[Artificial Intelligence in the Life Sciences 3 (2023) 100069](https://doi.org/10.1016/j.ailsci.2023.100069)

Contents lists available at [ScienceDirect](http://www.ScienceDirect.com/)

Artificial Intelligence in the Life Sciences

journal homepage: [www.elsevier.com/locate/ailsci](http://www.elsevier.com/locate/ailsci)

Pharmaceutical patent landscaping: A novel approach to understand patents from the drug discovery perspective

Yojana Gadiya[a](#_bookmark0),[b](#_bookmark1),[c](#_bookmark2),[∗](#_bookmark4), Philip Gribbon [a](#_bookmark0),[b](#_bookmark1), Martin Hofmann-Apitius[c](#_bookmark2),[d](#_bookmark3), Andrea Zaliani [a](#_bookmark0),[b](#_bookmark1)

a *Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP), Schnackenburgallee 114, Hamburg 22525, Germany*

b *Fraunhofer Cluster of Excellence for Immune-Mediated Diseases (CIMD), Theodor Stern Kai 7, Frankfurt 60590, Germany*

c *Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn 53113, Germany*

d *Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, Sankt Augustin 53757, Germany*

a r t i c l e i n f o a b s t r a c t

*Keywords:* Patent analysis Drug discovery

Patent landscape Rare disease Alzheimer’s disease Drug repurposing Target importance

Patents play a crucial role in the drug discovery process by providing legal protection for discoveries and in- centivising investments in research and development. By identifying patterns within patent data resources, re- searchers can gain insight into the market trends and priorities of the pharmaceutical and biotechnology indus- tries, as well as provide additional perspectives on more fundamental aspects such as the emergence of potential new drug targets. In this paper, we used the patent enrichment tool, PEMT, to extract, integrate, and analyse patent literature for rare diseases (RD) and Alzheimer’s disease (AD). This is followed by a systematic review of the underlying patent landscape to decipher trends and applications in patents for these diseases. To do so, we discuss prominent organisations involved in drug discovery research in AD and RD. This allows us to gain an understanding of the importance of AD and RD from specific organisational (pharmaceutical or university) perspectives. Next, we analyse the historical focus of patents in relation to individual therapeutic targets and correlate them with market scenarios allowing the identification of prominent targets for a disease. Lastly, we identified drug repurposing activities within the two diseases with the help of patents. This resulted in identify- ing existing repurposed drugs and novel potential therapeutic approaches applicable to the indication areas. The study demonstrates the expanded applicability of patent documents from legal to drug discovery, design, and research, thus, providing a valuable resource for future drug discovery efforts. Moreover, this study is an attempt towards understanding the importance of data underlying patent documents and raising the need for preparing the data for machine learning-based applications.

# Introduction

Patent documents are considered crucial assets within the drug dis- covery domain as they allow the inventor rights over an invention for typically 20 years post filing. In the biomedical domain, these inven- tions could include information on drug formulation, dosage or eﬃcacy, as well as information on the medicinal chemistry properties of leads or pre-clinical candidates. Moreover, the legal value of intellectual prop- erty in patent documents, and their distinctive function compared to scientific literature, makes patents crucial milestones in drug discovery and development [[1]](#_bookmark18). Despite this, patent documents have been rela- tively untapped as a source to assist in scientific discovery and are used in the latter parts of drug discovery to assist investors in drafting new patents for their innovation or extension of existing patents in case of “me-too” drugs [[2]](#_bookmark19). Also, the use of legal phraseology within patent doc- uments requires experts in both patent drafting and patent analysis to

evaluate and understand the underlying content. As a result, patent liter- ature mining has developed into a specialised field that is closely linked with chemoinformatic analysis and commercial evaluation, meaning the appearance of patent sourced information is less common in publicly available scientific data resources. In the current context, we use the word “patent” to indicate both patent applications and granted patents. Recently, there has been an increase in the attention given to patent documents as an aid to monitor advancements in drug dis- covery. This is seen in the field of oncology drug discovery, where patenting activity is intense, and patents offer a window into the latest techniques in translational cancer therapies [[3–5]](#_bookmark20). To support these efforts, oncology-specific patent datasets have been catalogued by established patent oﬃces such as the USPTO Cancer Moonshot Patent [Data (https://www.uspto.gov/ip-policy/economic-research/research- datasets/cancer-moonshot-patent-data). Moreover, there have been ef-](https://www.uspto.gov/ip-policy/economic-research/research-datasets/cancer-moonshot-patent-data) forts in mining and ingesting patent-related information into knowledge

∗ Corresponding authors at: Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP), Schnackenburgallee 114, Hamburg 22525, Germany.

*E-mail address:* [Yojana.Gadiya@itmp.fraunhofer.de](mailto:Yojana.Gadiya@itmp.fraunhofer.de) (Y. Gadiya).

