

# Where are the Original Modulators of New Targets and New Bioactive Molecular Frameworks Published? A Comparison between Patents and Scientific Literature

Stephanie Kay Ashenden

*ska35@cam.ac.uk*

28<sup>th</sup> April 2017

UK QSAR



# Contents

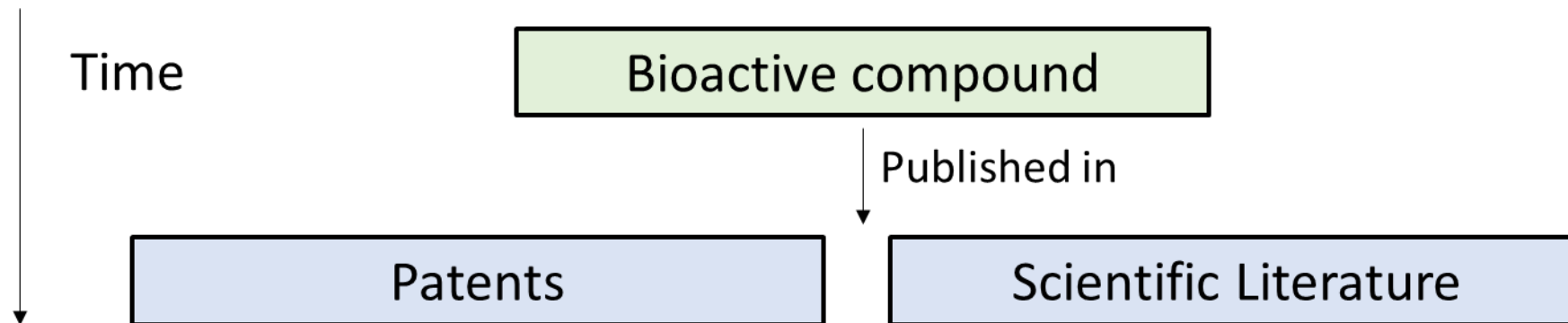
- Introduction
  - Patents and Scientific Literature
  - Drug Discovery
- Aim
- Methodology
- Results
  - Analysis of the data by time and target class
  - Where are novel compound-target annotations first published
  - Where are novel structures first published
- Conclusions

# Introduction – Drug Discovery

- Drug Discovery is an ever evolving environment



- Discovery of a broader range of druggable targets
- Improved technological systems



# Introduction – Patents and Peer-Reviewed Scientific Literature

## Patents

- Used to protect novel intellectual property in the drug discovery process
- Minimum time for disclosure is 18 months
- A compound can appear in multiple patents if it has been patented as part of a formulation or if it was patented for a different use

## Scientific Literature

- Novel findings to be available to the whole scientific community
- Citations reflect how useful the publication was found by others

# Aim

To analyse the sources of novel bioactive compounds over time as well as their target annotations

# Methodology

ChEMBL 21

GOSTAR Journal

GOSTAR Patent

Compounds  
Standardised using  
In House  
Methodology

First occurrence of compound–target annotation retained

Filtering applied:

- Human Targets Only
- Activities  $\leq 1\mu\text{M}$  “exact” value reported
  - $K_i$ ,  $K_d$ ,  $\text{IC}_{50}$  and  $\text{EC}_{50}$
- Molecular Weight  $\leq 900\text{Da}$

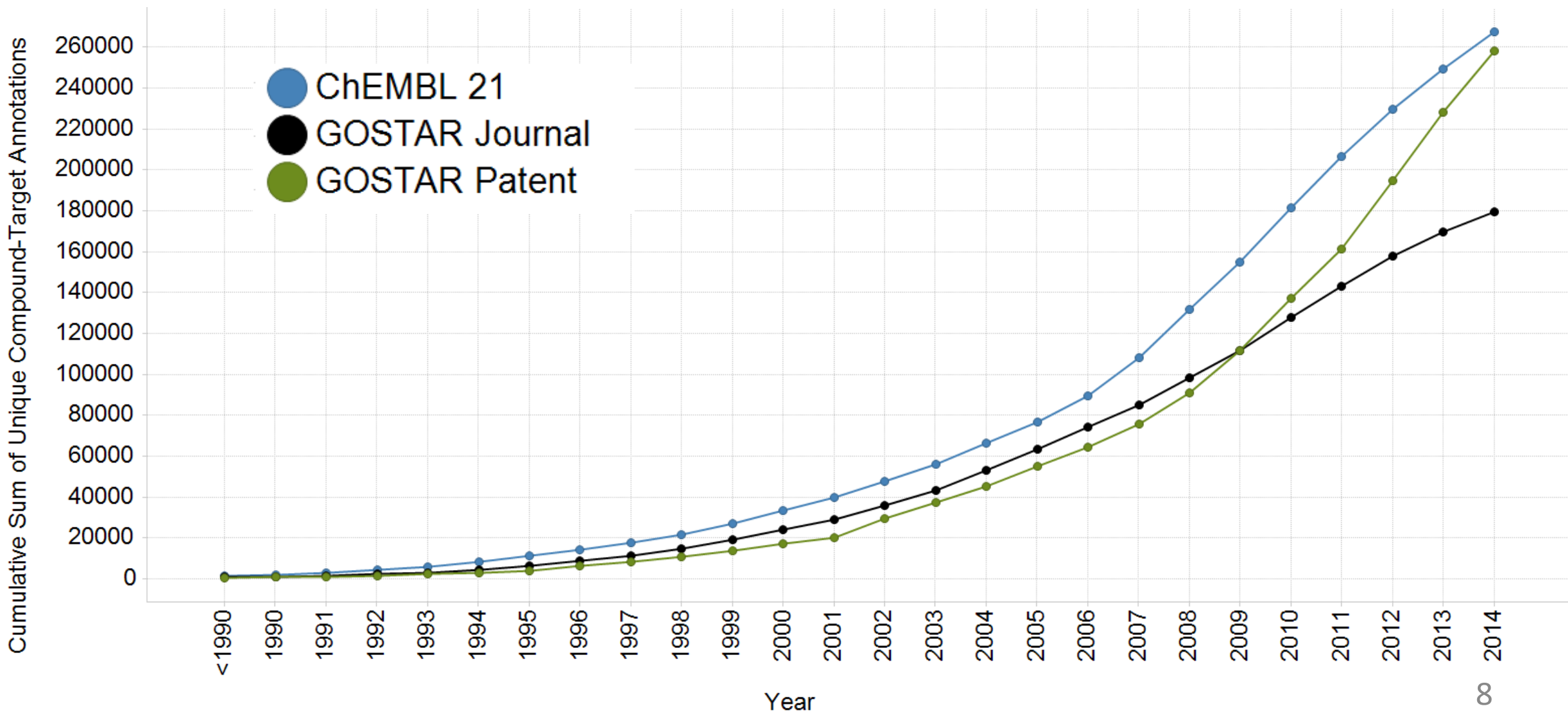
Software:

