

# RIsearch3 for fast RNA-RNA interaction predictions

Lorenzo Favaro, PhD student  
Center for non-coding RNA in Technology and Health (RTH)  
Faculty of Health and Medical Sciences  
University of Copenhagen

10/02/26

# Background



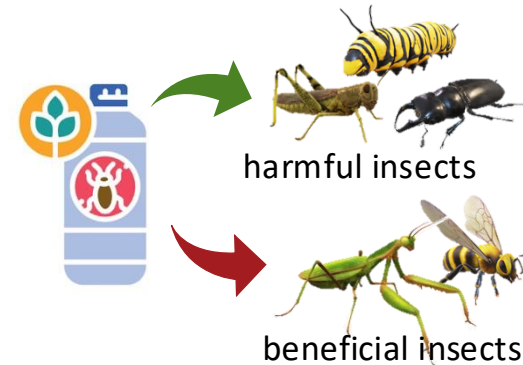
- **Why does off-target assessment of biopesticides matter?**



Increased global  
food demand



Effective crop  
protection



Novel biopesticides



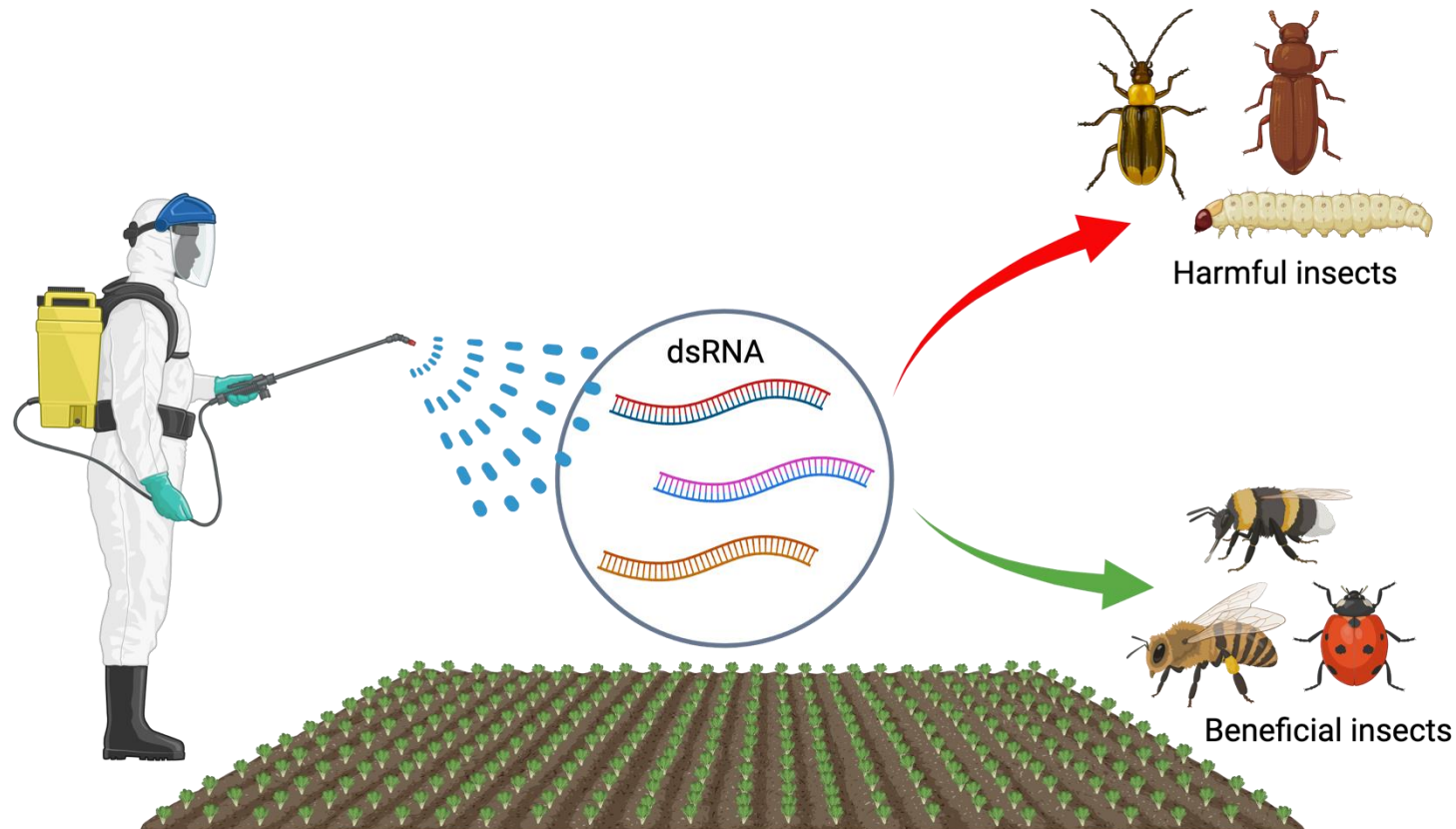
Lack of EU regulations

**novo nordisk  
foundation**

Novo Nordisk Foundation Challenge  
Grant no. NNF24OC0087112

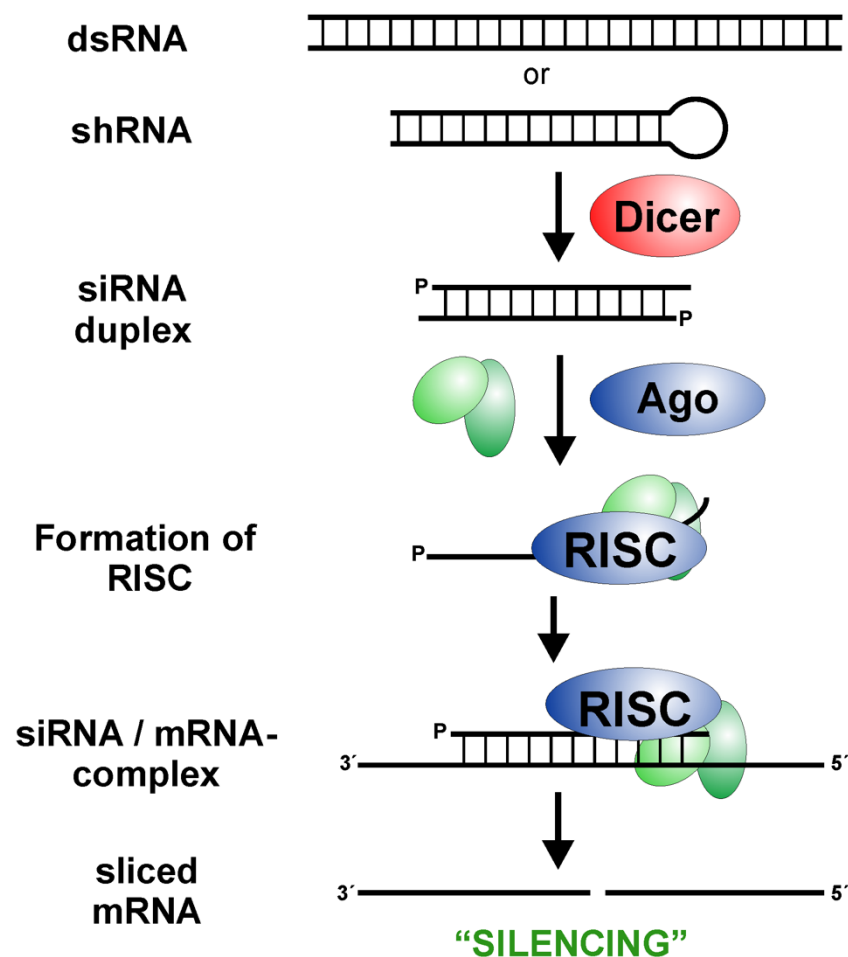
**ENSAFE** is an environmental risk assessment framework for siRNA- and peptide-based pesticides: integrating *in silico*, *in vitro*, and *in vivo* approaches.

