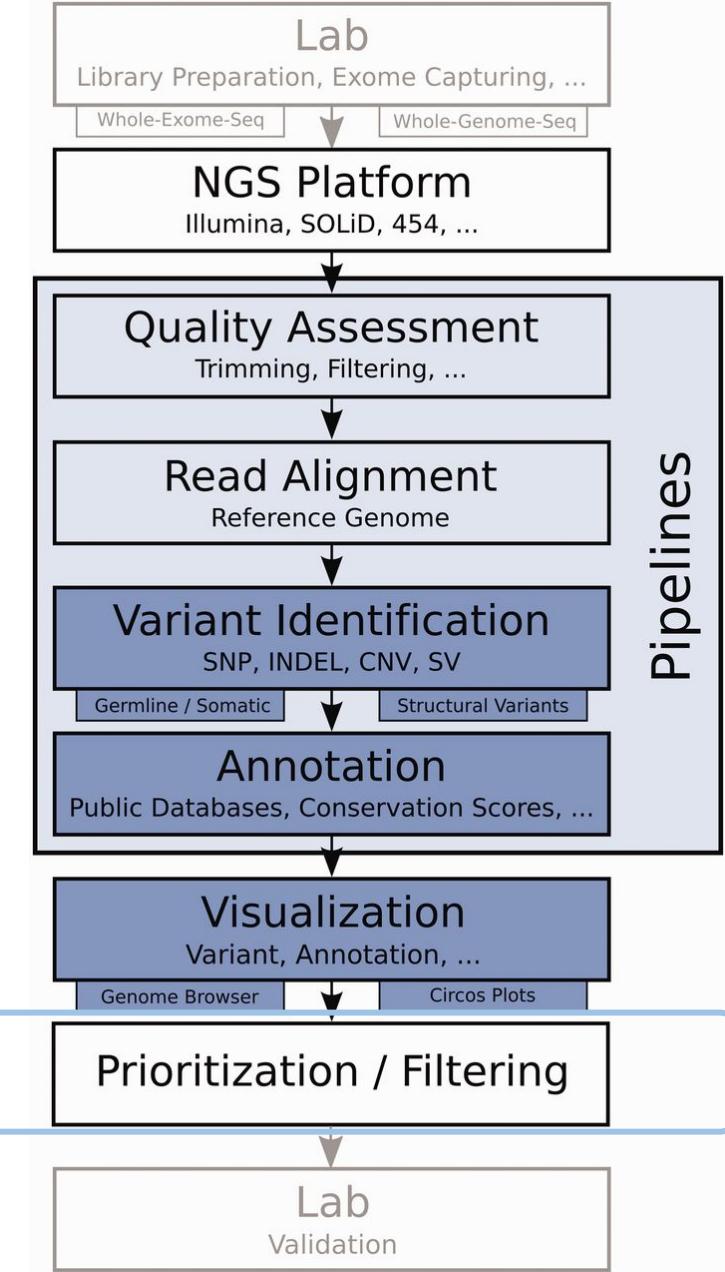


# PO21: Precision Oncology Course

## Variant Prioritization