KNIME (with extensions and community contributions), R  
and RStudio, SPOTFIRE and COMMAND LINE SCRIPTING

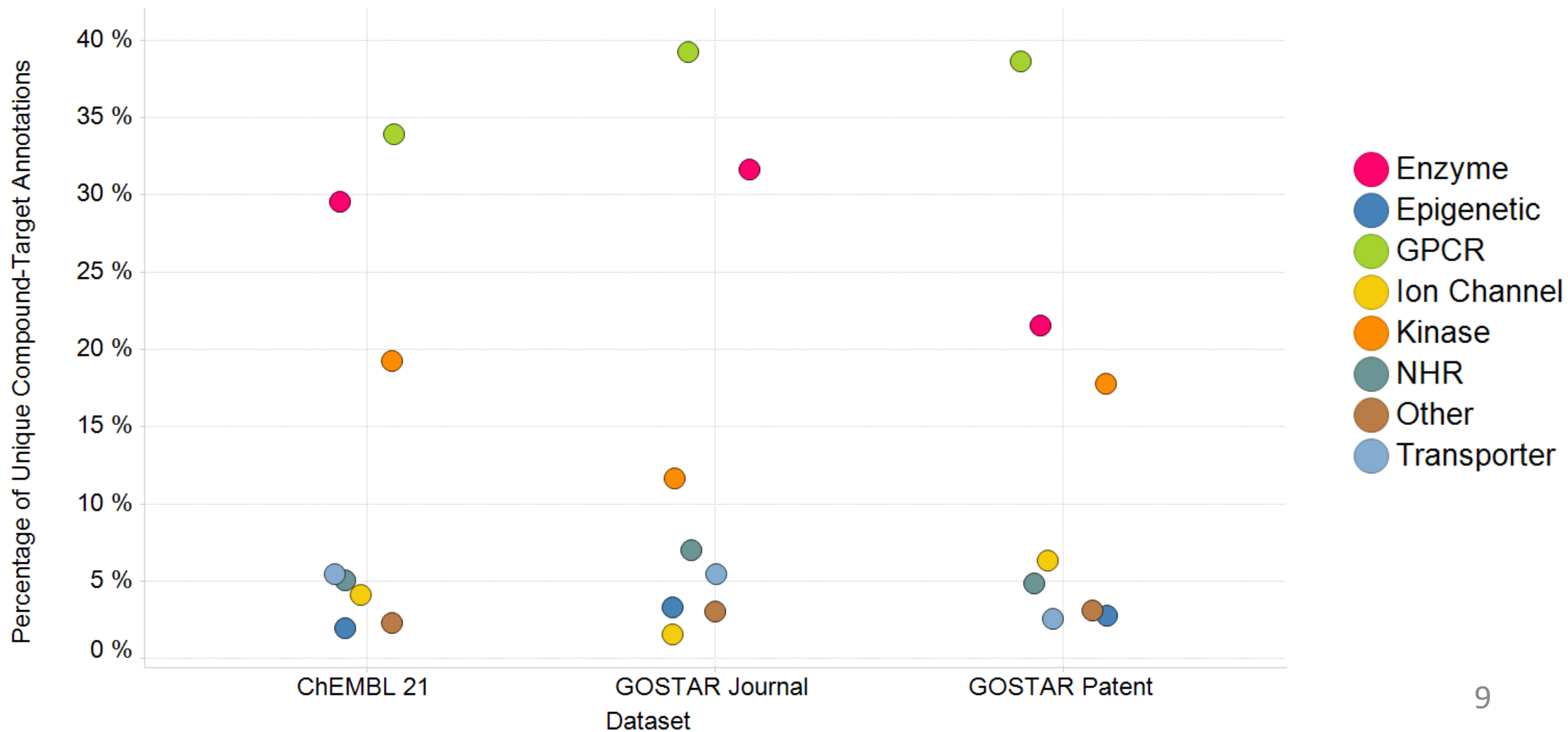
# Analysing the datasets as a function of time and target class

- Understanding the distribution of the data
- Differences between patented and public sources in terms of;
  - Target class
  - Time

