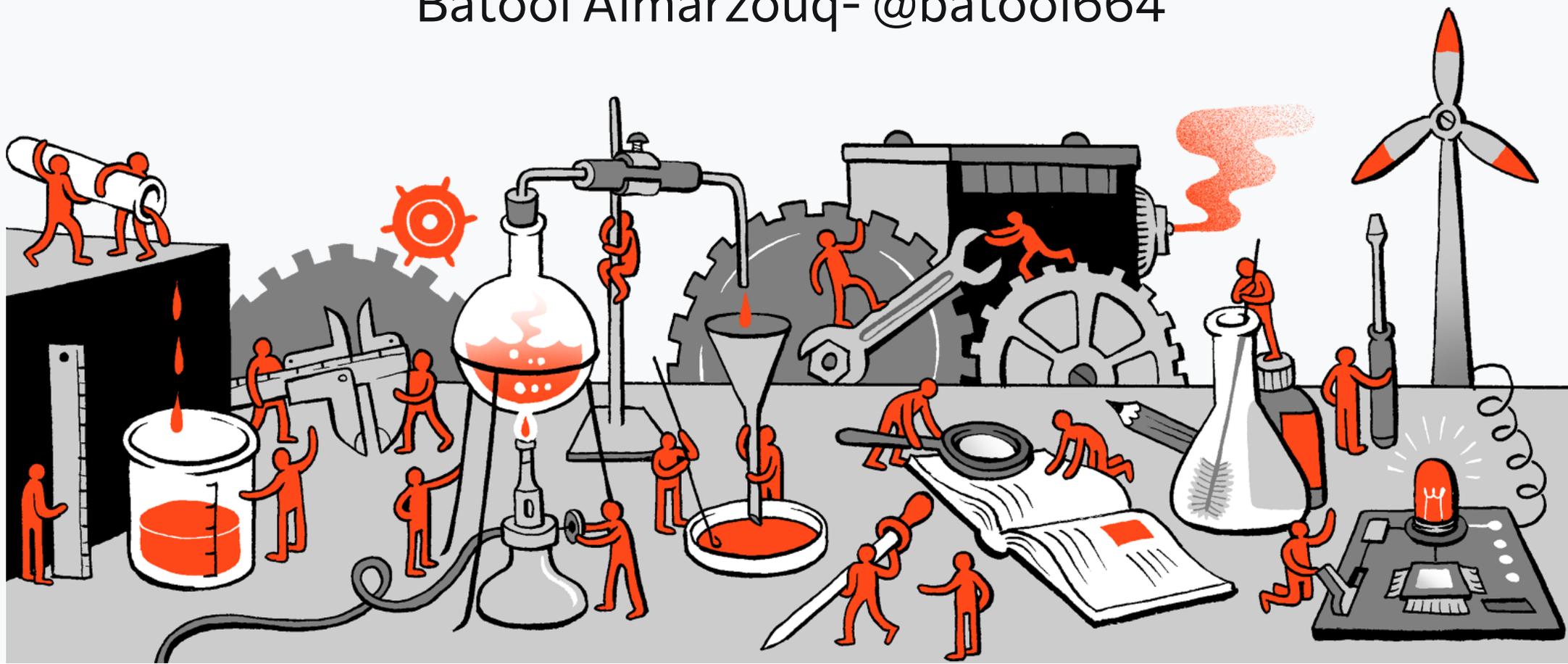


An Open Science Approach to Machine Learning in Biomedical Research

Batool Almarzouq- @batool664



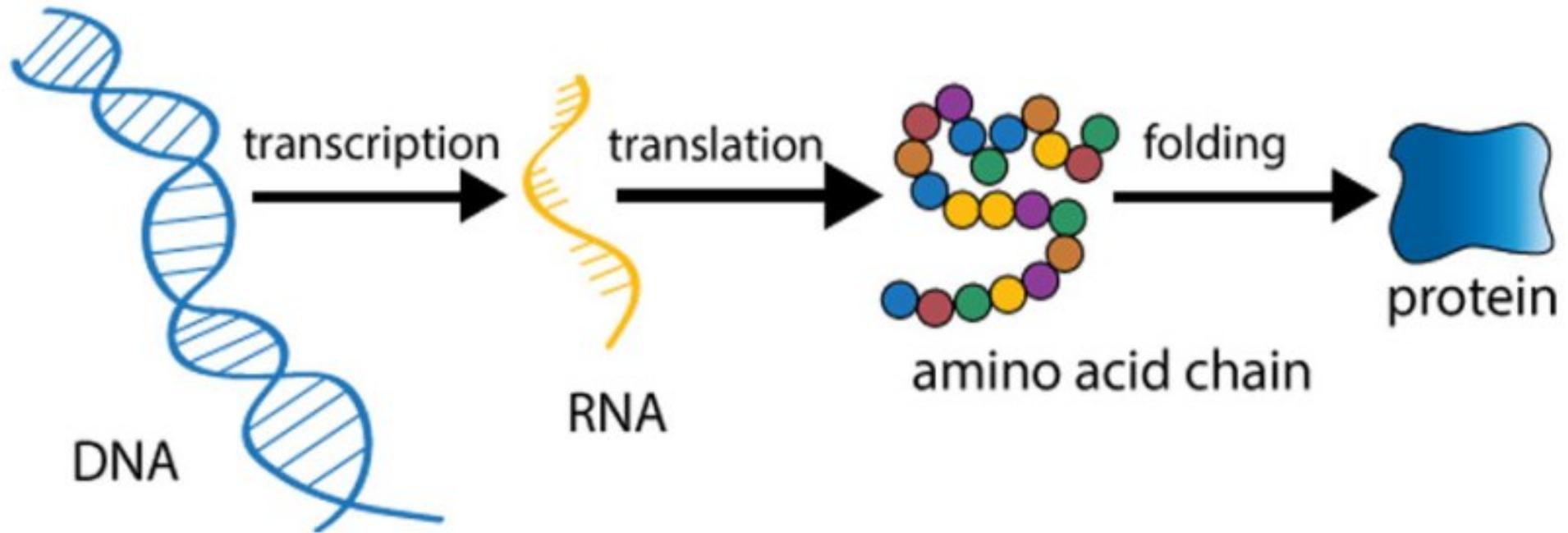
A little bit about me!

