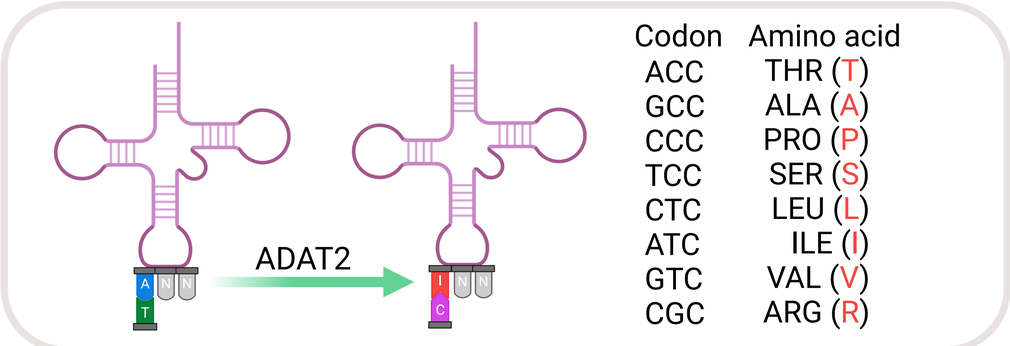
Deciphering **ADAT2**-sensitivity determinants using **machine learning**

Marine Leclercq, Ning An, Coralie Capron, Christine Grill, Arnaud Blomme, Francesca Rapino, Pierre Close

SCAN ME

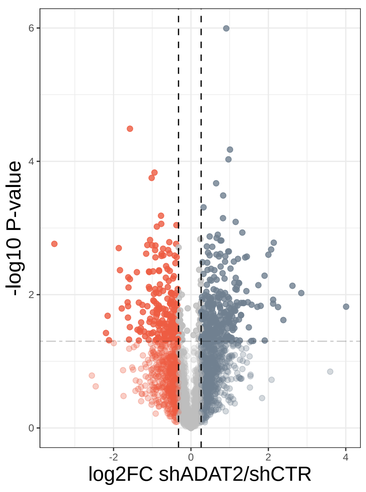
ADAT2 - Adenosine Deaminase tRNA-specific 2



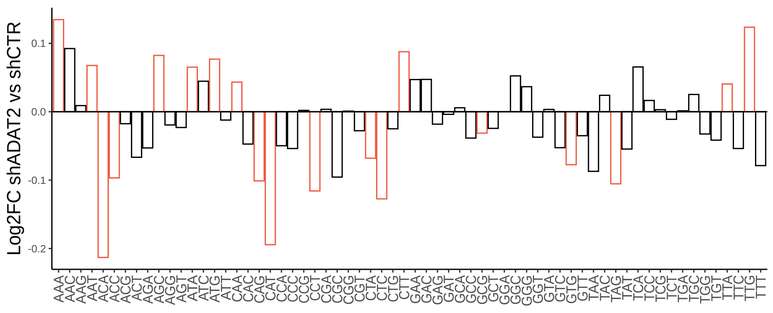
The **adenine-to-inosine modification expands the decoding repertoire** of select tRNAs thanks to wobble-pairing.
⇒ **decoding of C-ending codons (NNC)** that **do not have a native** GNN cognate **tRNA**.
What is the impact of ADAT2 on the proteome? What makes a protein more dependent on ADAT2?