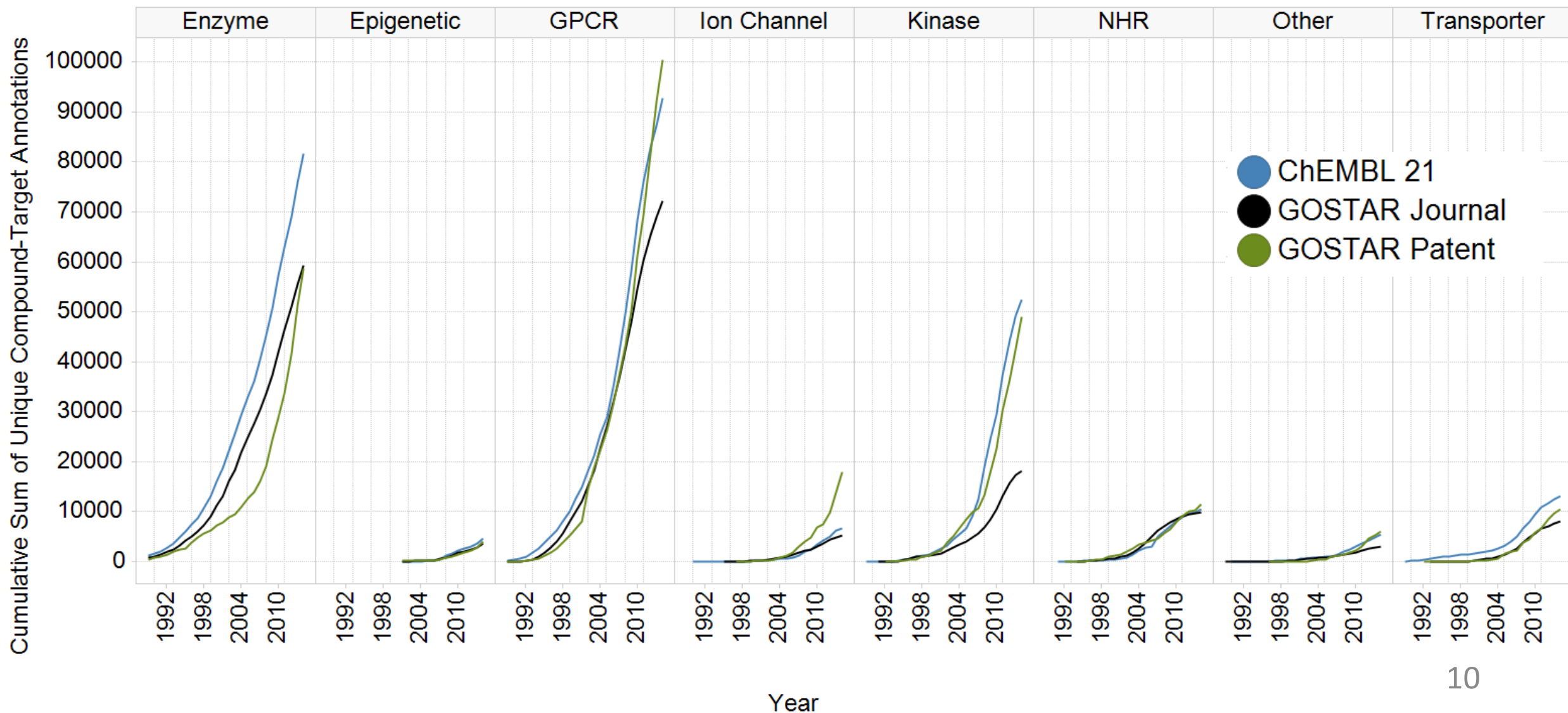
# The number of unique-compound target annotations increases over time



The highest percentage of compounds in each dataset are associated with GPCR, enzymes and kinases. Ion channels are represented less in GOSTAR Patents



Increases in unique compound-target annotations for a given target class in patents follows the evolution observed in scientific journals



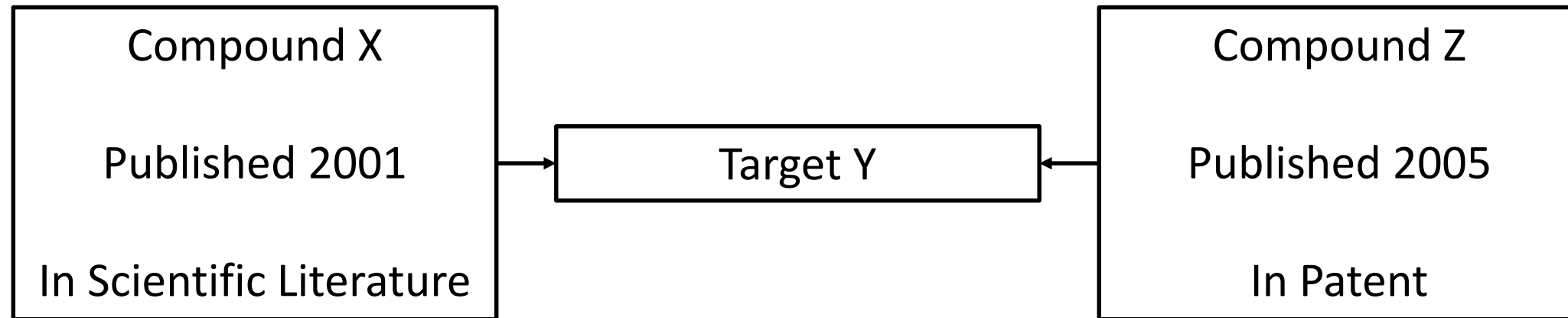
# In Summary – Section 1

- Compound-target annotations have increased over the years for all datasets
- The gap between patenting and publishing in scientific literature has decreased
- Target classes of interest vary between academia and industry resulting in differences in target classes of interest
- Those patenting tend to follow interests that were or are being published in literature - as seen in the number of compounds published each year by target class

