MAYO CLINIC

CENTER FOR INDIVIDUALIZED MEDICINE

# Translational use of multifaceted RNA-Seq bioinformatics analysis in genetic disease investigation

**Eric W. Klee, PhD**

*Director of Bioinformatics, Center for Individualized Medicine*

*Associate Professor of Biomedical Informatics*

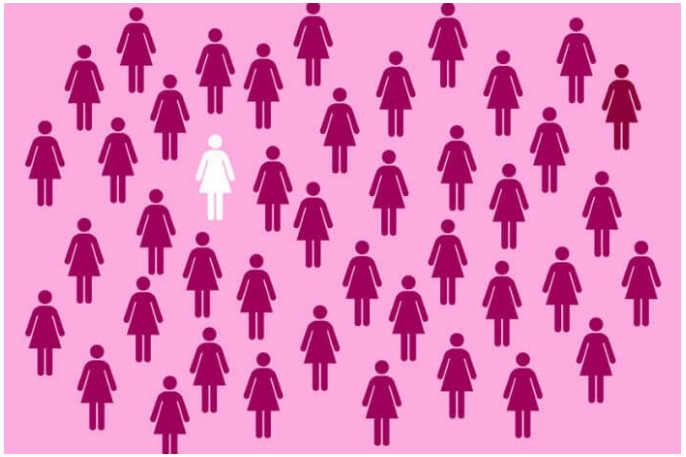
# Outline

- Discussing the role of RNA-sequencing in rare and undiagnosed disease
- Unit consists of four lectures:
  - Introduction to Rare and Undiagnosed Disease
  - Expression Analysis for Outlier Observations
  - Fusion analysis in RNA-sequencing data
  - Splicing analysis
- Lectures given by Gavin Oliver and Eric Klee

# Introduction

- What is rare genetic disease?
- A common problem - when rare isn't rare
- Rare genetic disease diagnosis in the era of next-generation sequencing
- The promise of RNA-Seq in improving rare genetic disease diagnosis

# Rare Genetic Disease



In the United States, a rare disease is defined as a condition that affects fewer than 200,000 people in the US. This definition was created by Congress in the Orphan Drug Act of 1983. There may be as many as 7,000 rare diseases. *rarediseases.info.nih.gov*

The European Union defines a disease or condition as rare if it affects fewer than 1 in 2,000 (1) people within the general population. Currently, there are over 6,000 (2) known rare diseases. *raredisease.org.uk*

**80% have a genetic component**

*Rare Genomics Institute*

# Faces of Rare Genetic Disease

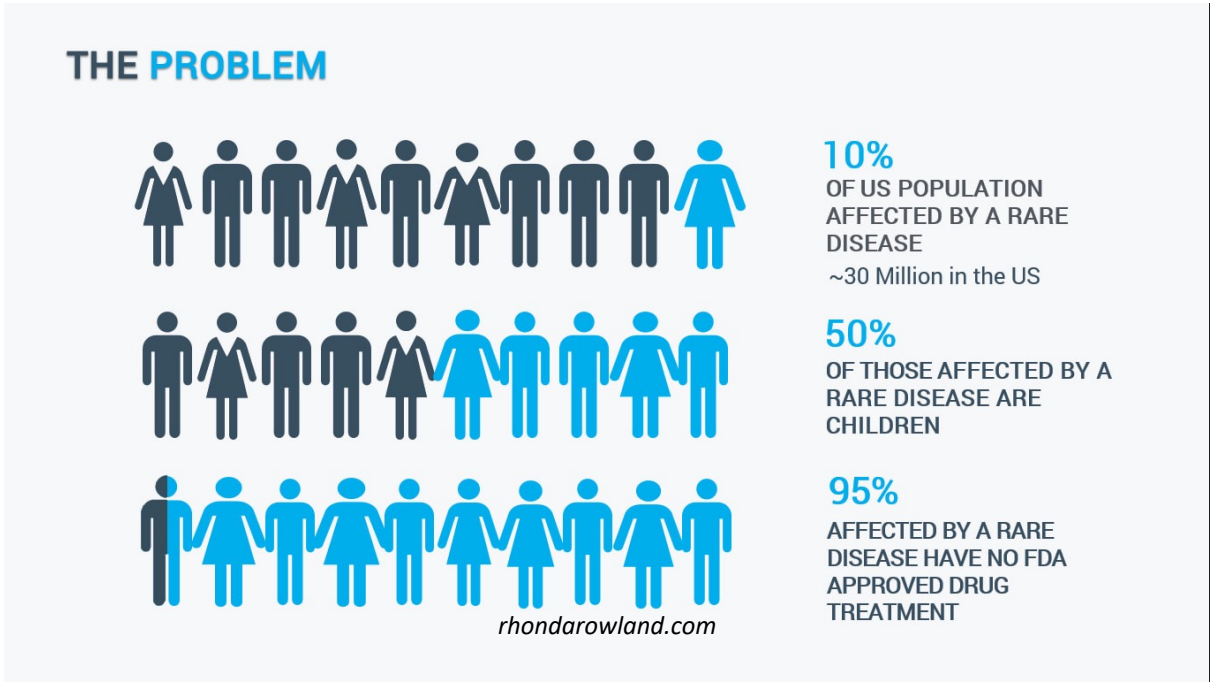



# A common problem - when rare isn't rare

An estimated **300 million people** worldwide are affected by a rare disease *globalgenes.org*