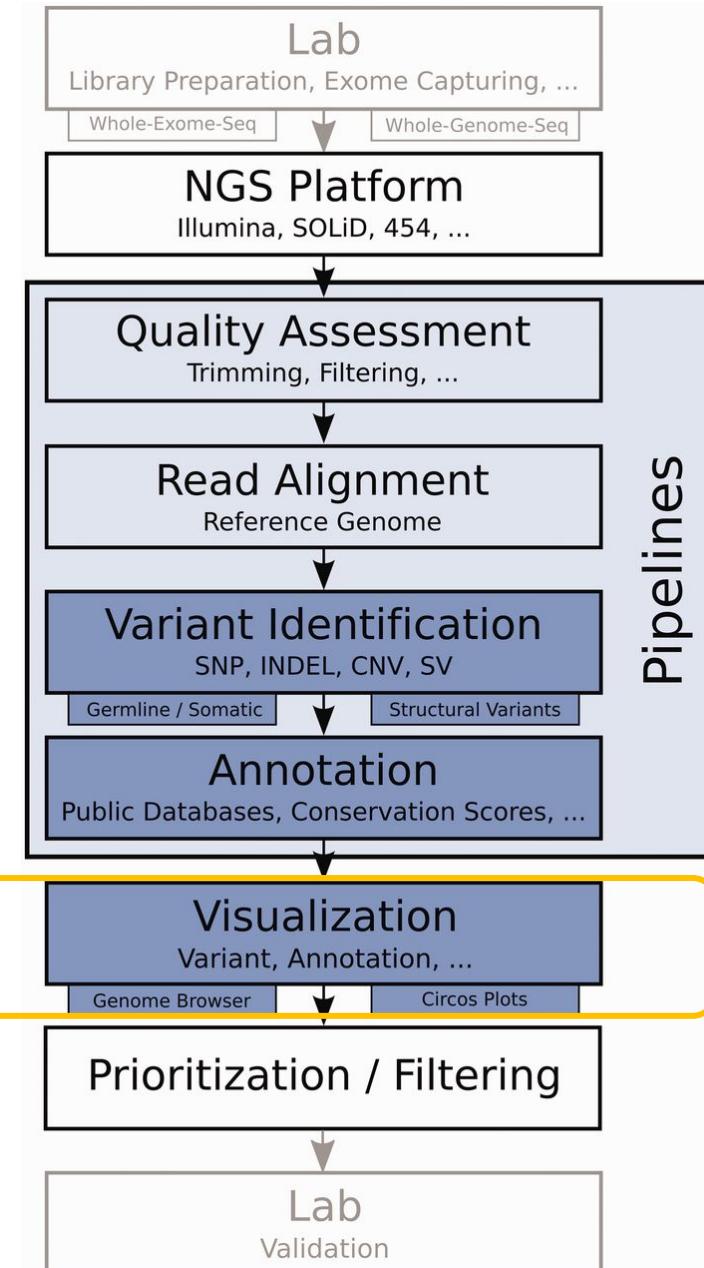
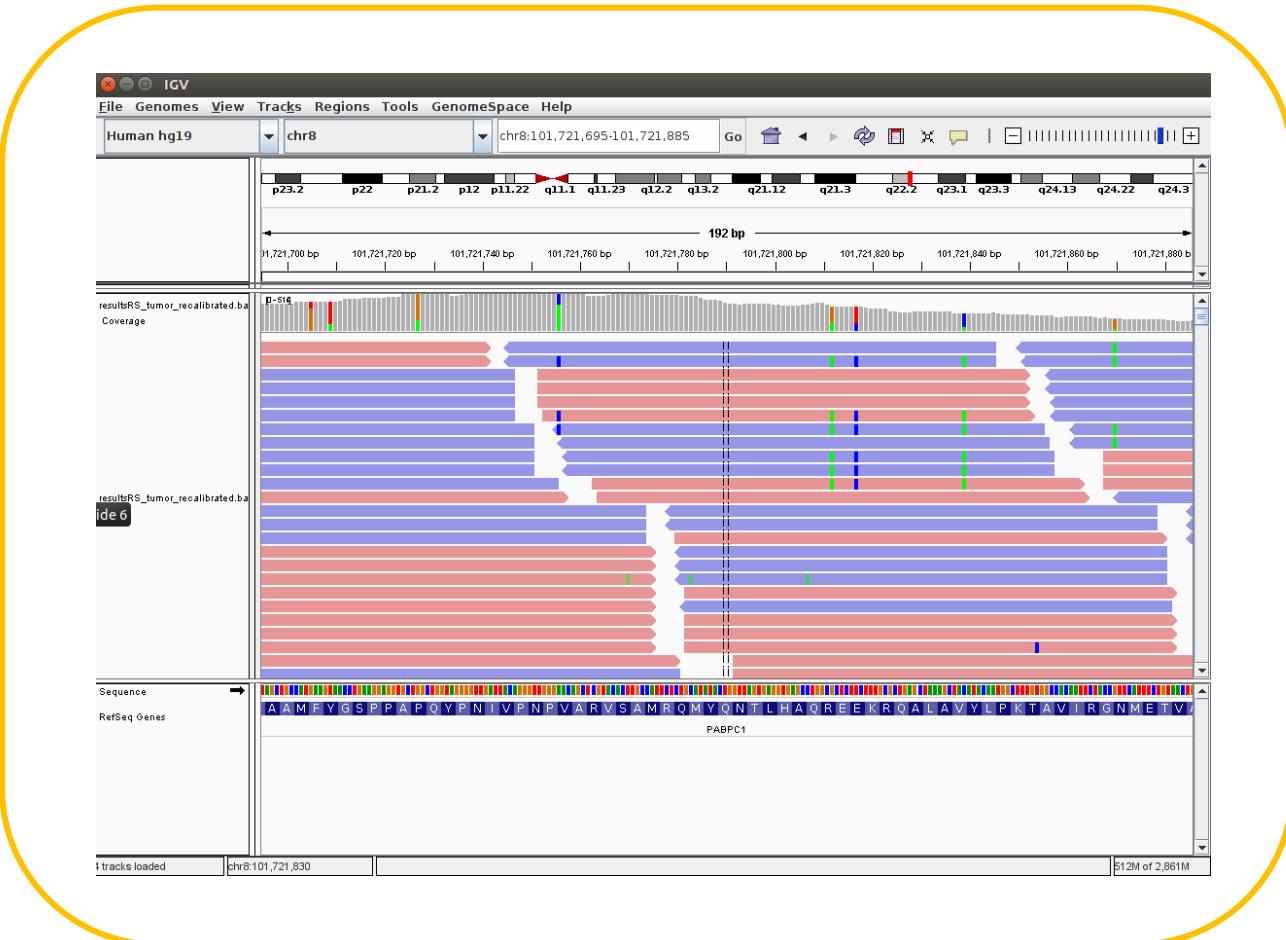
# Prioritization