# Analysing the publication of novel compound-target annotations

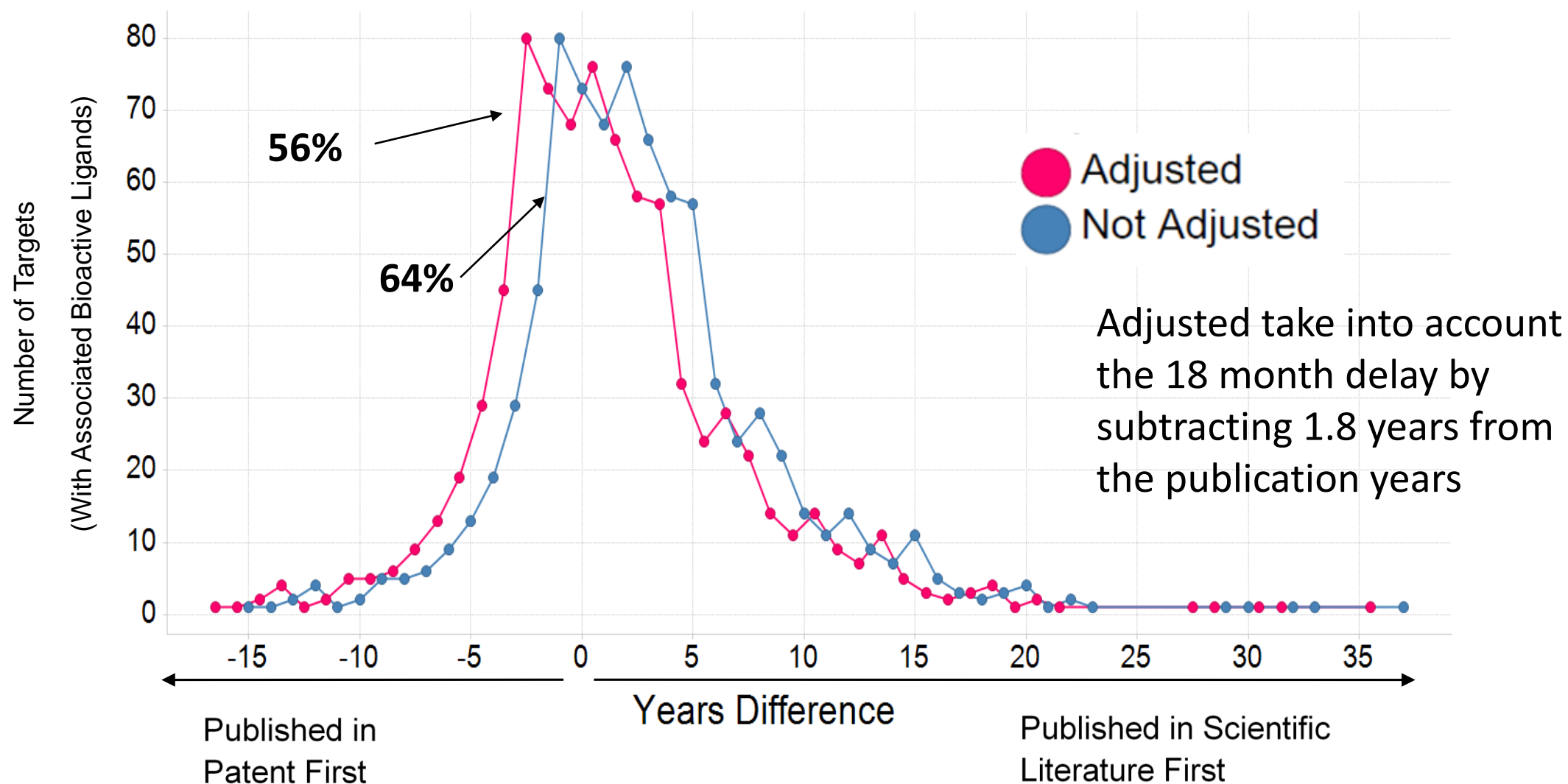
- Where are novel compound-target annotations published
- Are the results affected by the filtering methods applied?
- Does the target class / year affect when and where something is published
- Example of targets that are published first in scientific literature or a patent