Preliminary data

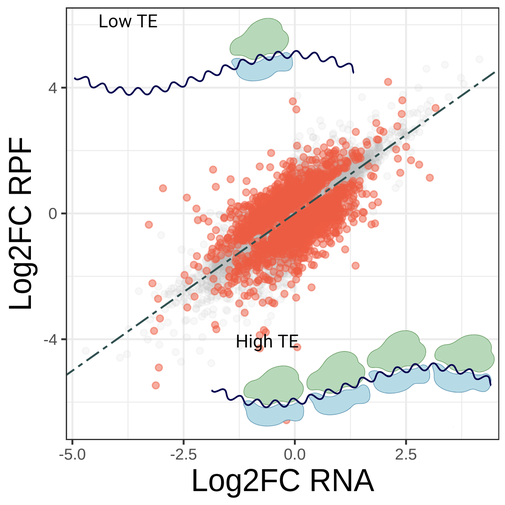
Proteomics ↓



RiboSeq - Codon occupancy ↓



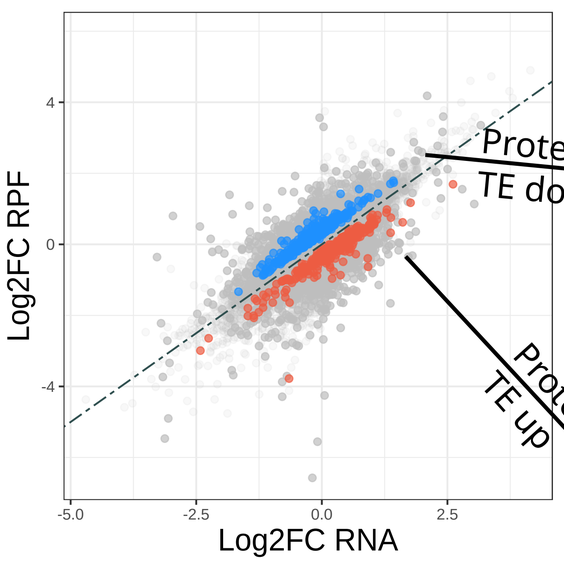
RiboSeq - Translation efficiency ↓

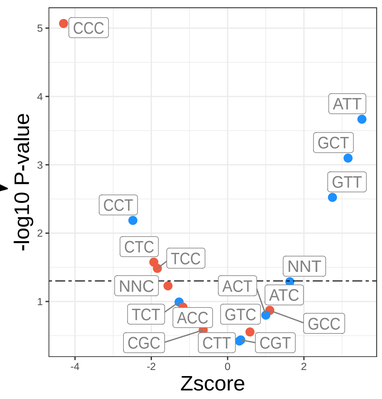


ADAT2-depletion has a **strong global impact** on **translation.**

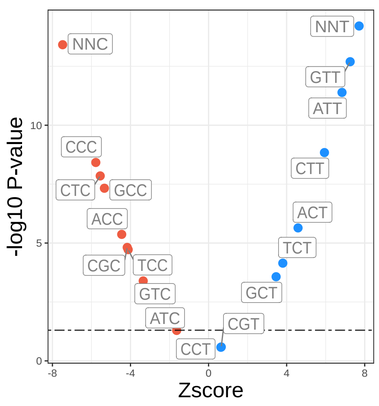
But it cannot be fully explained by NNC codons alone.

Crossing proteomics and RiboSeq ↓





↕ Codon enrichment ↕

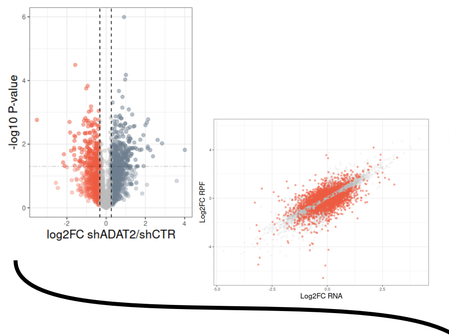


Codon enrichment comparison:
* Protein down TE up group shows **strong enrichment of NNT codons** and strong depletion of NNC codons.

⇒ **Cohesive group of proteins to study ADAT2-sensitivity determinants**

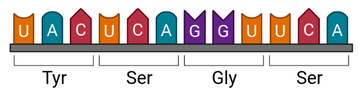
Feature selection pipeline

Omics data



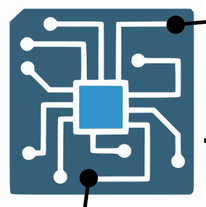
+

Sequence features



- GC content,
- Aromaticity,
- Secondary structures,
- ...

Machine learning

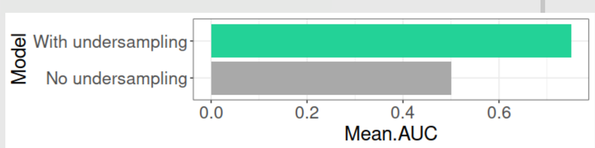


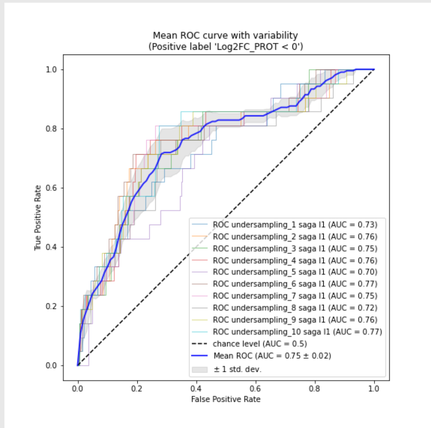
Model information

* **Logistic regression**
* **249 variables**
  * Scaled into **Zscores**
* **No pre-filtering** based on correlation
* **L1** regularization
* Hyperparameters tuned by k-fold crossvalidation

Dealing with imbalanced data

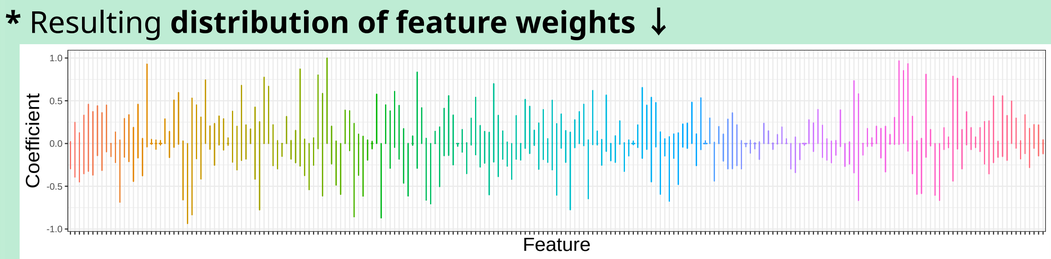
* **Only 2.6%** (106/4063) **of the dataset is in the positive class** (prot down TE up)
  ⇒ **Model is incapable of learning**
* **Solution: undersampling 1:1**
  ⇒ **Boosts model performance**





Feature selection

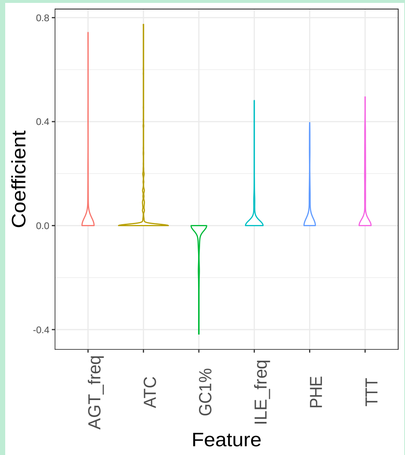
* We used a **nested Monte Carlo crossvalidation to select a group of most relevant features:**
  Randomization of **undersampling** x 10
  Randomization of **validation** set x 10
  ⇒ **Train 100 models**
* Resulting **distribution of feature weights** ↓



* **Restricting to features** that
  a) have **only positive or negative weights**
  b) have a **nonzero weight in >=33% of models**

⇒ Results in a final set of 6 core features

→



Next steps

Validation of the 6-feature set:
a) Biologically (mutant experiments)
b) in-silico (independant datasets)