

# Replicability crisis in science?

*Replicability in biomedical research: limitations, risks and solutions*

# Before we start

- This is the way I perceived the replability crisis, my point of view (from 2005)
- Replicability: if I follow the set of instructions (protocol), using the same biological system and compounds (reagents), do I get the same results, previously reported by others?

# The beginning

*Over the past decade, before pursuing a particular line of research, scientists in the haematology and oncology department at the biotechnology firm Amgen in Thousand Oaks, California, tried to confirm published findings related to that work. Fifty-three papers were deemed 'landmark' studies. It was acknowledged from the outset that some of the data might not hold up, because papers were deliberately selected that described something completely new, such as fresh approaches to targeting cancers or alternative clinical uses for existing therapeutics. Nevertheless, scientific findings were confirmed in only 6 (11%) cases. Even knowing the limitations of preclinical research, this was a shocking result.*

(Begley, C., Ellis, L. Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012))

My reaction: Shocked and scandalised



# Why?

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

*Only 5% of agents that have anticancer activity in preclinical development are licensed after demonstrating sufficient efficacy in phase III testing, which is much lower than, for example, 20% for cardiovascular disease.*

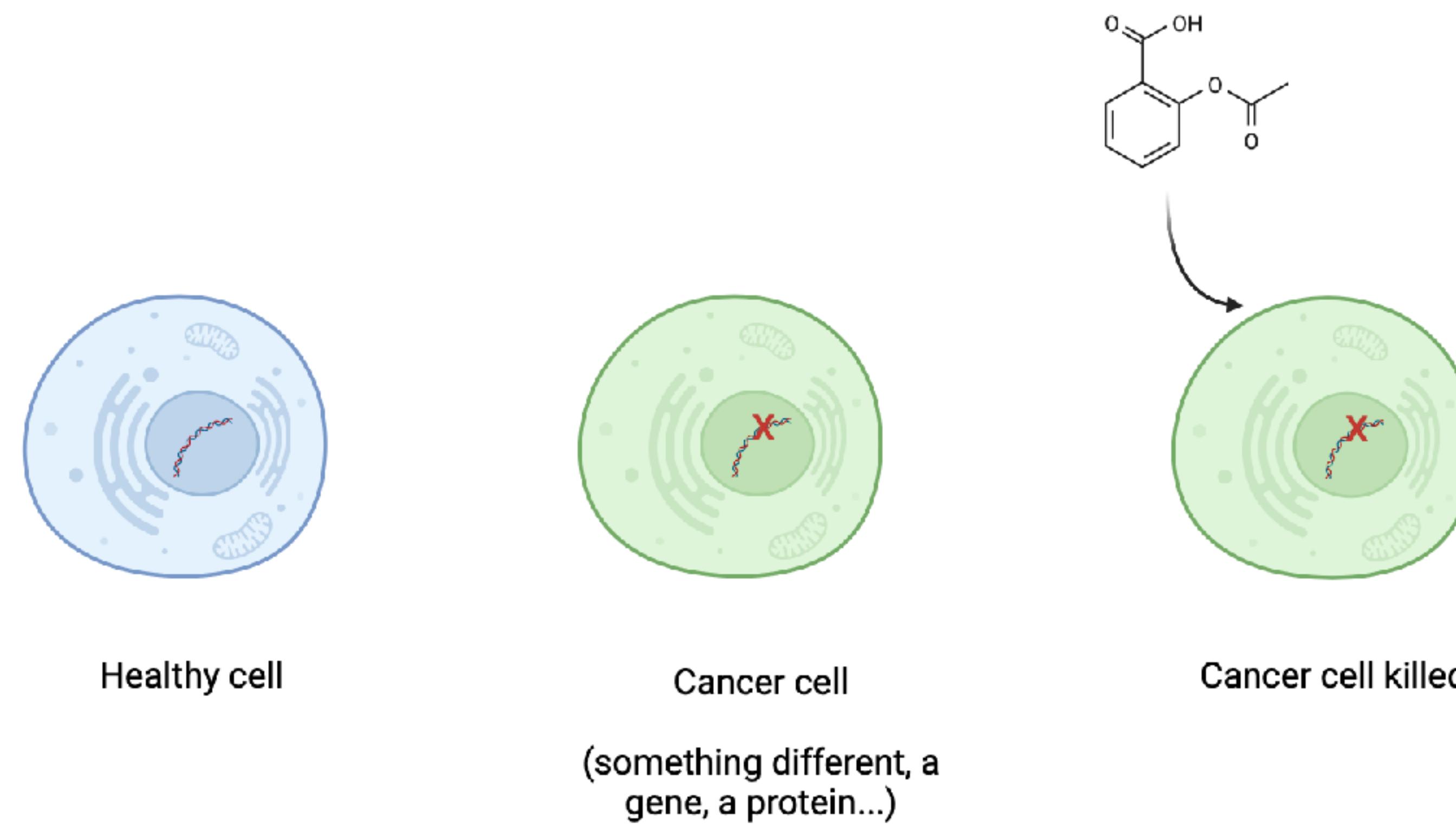
*(Hutchinson, L., Kirk, R. High drug attrition rates—where are we going wrong?. Nat Rev Clin Oncol 8, 189–190 (2011))*

- Literature flooded with wrong studies?
- Cancer research should be aimed at the wellbeing, not just at publishing
- Research is very expensive

# But

- 11%? how can science make any progress? better making random assumptions, or tossing a coin, than reading papers? ( $50\% > 11\%$ )
- My first-hand experience, as an experimentalist, is that the first attempt at replicating a published protocol often fails.
- It was, and still is, a common practice to collaborate or contact the group that developed a technique (e.g. reprogramming), while setting it up.

# What is cancer research?



- Find something different in cancer cells vs healthy ones
  - mutated gene
  - more/less active gene
  - new mechanism / behaviour
- Exploit such difference to design a treatment selectively affecting cancer cells

# Back to the story



## Project Overview

The *Reproducibility Project: Cancer Biology* was an 8-year effort to replicate experiments from high-impact cancer biology papers published between 2010 and 2012. The project was a collaboration between the [Center of Open Science](#) and [Science Exchange](#) with all papers published as part of this project available in a [collection at eLife](#) and all replication data, code, and digital materials for the project available in a [collection on OSF](#). When preparing replications of **193 experiments** from **53 papers** there were a number of challenges.

- High profile papers from 2010-2012 (top citations, altmetric...)
- Replication efforts from 2014 to 2022
- A total of 50 experiments from 23 papers were repeated, generating data about the replicability of a total of 158 effects. Most of the original effects were positive effects (136), with the rest being null effects (22). A majority of the original effect sizes were reported as numerical values (117), with the rest being reported as representative images (41).
- Several papers in eLife reporting the outcome of individual studies + 2 final outputs