# Comparing the first published compounds (associated with a particular target) from patents and scientific literature

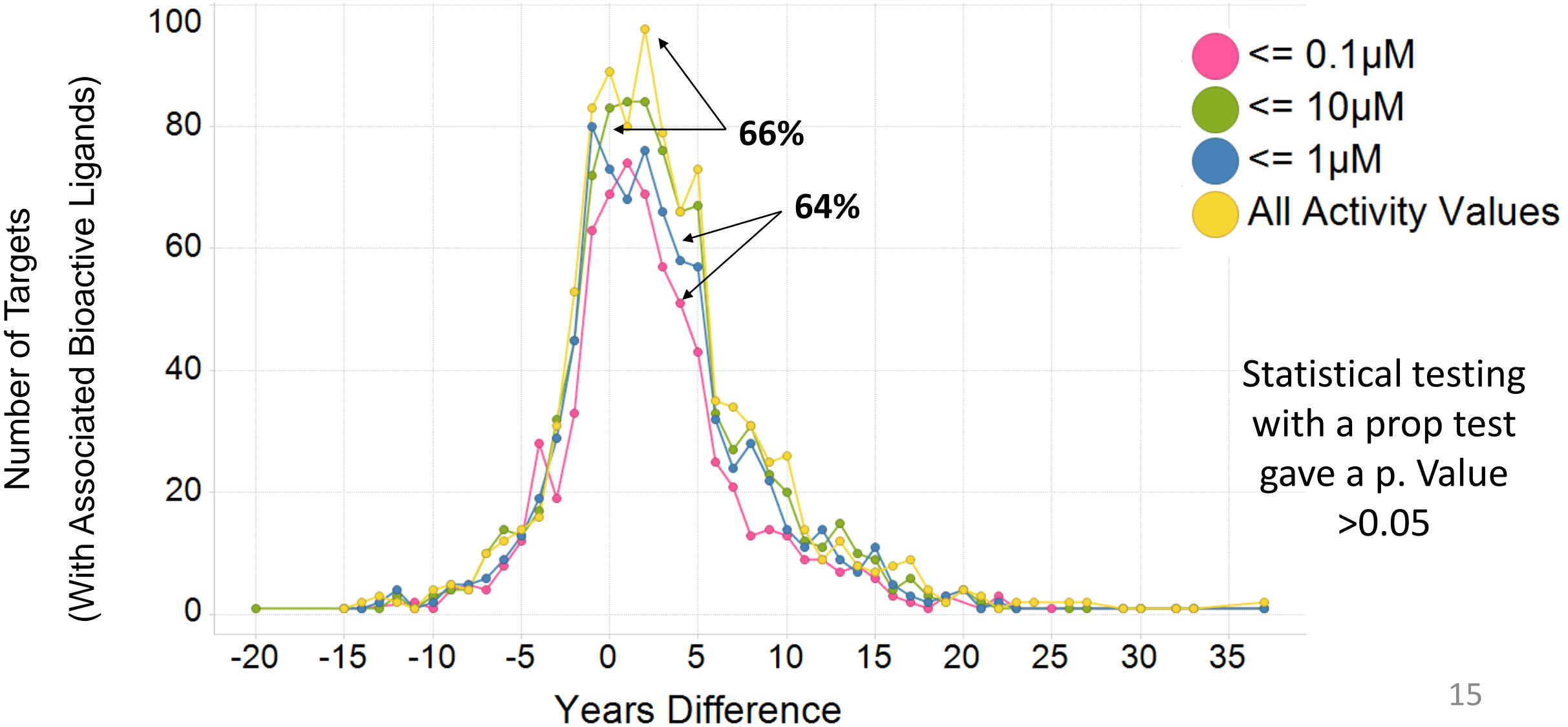


$$\text{Patent Year} - \text{Scientific Literature Year} = \text{Year Difference}$$

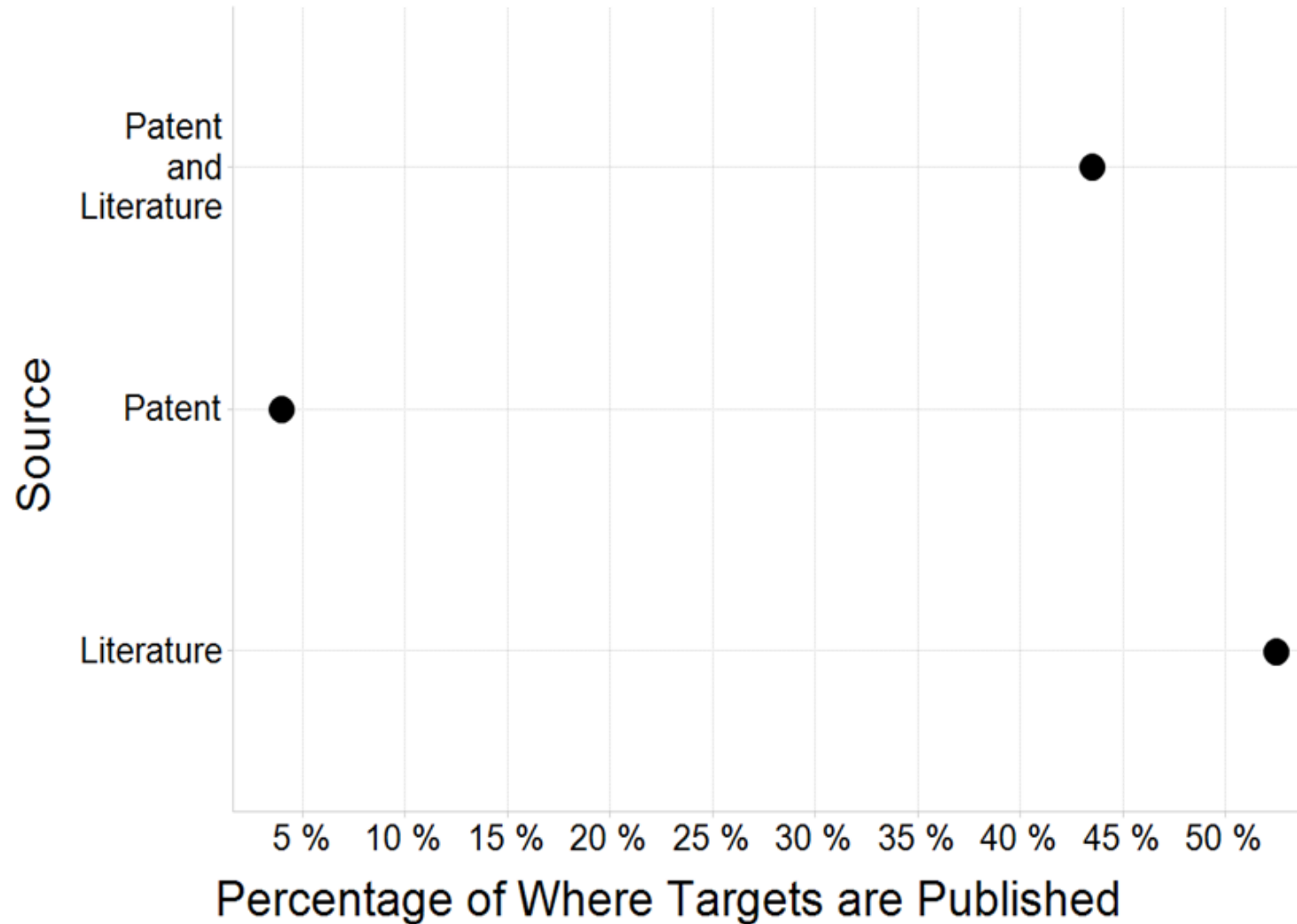
# Compound-target annotations tend to be published in literature first regardless of the 18 month delay in patenting