Help in the identification of:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5678989/>

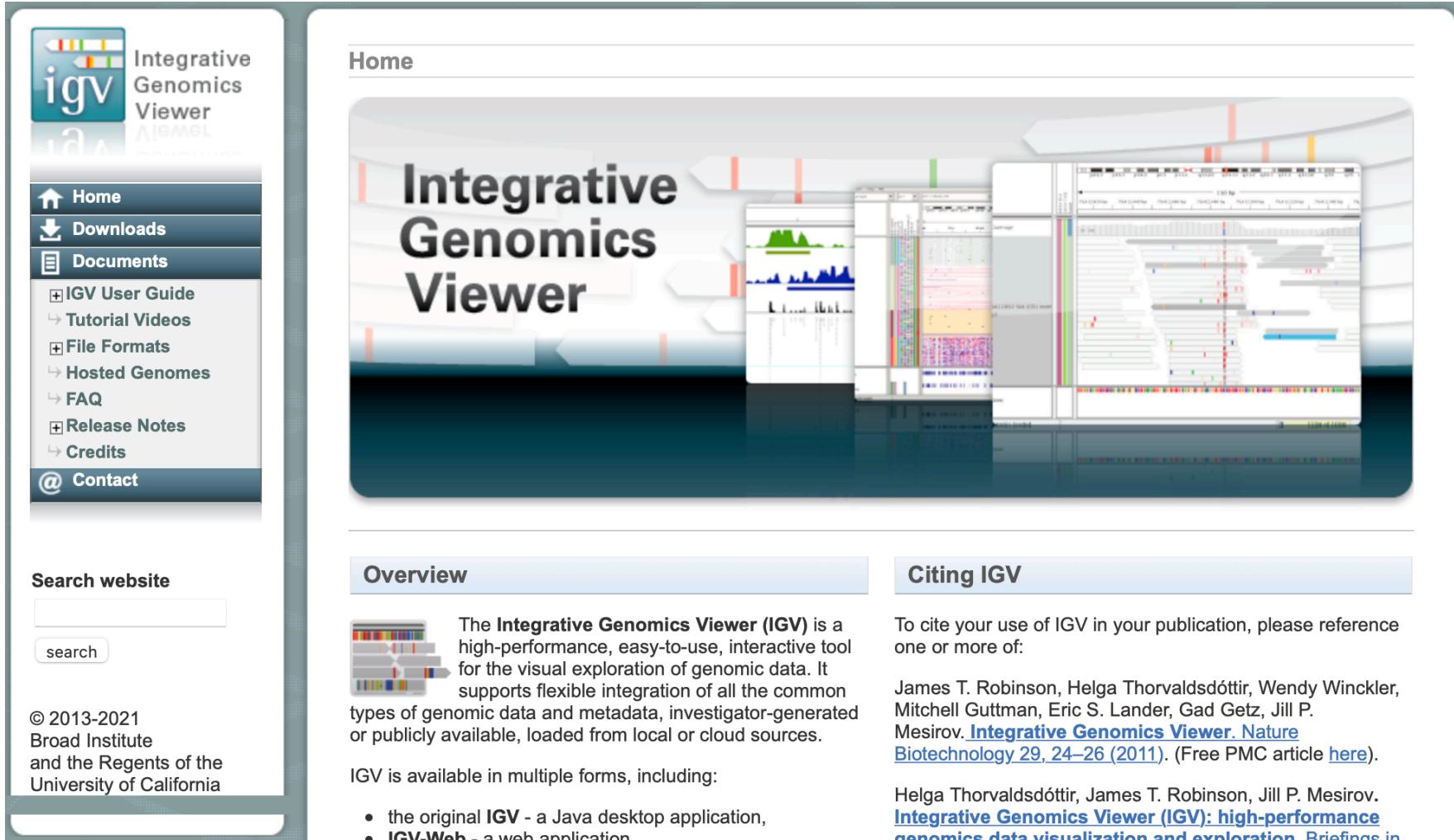
- **False Positives (artifacts)**: sequencing and processing errors
- **False Negatives**: in low depth of coverage regions or low frequency variants