- A computational biologist affiliated with the University of Liverpool.
- Founder of RLadies Chapter in Saudi Arabia (Dammam).
- A curator in the R Weekly team.
- Member of MiR accessibility committee.
- Member in the turing way community.
- Working on establishing an Open Science community in Saudi Arabia.

Acknowledgment

- Anelda Van der
- Malvika Sharan, Kirstie Whitaker and Martina G. Vilas
- The Turing Way Community
- Alison Presmanes Hill (slides)

Why do we use ML in Biomedical
Research?



DNA

- DNA sequence alignment
- DNA sequence classification
- DNA sequence clustering
- DNA pattern mining

Algorithms includes fuzzy sets, neural networks, genetic algorithms.

Image Credit: [ABC Science](#)

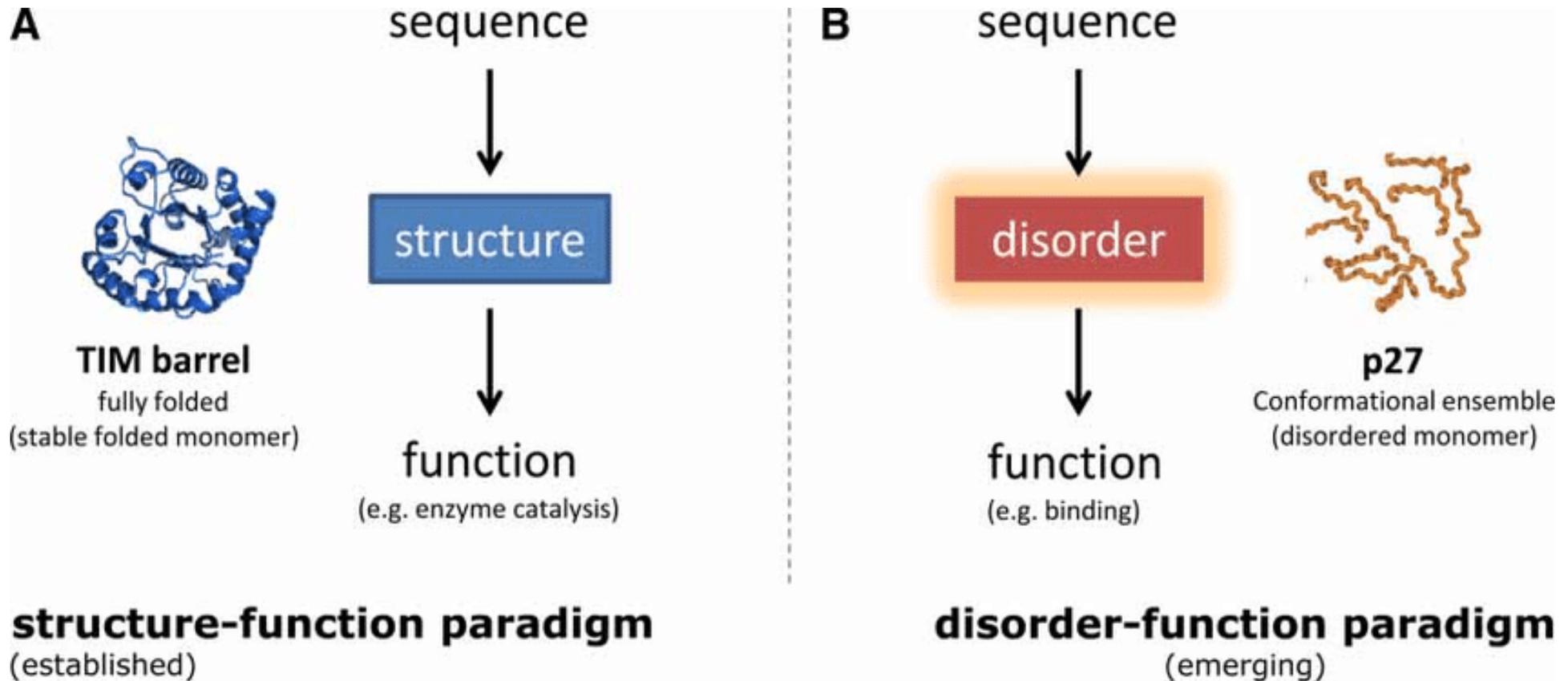
RNA

- Mainly RNA-sequencing (RNA-seq)
- Differentially expressed genes (DEGs)
- Alternative splicing
- Small RNA expression

Algorithms include Logistic Regression, Random Forest, LMT, Random Subspace.