# How do we identify off-target organisms? ————— ENSAFE

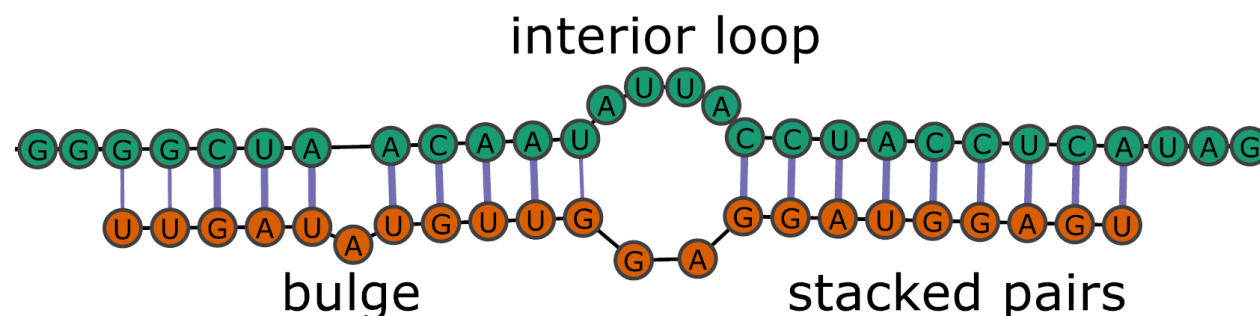


- *In silico* pipeline to quickly scan thousands of siRNAs through hundreds of genomes/transcriptomes
- Temperature-dependent RNA interactions (insect cell lines grow at 25-27°C)

# The complex scenario of RNAi



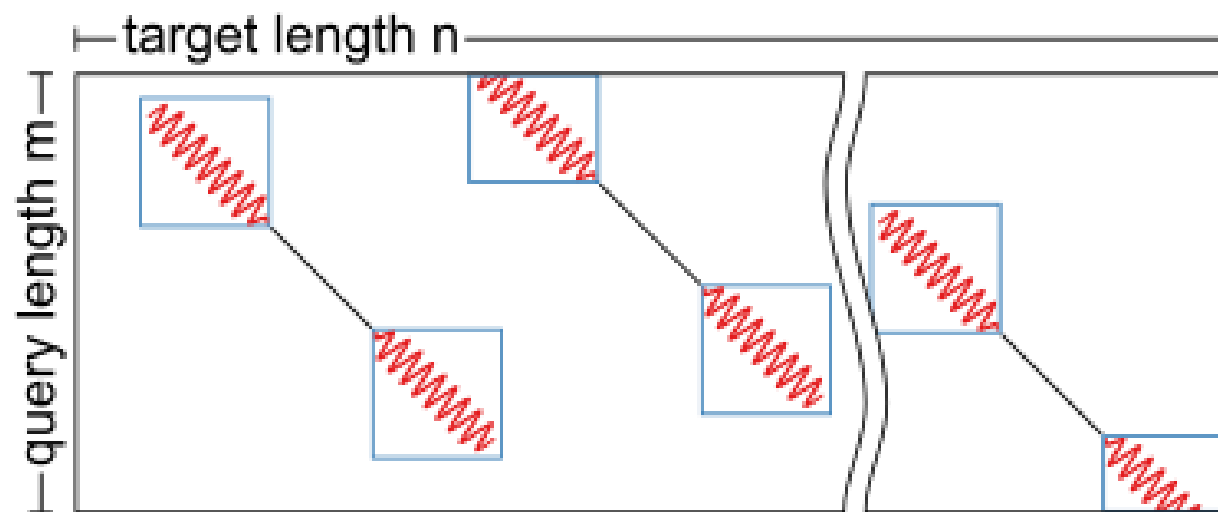
- Hundreds of genomes/transcriptomes → huge search space
- RNA binding is complex, complementary is not enough
- Purely thermodynamic algorithms scale poorly



