from these doppelgängers.

Table 1: Protein domains, stabilizing signal status, and percent identity to DUX4FL (full length) across candidate D4Z4-like sequences.

Locus	**Protein length (amino acid)**	**Homeodomain 1**	**Homeodomain 2**	**Transactivation Domain**	**Stabilizing signal**	**% identity to DUX4FL**
chr10	422 aa	✓	✓	✓	✗	100%*
chr1	460 aa	✓	✓	✓	✗	80.6%
chr14 (locus 1)	207 aa	✓	✗	✗	✓	74.9%
chr14 (locus 2)	379 aa	✓	✓	✓	✓	80.0%
chr22 (locus 1)	158 aa	✓	✓	✗	✓	82.3%
chr22 (locus 2, frame 1)	129 aa	✗	✗	✗	✓	85.3%

Locus	Protein length (amino acid)	Homeodomain 1	Homeodomain 2	Transactivation Domain	Stabilizing signal	% identity to DUX4FL
chr22 (locus 2, frame 2)	231 aa	✗	✗	✗	✓	78.0%

Current evidence suggests that these D4Z4-like doppelgängers are mostly bystanders. None produced a stable DUX4-like transcript. However, chromosomes 1, 14, and 22, come close as shown in table 1.

The authors urge a shift to a "repeat-aware" framework anchoring assays to unique flanking sequence bordering the repeat for more accuracy. Using long-read technologies offers a two-fold advantage: they can span the full length of repetitive regions unlike short-reads, and they natively detect methylation status.

T2T-CHM13 was built from a single individual's genome, so it can't capture the full structural variability across populations. This matters especially for those historically underrepresented in genomic reference data, such as African populations. T2T is still a valuable foundation, and newer pangenome initiatives are revealing structural variation in these repetitive regions. Broadening reference data matters just as much as filling in the gaps.

High-fidelity D4Z4 mapping, across more populations, is a real step toward explaining incomplete penetrance and variable expressivity, which can lead to better prognosis, treatment, and management for FSHD.

References:

Salsi V, Losi F, Pini S, Chiara M, et al. Rethinking genomics of facioscapulohumeral muscular dystrophy in the telomere-to-telomere era: pitfalls in the hidden landscape of D4Z4 repeats. *European Journal of Human Genetics*. 2026;34:357-367. https://doi.org/10.1038/s41431-025-02000-x

Giardina E, Camaño P, Burton-Jones S, Ravenscroft G, Henning F, Magdinier F, et al. Best practice guidelines on genetic diagnostics of facioscapulohumeral muscular dystrophy: update of the 2012 guidelines. *Clinical Genetics*. 2024;106:13-26. https://doi.org/10.1111/cge.14533