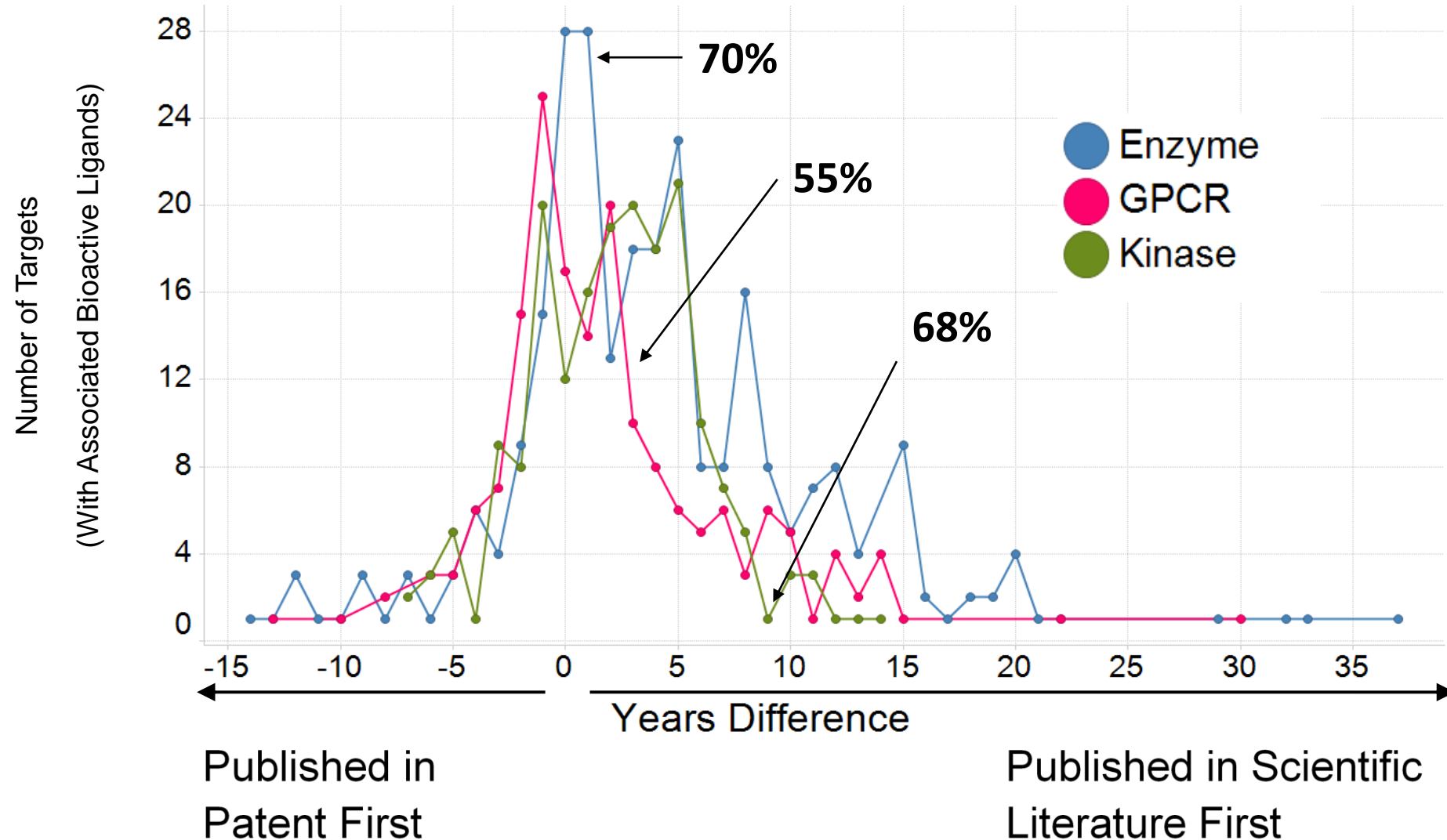
# Activity filtering does not affect the result



Targets (and their first associated ligand) tend to be published in either literature only or in both patent and literature rather than patents only



# The target class affects when and where compound-target annotations are published



# The year bin, as well as the target class, affects when and where compound-target annotations are first published



# Examples of targets that have been analysed and their first associated ligand

	Target		
	<b>BACE1</b>	<b>GSK3b</b>	<b>LRRK2</b>
<b>Target Class</b>	Enzyme	Kinase	Kinase
<b>First published location associated with a compound</b>	Literature	Literature	Patent
<b>1st Source</b>	Journal of the American Chemical Society	Bioorganic & Medicinal Chemistry Letters	WO 2010/085799 A2
<b>2nd Source</b>	WO 01/29563 A1	WO 03/076398 A2	Nature Biotechnology
<b>1st Source Year</b>	2000	2002	2010
<b>2nd Source Year</b>	2001	2003	2011