# Some examples - a bad one

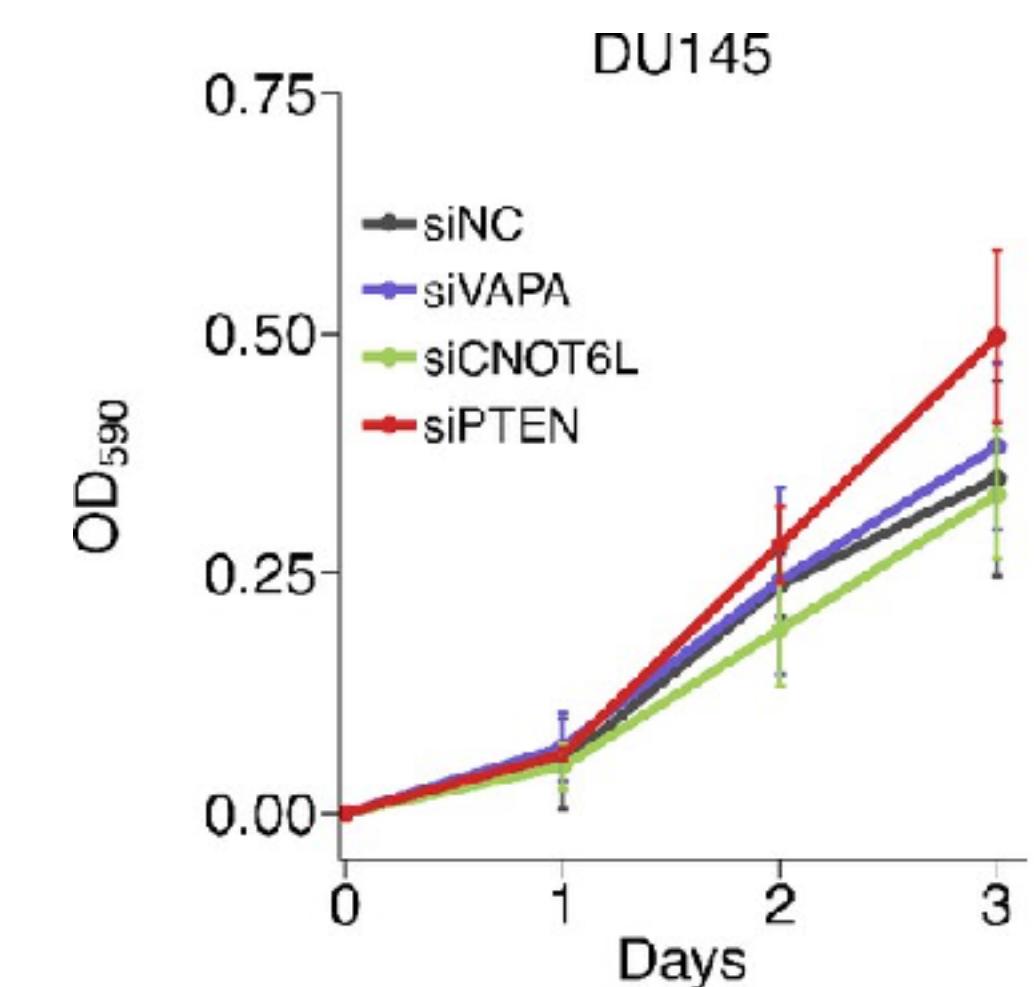
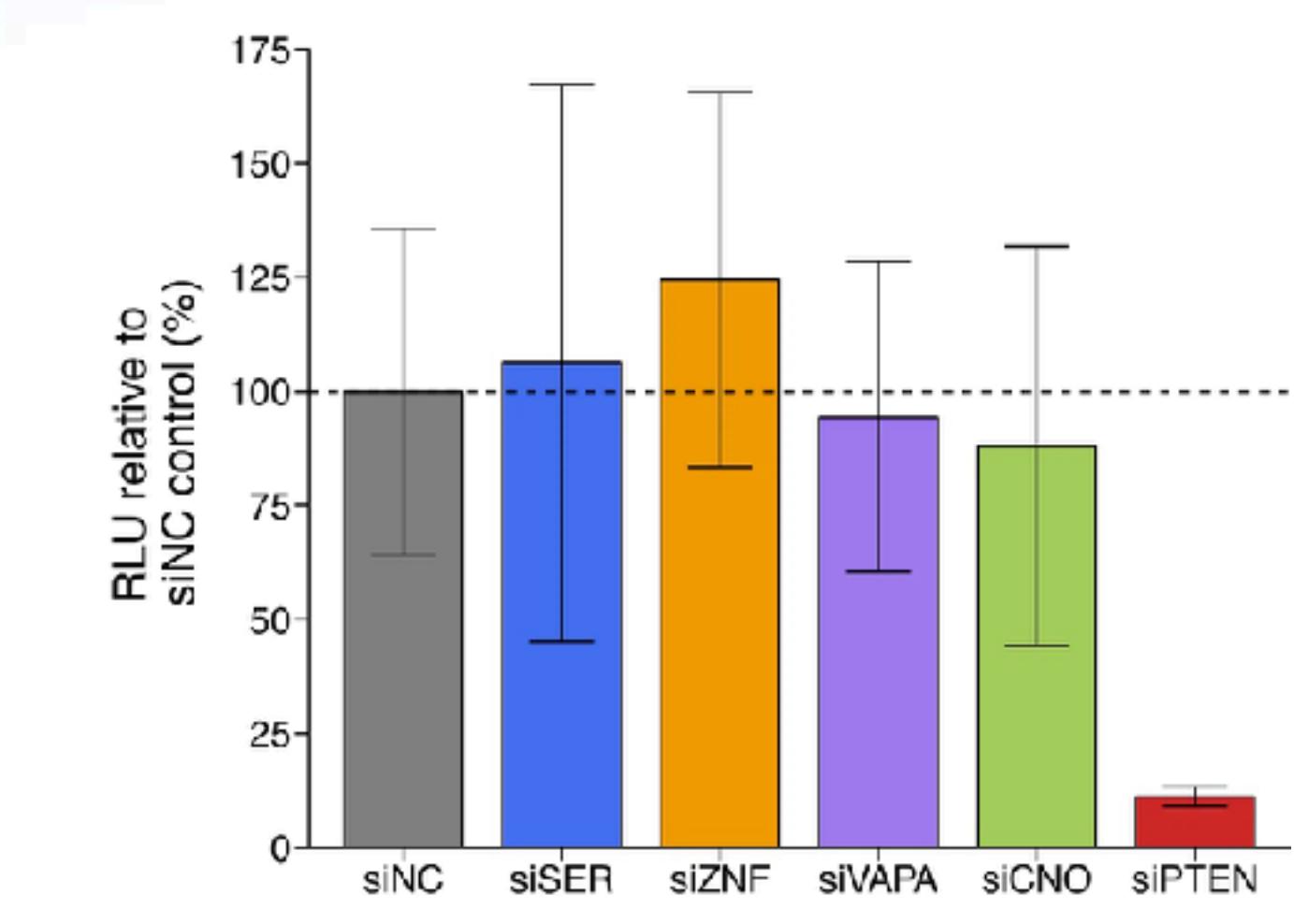
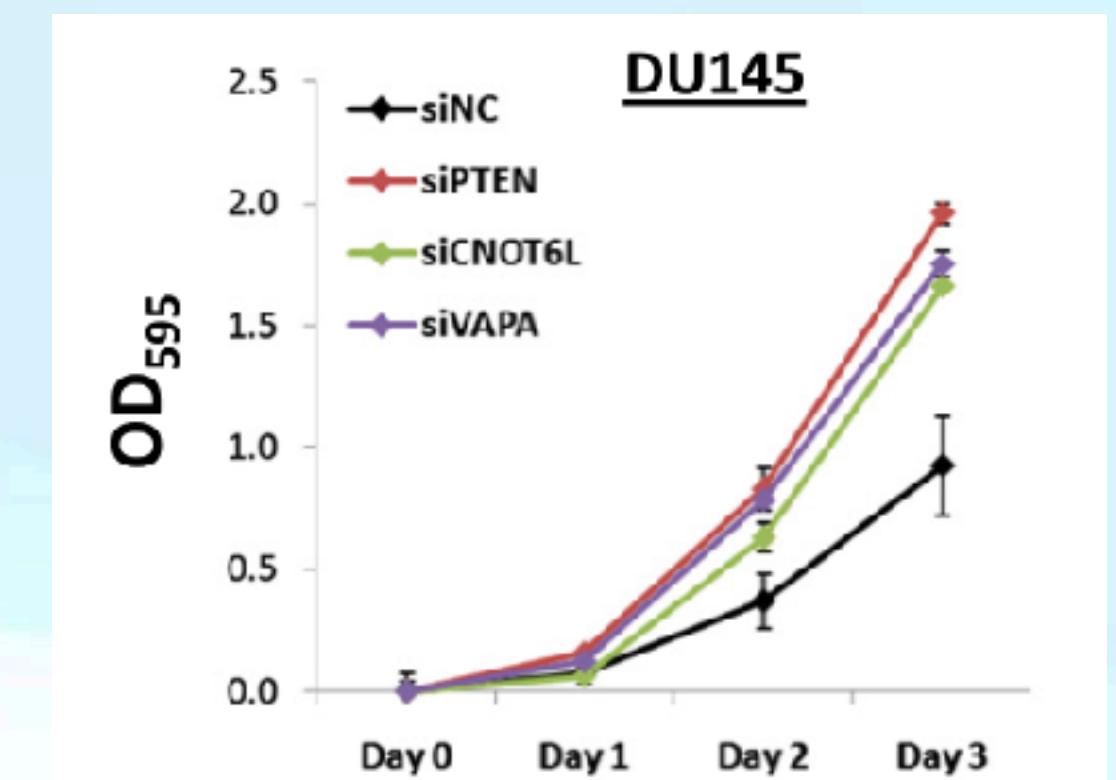
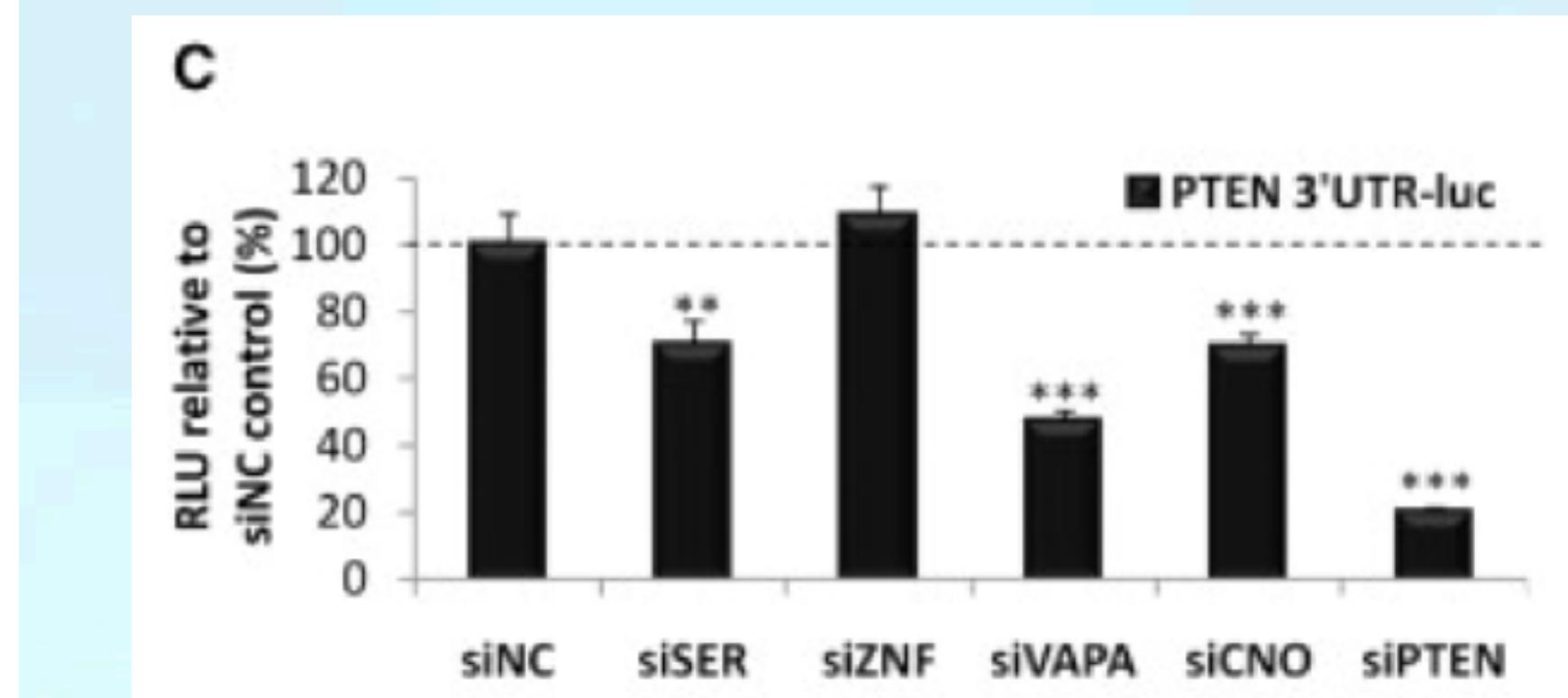
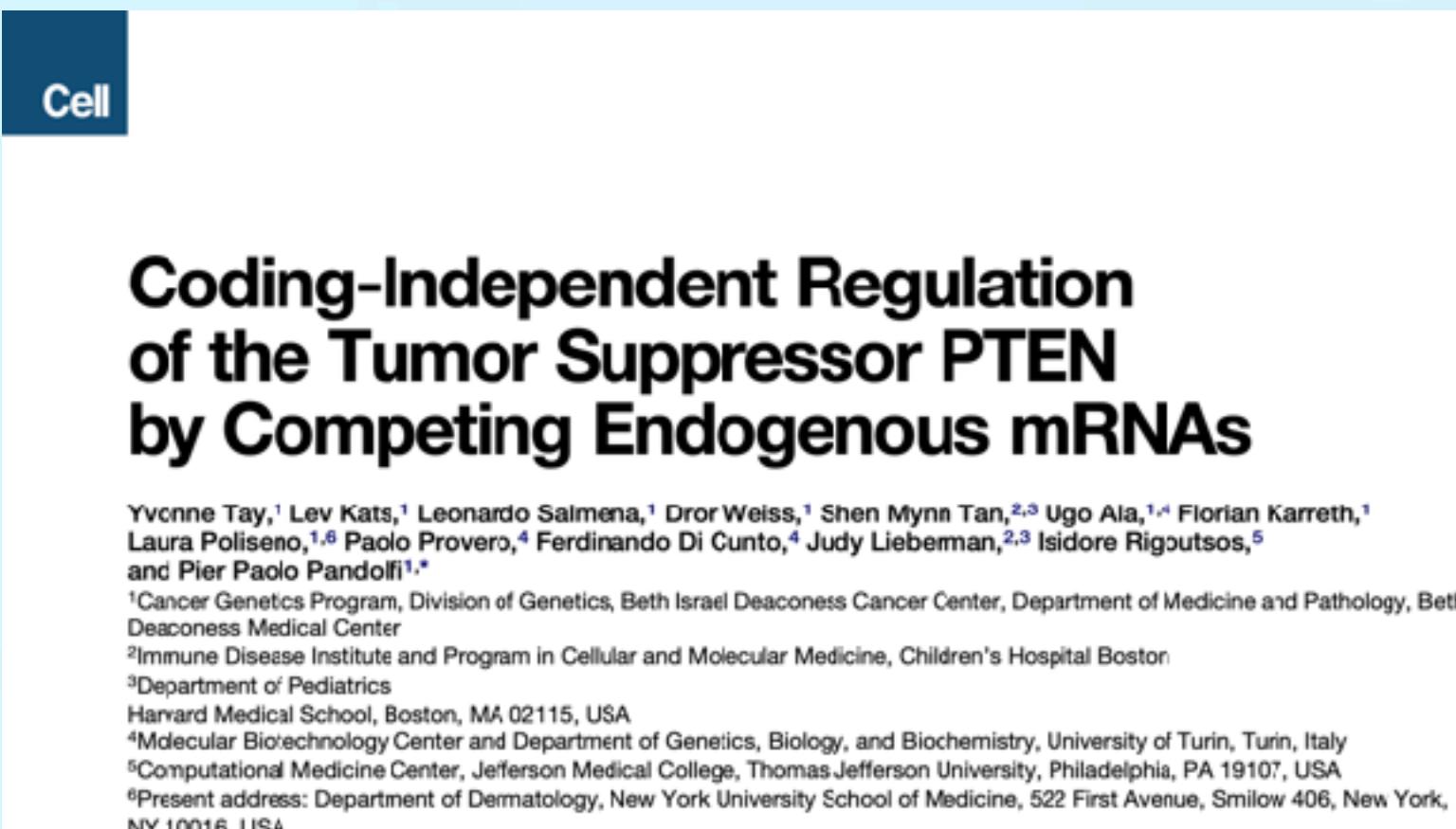