# RIsearch2



- Seed-and-extend strategy
- Suffix-array data structure for seed lookup (includes G-U wobble)
- Extends seeds on both ends with dynamic programming using a nearest neighbour energy model (simplified Turner) of RIsearch<sup>†</sup>
- Allows flexible parameter settings for different classes of RNA-RNA interactions
- **Outputs raw binding energy**



<sup>†</sup>Wenzel, Akbasli, Gorodkin, Bioinformatics, 2012.

# siRNA off-target discovery pipeline



**Objective:** To predict individual off-target transcripts and quantify the overall off-targeting potential (POFF) of an siRNA at a transcriptome-wide scale.

- Partition function<sup>†</sup> approach
- All putative siRNA-target interactions (RIs<sub>search2</sub><sup>\*</sup>) represent a microstate variables
- State variables:
  - hybridization energy (RIs<sub>search2</sub>)
  - opening energy (precomputed accessibilities with RNAplfold<sup>‡</sup>)
  - abundance (expression data)

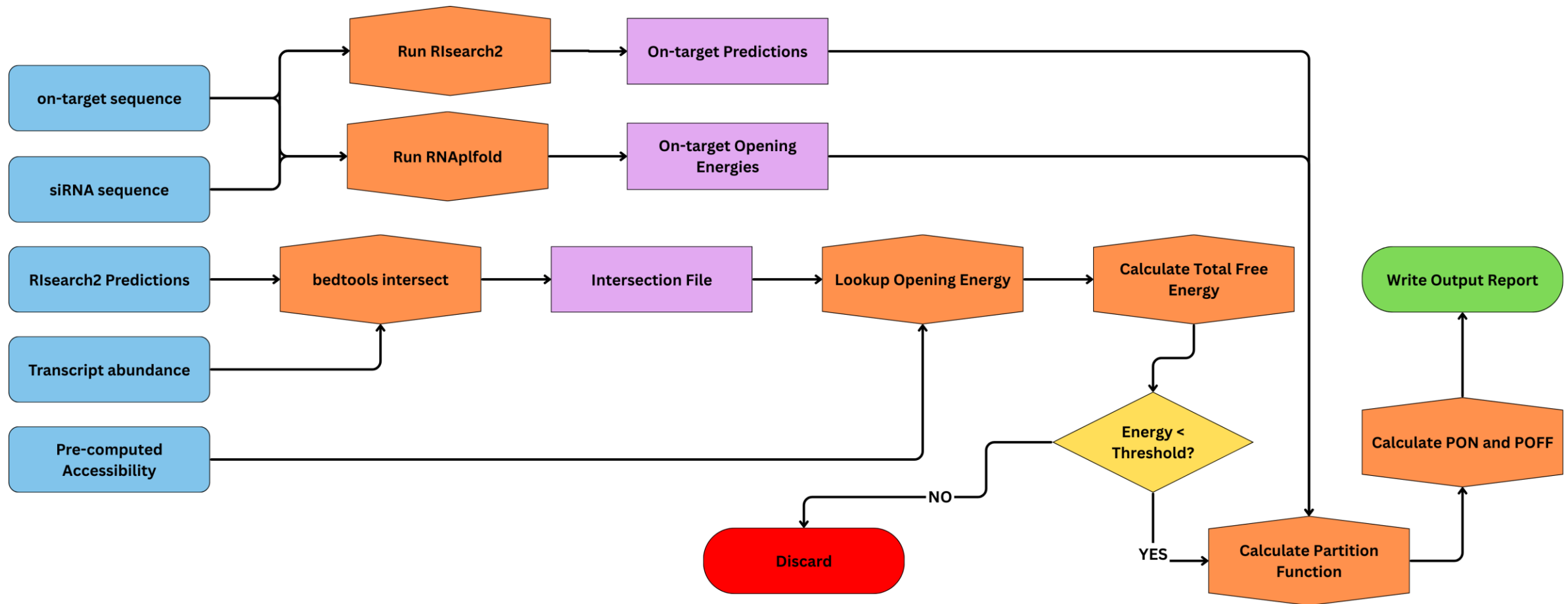
$$Z = \sum_{i \in \mathcal{I}} A_i e^{-\beta(E'_i + O_i)}$$