# In Summary – Section 2

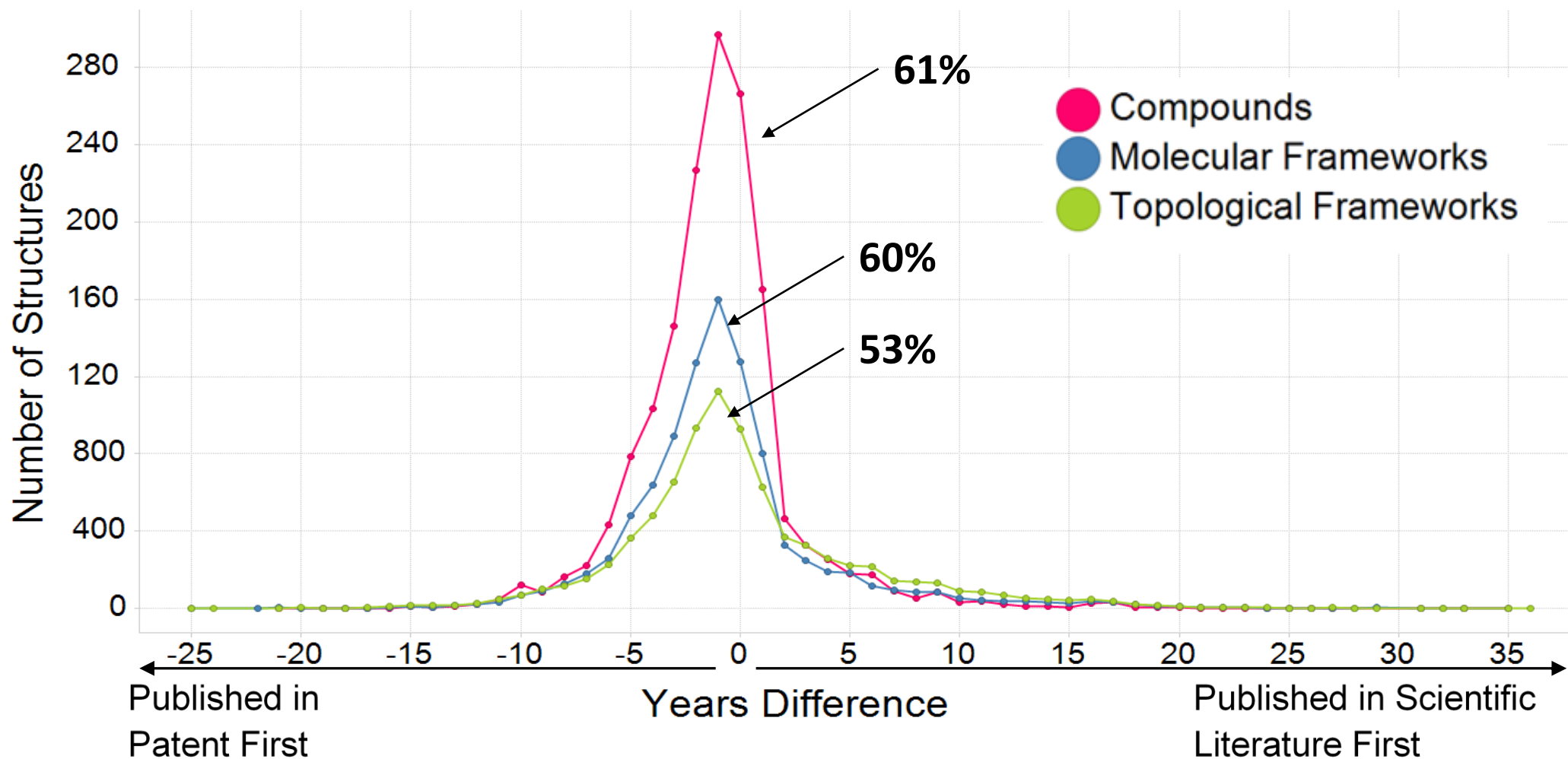
- Compound-target annotations tend to be published in literature prior to being published in a patent
- Activity Filtering does not significantly effect the result
- Compound-target annotations are published in either patents / literature or literature only
- The target class significantly affects when and where compound-target annotations are published
- Some target classes (such as GPCRs) tend to have their compounds published in patents first
- The number of novel targets increases for some target classes over years but decreases for others

# Analysing the publication of novel structures

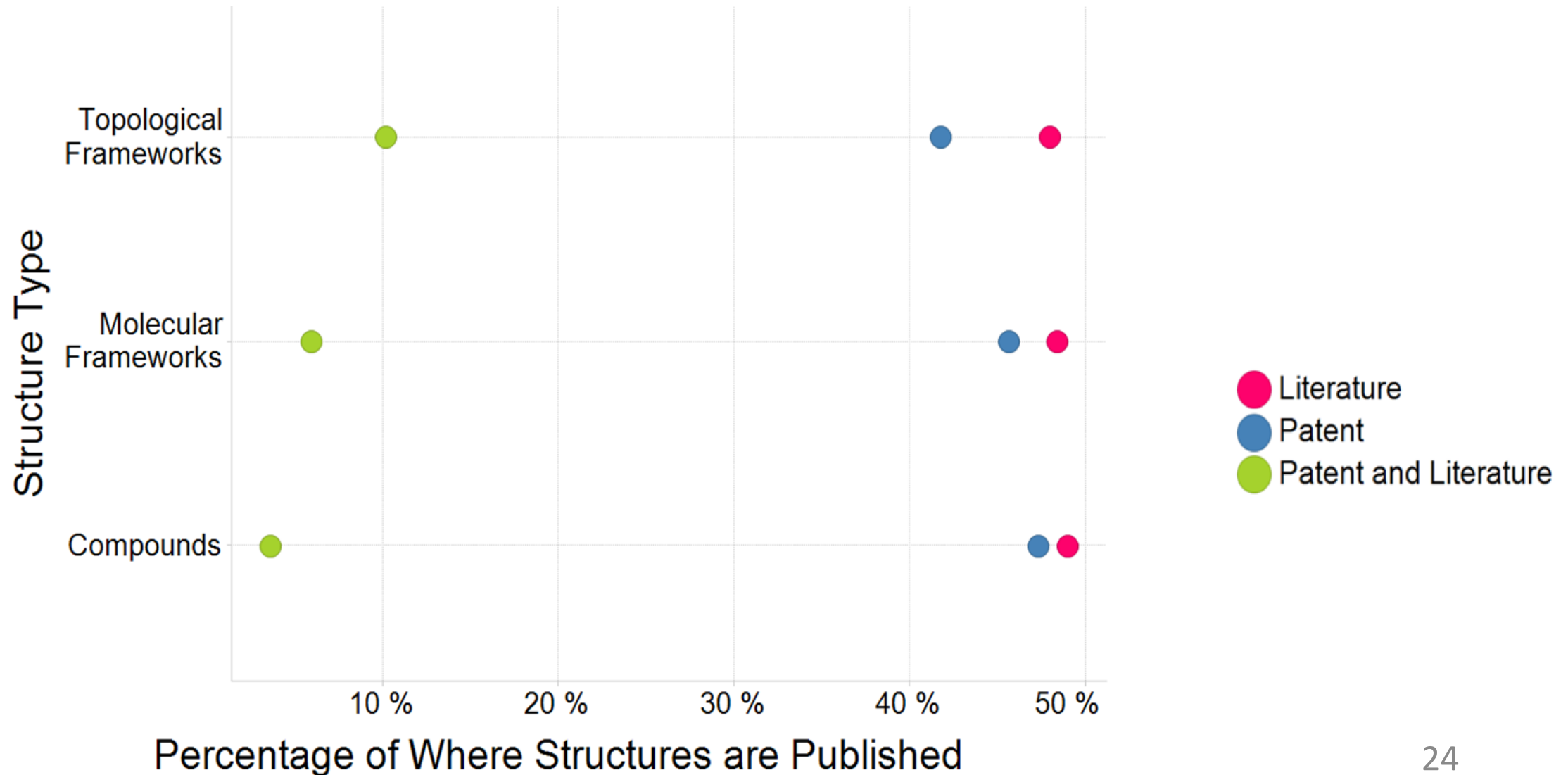
- Where are novel compounds/ molecular and topological frameworks first published
- Quantifying the scaffold novelty?