Figure 3: Putative PTEN mRNA variants modulate PTEN expression.  
 (A) Western blot for PTEN variants in DU145 cells transfected with siRNAs against putative PTEN mRNA variants. ERIN116BEP, ZF-250, siVAPA, siWAP, and siCNOT6L reduce PTEN protein levels. siSER, siZNF, and siCNO do not significantly affect PTEN protein levels as measured by Western blot.  
 (B) Luciferase activity in DU145 cells cotransfected with siRNAs against PTEN variants and a luciferase-PTEN 3'UTR reporter construct.  
 (C) Luciferase activity in DU145 cells cotransfected with PTEN 3'UTR 2'UTPs and a luciferase-PTEN 3'UTR reporter construct.  
 (D) Western blot showing PTEN protein in response to overexpression of siWAP, siZNF, and siCNOT6L in DU145 cells.  
 (E) Quantitation of PTEN protein above in (D).  
 (F) Western blot for PTEN in HCT116-DK cells and HCT116-DK cells transfected with siRNAs against PTEN variants.  
 (G) Quantitation of PTEN protein above in (F).  
 (H) Quantitation of PTEN protein above in (G).  
 (I)  $p < 0.05$ ;  $p < 0.01$ ;  $p < 0.001$ . See also Figure 3.

# Some examples - a quite good one

nature

Explore content ▾ About the journal ▾ Publish with us ▾

nature > letters > article

Published: 02 October 2011

## Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia

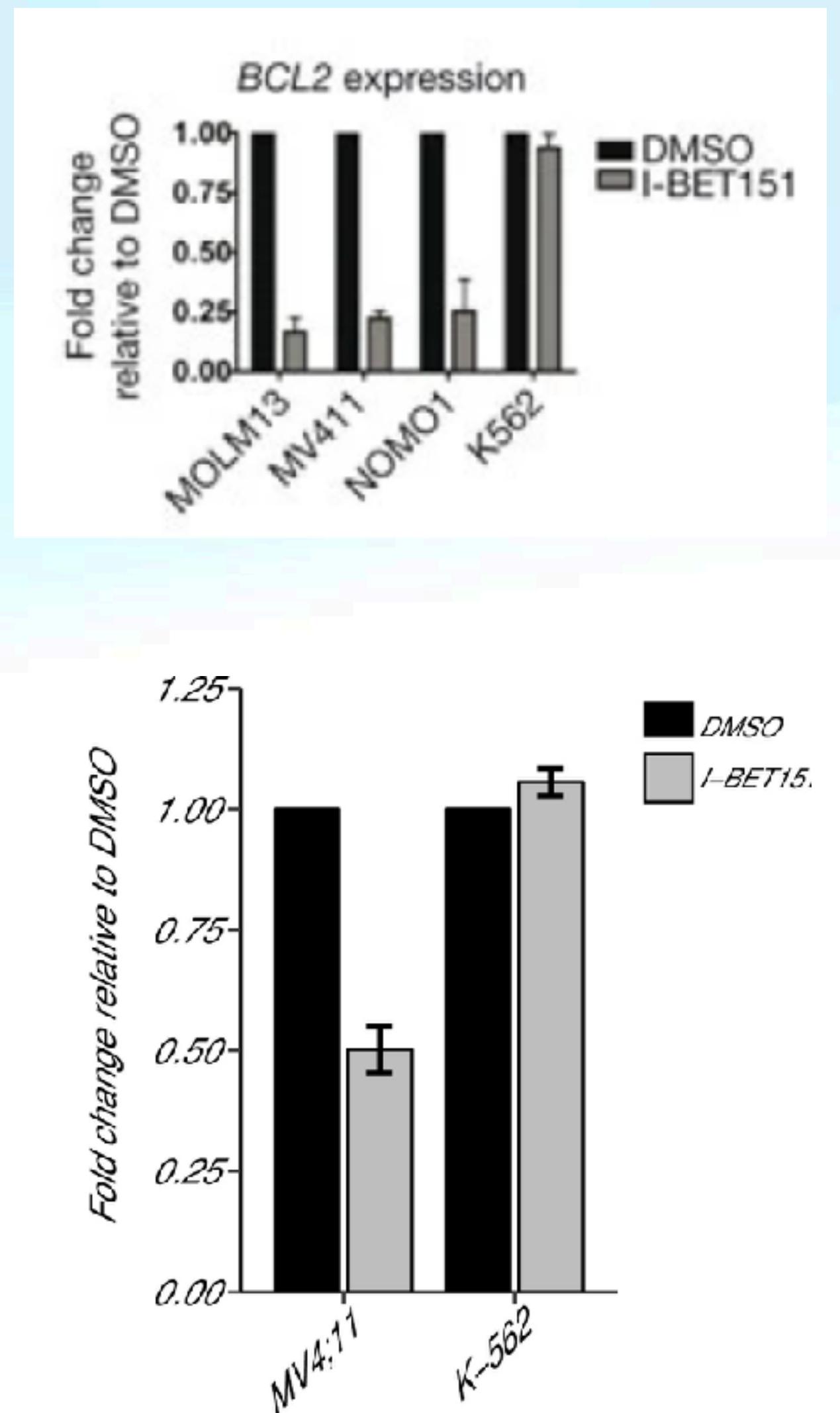
Mark A. Dawson , Rab K. Prinjha, Antje Dittmann, George Giopoulos, Marcus Bantscheff, Wai-In Chan, Samuel C. Robson, Chun-wa Chung, Carsten Hopf, Mikhail M. Savitski, Carola Huthmacher, Emma Gudgin, Dave Lugo, Soren Beinke, Trevor D. Chapman, Emma J. Roberts, Peter E. Soden, Kurt R. Auger, Olivier Mirguet, Konstanze Doechner, Ruud Delwel, Alan K. Burnett, Phillip Jeffrey, Gerard Drewes, ... Tony Kouzarides  + Show authors

Nature 478, 529–533 (2011) | Cite this article

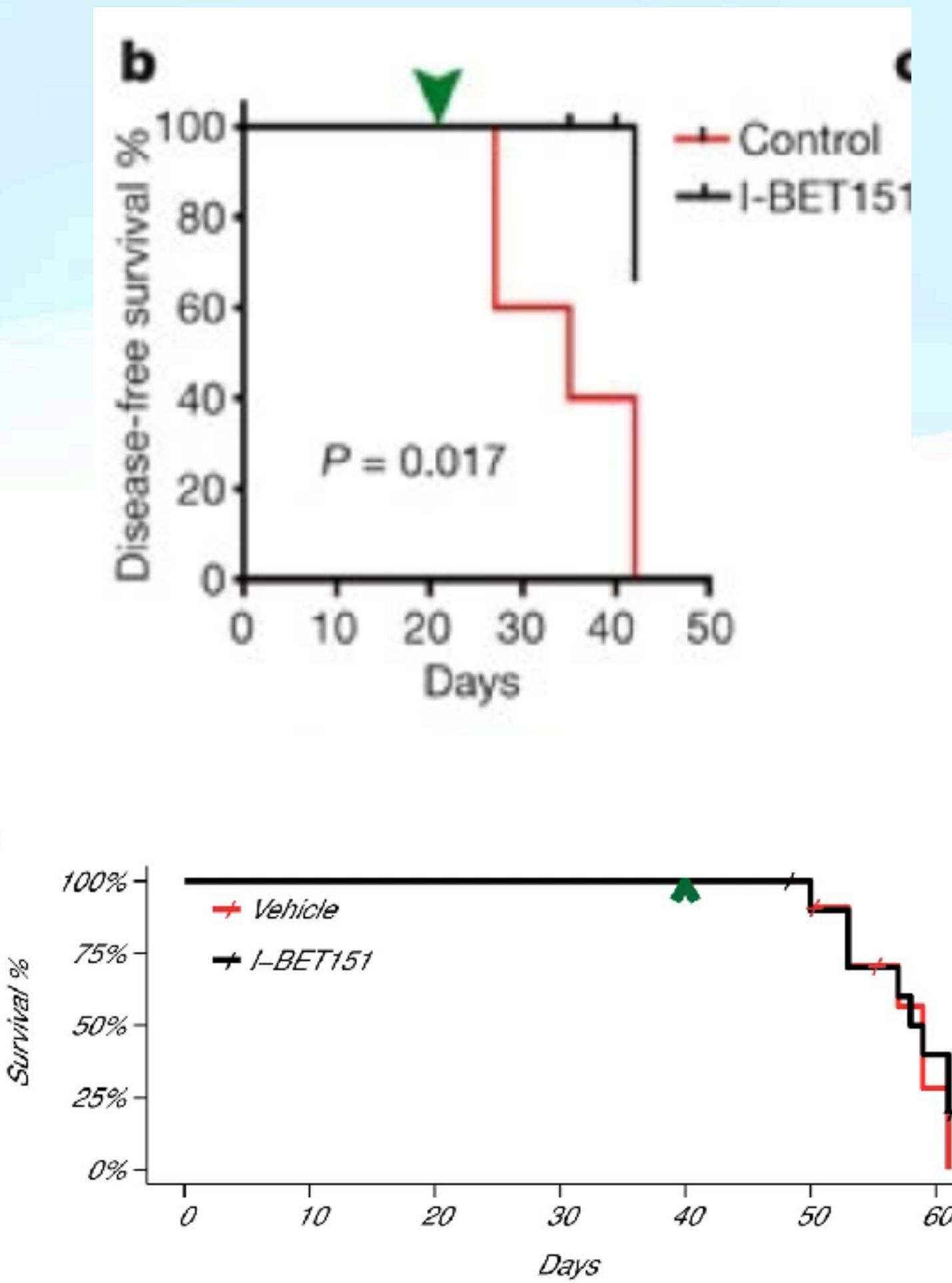
Some forms of Leukemia are linked to alterations in an epigenetic mechanism.  
So a compound acting on such mechanism should kill those specific leukemic cells.