# Identify possible artifacts

- **Using IGV visualization**

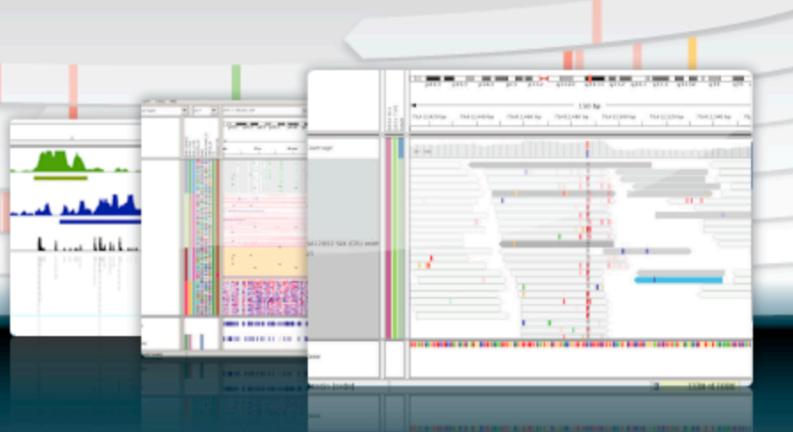
[Variant Review with the Integrative Genomics Viewer \(IGV\). Cancer Research 77\(21\) 31-34 \(2017\)](#)



The screenshot shows the official website for the Integrative Genomics Viewer (IGV). The left sidebar contains a navigation menu with links to Home, Downloads, Documents, and various informational sections like User Guide, Tutorial Videos, File Formats, Hosted Genomes, FAQ, Release Notes, Credits, and Contact. Below the menu is a search bar labeled "Search website" with a "search" button. At the bottom of the sidebar, there's a copyright notice: "© 2013-2021 Broad Institute and the Regents of the University of California". The main content area features a large banner with the text "Integrative Genomics Viewer" and a screenshot of the software's graphical user interface, which displays multiple tracks of genomic data. Below the banner are two sections: "Overview" and "Citing IGV". The "Overview" section includes a small icon of the IGV logo and a detailed description of what the tool is and how it works. It also lists the available forms of the software. The "Citing IGV" section provides instructions for citing the tool in publications, mentioning specific articles and PMC articles.

**Home**

# Integrative Genomics Viewer



**Overview**

The **Integrative Genomics Viewer (IGV)** is a high-performance, easy-to-use, interactive tool for the visual exploration of genomic data. It supports flexible integration of all the common types of genomic data and metadata, investigator-generated or publicly available, loaded from local or cloud sources.

IGV is available in multiple forms, including:

- the original **IGV** - a Java desktop application,
- IGV-Web** - a web application

**Citing IGV**

To cite your use of IGV in your publication, please reference one or more of:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 \(2011\)](#). (Free PMC article [here](#)).

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration Briefings in](#)

# Visual review of alignments

Review | Open Access | Published: 26 October 2020

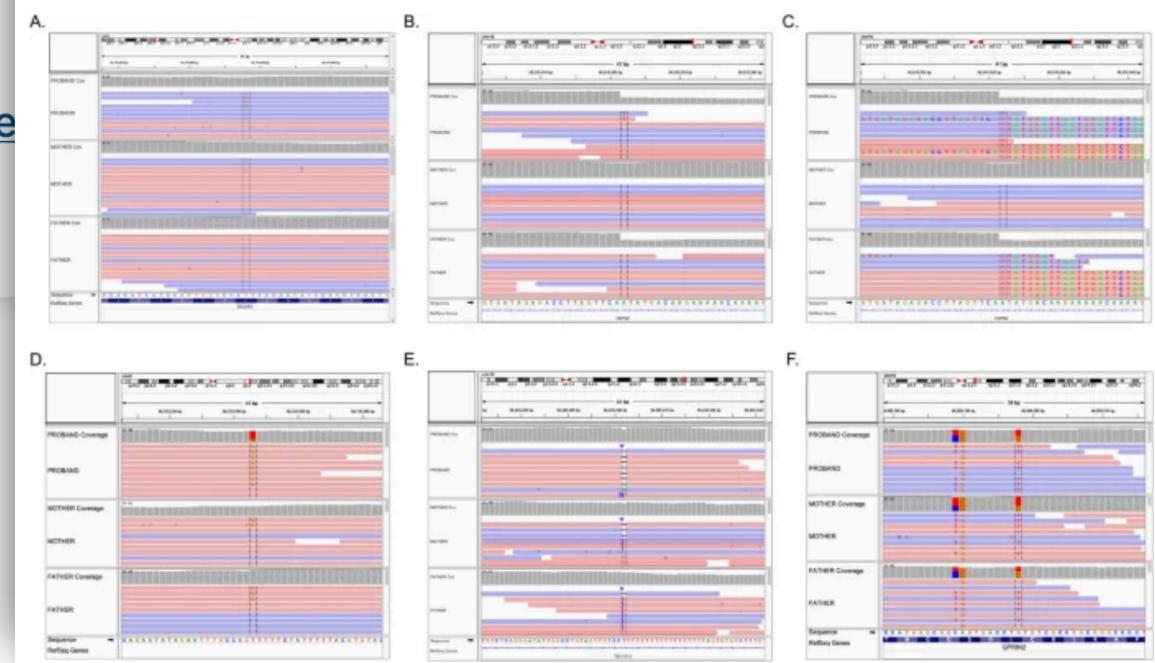
## Best practices for variant calling in clinical sequencing