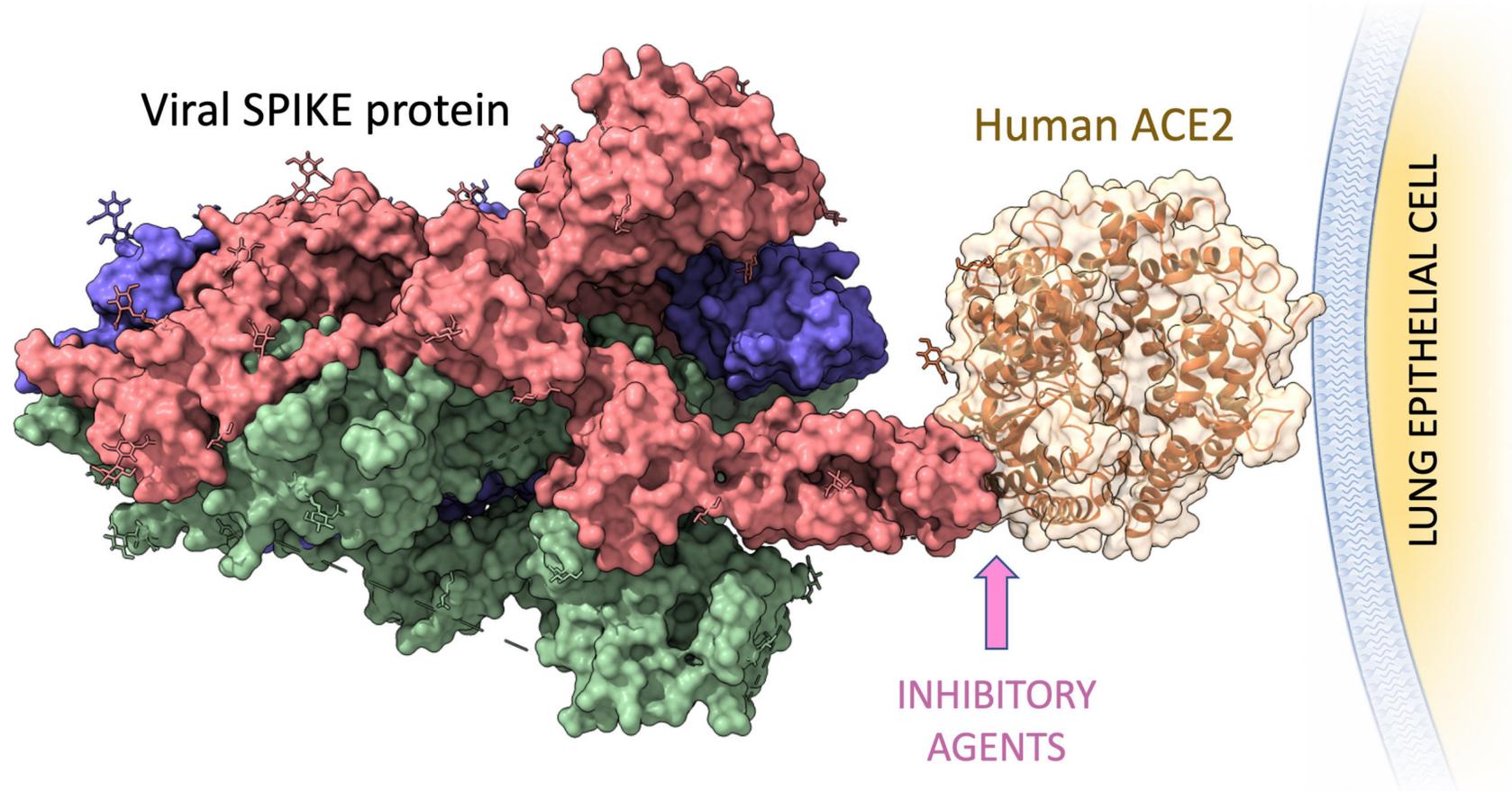
Solving the sequence is not enough!

We need to know the structure and function of the protein!