$$p_i = \frac{A_i e^{-\beta(E'_i + O_i)}}{Z}$$

<sup>†</sup>The function of thermodynamic state variables describing the statistical properties of a system in thermodynamic equilibrium.

<sup>\*</sup>Alkan et al., Nucl. Acids Res. 2017. <sup>‡</sup>Bernhart et al., Bioinformatics, 2006.

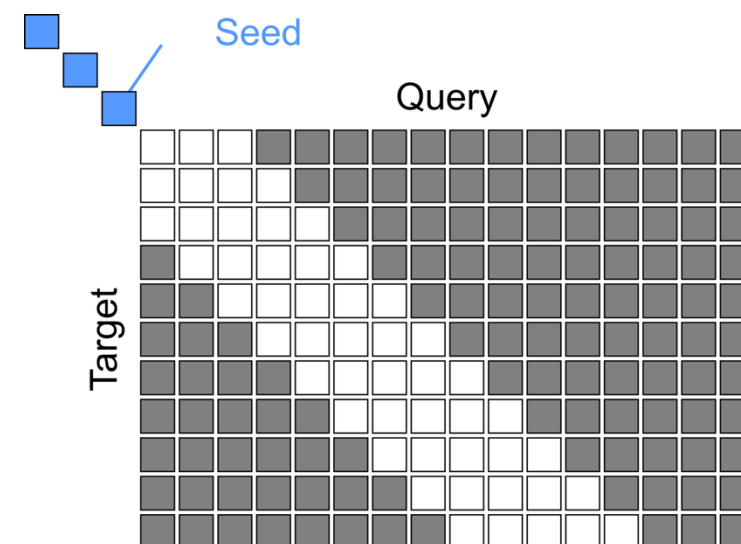
# Pipeline workflow



# What's new in RIssearch3?



- **Temperature Dependence:** The tool now includes temperature-dependent energy parameters for **RNA/RNA**, **RNA/DNA**, and **DNA/DNA** interactions.
- Dedicated **CRISPR mode** with specific features for guide RNA (gRNA) target prediction: Bulged Seed Search, PAM Filtering, Repeat Filtering.
- **Banded Alignments:** A "banded alignment" technique has been implemented to speed up dynamic alignment by only calculating values around the diagonal.