Confirmed the differential sensitivity and mechanism of action in cell lines (in vitro).

But...



the compound prolongs the life of mice with leukemia



# Overall..

Research Article  
Cancer Biology, Computational and Systems Biology

## Investigating the replicability of preclinical cancer biology

Timothy M Errington , Maya Mathur, Courtney K Soderberg, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, Brian A Nosek

Center for Open Science, United States; Quantitative Sciences Unit, Stanford University, United States; Science Exchange, United States; University of Virginia, United States

Dec 7, 2021 · <https://doi.org/10.7554/eLife.71601>  

We employed seven methods to assess replicability, and some of these methods were not suitable for all the effects in our sample. One method compared effect sizes: for positive effects, the median effect size in the replications was 85% smaller than the median effect size in the original experiments, and 92% of replication effect sizes were smaller than the original. The other methods were binary – the replication was either a success or a failure – and five of these methods could be used to assess both positive and null effects when effect sizes were reported as numerical values. For positive effects, 40% of replications (39/97) succeeded according to three or more of these five methods, and for null effects 80% of replications (12/15) were successful on this basis; combining positive and null effects, the success rate was 46% (51/112).

BUT:

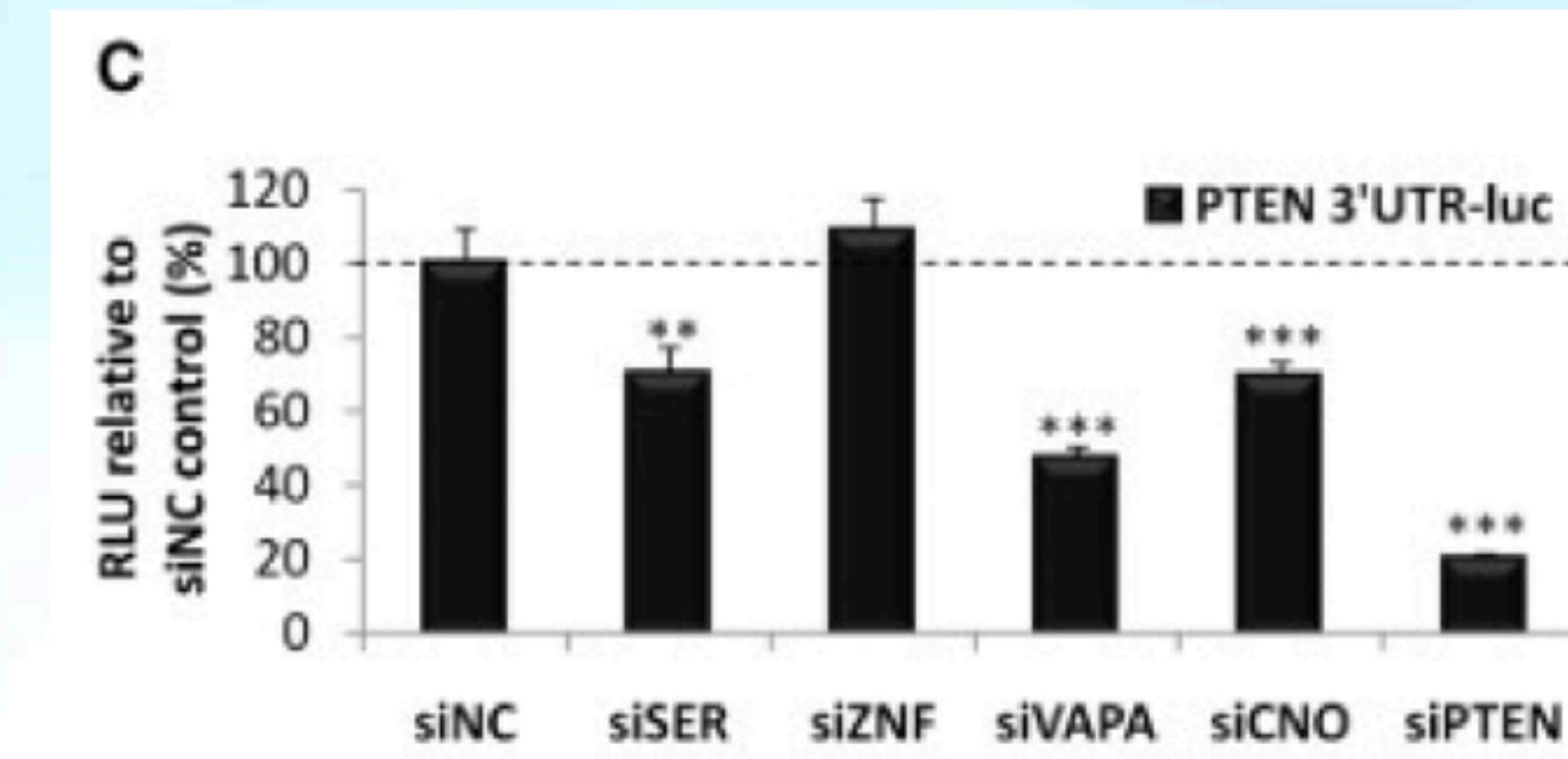
Overall same direction: 80.1 %

Overall direction and statistical significance: 68.0%

	Papers	Experiments	Effects	All outcomes
Total number	23	50	158	188
<strong>ORIGINAL POSITIVE RESULTS</strong>				
<em>Numerical results</em>				
Same direction	17 of 19 (89%)	26 of 35 (74%)	80 of 101 (79%)	95 of 116 (82%)
Direction and statistical significance	8 of 19 (42%)	17 of 33 (52%)	42 of 97 (43%)	44 of 112 (39%)
<em>Representative images</em>				
Same direction	9 of 10 (90%)	12 of 16 (75%)	28 of 35 (80%)	34 of 45 (76%)
Direction and statistical significance	3 of 8 (40%)	7 of 12 (58%)	14 of 22 (64%)	14 of 22 (64%)
<strong>ORIGINAL NULL RESULTS</strong>				
<em>Numerical results</em>				
Same direction	N/A	N/A	N/A	N/A
Direction and statistical significance	9 of 11 (82%)	10 of 12 (83%)	11 of 15 (73%)	10 of 20 (50%)
<em>Representative images</em>				
Same direction	N/A	N/A	N/A	N/A
Direction and statistical significance	3 of 3 (100%)	3 of 3 (100%)	4 of 5 (80%)	4 of 5 (80%)

# Replicability still below 100% and the effects are milder than those reported, why?

- 1. Bad statistics (n=3, what test?, single data points?, what kind of replicates? raw data?)



Mean  $\pm$  SD; n  $\geq$  4; \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001

Very hard to know what we are looking at.  
How do you replicate something that you do not fully understand?

# Replicability still below 100% and the effects are milder than those reported, why?

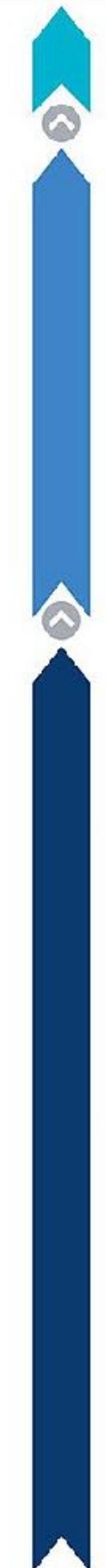
- 2. Sloppy protocols
- (missing details, chain of references...)
- “Risotto effect”