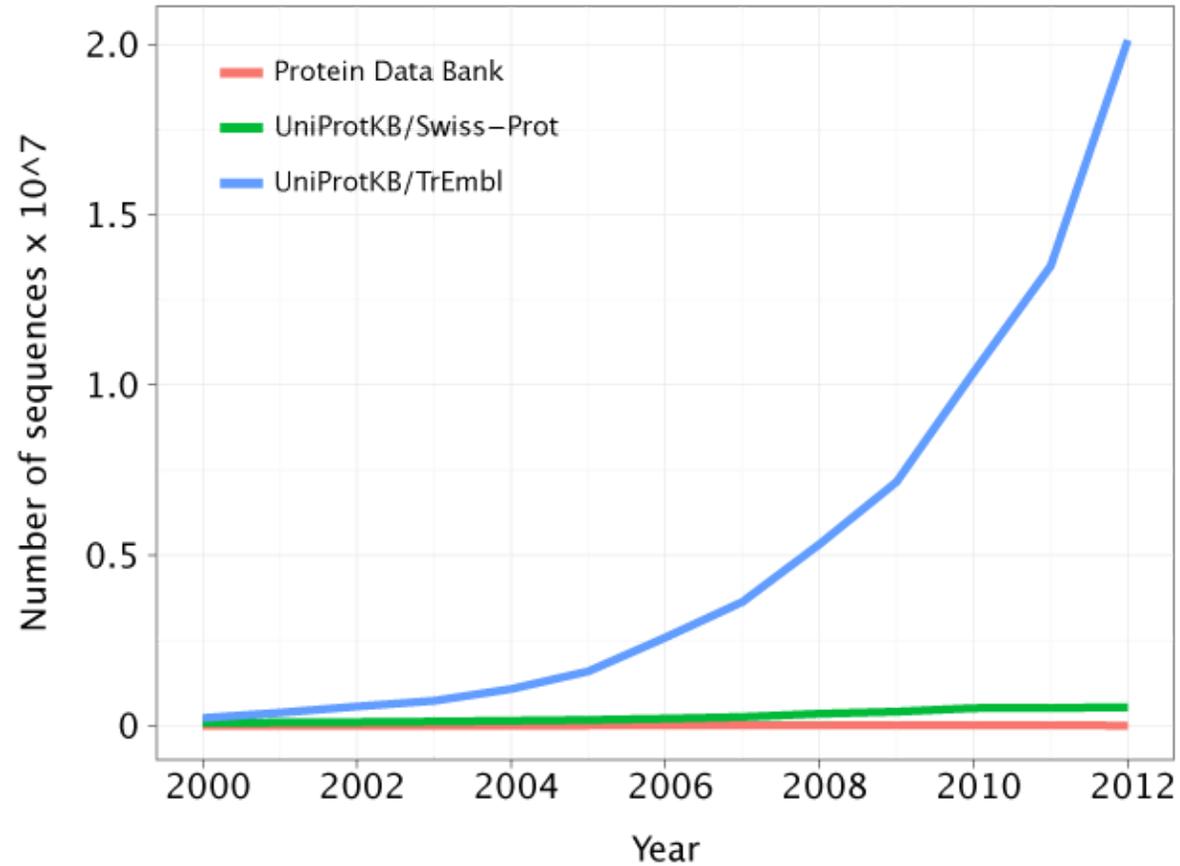
HOW CAN WE PREDICT FUNCTION FROM STRUCTURE?

To predict the function from the structure, scientists use different approaches including machine learning (ML) and deep learning algorithms .



Prediction of protein structure is important to develop small molecules and targeted therapy for diseases.

Why not only rely on Experimental Methods?



Because of the growing gap between the newly-sequenced and characterized sequences in the genome databases, computational methods in gene functional annotation are indispensable. Moreover, given the drop in the genome sequencing techniques' cost, this gap is only destined to grow.

Biology has become a highly data-intensive science, dependent on complex, computational, and statistical methods!

So, how can we make these methods available and accessible for researchers, while ensuring that scientific results remain reproducible?

What is the percentage of reproducible research?

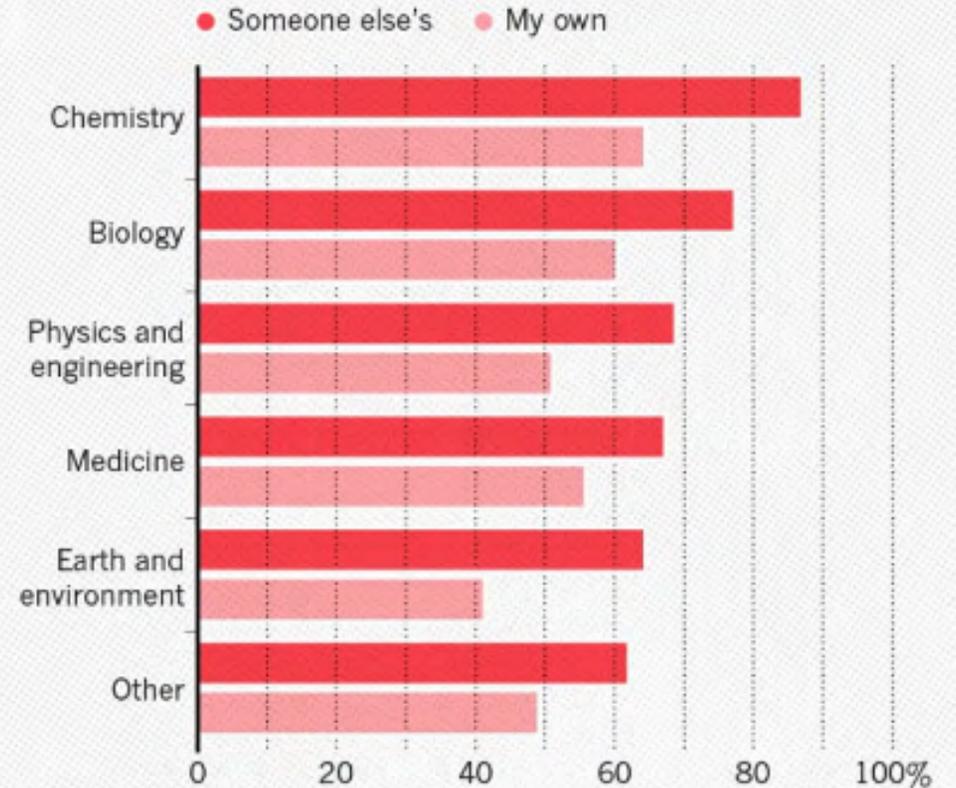
IS THERE A REPRODUCIBILITY CRISIS?