1 in 10 Americans have a rare disease *raregenomics.org*

**6% to 8% of the population** of the European Union is affected by a rare disease *eurodis.org*

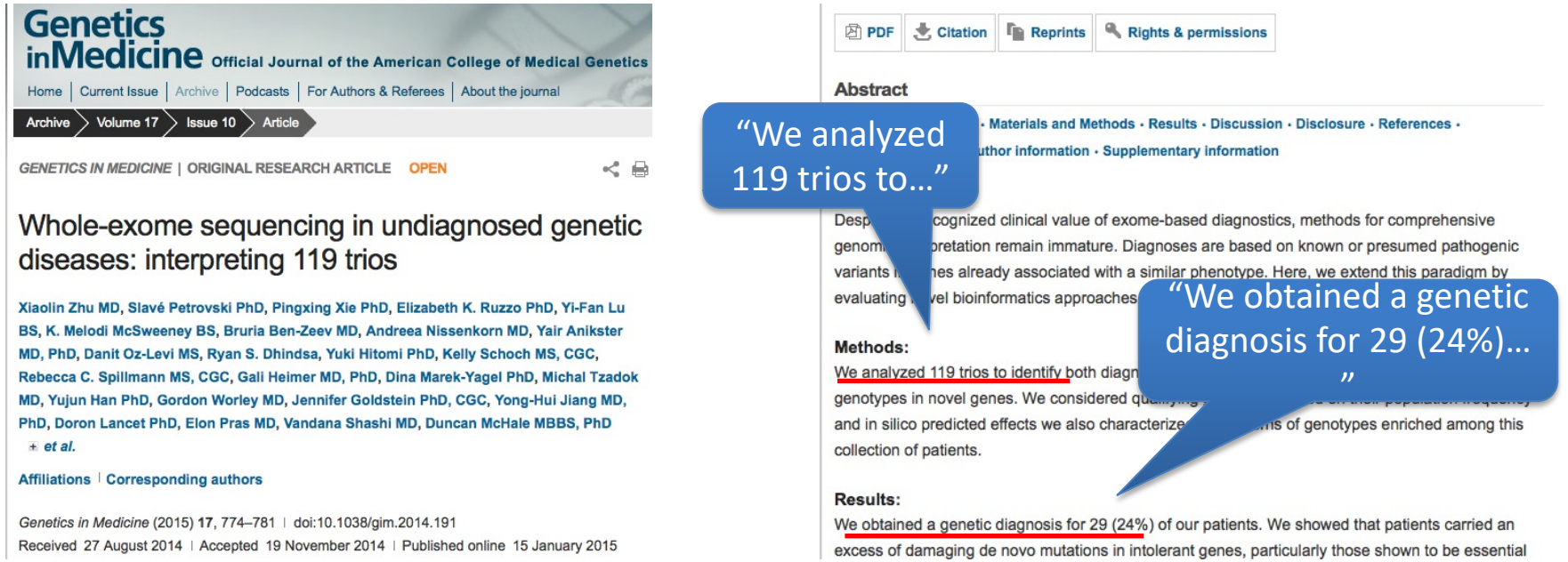



**30% of patients with rare disease will not live to see their 5th birthday**

**Rare diseases are responsible for 35% of deaths in the first year of life**

*raregenomics.org*

# Proliferation of Exome Diagnostic Testing



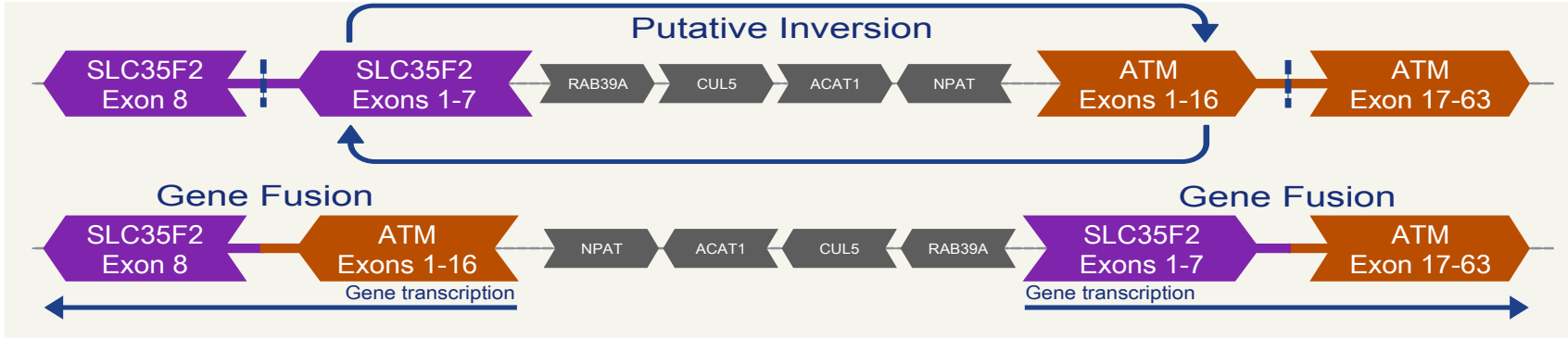

**Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders - JAMA**

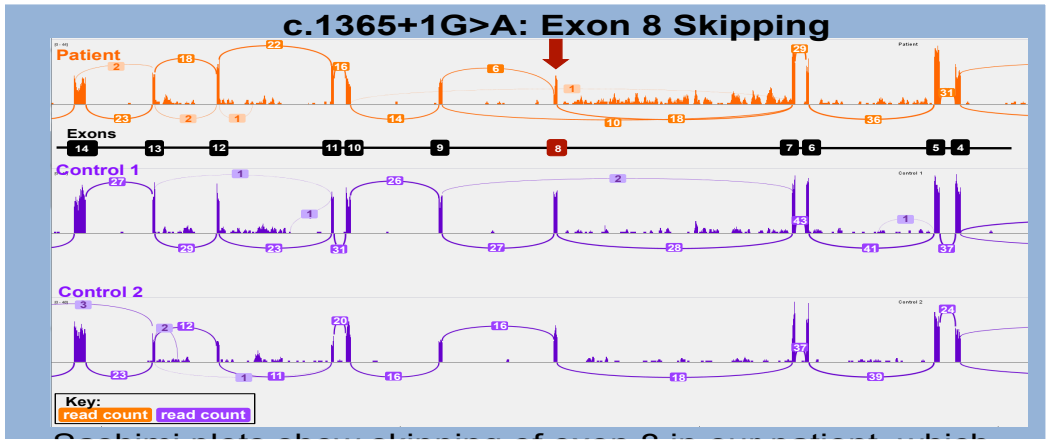
"**Results—**Of the 814 cases, the overall molecular diagnosis rate was 26%"

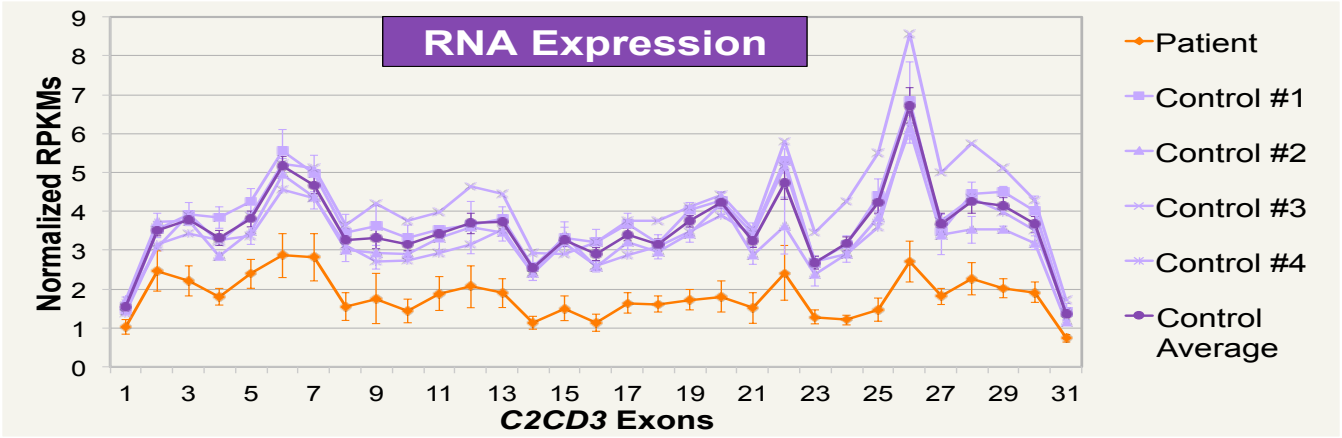
**Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation – NEJM**

"A molecular diagnosis was rendered for 2076 of 7374 patients (28.2%)"

# RNA Sequencing to Improve Diagnostic Rate



# Case Examples

1. Patient #1: GNPTAB – cryptic splicing
2. Patient #2: ATM – expressed fusion
3. Patient #3: CASK – outlier expression
4. Patient #4: SGSH – allele-specific expression

# Patient Example Case #1

Age: 9y female

Reason for Referral: Mild global developmental delay, brain MRI abnormal joint contractures, slightly distinctive facial features

Clinical Testing reported compound heterozygous variants in MEGF10, which upon further review did not seem a good fit.