Daniel C. Koboldt 

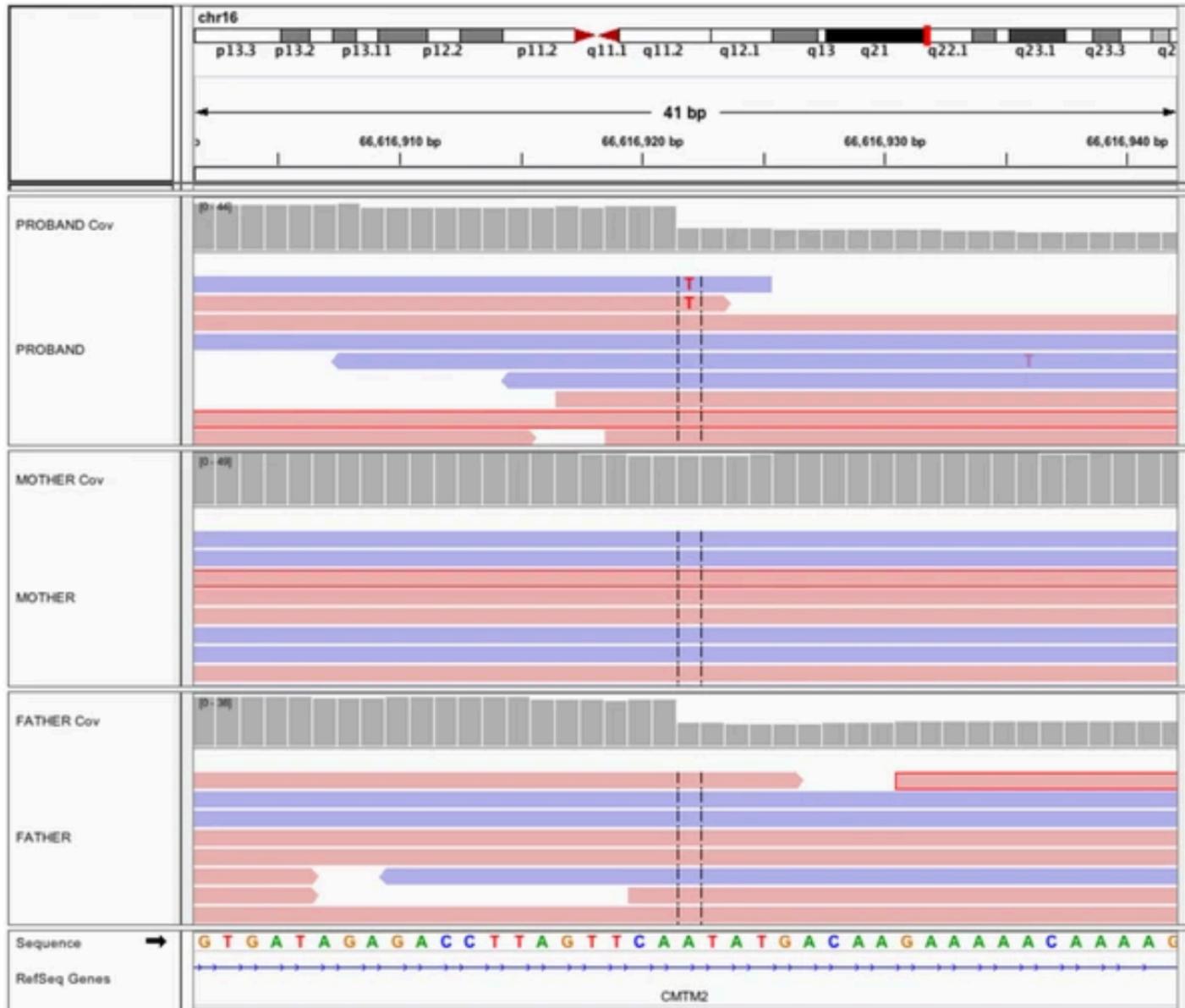
*Genome Medicine* 12, Article number: 91 (2020) | [Cite this article](#)

32k Accesses | 10 Citations | 15 Altmetric | [Metrics](#)

