

RNU2-2: snRNA genes removed by a protein-coding filter

Drug-resistant epilepsy with regression, whole-genome VCF. Required answer: four heterozygous **RNU2-2** variants (chr11:62841621, 62841647, 62841748, 62841767) reported as one combined candidate event, uncertain significance, candidate_finding. RNU2-2 encodes a small nuclear RNA and has no protein product. Neither model passed, so the right column shows what the truth required at Fable's decisive step.

Claude Fable 5.1 (xhigh)

FAIL · 0.182

Required by the truth

VUS · candidate · one event

BEFORE · EVENT 295 · TOOL OUTPUT

Gene models include snRNA genes

```
genes 78733 tx 644280 exons 5078384 cds 3214876
[('lncRNA', 34866), ('protein_coding', 20107), ...
 ('snRNA', 1901), ('miRNA', 1878), ...]
```

Builds gene, transcript and exon tables from the Ensembl 116 GFF3. The snRNA class, which contains RNU2-2, is loaded.

DURING · EVENTS 331-346 · REASONING, CODE, OUTPUT

Annotates protein-coding transcripts only

GFF tables and CDS FASTA are set—now I'll build a local consequence annotator using Ensembl canonical transcript mapping to run genome-wide, so only a narrow final candidate set needs external lookups.

```
if "Ensembl_canonical" not in t[8]: continue
if t[6]!="protein_coding": continue
canonical protein-coding tx 19749
variants scanned 5115881 annotated rows 80980
```

Error: one line in the annotator removes every non-coding gene, including all 1,901 snRNA genes. The four RNU2-2 calls never enter the working set, and no later step adds non-coding genes back.

AFTER · EVENT 1157 · FINAL REPORT

Reports two unrelated missense VUS

```
"status": "candidate_finding"
rank 1 ASH1L chr1:155346383 C>G
uncertain_significance
rank 2 GRIN2A chr16:9763191 T>A
uncertain_significance
"canonical-transcript coding/splice annotation of
all 5.2M records"
"non-coding regulatory variants are not directly
callable from this VCF"
```

The report describes a protein-coding scan as covering all records. The strings RNU2 and 62841 never occur in the transcript.

BEFORE · GRADER/TRUTH.JSON · REQUIRED EVENT

Four heterozygous members in one gene

```
gene RNU2-2 genotype 0/1 for each member
E1 chr11:62841621 A>G E2 chr11:62841647 C>CA
E3 chr11:62841748 A>C E4 chr11:62841767 A>C
```

All four lie within 146 bp. Exact positions are required (pos_tolerance_bp: 0).

DURING · TASK/INPUTS CASE.VCF.GZ · CALLER ROWS

The calls are ordinary rows in the VCF

```
chr11 62841621 A G QUAL 222.4 GT 0/1 AD 18,16
chr11 62841647 C CA QUAL 222.0 GT 0/1 AD 17,18
chr11 62841748 A C QUAL 222.0 GT 0/1 AD 14,27
chr11 62841767 A C QUAL 221.9 GT 0/1 AD 12,24
```

Needed at this step: keep non-coding gene biotypes in the annotation set (snRNA at minimum), so that high-quality heterozygous calls in RNU2-2 stay in view. Seven other heterozygous calls sit in the same 330 bp window, so the event still has to be picked out.

AFTER · GRADER/TRUTH.JSON · CLASSIFICATION AND NOTES

One combined candidate event

```
accepted_classifications: uncertain_significance
accepted_relevance: candidate max_case_status:
candidate_finding
[E1;E2] in cis on one haplotype, [E3;E4] on the
other
```

The four loci must be reported together as one event. The VCF is unphased, so the haplotype structure comes from the source study.

WHAT HAPPENED

Fable 5.1 loaded every Ensembl gene class, then wrote an annotator that keeps only canonical protein-coding transcripts. That single filter removed RNU2-2 before any phenotype reasoning began. The panel screens, recessive checks and copy-number scans that followed were thorough within the protein-coding set and found only weak missense candidates.

TRAINING TARGET

When protein-coding analysis of a severe, unexplained phenotype yields only weak candidates, extend the search to non-coding genes before reporting. Describe the search scope in the report as it was actually run.