**Requested raw data from the testing provider and reanalyzed**

| Label | ID | Meta | ID |
|---|---|---|---|
| Wide pubic symphysis | HP:0003183 | | |
| Vertebral hypoplasia | HP:0008417 | | |
| Short stature | HP:0004322 | | |
| Posterior scalloping of vertebral bodies | HP:0005121 | | |
| Platyspondyly | HP:0000926 | | |
| Periorbital fullness | HP:0000629 | | |
| Pectus carinatum | HP:0000768 | Mild | HP:0012825 |
| Narrow forehead | HP:0000341 | | |
| Mitral valve prolapse | HP:0001634 | Mild | HP:0012825 |
| Mitral regurgitation | HP:0001653 | Mild | HP:0012825 |
| Lumbar hyperlordosis | HP:0002938 | | |
| Intellectual disability | HP:0001249 | | |
| Hypoplastic distal radial epiphyses | HP:0006386 | | |
| Global developmental delay | HP:0001263 | Mild | HP:0012825 |
| Gastroesophageal reflux | HP:0002020 | | |
| Flexion contracture | HP:0001371 | | |
| Flattened humeral heads | HP:0003888 | | |
| Epicanthus | HP:0000286 | | |
| Dysarthria | HP:0001260 | | |
| Coarse facial features | HP:0000280 | Mild | HP:0012825 |
| Cafe-au-lait spot | HP:0000957 | | |
| Broad nasal tip | HP:0000455 | | |
| Beaking of vertebral bodies | HP:0004568 | | |
| Asymmetry of the ears | HP:0010722 | | |
| Aortic regurgitation | HP:0001659 | Mild | HP:0012825 |
| Abnormality of the skeletal system | HP:0000924 | | |
| Abnormality of the glenoid fossa | HP:0011912 | | |

# Candidate Splicing Variant Identified

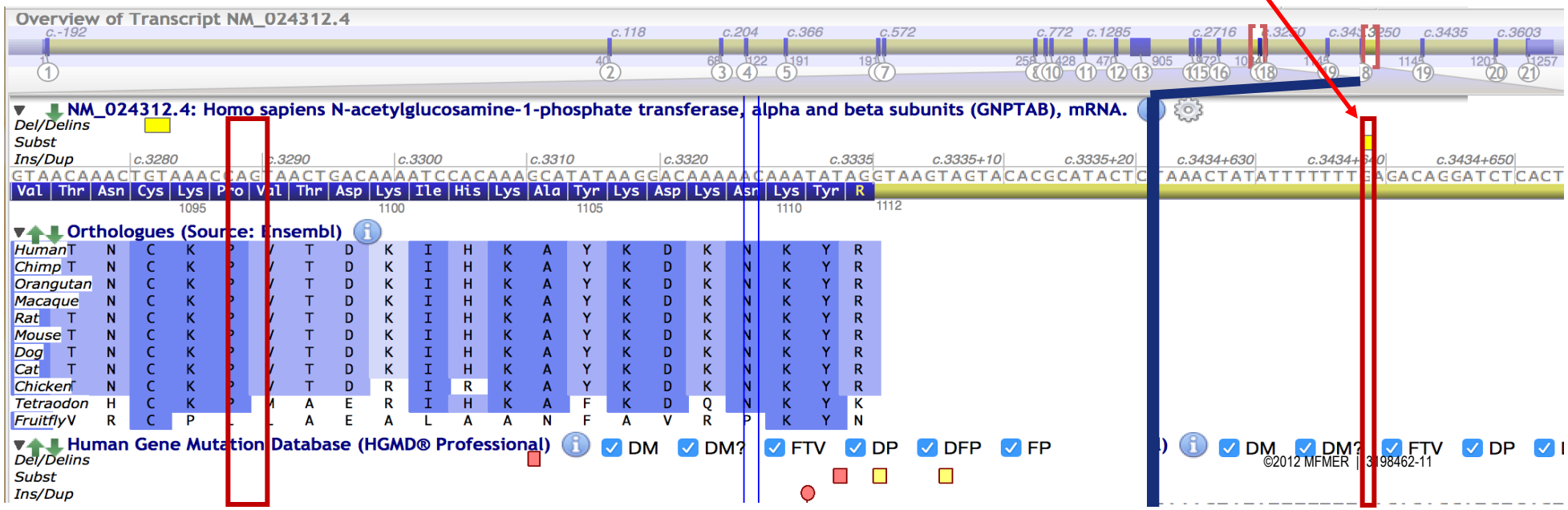
***GNPTAB*** **TWO HET c.3281_3282delGT, p.C1094fs* – mother is het, father is neg**
**c.3434+639G>C, p.?splice – de novo, parents are neg**

**Disease:** Mucolipidosis alpha/beta **AR** type II (MIM:252500) or III (MIM:252600)
**Comments:** Encodes two of three subunit types of the membrane-bound enzyme N-Acetylglucosamine-1-Phosphate Transferase Alpha And Beta Subunits. a heterohexameric complex composed of two alpha, two beta, and two gamma subunits. The encoded protein is proteolytically cleaved at the Lys928-Asp929 bond to yield mature alpha and beta polypeptides while the gamma subunits are the product of a distinct gene (GeneID 84572). In the Golgi apparatus, the heterohexameric complex **catalyzes the first step in the synthesis of mannose 6-phosphate (M6P) recognition markers on certain oligosaccharides of newly synthesized lysosomal enzymes**, which mediate vesicular transport of lysosomal enzymes to the endosomal/prelysosomal compartment.