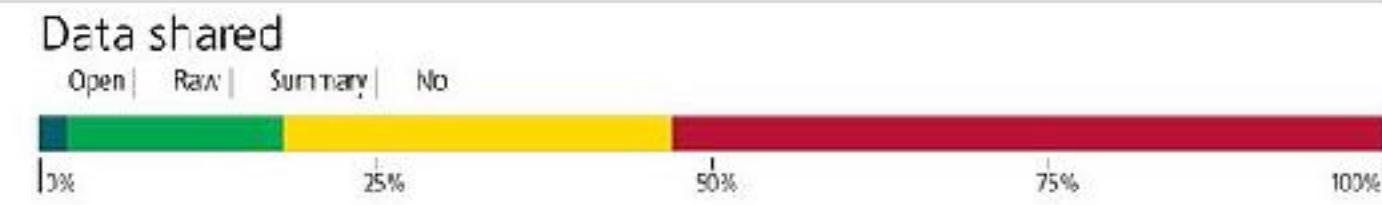
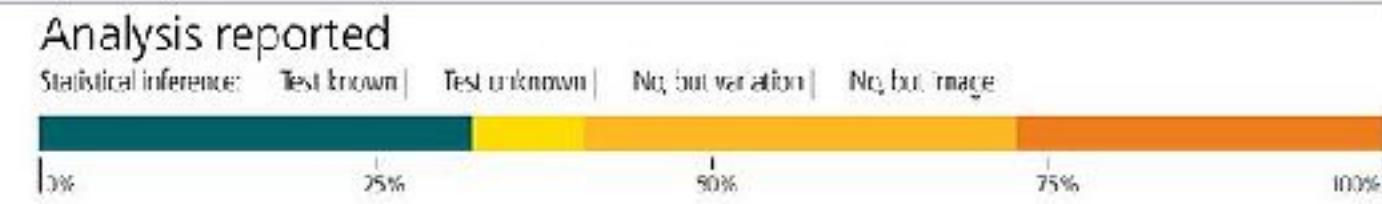
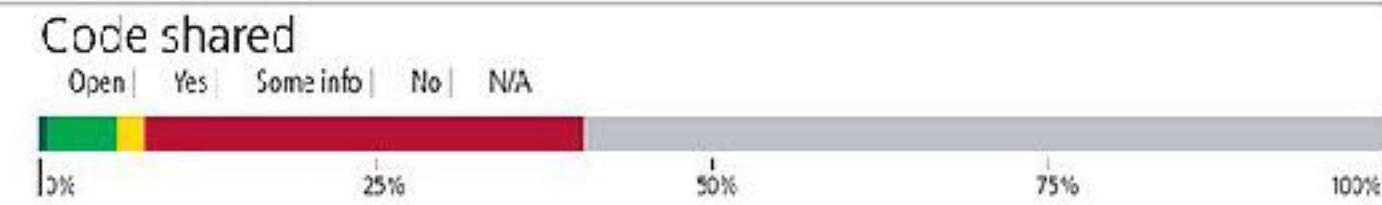
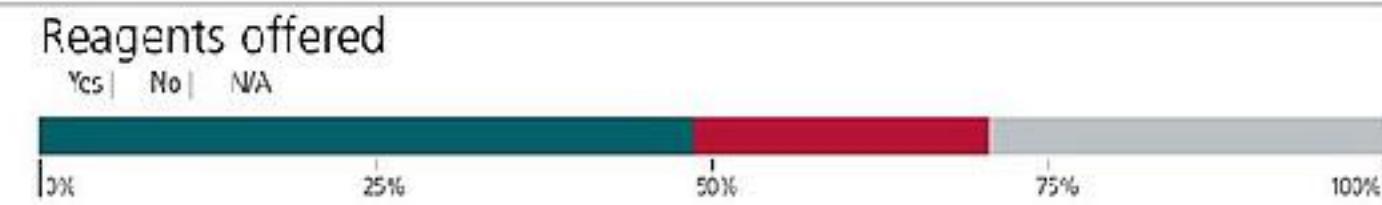
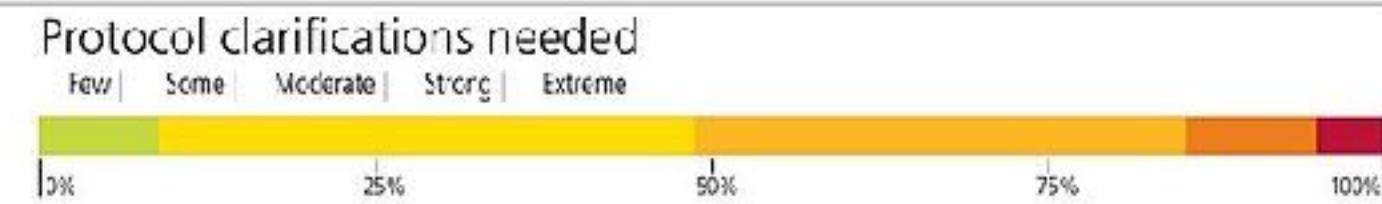
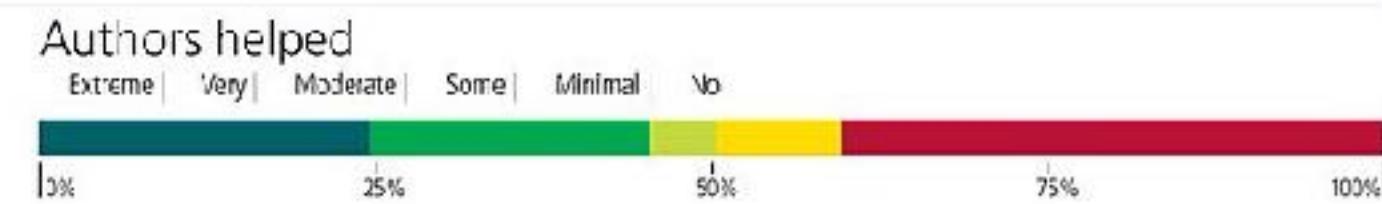
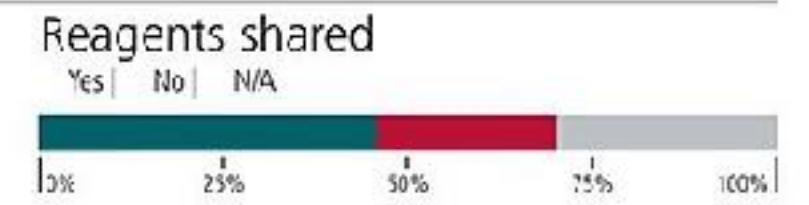
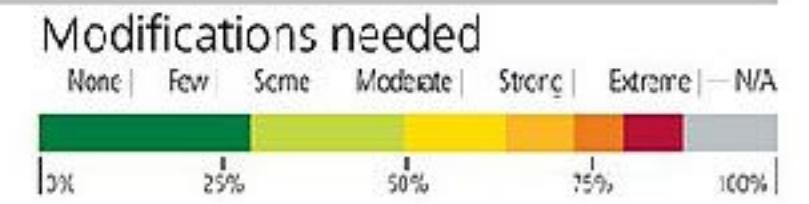
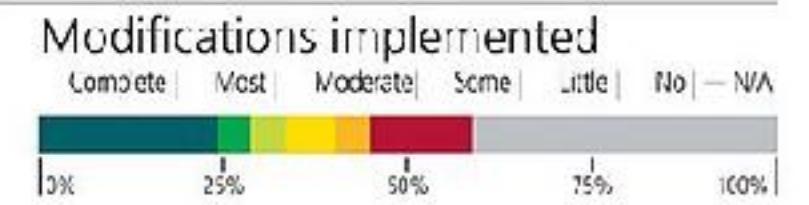
**DESIGNED**  
193 experiments

**COMPLETED**  
50 experiments

**INITIATED**  
87 experiments



## BARRIERS



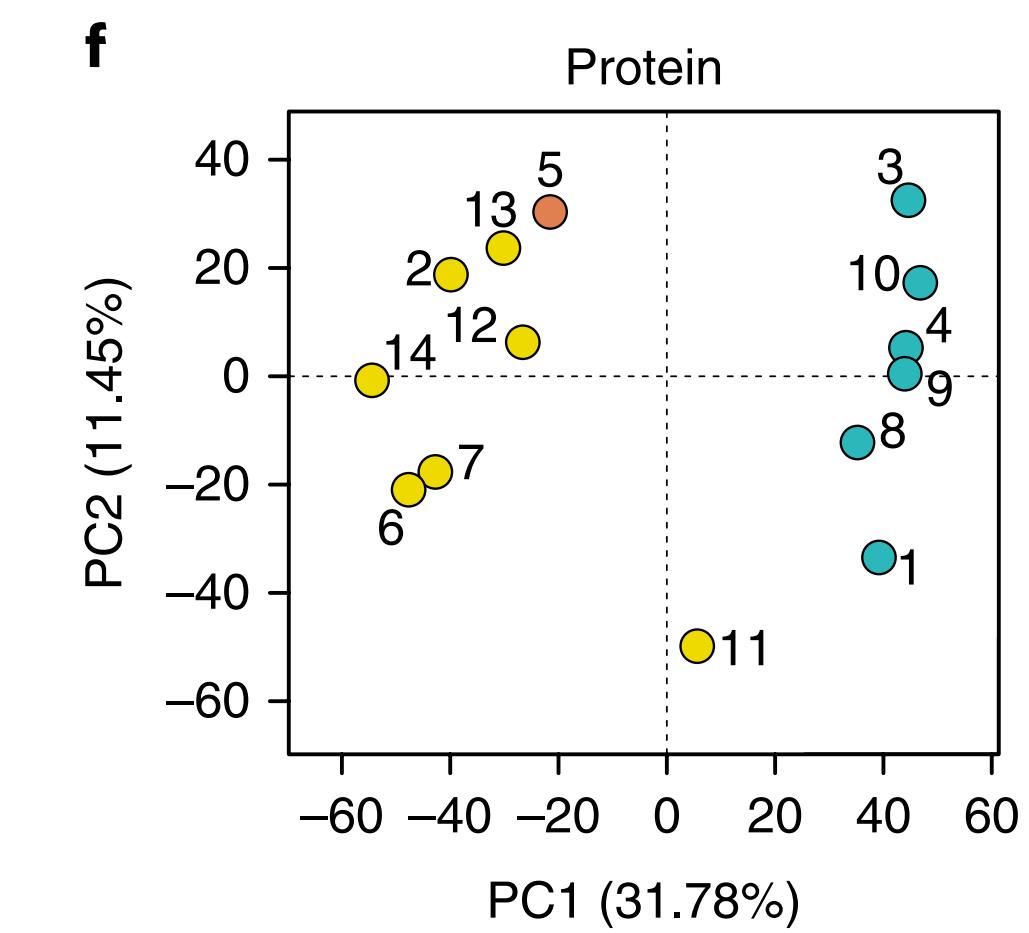
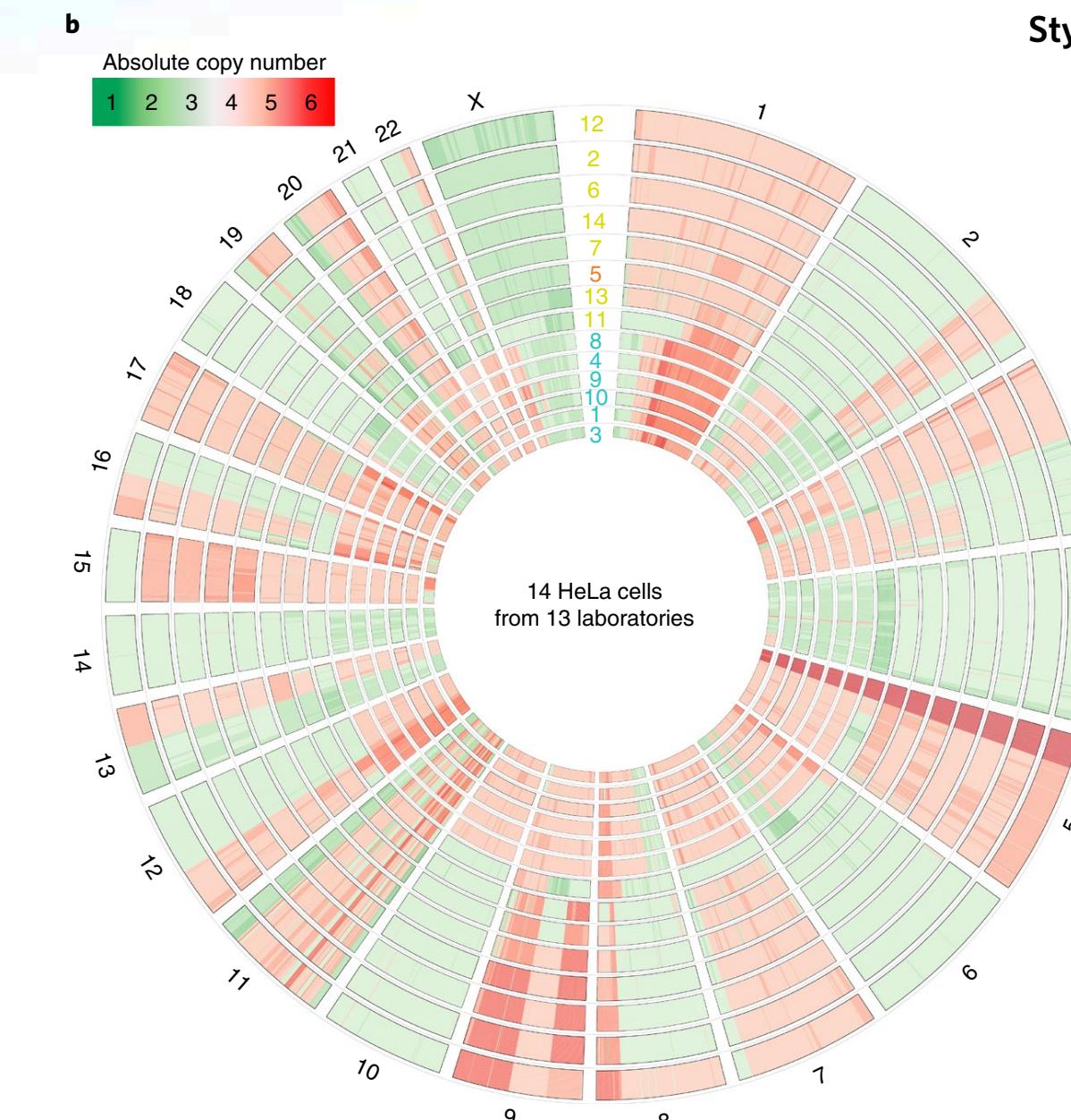
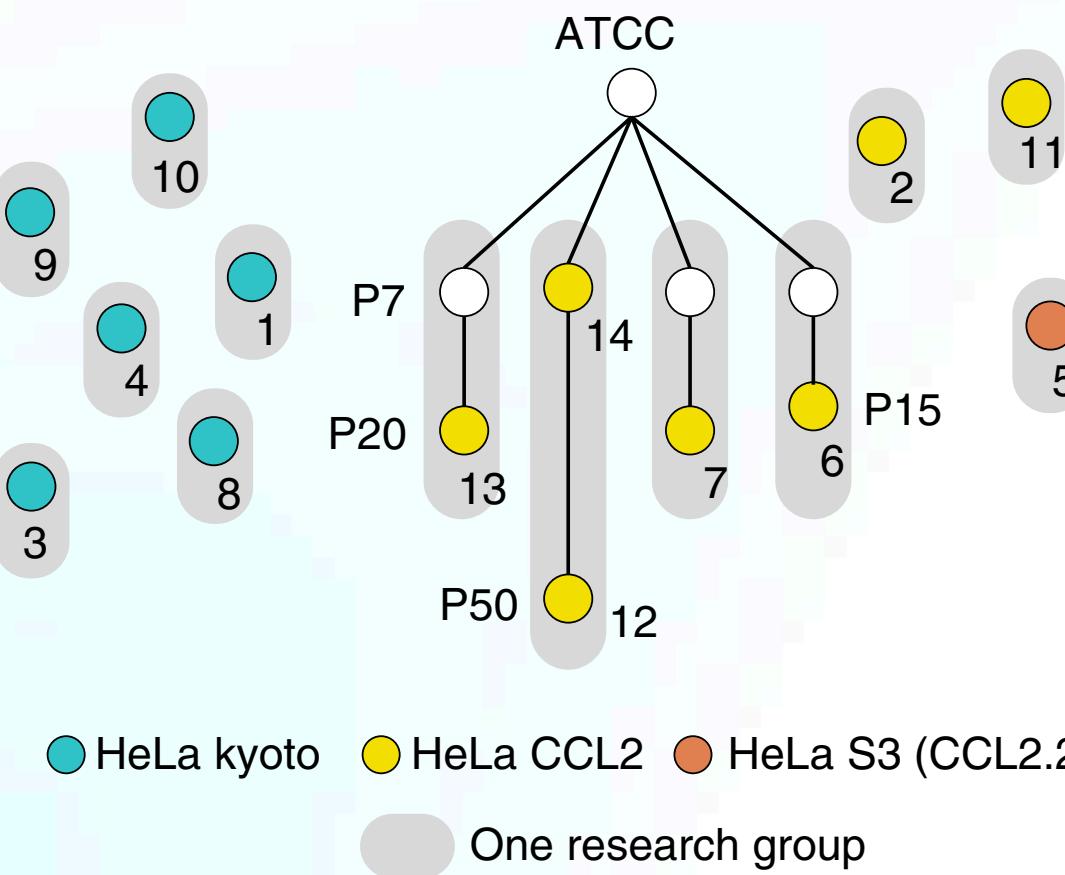
# Replicability still below 100% and the effects are milder than those reported, why?