**Fig. 2**



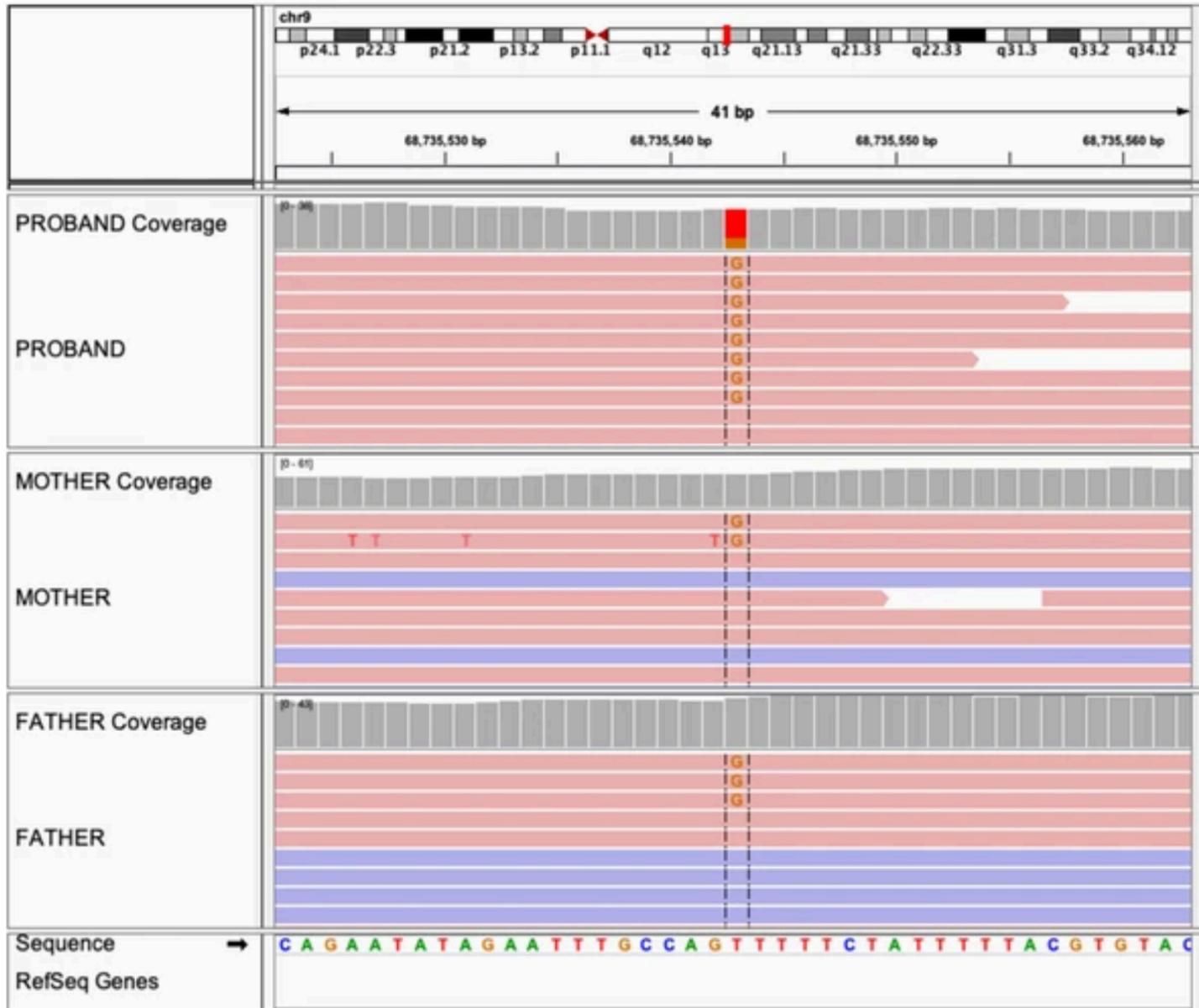
# Visual review of alignments



**False positive due to misalignments near the start or end of reads**

alternate allele is only observed at the start/end of reads in the proband

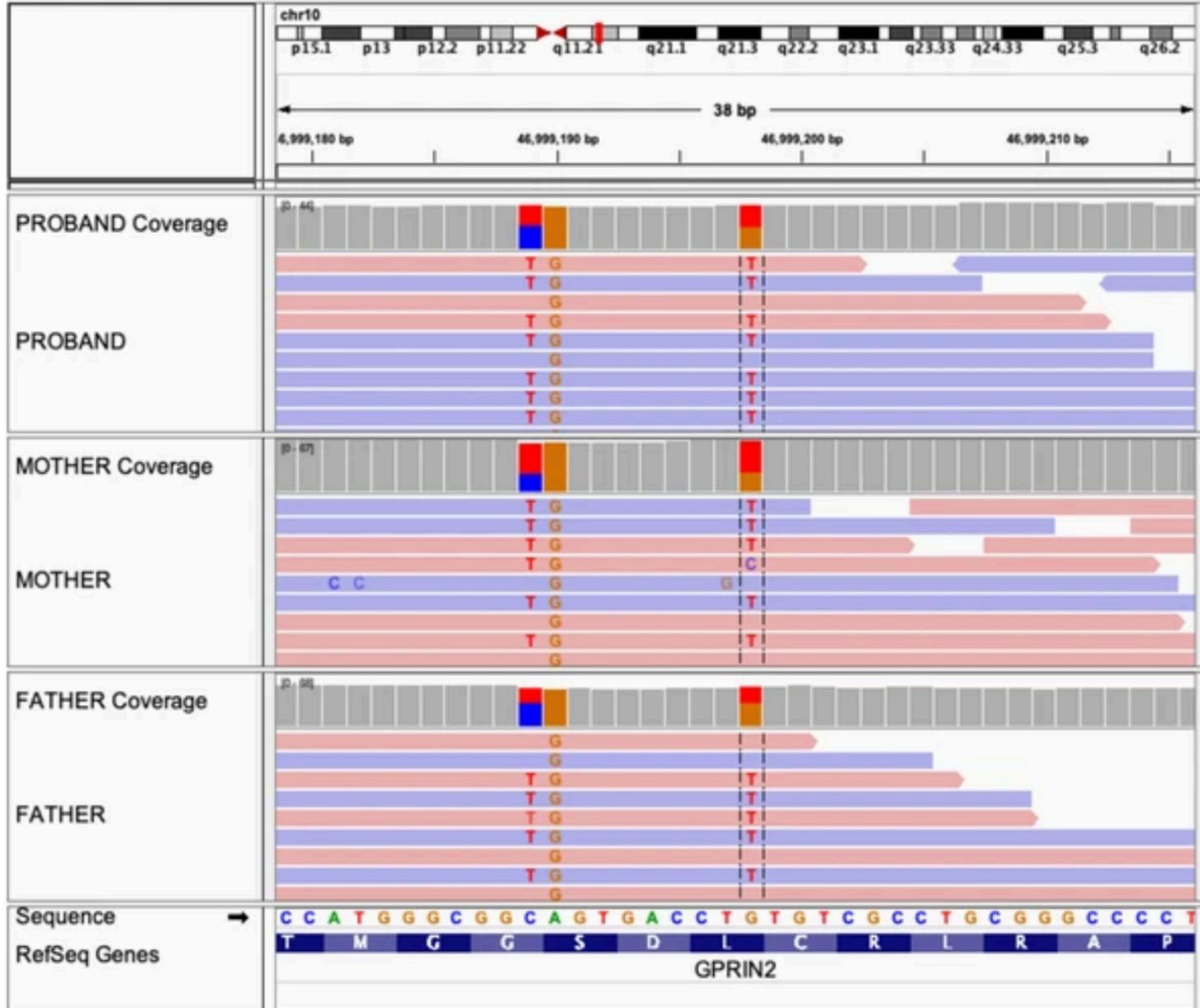
# Visual review of alignments



**False positive  
associated with  
strand bias**

All but one variant-supporting reads in the proband are on the reverse strand, whereas reference-supporting reads are equally represented on both strands

# Visual review of alignments



**False positives due to paralogous alignments of reads from regions not well represented in the reference.**

Alignments for proband include reads with several substitutions relative to the reference sequence within the 41-bp viewing window. This typically occurs when reads from sequences not represented in the reference are mapped to the closest paralog.

# Identify possible artifacts

- **Using IGV visualization**
- **Variants repeated in other samples sequenced with the same technology:**
  - Repetition can indicate a polymorphism if it is present in at least a 1% of the population
  - Repetition can indicate a frequent alteration in the phenotype if its presence is validated in other studies
  - Otherwise, it can be an artifact

# Identify possible artifacts

- **Using IGV visualization**
- **Variants repeated in other samples sequenced with the same technology:**
  - Repetition can indicate a polymorphism if it is present in at least a 1% of the population
  - Repetition can indicate a frequent alteration in the phenotype if its presence is validated in other studies