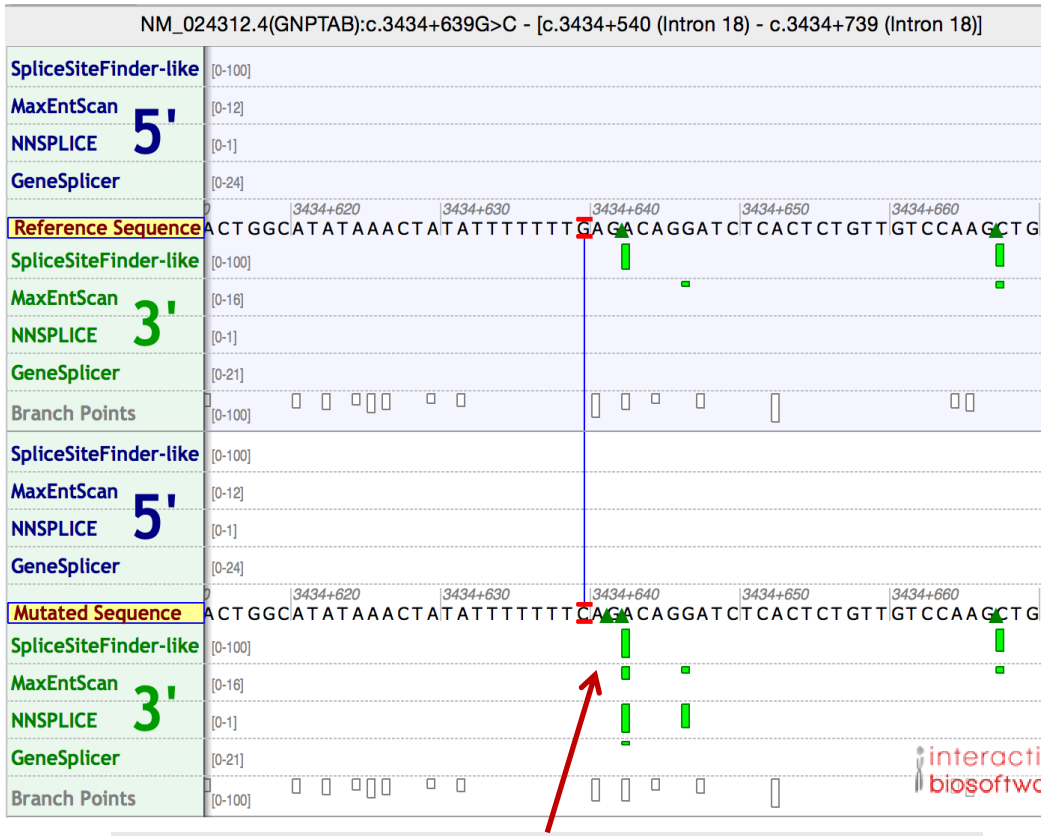
| | c.3281_3282delGT, p.C1094fs* - mat | c.3434+639G>C, p.?splice – de novo |
|---|---|---|
| gnomAD: | NR | NR |
| In silico | NA | NA |
| Location: | Exon 17 of 21 | Intron 18 |

A deep intronic variant was found in trans with a loss of function frame-shift variant.

If the intronic variant impacts gene splicing it may explain the patients phenotype



# Splicing predictor identifies putative donor site




Patient variant creates spice acceptor

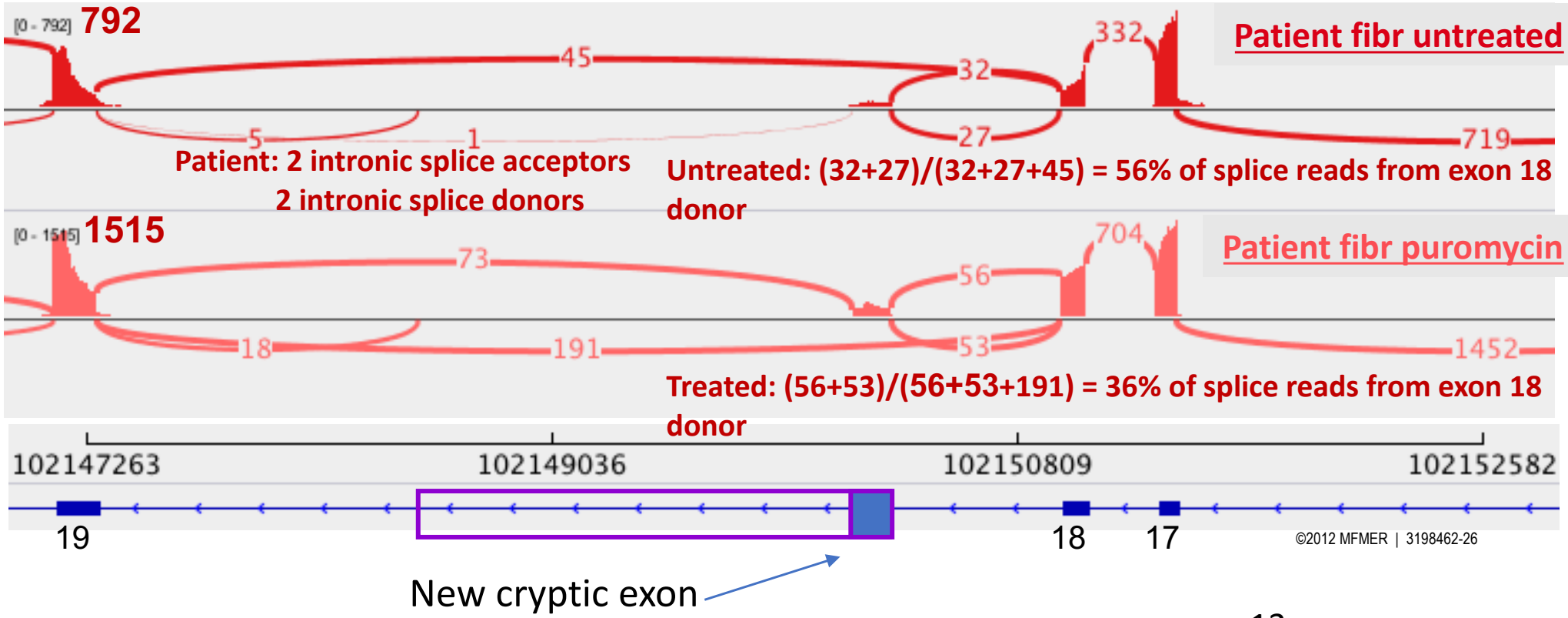

Cryptic splice donor

The intronic variant creates a predicted splice acceptor at c.3434+642 and there is a cryptic splice donor at 3434+1244

# RNAseq confirms presence of a cryptic exon

***GNPTAB* TWO HET c.3281_3282delGT, p.C1094fs* – mother is het, father is neg**
**c.3434+639G>C, p.?splice – de novo, parents are neg**
**Disease:** Mucolipidosis alpha/beta **AR** type II (MIM:252500) or III (MIM:252600)