- 3. Variability in reagents (cell lines, batch effect of molecules...)

Cancer cell lines: immortal cell derived from cancers that are kept in culture for long time leading to potential accumulation of genetic mutations and contaminations

HeLa cells, a “classic” cell line, referenced in 100,000 papers derived from cervical cancer of Henrietta Lacks in 1951

# Are all HeLa cells the same?



ANALYSIS

<https://doi.org/10.1038/s41587-019-0037>

**nature  
biotechnology**

# **Replicability still below 100% and the effects are milder than those reported, why?**

- 1. Bad statistics (n=3, what test?, single data points?, what kind of replicates? raw data?)
- 2. Sloppy protocols (missing details, often refer to other papers...)
- 3. Variability in reagents (cell lines, batch effect of molecules...)
- 4. Secretive attitude
- 5. “Funny” approach to negative results / cherry picking.
- 6. Plain fraud or mistakes (wrong tube or cell line)

# However, do not feel too sad



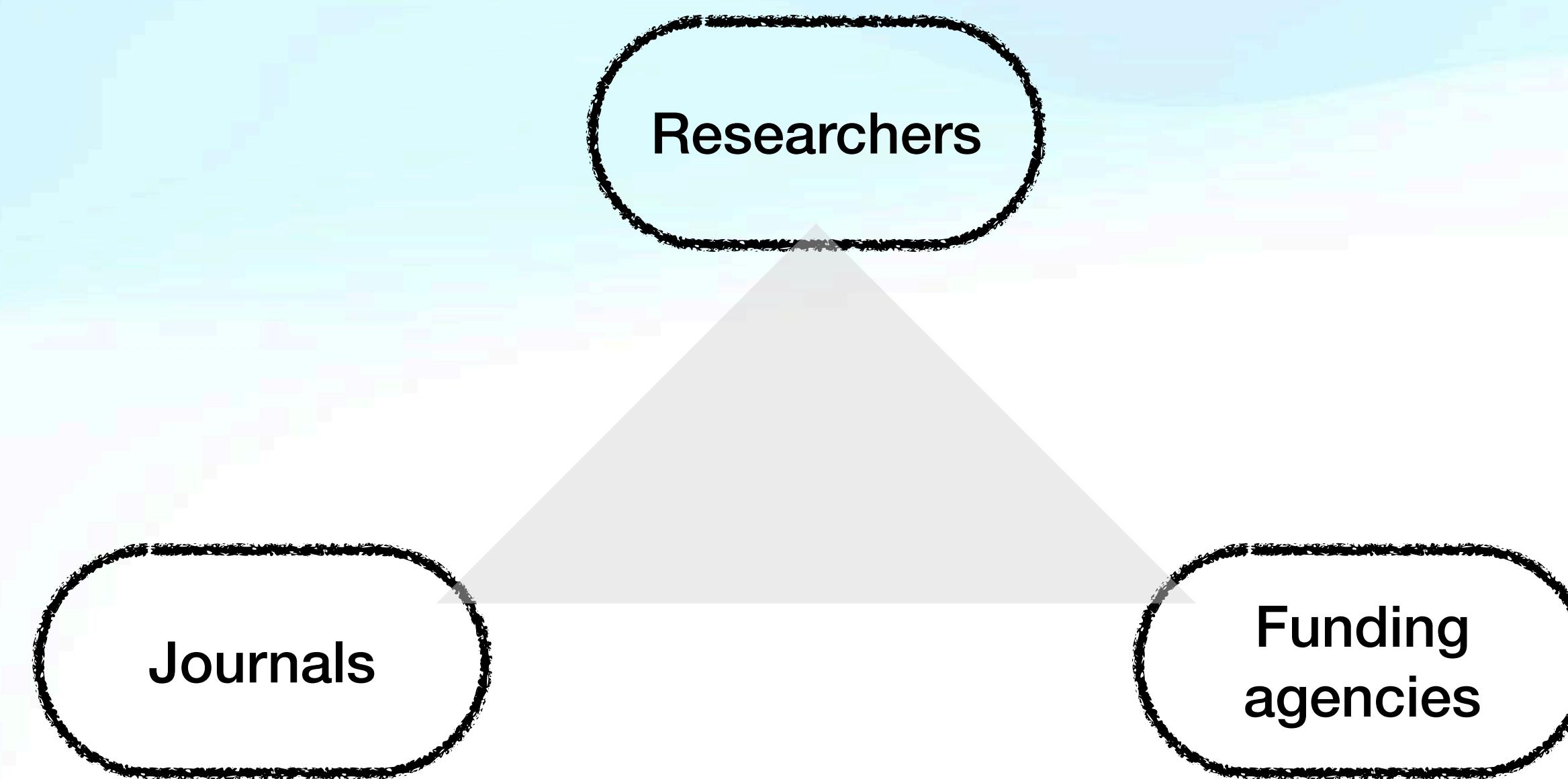
*In labs where knowledge seeks its way,  
Amidst the doubts that cloud the day,  
Scientific minds in deep despair,  
Confronted the replicability crises, unaware.*

*The data skewed, the hypotheses fell,  
But in those moments, they learned so well.  
From setbacks and doubts, they gained insight,  
To strengthen their methods, to see the light.*

*Scientific hearts, resilient and bold,  
Emerging from crisis, their stories told.  
For in every challenge, they did discern,  
The lessons they needed to truly learn.*

*(Obviously generated with ChatGPT)*

# Things are getting much better



# Things are getting much better

- 1. Statistics
- Currently mandatory in several high profile journals to include raw data, show single datapoints, indicate the p-value, the test used (tails, normality, correction...).
- For example, as a standard in my lab, before we put together the paper, we collect and organise all raw data, repeat analyses and graphs, then make final figures and write the paper.