©nature

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



Credit: Key results of the survey on reproducibility conducted by Nature in 2016

How can we overcome the reproducibility
crisis?

How can you improve the reproducibility of your data science project?

OPEN SOURCE SOFTWARE

SHARE CODE/ANALYSIS

SHARE COMPUTATIONAL ENVIRONMENT

VERSION CONTROL

TESTING

DOCUMENTATION

OPEN DATA/FAIR DATA

OPEN ACCESS

This is called Open Science.

Open Science is about extending the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done

-- [FOSTER Plus](#)

What are the FAIR principles?



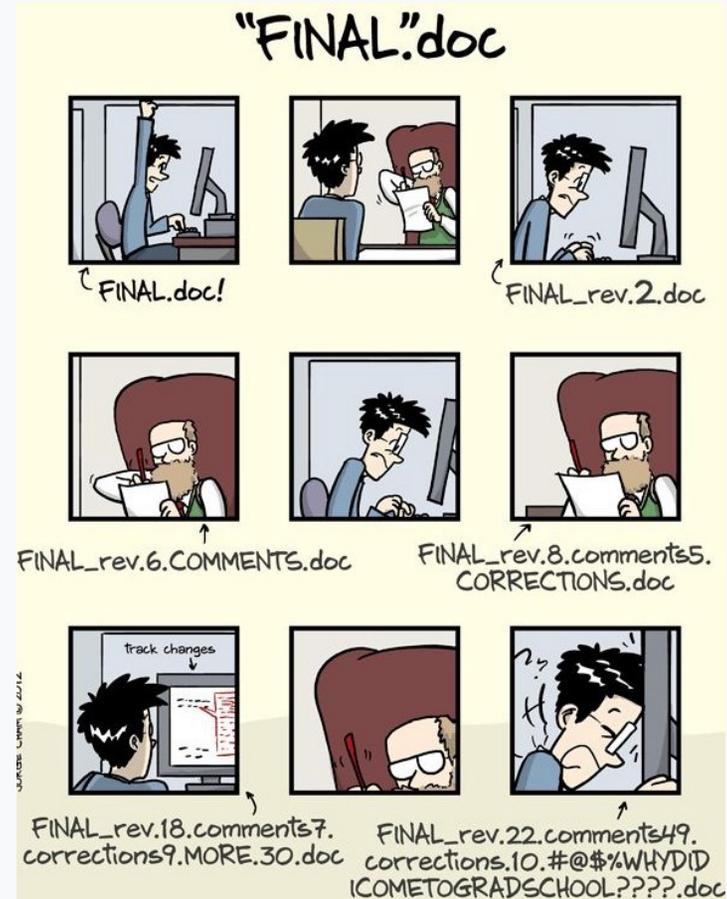
Scriberia 

* The Turing Way project illustration by Scriberia. Zenodo. <http://doi.org/10.5281/zenodo.3332807>

Why do we use version control (git)?

Version Control in the Old Days ..

FINAL
FINAL.FINAL
final.FOR REAL
FINAL.version 2
absolutely.FINAL
FINAL.2
FINAL.3
FINAL.3.01
FINAL.3.02
FINAL.working



Real Version Control (including backup)

BatoolMM / [An-Open-Science-Approach-to-Machine-Learning](#) Unwatch 1 Star 0 Fork 0

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

adding the first part of the talk

main [Browse files](#)

BatoolMM committed 10 hours ago 1 parent [d4b57f7](#) commit [b035487ef1f2deae91a8c58149a30b83ff622eac](#)

Showing **2 changed files** with 89 additions and 1,154 deletions. Unified Split

569 bin/presentation.Rmd

```
...  ...  @@ -1,6 +1,6 @@
1  1  ---
2  2  - title: "An Open Science Approach to Machine Learning in Medical and Biological Research"
3  3  - subtitle: "Saudi Data Community"
4  4  + title: "An Open Science Approach to Machine Learning in Biomedical Research"
5  5  + subtitle: "Talk @ Saudi Data Community"
6  6  author: Dr. Batool Almarzouq
7  7  date: "`r Sys.Date()`"
8  8  output:
9  9  @@ -22,9 +22,8 @@ output:
10 10  ---
11 11  ---
```

In the pandemic, some publishers have “opened” their journals to make certain articles freely available.

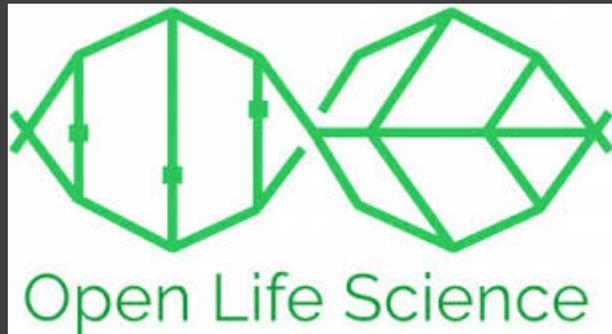
Databases have been created that are completely open access, such as the Open COVID Pledge.