  - Otherwise, it can be an artifact
- **Variants in positions with very low depth of coverage**
- **Variants at very low allele frequency**

# Prioritization strategies

1. Prioritization is done using a **set of evidences based on the annotations**.
2. Annotations used in the prioritization **vary with the pathology or condition** under study.
3. Criteria may **vary depending on the objective** (somatic variants, common variants in population, ...)

# Some evidence for clinically relevant variants

- **High or moderate impact sequencing consequences:**

transcript\_ablation | splice\_donor\_variant | splice\_acceptor\_variant | stop\_gained | frameshift\_variant | stop\_lost | start\_lost | transcript\_amplification | inframe\_insertion | inframe\_deletion | missense\_variant | protein\_altering\_variant | splice\_region\_variant | incomplete\_terminal\_codon\_variant | stop\_retained\_variant

- **Impact in the protein function:**

damaging prediction, affecting protein domains

- **Attached clinical significance:**

pathogenic ClinVar

- **Relevance in the pathology:**

gene with a role in processes involved in the disease, affected gene or variant frequent in the disease

# Some evidence of non-relevance:

- Variants located in **no functional genes**: BACs, pseudogenes,...
- Variants in **no relevant or poor supported transcripts**
- **Polymorphisms** (variants present at least in a 1% of the population)  
Population frequency in 1000 Genomes project, gnomAD, EVS ...  $\geq 1\%$   
! unless it is associated with predisposition, prognosis, drug response,  
...