# Things are getting much better

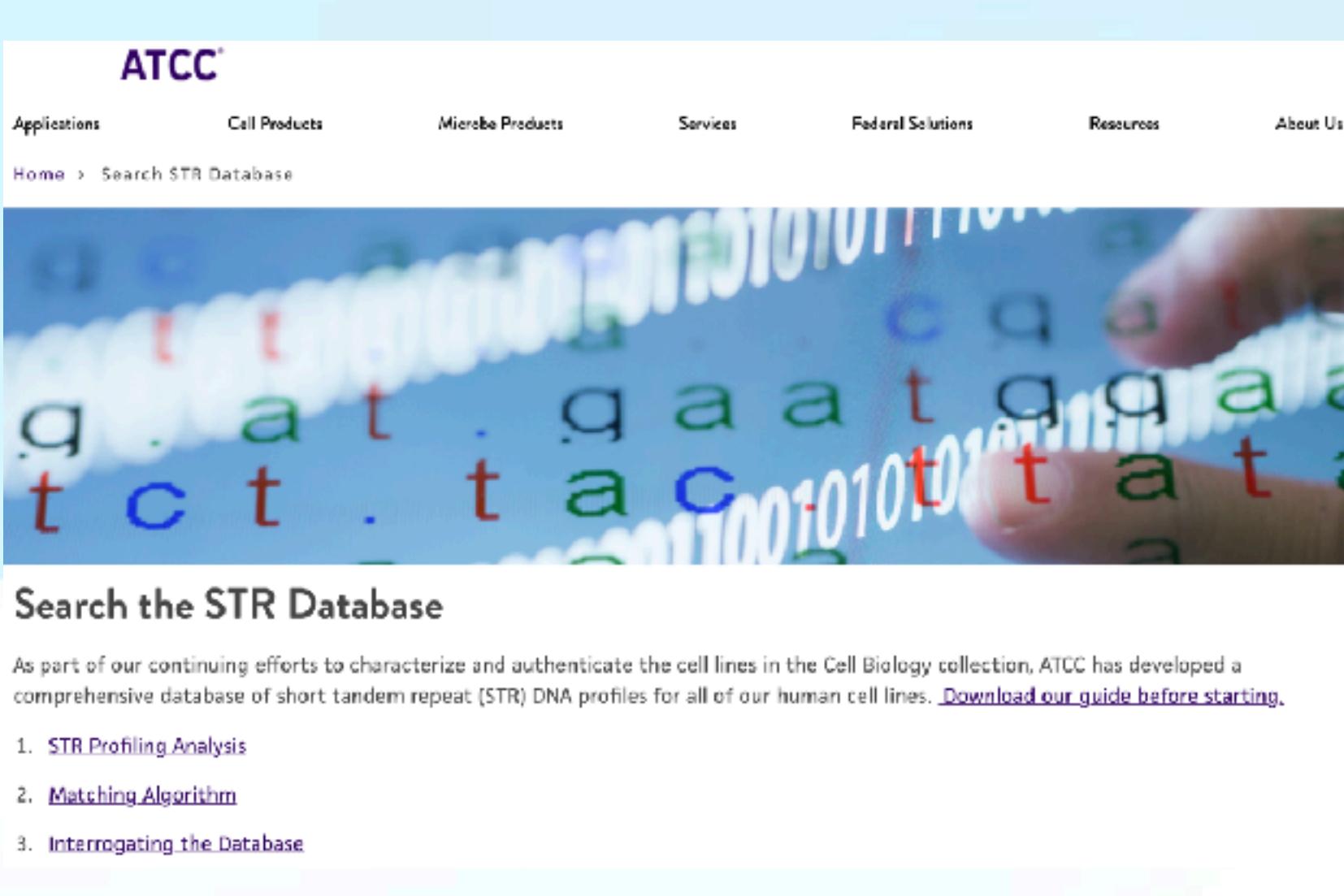
- 2. Sloppy protocols
- The Methods section of biological papers is now much more detailed (e.g. Star protocols) and often a detailed protocol counts as a separate manuscript (a good incentive).

The screenshot shows the homepage of the **nature protocols** website. At the top, there are navigation links for [View all journals](#), [Search](#), [Explore content](#), [About the journal](#), and [Publish with us](#). Below this, a breadcrumb trail shows the path: [nature](#) > [nature protocols](#). A prominent feature on the left is a box titled **This Week's Featured Protocol:** [Template-assisted membrane crystallization for monoclonal antibodies purification](#). To the right of this box is a colorful 3D ribbon model of a protein structure.

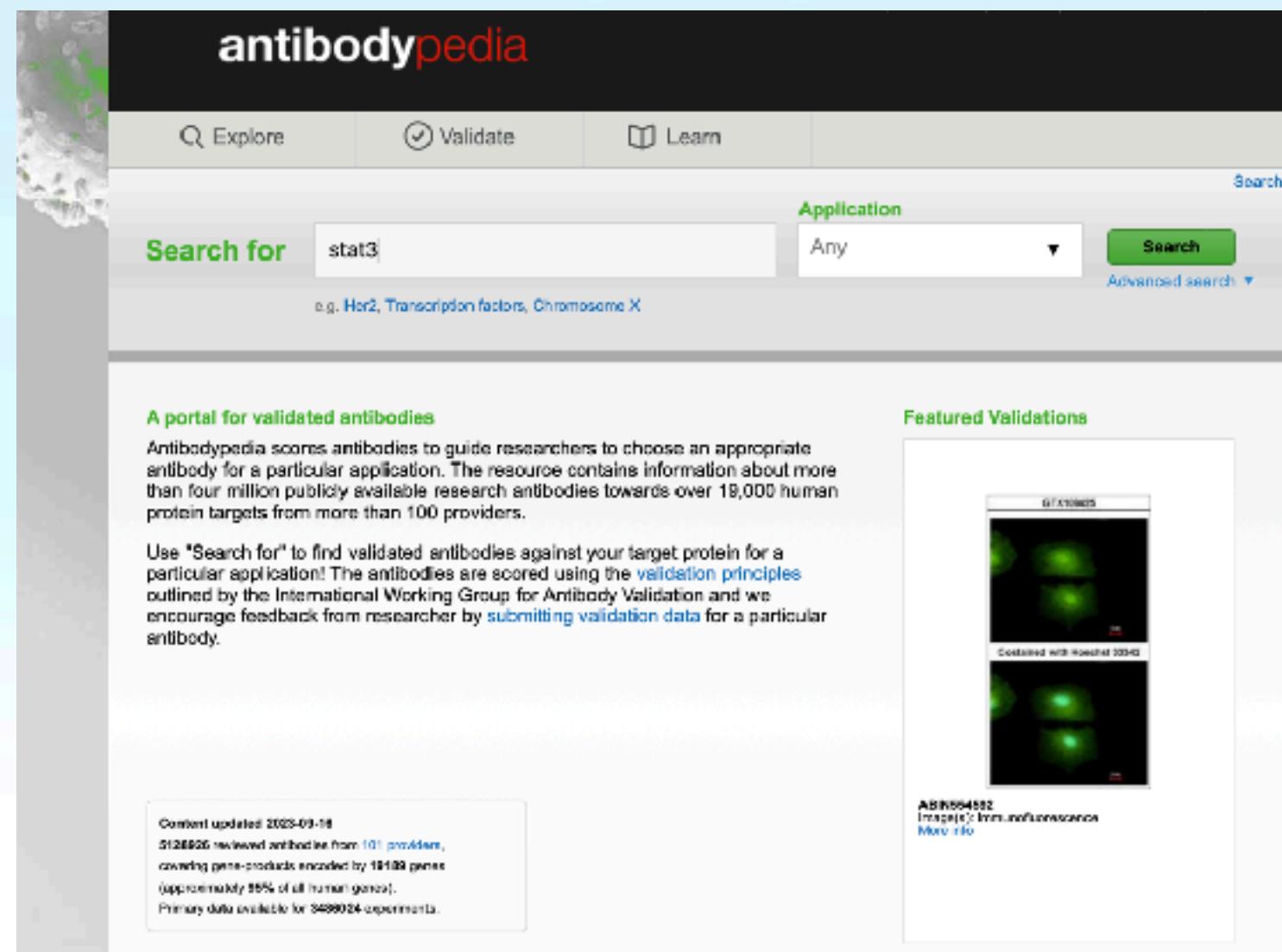
The screenshot shows the homepage of the **STAR Protocols** website. The header features the journal name in large white letters on a blue background, with the text "Open access" below it. On the left, there is a thumbnail image of a brain organoid. To the right of the thumbnail are links for [Current issue](#), [Table of contents](#), and [Submit article](#). Below this, a section titled **Featured articles** displays two protocol entries. The first article is titled "Protocol for human brain organoid transplantation into a rat visual cortex to model neural repair" by Jgarnadze et al., with a thumbnail showing various experimental steps. The second article is titled "Characterizing interaction between the juxtamembrane region of the single transmembrane protein and membrane using chemically synthesized peptides" by Sato and Shinohara, also with a thumbnail showing experimental details.

# Things are getting much better

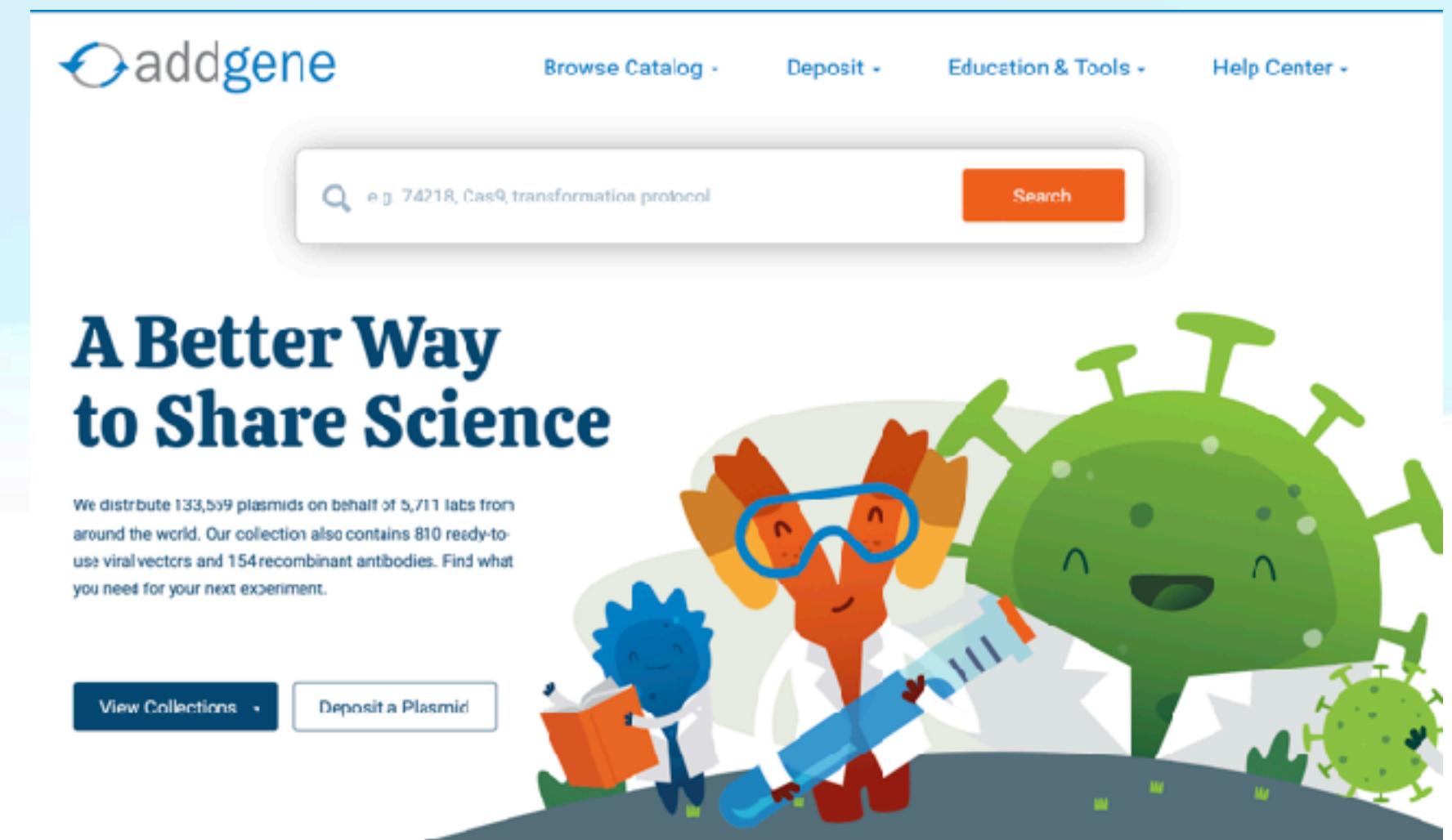
- 3. Variability in reagents



The screenshot shows the ATCC STR Database homepage. At the top, there's a navigation bar with links for Applications, Cell Products, Microbe Products, Services, Federal Solutions, Resources, and About Us. Below the navigation is a breadcrumb trail: Home > Search STR Database. The main content features a large image of a DNA sequence with red and blue highlights. Below the image is a section titled "Search the STR Database". It explains that ATCC has developed a comprehensive database of STR DNA profiles for all human cell lines. It includes links for "STR Profiling Analysis", "Matching Algorithm", and "Interrogating the Database".