UNESCO is launching international consultations aimed at developing a Recommendation on Open Science for adoption by member states in 2021

There is a network of Open Science Communities in Netherlands, Sweden, Germany, UK and others

In line with vision 2030, we are starting an Open Science Community in Saudi Arabia.

It's created and developed with the help of the "Open Life Sciences"



Open Life Sciences (OLS3) program helps individuals and stakeholders in research to become Open Science ambassadors.

We want to provide a place where newcomers and experienced peers interact, inspire each other to embed open science (research) practices and values in their workflows and provide feedback on policies, infrastructures and support services. Together working to make Open Science the norm. So we are calling out to researchers and colleagues in Saudi Arabia.



Batool Almarzouq
The University of
Liverpool



Founder and director
of Talarify, Mentor
OLS3



Paula Moraga,
Assistant Professor in
Statistics for Public
Health (KAUST)

Join me on the 24th
of Feb for a
workshop titled
"Collaborating on
Open Data Science
Projects" as part of
the Datathon for
WiDS2021.

How can you start learning about Open Science?



Kirstie Whitaker,
Project Lead



Malvika Sharan,
Community Manager



So, What is the turing way?



Book:
the-turing-way.netlify.app/

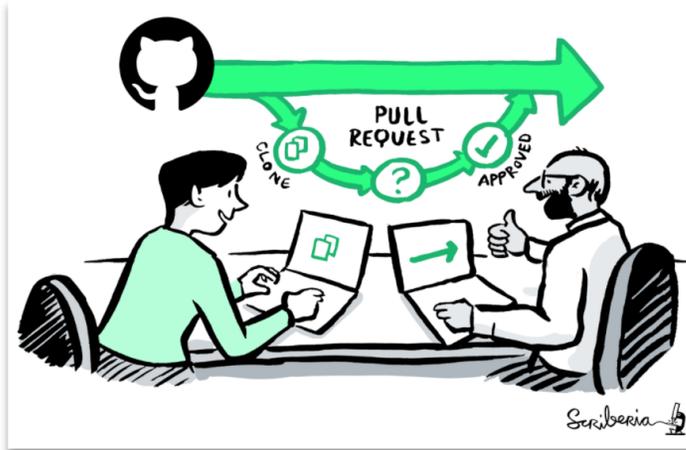
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

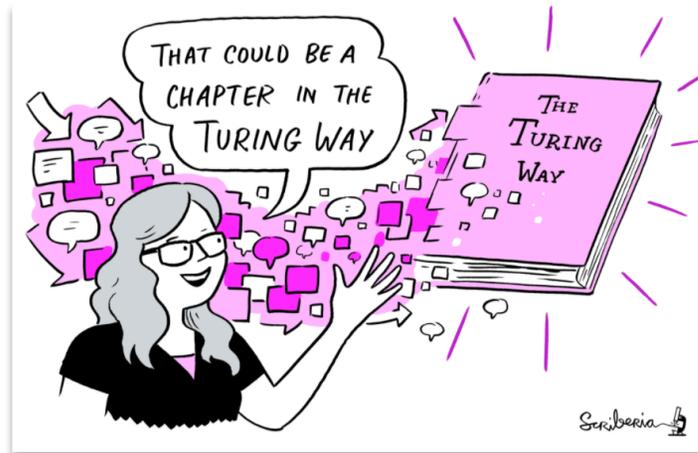
An Open Source Project



A Community



A Book



A Culture of Collaboration



Moonshot Goal: Reproducibility “too easy not to do”



The Turing Way

🔍 Search this book...

Welcome

- Guide for Reproducible Research
- Guide for Project Design
- Guide for Communication
- Guide for Collaboration
- Guide for Ethical Research
- Community Handbook
- Afterword

Visit our [GitHub Repository](#)

This book is powered by [Jupyter Book](#)

Welcome

The Turing Way is an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

Our goal is to provide all the information that data scientists in academia, industry, government and the third sector need at the start of their projects to ensure that they are easy to reproduce and reuse at the end.

The book started as a guide for reproducibility, covering version control, testing, and continuous integration. However, technical skills are just one aspect of making data science research “open for all”.

In February 2020, *The Turing Way* expanded to a series of books covering reproducible research, project design, communication, collaboration, and ethical research.