# Stratification in Tiers

A common approach for the prioritization is to **classify the variants into tiers** depending on whether their annotations fulfill some conditions

Tier 1 > Tier 2 > Tier 3 > Tier n



Evidence of relevance

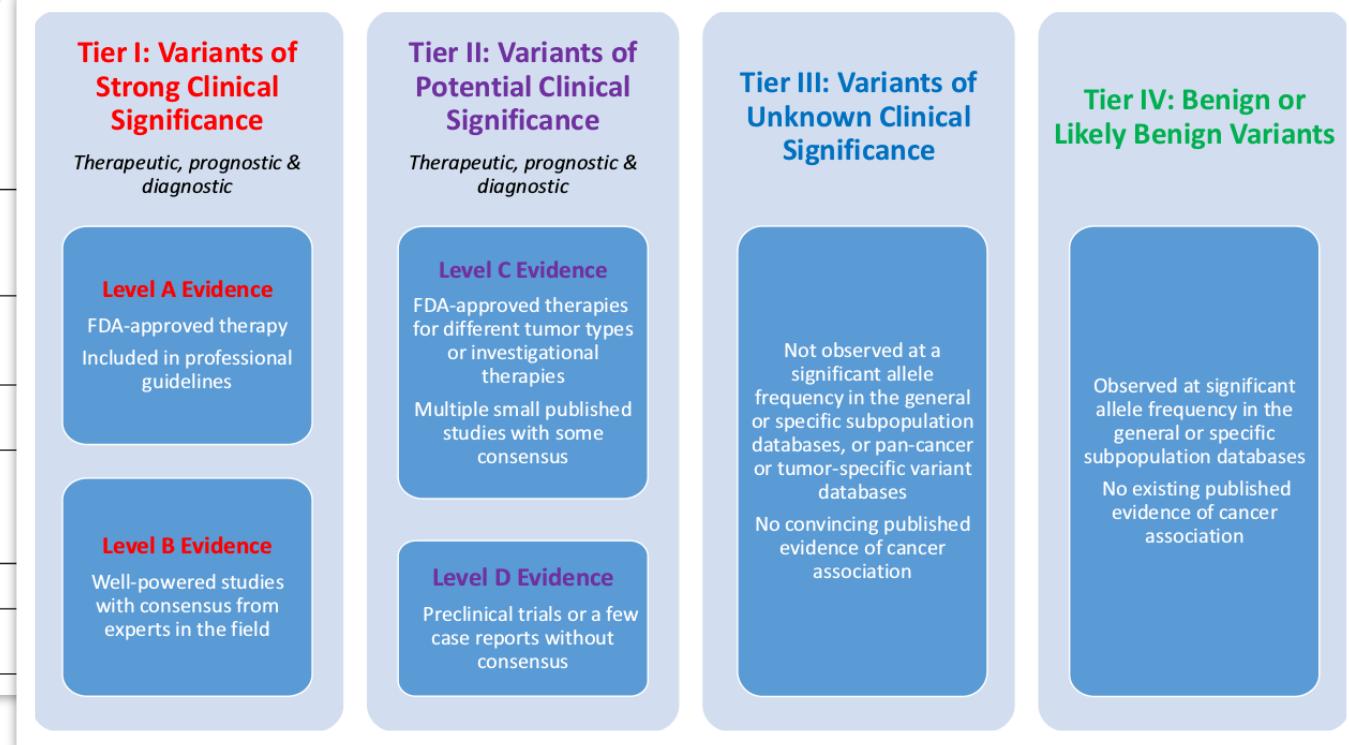
# Stratification in tiers according to the number and type of evidences

## Mendelian disorders

	Evidence scale from Benign to Pathogenic					
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2		Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4		
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4  Missense in gene where only truncating cause disease BP1  Silent variant with non predicted splice impact BP7  In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene/gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5  Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2  Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

[Genet Med. 2015 May;17\(5\):405-24](#)

## Cancer disease



[J Mol Diagn. 2017 Jan;19\(1\):4-23](#)

# Score calculation

**Construction of an evidence-based-score** computed from selected annotations.

This **provides a ranked list of variants** with those with more evidence at the top.

Higher >>>>>> Lower



Evidence of relevance

Elena Piñeiro-Yáñez

[epineiro@cnio.es](mailto:epineiro@cnio.es)

Coral Fustero-Torre

[cfustero@cnio.es](mailto:cfustero@cnio.es)

María José Jiménez-Santos

[mjjimenez@cnio.es](mailto:mjjimenez@cnio.es)

Fátima Al-Shahrour

[falshahrour@cnio.es](mailto:falshahrour@cnio.es)