The screenshot shows the antibodypedia homepage. The header includes a search bar with placeholder text "Search for stat3", a dropdown menu for "Application" set to "Any", and a "Search" button. Below the search bar is a section titled "A portal for validated antibodies" with a brief description. To the right is a "Featured Validations" section showing a thumbnail of a fluorescence microscopy image. At the bottom left is a small box stating "Content updated 2023-09-18" and "2128825 reviewed antibodies from 101 providers, covering gene-products encoded by 18100 genes (approximately 95% of all human genes). Primary data available for 3489024 experiments."



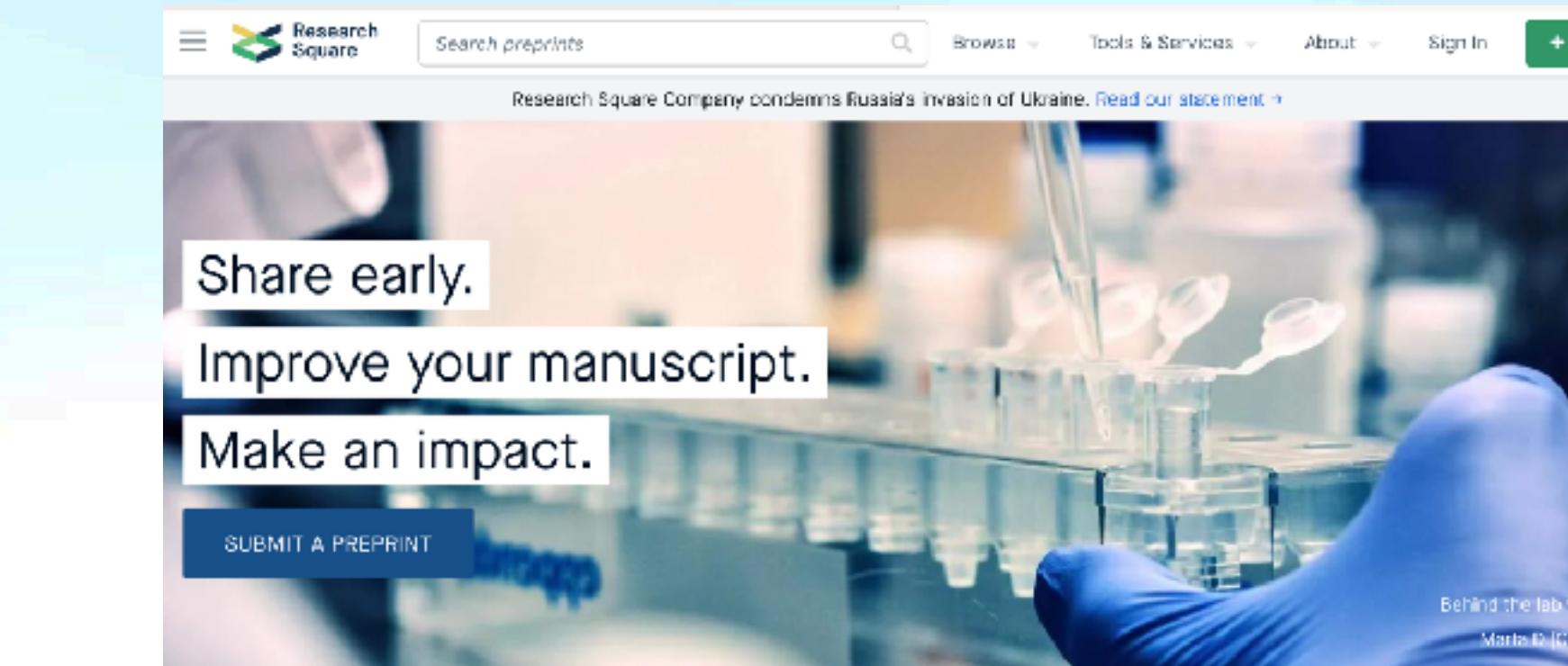
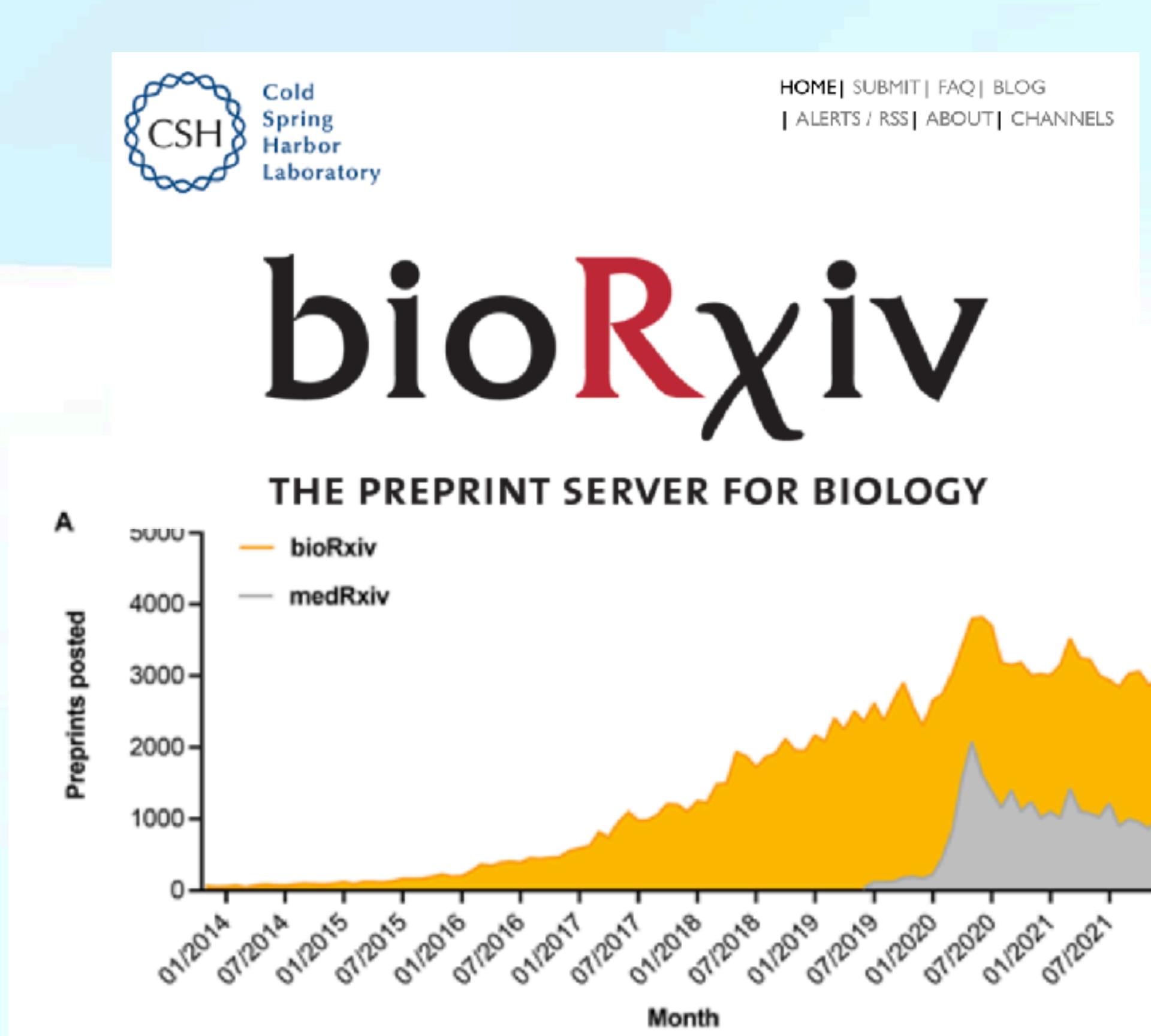
The screenshot shows the addgene homepage. The header includes links for Browse Catalog, Deposit, Education & Tools, and Help Center. Below the header is a search bar with placeholder text "e.g. 7421R, Cas9, transformation protocol" and a "Search" button. The main content features a large heading "A Better Way to Share Science" above a cartoon illustration of a green virus-like character and two smaller characters in lab coats. Text below the heading states "We distribute 133,559 plasmids on behalf of 5,711 labs from around the world. Our collection also contains 810 ready-to-use viral vectors and 154 recombinant antibodies. Find what you need for your next experiment." There are also "View Collections" and "Deposit a Plasmid" buttons.

Now it is compulsory to validate the cell lines used, to validate antibodies, to indicate what batch/lot of every reagent have been used and to share new reagents.

By sharing you get more citations! (a good incentive)

# Things are getting much better

- 4. Secretive attitude
- Over the last 10 years preprints became very common in biological research.
- All major publishing groups developed their own preprint servers and they “transfer” from bioRxiv.



The Cell Press Sneak Peek page is displayed. It includes the CellPress logo, a 'Sneak Peek A PREVIEW OF PAPERS UNDER REVIEW' section, and a 'Cell Stem Cell' preview. A 'Cell Press Sneak Peek Statistics' box on the right provides the following data:

- 1,985 Authors
- 151 Papers
- 5,525 Downloads
- 103,439 Abstract Views

# Things are getting much better

- 5. and 6. Questionable behaviour
- Transparency and more open science make it harder to misbehave or to make mistakes
- You build your scientific reputation by publishing good, robust papers AND by sharing protocols, data, reagents.

# In sum

- Replicability is a crucial element of the scientific method, which sets Science apart from pseudoscience, urban legends etc etc
- An experiment that can be replicated is not necessarily a “good” experiment.
- Science “fixes” itself in many ways, and we are witnessing it.

# Thanks to



# You!