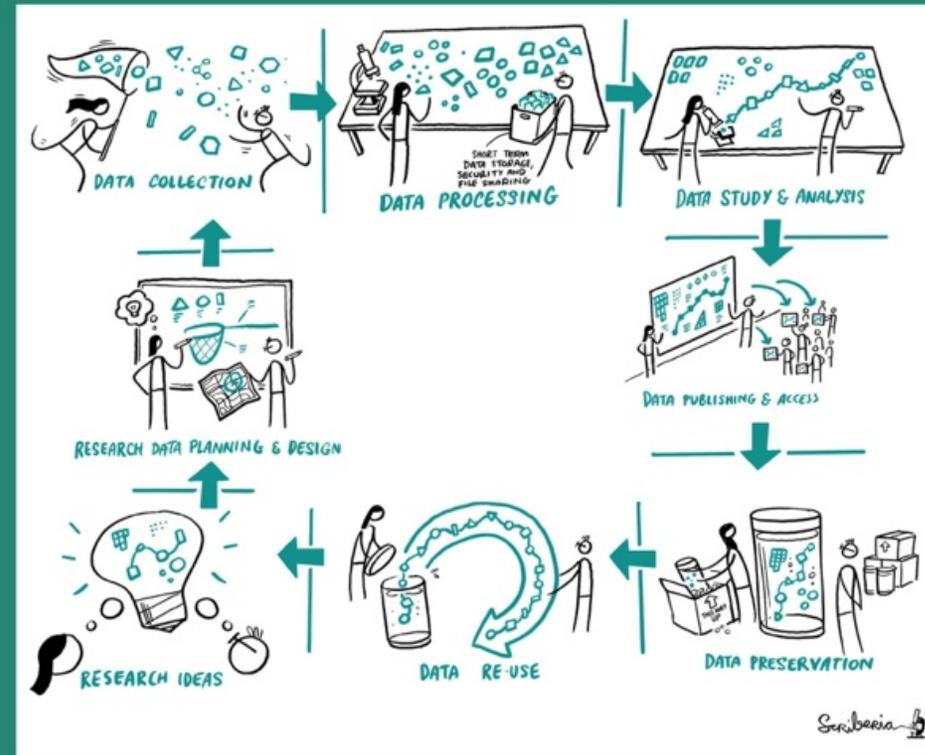


The Turing Way book with five guides: <https://the-turing-way.netlify.app/welcome>, @turingway, DOI: 10.5281/zenodo.4650221

A book on reproducibility

Same analysis steps on the same dataset produces same answer

The moonshot goal of the project is to make reproducible research "too easy not to do"



Pathways for Collaboration



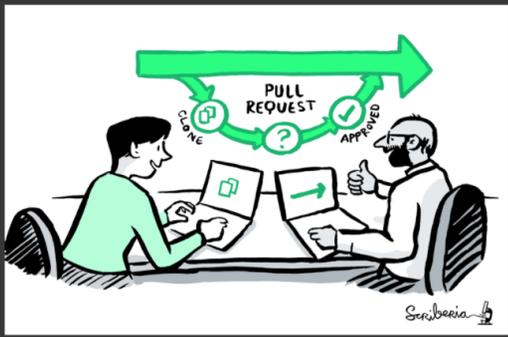
Connect with us



Start where you can



Discuss your ideas



Edit, review, update



Help make it global



Join the community

Join the next book dash event!

BOOK DASH NOVEMBER 2020

Review README.md Arabic translation

Open Translate README.md to Arabic #1586 Changes from all commits File filter... Jump to... 0 / 1 files viewed Review changes

171 README-translated/README-ARABIC.md <> Viewed ...

31 + **## المشروع**

32 + البحث القابل للتكرار ضروري لضمان الوثوق بالعمل العلمي. بدأ الممولون والناشرون في المطالبة بأن تتضمن المنشورات إمكانية الوصول إلى البيانات الأساسية وكود التحليل.

33 + أي أن مشاركة مخرجات البحث مهمة في فهم إدارة البيانات، وعلوم المكتبات، وتطوير البرامج وتقنيات التكامل المستمر، وهي المهارات التي لم يتم تدريسها على نطاق واسع، أو لم يتم توقعها من الباحثين الأكاديميين وعلماء البيانات. نظرًا لأن هذه الأنشطة لا يتم تدريسها بشكل شائع، فإننا ندرك أن عبء المتطلبات واكتساب المهارات الجديدة يمكن أن يكون مخيفًا للأفراد الجدد في هذا العالم.

34 +

35 + أي أن مشاركة مخرجات البحث مهم في فهم إدارة البيانات وعلوم المكتبات وتطوير البرامج وتقنيات التكامل المستمر، وهي المهارات التي لم يتم تدريسها على نطاق واسع أو لم يتم توقعها من الباحثين الأكاديميين وعلماء البيانات، نظرًا لأن هذه الأنشطة لا يتم تدريسها بشكل شائع، فإننا ندرك أن عبء المتطلبات واكتساب المهارات الجديدة يمكن أن يكون مخيفًا للأفراد الجدد في هذا العالم.

36 + طريقة تورنج هي عبارة عن كتيب لدعم الطلاب ومشرفيهم ومموليهم ومحرري المجلات في ضمان أن علم البيانات القابل للتكرار "من السهل جدًا عدم القيام به" حتى بالنسبة للأشخاص الذين لم يعملوا بهذه الطريقة من قبل، ويشمل مواد تدريبية حول التحكم في الإصدار واختبار التحليل والتواصل المفتوح والشفاف مع المستخدمين المستقبليين، والبناء على دراسات الحالة وورش العمل الخاصة بمعهد تورنج.

37 + تم تطوير هذا المشروع بشكل مفتوح ، نرحب بأي وجميع الأسئلة والتعليقات والتوصيات في بوابة جيت هب،

38 + الخاص بنا

39 +

40 + **## الفريق**

41 +

42 + (Turing Institute)

43 + فريق العمل من معهد تورنج

44 +

45 + لمزيد من المعلومات حول كيفية الاتصال بنا ، راجع القسم الأخير.