Read Depth



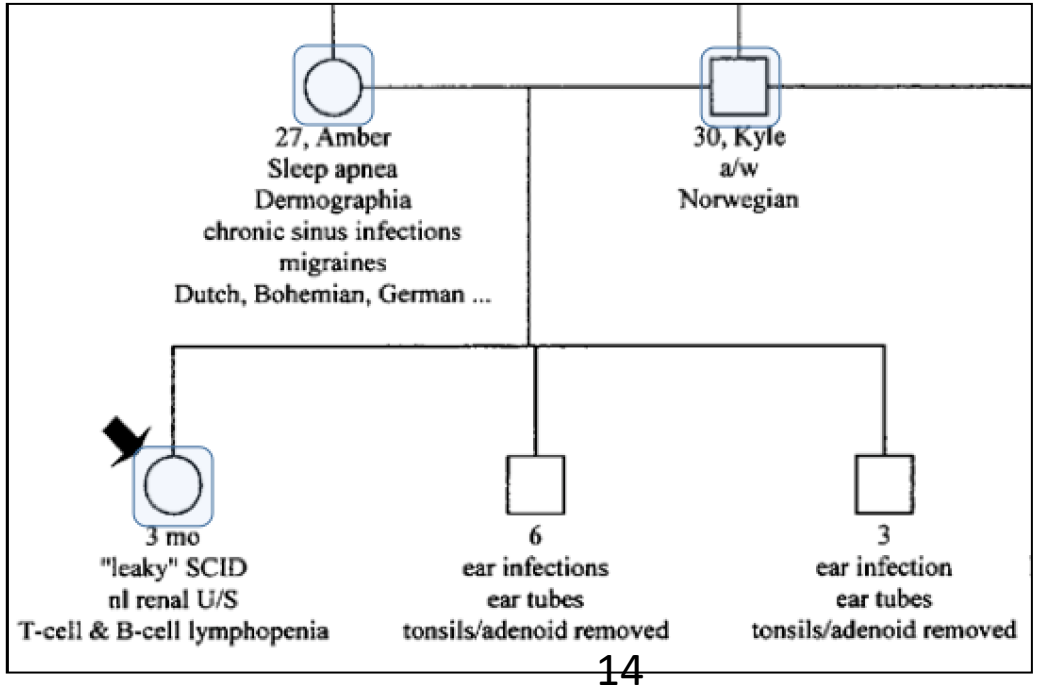
# Patient Example Case #2

**Patient Symptoms:**

- Tested positive for SCID by newborn screening
- Evidence of radiation sensitivity (partial)
- B and T-cell lymphopenia (T-B-NK+)
- TCR vbeta spectratyping is suggestive of polyclonal gaussian repertoire which is most likely suggestive of a combined immune deficiency
- Thrombocytosis
- Some ataxia-related phenotype potentially manifesting recently
- Elevated AFP

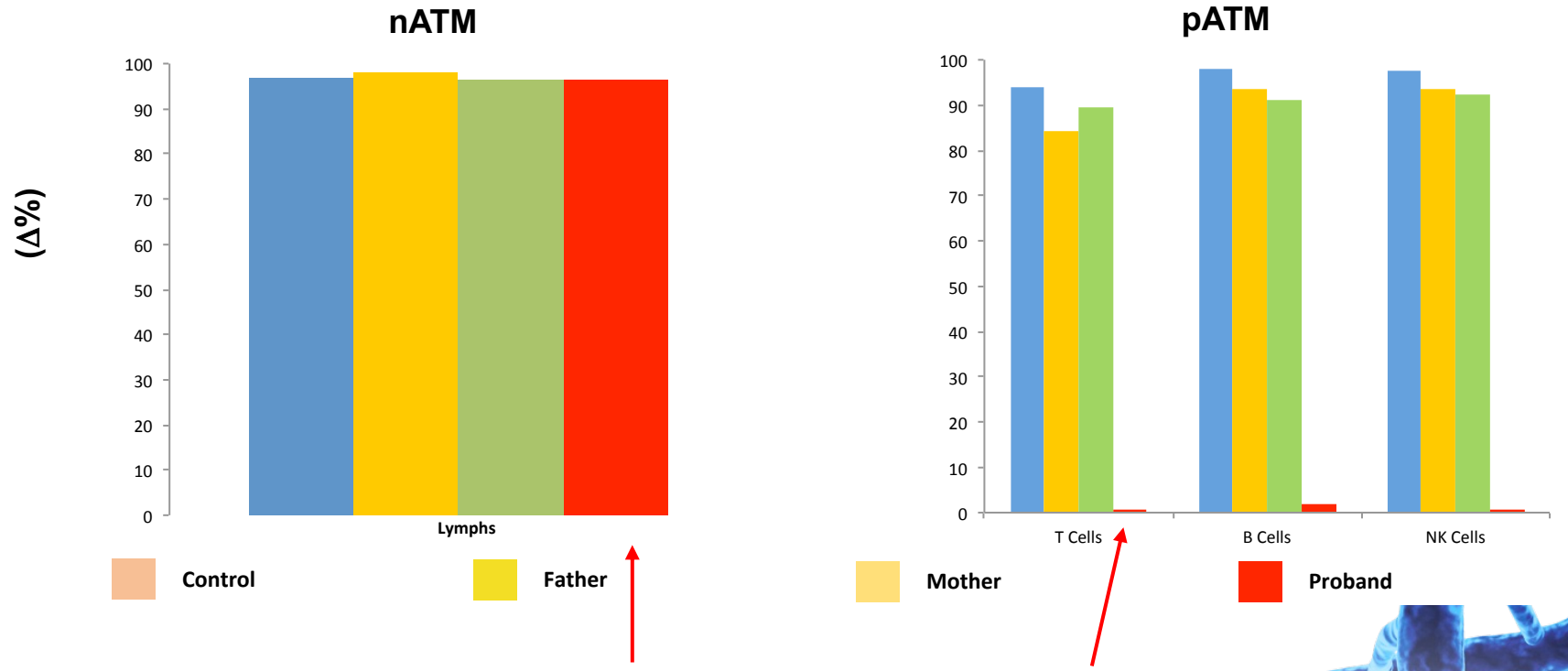
**Single pathogenic variant in ATM (recessive condition) that would explain patient symptoms**

| Disease | Inher. pattern | Gene | Location | variant | inherited from |
|---|---|---|---|---|---|
| Ataxia Telagiectasia 208900 | AR | ATM | chr11:108143540_ 108143542 | c.3245_3247delinsTGAT p.His1082Leufs*14 | Dad het |



# Protein testing to evaluate diagnosis

Protein studies in proband and parents showed that the ATM protein was present in all three individuals, but non-functional in the proband