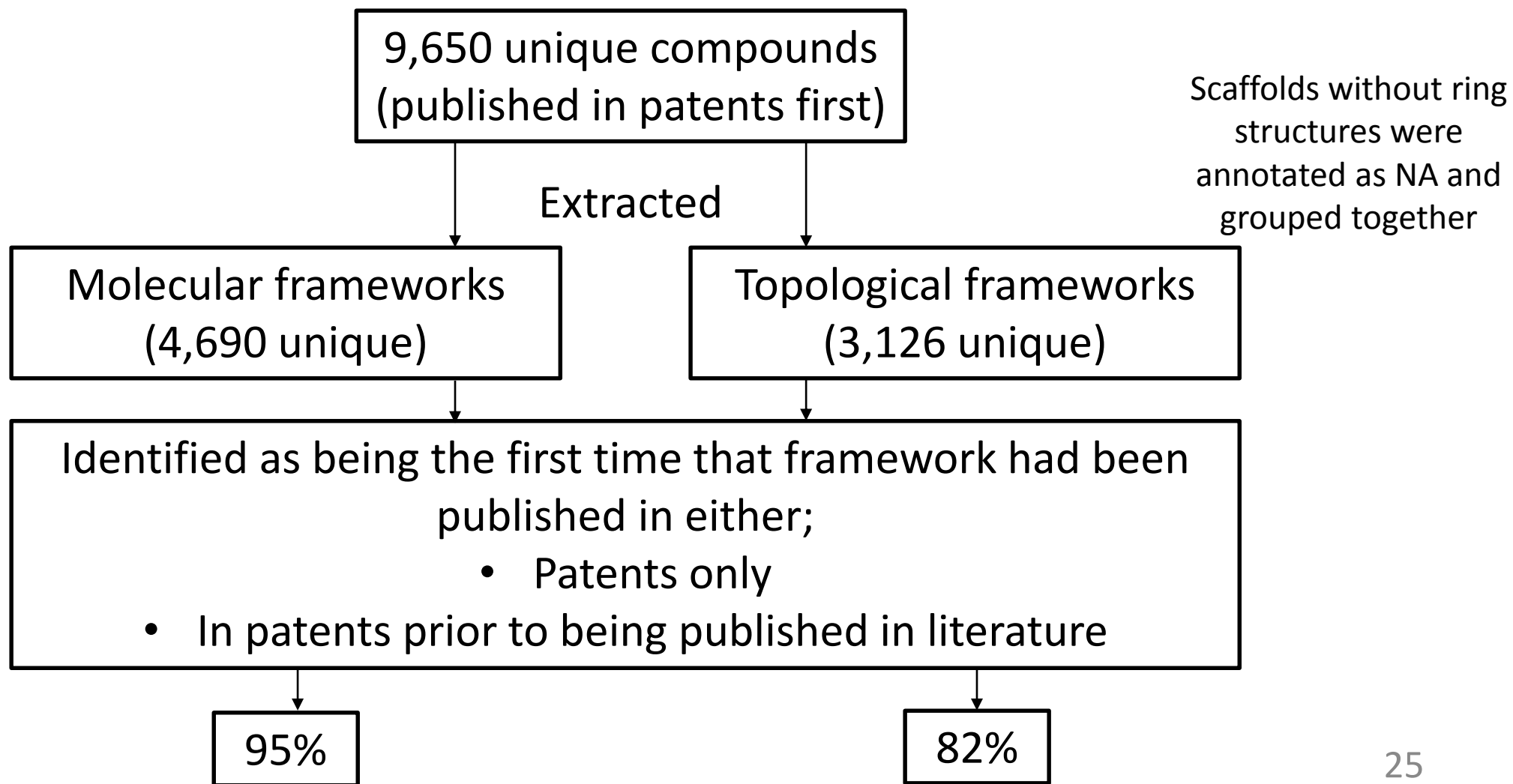
# Structures tend to be published in patents prior to being published in literature



Structures tend to be published in literature only or a patent only rather than in both patents and literature



# Compounds that are patented first usually represent novel scaffolds



# In Summary – Section 3

- Structures tend to be published in patents prior to being published in literature
- However, there is a statistically significant difference between the proportion of structures published in patents first
  - Suggesting novel chemistry
- Structures tend to be published in either patents only or literature only
- The chemistry of patented compounds is novel

# Conclusion

- The number of published novel compound-target annotations has grown cumulatively over the years
  - Varies by dataset and target class
- Compound-target annotations are primarily published in literature
  - tend to be published in both patents and scientific literature or only in scientific literature
- Target class and the year bin affects when and where a compound-target annotation is first published
- Novel compounds and their frameworks tend to be published in patents first
  - They tend to be published solely in scientific literature or patents

# Acknowledgements

- Andreas Bender
- Thierry Kogej
- Ola Engkvist
- Krishna Bulusu
- Kathryn Giblin
- Lewis Mervin
- Mike Firth
- Ian Barrett
- Rosa Buonfiglio
- Peter Varkonyi
- Michael Kossenjans
- Emma Rivers
- Nidhal Selmi
- Gary Pairaudeau
- Nick Tomkinson
- Noel O’Boyle
- Bender Group
- Discovery Sciences,  
Quantitative Biology  
and External  
Sciences