46 +

47 + **## المساهمة**

Wall of all Contributors

tagged by emojis

eventOrganizing

financial

fundingFinding

ideas

review

security

tool

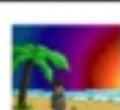
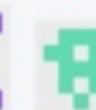
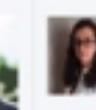
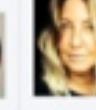
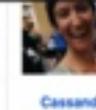
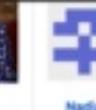
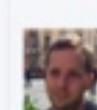
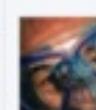
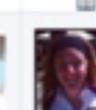
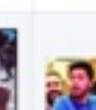
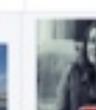
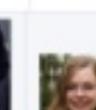
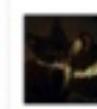
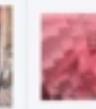
translation

test

tutorial

talk

Join our community of contributors & friends!

 Rachael Ainsworth 📖 📄 🗨️	 Tarek Altam 🌴 📖	 Tania Allard 😊 🗨️	 Diego Alonso Alvarez 😄 🗨️	 Kristijan Anneri 🥒	 Becky Arnold 🗨️ 📄 🗨️	 Louise Bowler 😊 📄 🗨️	 Paula Andrea Martinez 😊	 Lachlan Mason 😊 📄 🗨️	 Rohit Midha 🗨️	 Javier Molden 🗨️	 Beth Montague-Hellan 🗨️	 Alexander Morley 🗨️ 🗨️ 🗨️	 James Myatt 🗨️
 Alex Clarke 🗨️	 Jez Cope 🗨️	 Eric Deub 🗨️	 Stephen Druskat 🗨️	 Elizabeth DuPre 🗨️	 Eglen 🗨️	 Joe Fennell 🗨️	 Olumide 🗨️	 Martin O'Reilly 🗨️	 Pickering 🗨️	 Carole Poyart 🗨️	 Rash Rashid 🗨️	 Robinson 🗨️	 Ananya 🗨️
 Oliver Forrest 🗨️ 😊	 Pooja Gadige 🗨️	 Jason Gates 🗨️ 🗨️	 Sarah Gibson 🗨️ 📄 🗨️	 Oscar Giles 🗨️	 Richard Gilham 🗨️ 😊	 Cassandra Gould van Praag 🗨️	 Ali Seyhun Saral 🗨️	 Chenuki Bhushika Suresh 🗨️	 Nadia Soliman 🗨️	 Andrew Stewart 🗨️ ✅	 Sarah Stewart 🗨️	 Oliver Strickson 🗨️ 🗨️	 Natalie Thurby 🗨️ ⚠️
 Michael Grayling 🗨️	 Liberty Hamilton 🥒	 Tim Head 🗨️ 😊	 Patricia Hertenrich 🗨️ 🗨️ 🗨️ 🗨️	 Rosie Hignam 🗨️ 🗨️ 🗨️	 Ian Hinder 🗨️	 Hieu Hoang 😊	 Gertjan van den Burg 🗨️ 😊 🗨️	 Kerrie Whitaker 🗨️ 🗨️ 🗨️ 🗨️	 Tony Yang 🗨️	 Yo Yehudi 🗨️ 🗨️	 Malvika Sharon 🗨️ 🗨️ 😊	 Esther 🥒	 Anna Hadjiloff 🗨️
 Dan Hobbey 🗨️	 Chris Holdgraf 🗨️ 😊	 Will Hulme 🗨️	 James Kent 🥒	 Greg Kier 🗨️ 🗨️	 Danbee Kim 🗨️	 Anna Krystall 🗨️ 🗨️ 🗨️ 🗨️	 Kevin Kutzmann 🗨️ 😊	 Eric Leung 🥒	 Clare Liggins 🗨️	 Robin Long 🗨️	 Christopher Lovell 🗨️	 Eirini Malliaraki 🗨️	 Chris Markiewicz 🗨️

Upcoming Workshop by the turing way

Register for the free workshop '[Boost your research reproducibility with Binder](#)' run by Sarah Gibson from the Turing Way as part of our Research Software Camp on research accessibility.

The poster features a red background with white text and logos. At the top left is the Software Sustainability Institute logo. At the top right is the Research Software Camp logo. The central text reads: 'Sarah Gibson on "Boost your research reproducibility with Binder"'. Below this, it says 'Research Software Camp on Research Accessibility' and '22 February - 5 March 2021'. At the bottom, there is a 'Supported by:' section with logos for the University of Manchester, University of Oxford, University of Southampton, and several research councils including Arts and Humanities Research Council, Engineering and Physical Sciences Research Council, Medical Research Council, and Natural Environment Research Council.

Resources:

- [The Turing Way](#)
- [The CMU ML Blog](#)
- [Redesign open science for Asia, Africa and Latin America](#)
- [Open Science Beyond Open Access: For and with communities, A step towards the decolonization of knowledge](#)
- [Embracing science as it is: beyond Nobel-like research](#)
- [Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA](#)

Thank you so much!

batool@liverpool.ac.uk

Twitter: @batool664

Join RLadiesDammam: @RLadiesDammam