**Patient protein present, but not functional**

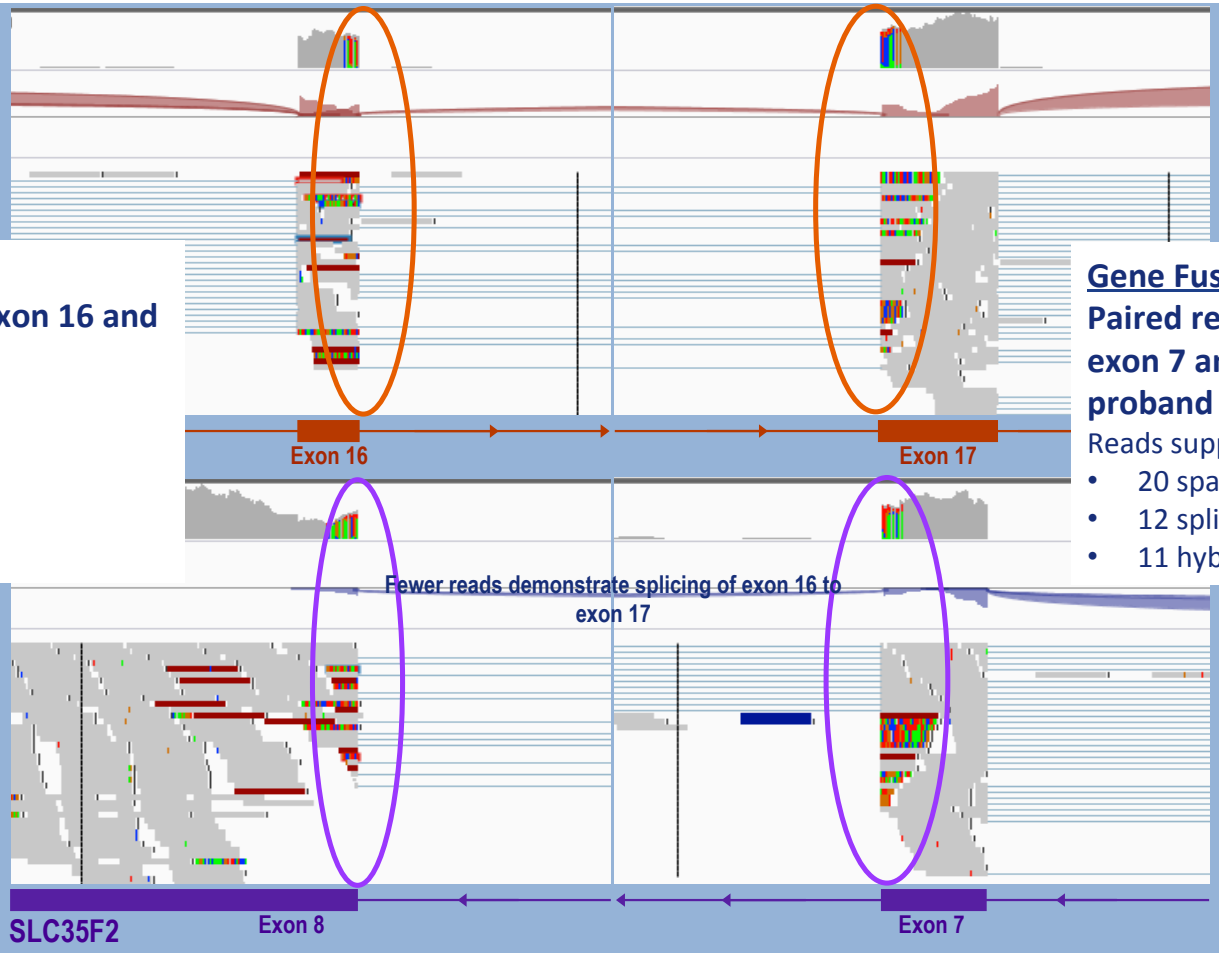
# RNAseq identifies a gene fusion in ATM

**Gene Fusion Analysis:**

**Paired reads mapping to ATM exon 16 and SLC35F2 exon 8 in proband RNA**

Reads supporting fusion:

- 7 spanning reads
- 3 split reads
- 4 hybrid reads



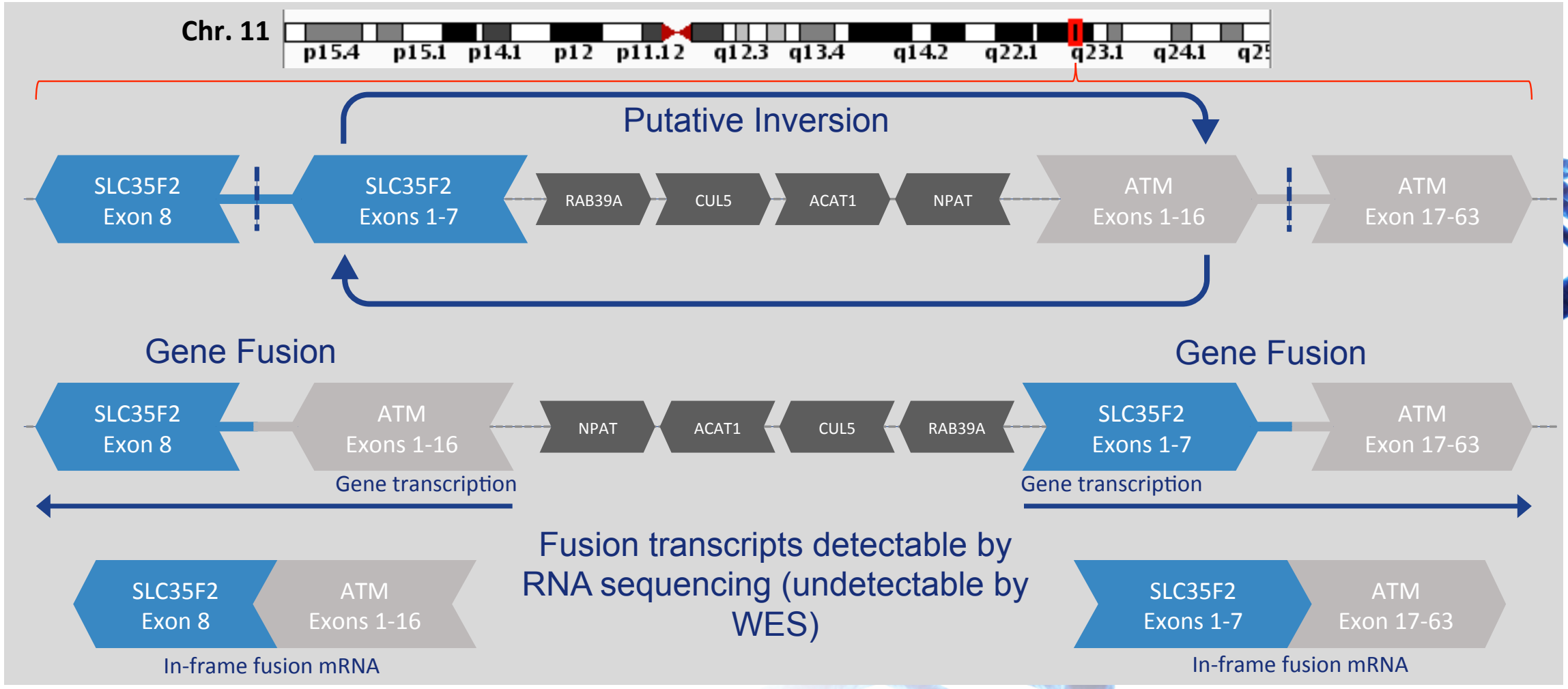

**Gene Fusion Analysis:**

**Paired reads mapping to SLC35F2 exon 7 and ATM exon 17 and in proband RNA**

Reads supporting fusion:

- 20 spanning reads
- 12 split reads
- 11 hybrid reads

# Fusion caused by DNA inversion

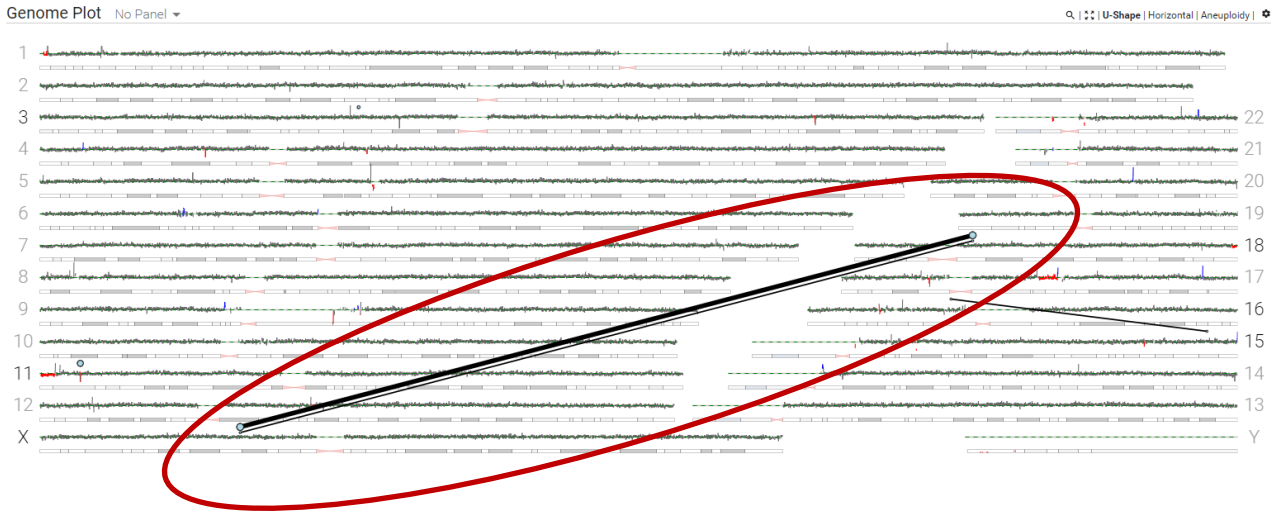


# Patient Example Case #3

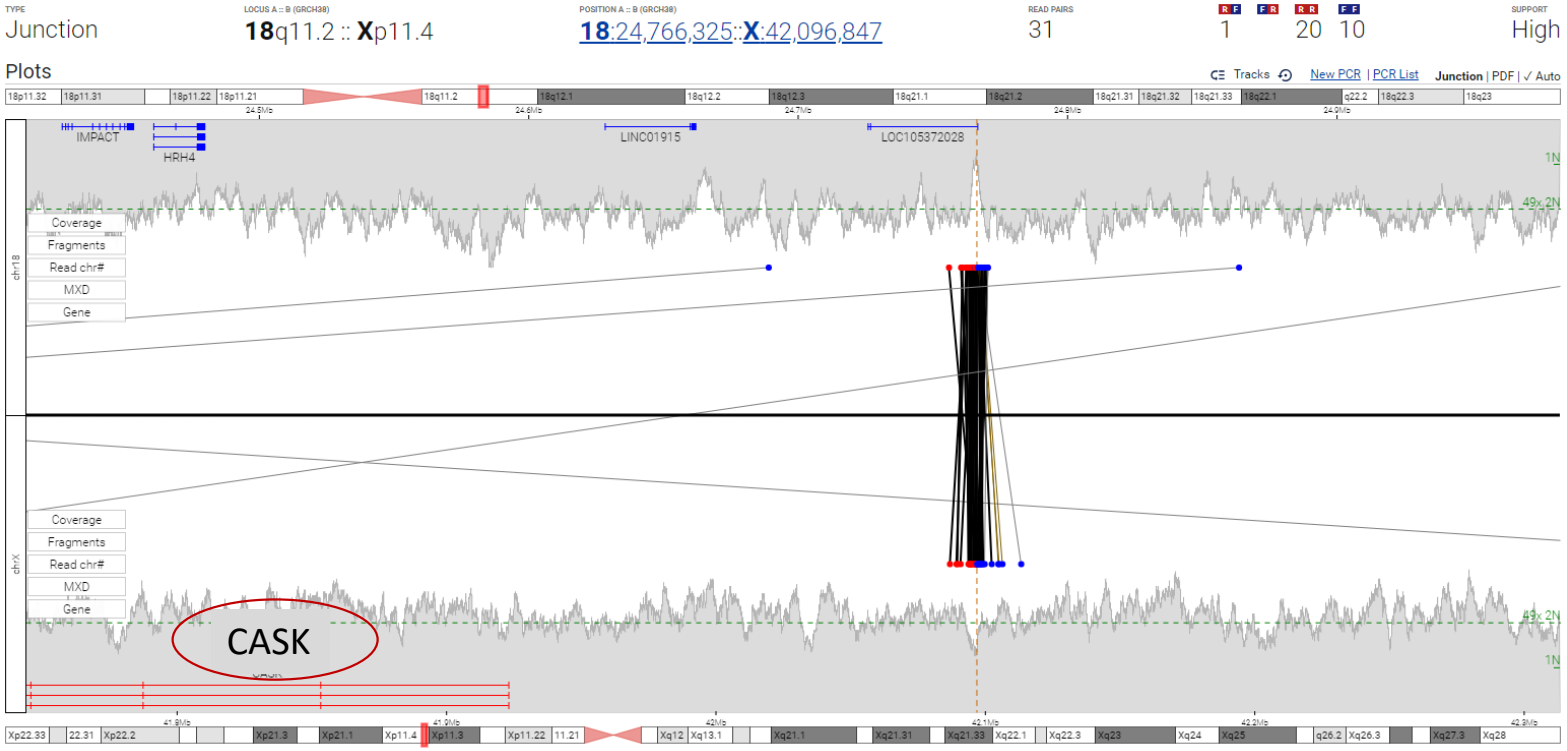
**Patient Symptoms:**

- 12 month old female
- GNAO1 related epileptic encephalopathy with cerebral and cerebellar atrophy
- corpus callosum dysgenesis
- left optic nerve hypoplasia
- Micropthalmia
- Cataract
- cleft soft palate
- ASD
- dysmorphic facies
- small size
- microcephaly

Mate-pair sequencing to investigate karyotype:
t(X;18)(p11.2;q11.2)



# Mate-pair characterizes t(X;18)(p11.2;q11.2)




**Chr18:**

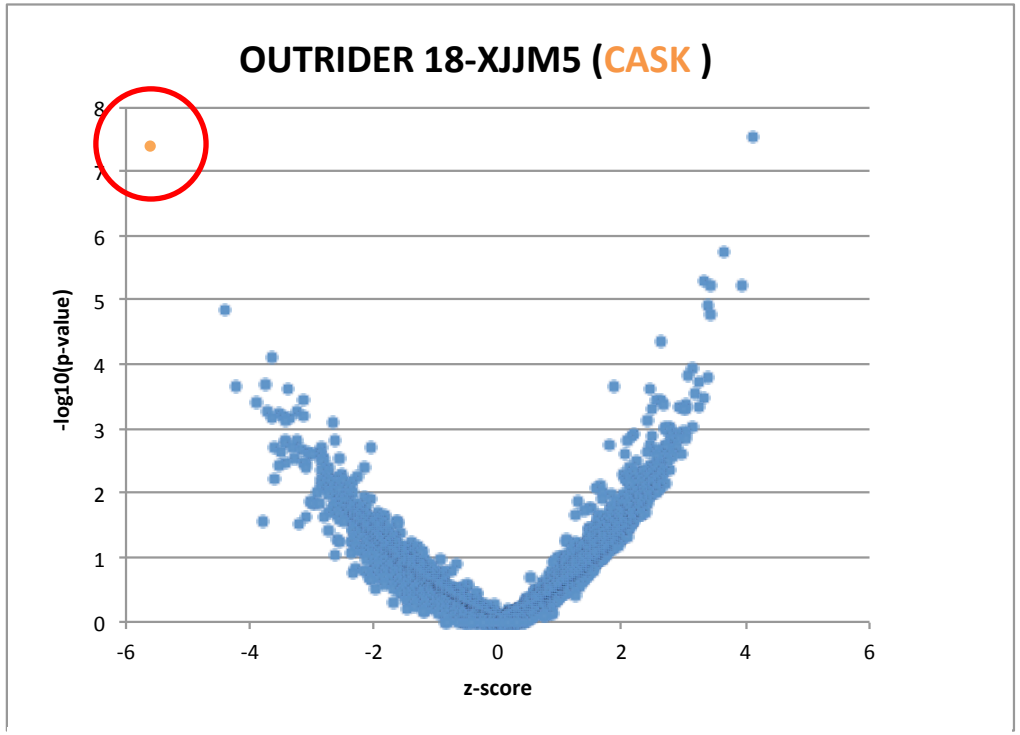
Disrupts LOC105372028 (non-coding RNA); haploinsufficiency not known to result in any abnormal phenotypes

