

COL6A1: causal variant removed by an exon ± 25 bp filter

Suspected muscular dystrophy, whole-genome VCF. Required answer: **COL6A1 chr21:45989967 C>T** (c.930+189C>T), heterozygous, pathogenic, diagnostic. The variant is 189 bp into intron 11 and activates a pseudoexon. Fable 5.1 had it in its working set, then removed it with a filter chosen to save annotation time.

Claude Fable 5.1 (xhigh) FAIL · 0.161 · no_finding

BEFORE · EVENTS 148-223 · TOOL CALLS / OUTPUT

Causal variant is in the working set

```
symbols = ""DMD EMD FHL1 LMNA ... LAMA2 COL6A1
COL6A2 ...
DONE 30949 rows in panel regions
```

Builds a 186-gene muscle panel including COL6A1 and extracts every VCF row in those gene bodies ± 5 kb. Still on track.

DURING · EVENTS 328-332 · REASONING, CODE, OUTPUT

Narrows to exons to save time

I'll verify it, build an exon table, filter variants to exonic/splice-region sites (needed since VEP takes ~22s per 30 variants), then kick off VEP annotation in the background.

```
PAD = 25
if pos + reflen - 1 >= a - PAD and pos <= b + PAD:
    hit = True
exonic/splice-region rows: 1165
```

Error: 30,949 rows become 1,165. The causal variant, 189 bp from the nearest exon, is dropped. The cut is made for throughput, not on biological grounds, and it is never revisited for the panel's top genes.

AFTER · EVENT 990 · FINAL REPORT

Reports no finding and calls the review full-VCF

```
"status": "no_finding", "review_basis": "full-VCF"
"No rare LoF or canonical splice variant in ...
COL6A1/2/3 ..."
"Deep intronic ... variants outside exon +/-25 bp
... were not annotated."
```

It lists the limitation, yet still reports no_finding for the case.

Required by the truth

DIAGNOSTIC

BEFORE · CASE.VCF.GZ · CALLER ROW

The causal variant is in the input

```
chr21 45989967 C T QUAL 222.4 GT 0/1 AD 14,12
```

A heterozygous, well-supported call in COL6A1 intron 11, 189 bp from the nearest exon (NM_001848.3:c.930+189C>T).

DURING · WHAT THE STEP REQUIRED

Keep intronic variants in high-prior genes

COL6A1 was already in the panel. Collagen VI dystrophy is a known cause of this phenotype, and deep-intronic pseudoexon variants are an established mechanism in COL6A1. Any narrowing for speed had to keep, or later revisit, intronic variants in the panel's top genes, or check them against ClinVar genome-wide. ClinVar lists this variant as pathogenic/likely pathogenic (VCV000542998).

AFTER · GRADER/TRUTH.JSON

The answer the report should give

```
COL6A1 chr21:45989967 C>T heterozygous
classification: pathogenic (likely pathogenic
accepted)
case status: diagnostic
```

Rank 1, de novo dominant collagen VI-related muscular dystrophy caused by pseudoexon activation.

WHAT HAPPENED

Fable 5.1 followed a sound plan (panel genes, then annotation) but cut its candidate set to exon ± 25 bp because annotation was slow. That step removed the causal variant, and the rest of the run worked on a set that no longer contained it. The final report notes that deep-intronic variants were not examined, yet concludes no_finding.

TRAINING TARGET

Do not narrow the search space for speed in high-prior genes without revisiting it. Report a negative result as diagnostic only when the search actually covered the variant classes that cause the disease.