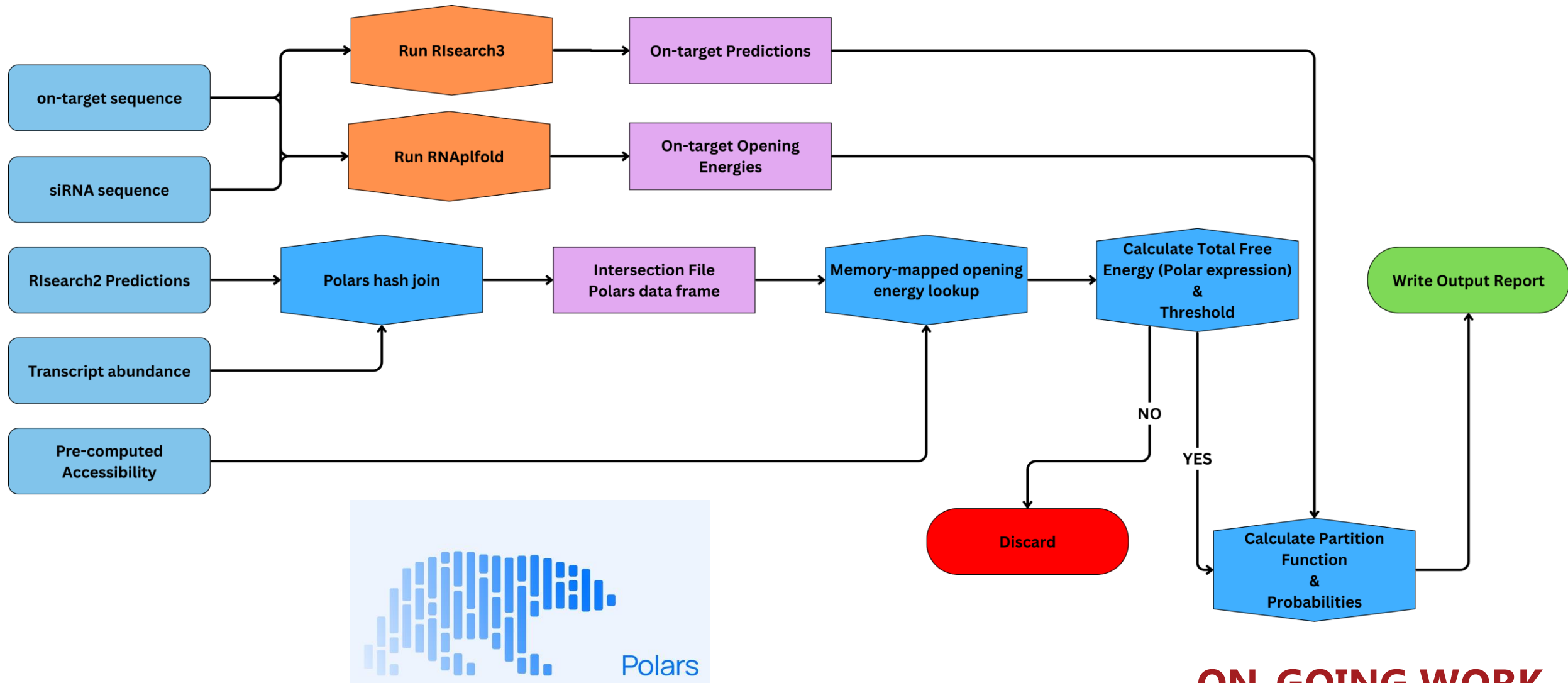
# New siRNA off-target pipeline implementation — ENSAFE

- Current: Python 3.14+, Polars, vectorized operations, lazy execution, memory-mapped accessibility profiles.
- Old: Python 2, heavy file I/O and shelling out (e.g., bedtools, sort, cut, RNAplfold, Rsearch2), manual parsing and lists in memory.

| # siRNAs    | Old                   | New                   | Change                                                                                                |
|-------------|-----------------------|-----------------------|-------------------------------------------------------------------------------------------------------|
| <b>100</b>  | 104.553 s $\pm$ 1.099 | 52.231 s $\pm$ 0.370  | ~2 times faster    |
| <b>500</b>  | 246.238 s $\pm$ 0.182 | 136.500 s $\pm$ 1.007 | ~2 times faster   |
| <b>1000</b> | 523.103 s $\pm$ 0.584 | 286.820 s $\pm$ 1.703 | ~2 times faster  |
| <b>2000</b> | 902.732 s $\pm$ 3.223 | 499.600 s $\pm$ 1.501 | ~2 times faster  |