**ChrX:**

No gene was disrupted at the Xp11.4 breakpoint, however CASK is located ~173 kb distal

**CASK deficiencies are a strong phenotypic fit for the patient's symptoms**

# RNAseq Outlier Expression (CASK)




*CASK:*

p-value (unadjusted) = $4.08 \times 10^{-8}$

p-value (adjusted) = 0.0003

Z-score = -5.61

$Log_2$(fold change) = -0.99

## Epigenetic Profiling

Hypothesis: epigenetic gene repression on chromosome 18 near the breakpoint that has been put upstream of CASK in this patient from the t(X;18) that is now causing repression of CASK expression.

Evaluate: H3K9me, H3K27me, H3K4me and H2K9Ac (activating mark) to see if these features are present near the breakpoints in blood lineage cells of unaffected individuals.

ChIP-seq experiment on the patient's blood sample to determine if there is a difference in the patient.

OUTRIDER: Brechtmann F, et al. Am J Hum Genet. 2018 Dec 6;103(6):907-917.
PMID: 30503520

# Patient Example Case #4

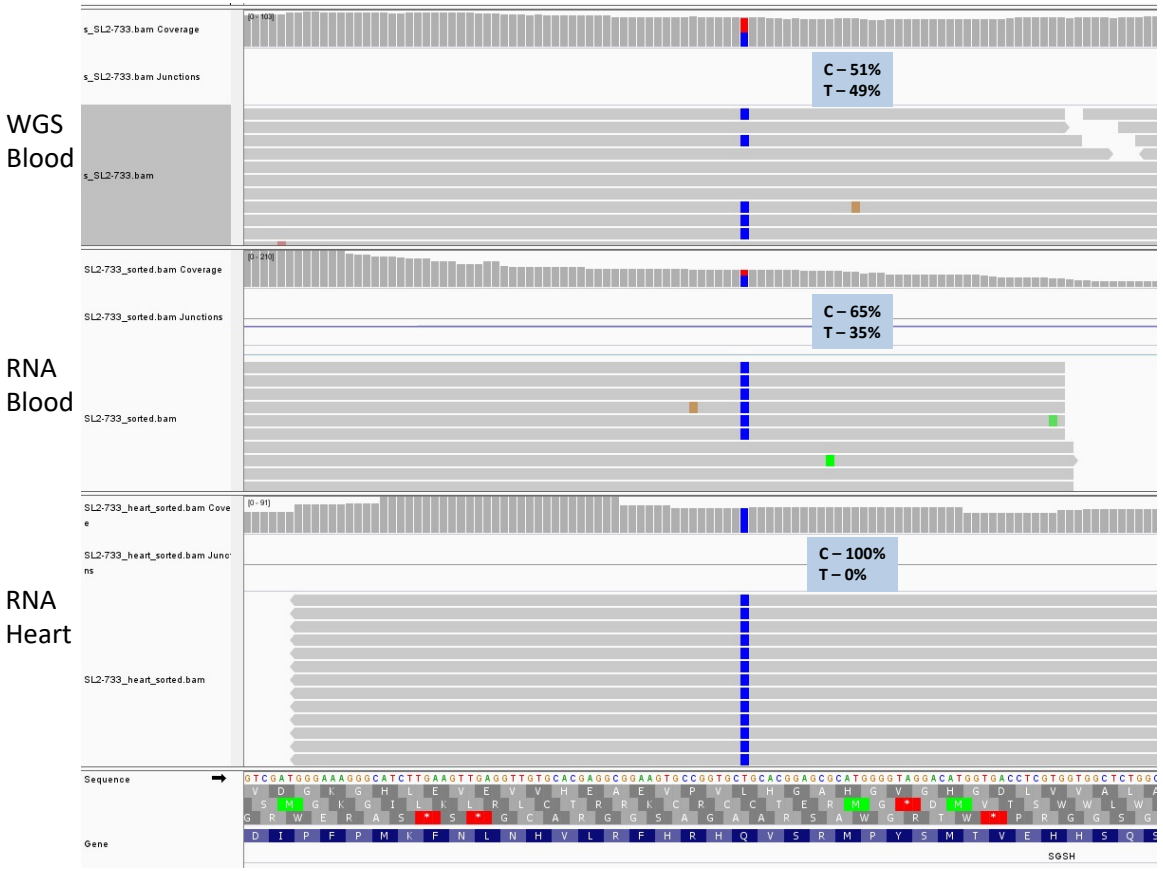
## Phenotype

- Adult-onset hepatomegaly
- Hypertrophic cardiomyopathy
- Stroke (2014)
- A. Fib
- Non-dysmorphic

**Glycosaminoglycans in Cerebrospinal Fluid:**
**GAGs in the CSF are 50-100 times higher than controls in others MPS with neurological compromise**

## Mucopolysaccharidosis type IIIA

- Mucopolysaccharidosis type III A (Sanfilippo syndrome type A; MPS IIIA) is characterized by psychomotor and speech delay, neurological regression, and behavioral disturbances.
- Somatic changes are usually milder than other MPSs, and include mild coarsening, mild dysostosis multiplex, and contractures. About half of patients have hepatomegaly but splenomegaly is infrequent.
- MPS IIIA is caused by biallelic pathogenic variants in *SGSH* resulting in a deficiency of sulfamidase, a lysosomal enzyme. The inability to degrade heparan sulfate leads to cellular accumulation of this glycosaminoglycan and increased excretion in urine.
- There is only one report of a patient with sulfamidase deficiency, increased excretion of heparan sulfate, and late-onset cardiomiopathy without neurological phenotype. However, focused genetic analysis failed to identify variants in *SGSH*. (Van Hove et al, 2003)




# RNA Sequencing Reveals Allele-Specific Expression

ASE was noted initially in the blood, with a modest 65%-35% skew. Expression studies in a heart biopsy revealed complete skew 100%-0%.

# Conclusions

Rare genetic disease can be individually rare, but collectively is quite common

NGS has transformed how rare genetic disease is tested, enabled considerably higher diagnostic rates and novel disease gene discovery

RNAseq can increase overall diagnostic yield in this patient population

RNAseq analysis is complex and involves looking at multiple event types

# Questions