**ON-GOING WORK**

# Pipeline workflow



**ON-GOING WORK**

# New features



- **Batch processing** for multi-siRNA workflows
- 'uv' package manager
- **Better maintainability** through service-oriented architecture
- **Modern CLI** with intuitive long-form flags and auto-generated help
- **YAML configuration** support for reproducible analyses

```
research-pipeline --help

Usage: research-pipeline [OPTIONS] COMMAND [ARGS]...

siRNA off-target discovery pipeline – analyze RIssearch2 predictions.

Options
--config -c PATH Path to YAML config file. Runs the command specified in the config.
--verbose -v Enable verbose logging (DEBUG level).
--help Show this message and exit.

Commands
accessibility Pre-compute accessibility profiles for a whole genome.
off-targets Analyze siRNA off-target predictions.
orthologs Download orthologs (OrthoDB) and transcriptomes (NCBI) for a list of species.
```

**ON-GOING WORK**

# What are the next steps?

---



- RIssearch3 python bindings
- Implementation of ViennaRNA python API in the pipeline
- Further benchmark of banded version and temperature-dependent energy parameters

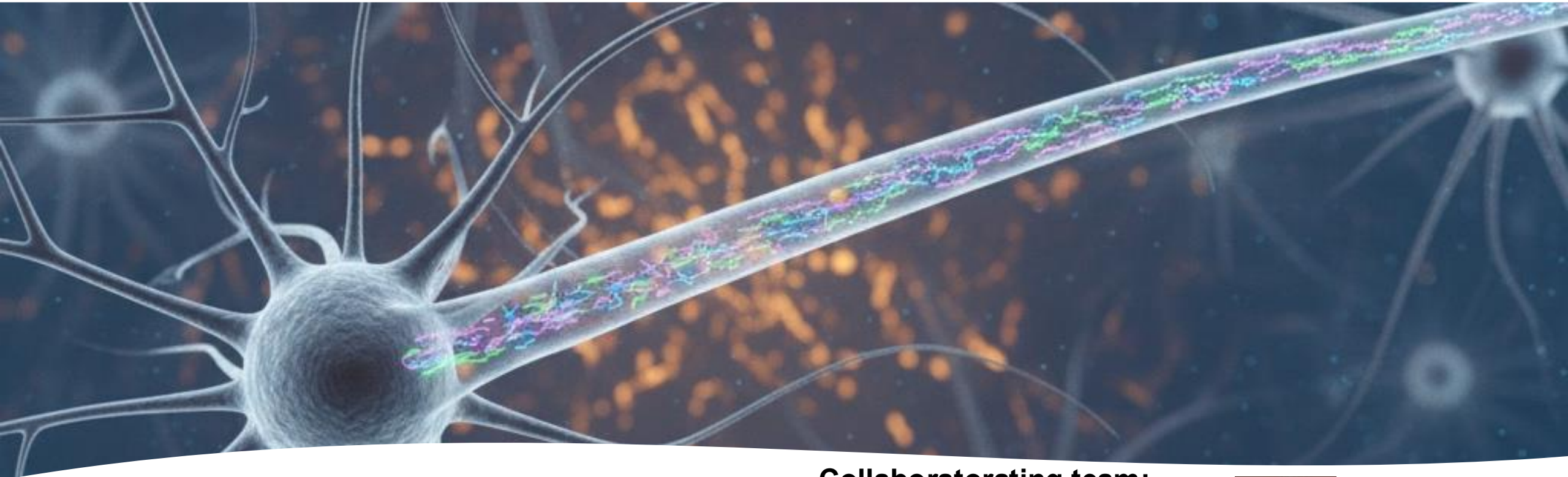
# Acknowledgement

- **Christian Anton**
- **Stefano Roncelli**
- Gül Demircan
- Ying Sun
- Buwen Liang
- Vi Tran
- Mikkel Wolfhagen
- Nadezhda Doncheva
- Stefan Seemann
- Jakob Hull Havgaard
- **Jan Gorodkin**



# Upcoming PhD or postdoc position at RTH

## Talk to Jan



**Decoding RNA journey:**

**Predictive models for the  
subcellular RNA fate in neurons**

**Collaboratorating team:**

**Chekulaeva (UCPH / Berlin)**



**Ulitsky (Weizmann Institute)**



**Gorodkin (UCPH)**