<https://doi.org/10.1016/j.ailsci.2023.100069>

Received 10 February 2023; Received in revised form 29 March 2023; Accepted 29 March 2023

Available online 31 March 2023

2667-3185/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

graphs. One such example is the Chinese patent medicine (CPM) knowl- edge graph, where a natural language model (BERT) was used to extract biological entities (specifically chemicals, diseases, and conditions) from patent documents [[6]](#_bookmark21). The resultant graph assisted in the identification of disease gaps as well as summarising drug prescription trends in China. In addition to the extraction of patent-related information, the importance of exploring patent literature has attracted new market interest from scientific information vendors. The value of an exhaus- tive patent exploration process is significant for the pharmaceutical industry, resulting in the establishment of commercial and open-source patent databases besides institutional oﬃces such as the United States Patent and Trademark Oﬃce (USPTO), European Patent Oﬃce (EPO), and Japan Patent Oﬃce (JPO). The dominant commercial providers [in this market are Clarivate (https://clarivate.com/products/ip- intelligence/ip-data-and-apis/derwent-world-patents-index/) and](https://clarivate.com/products/ip-intelligence/ip-data-and-apis/derwent-world-patents-index/) [CAS-Scifinder (https://www.cas.org/solutions/cas-scifinder-discovery- platform/cas-scifinder). Simultaneously, public chemical databases](https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder) have also incorporated patent information for their catalogued chemi- cal collections. PubChem [[7]](#_bookmark22) has historically incorporated compound links to USPTO, while the European counterpart, ChEMBL, has de- veloped a stand-alone patent database (SureChEMBL) [[8]](#_bookmark23) based upon annotated biomedical entities which allow for searching a larger chemical space to inform drug discovery. Within the drug develop- ment domain, SureChEMBL represents one of the largest open-access resources of patents with its ability to extract data from independent patent oﬃces. Such databases have enabled researchers to systemat- ically assess, identify, and explore patterns within patent documents

[[9](#_bookmark24),[10](#_bookmark25)].

Patent documents can serve as a vast data source for machine learn- ing (ML) and artificial intelligence (AI) applications [[11]](#_bookmark26). ML-based al- gorithms can be used to predict the success rate of patented compounds by incorporating clinical trial results. This could be also beneficial for patent applicants in making R&D strategic decisions related to competi- tors. However, the unstructured nature of patent documents and their actual digital media, which are often in PDF format when downloaded from public repositories, also poses a challenge and as a result, manual curation is required to extract valuable information, such as bioactivity data, from these documents [[12]](#_bookmark27). Therefore, there is an increasing need for tools for patent automatic analysis utilising language models (LM) to identify and extract relevant metadata for exhaustive patent data search. In 2022, we introduced PEMT, a patent enrichment tool that al- lows extraction of pharmaceutically relevant entities such as targets and chemicals of interest from patent documents [[13]](#_bookmark28). This tool annotates modulators with patent information rather than providing an overview of the usability of the patents in drug discovery. This manuscript aims to guide readers through the tool’s implementation and, in doing so, highlight the role patent literature can play in driving decisions in the drug discovery process. We utilised knowledge graphs (KGs), which are graphical databases composed of aggregated literature and experimen- tal data, as the platform for demonstrating the utility of our analyses. For rare diseases, we utilised OrphaNet [[14]](#_bookmark29), while for Alzheimer’s disease, we used the Human Brain Pharmacome [[15]](#_bookmark31). By means of these KGs, we extracted data from the past two decades of chemical patent space and performed patent landscaping, which involved the identification of patterns within patent documents. We start by identifying the commer- cial and non-commercial pharmaceutical originators based on the patent activity of their aﬃliated owners. Next, we performed a retrospective overview of targeted proteins from patent documents, to identify their importance in the drug development process at specific periods. Addi- tionally, we examined the impact clinical trials have had on the corre- sponding target prioritisation. Lastly, we examined the drug repurpos- ing aspect of pre-clinical small molecules and clinical stage drugs, high- lighting potential novel disease treatment options. Overall, this patent landscaping approach highlights the importance of analysing patents and their usefulness in making decisions during drug development

efforts.

# Methods

In the initial sections, we discuss the KGs employed as data sources, explain the functioning of the PEMT tool, and outline the curation pro- cess necessary for standardising information from patent documents. Fi- nally, we provide information on the implementation details pertinent to the analysis.

* 1. *Knowledge graph selection*

We extracted data from two indication-specific semantically organ- ised KGs: OrphaNet and Human Brain Pharamcome (HBP). OrphaNet (accessed on 2021–11–01) is an open-source network that focuses on rare diseases and includes information on biological entities such as or- phan gene targets and drugs, as well as clinical entities such as clinical trials and biobanks [[14]](#_bookmark29). In contrast, the Human Brain Pharmacome (HBP) is a publicly available biomedical knowledge graph representing neurodegenerative diseases with a particular focus on Alzheimer’s dis- ease [[15]](#_bookmark31). The data in the graph is built on the Biological Expression Language (BEL) framework and includes information on proteins, bio- logical processes, and other relevant features [[16]](#_bookmark32).

* 1. *Experimental and patent data extraction*

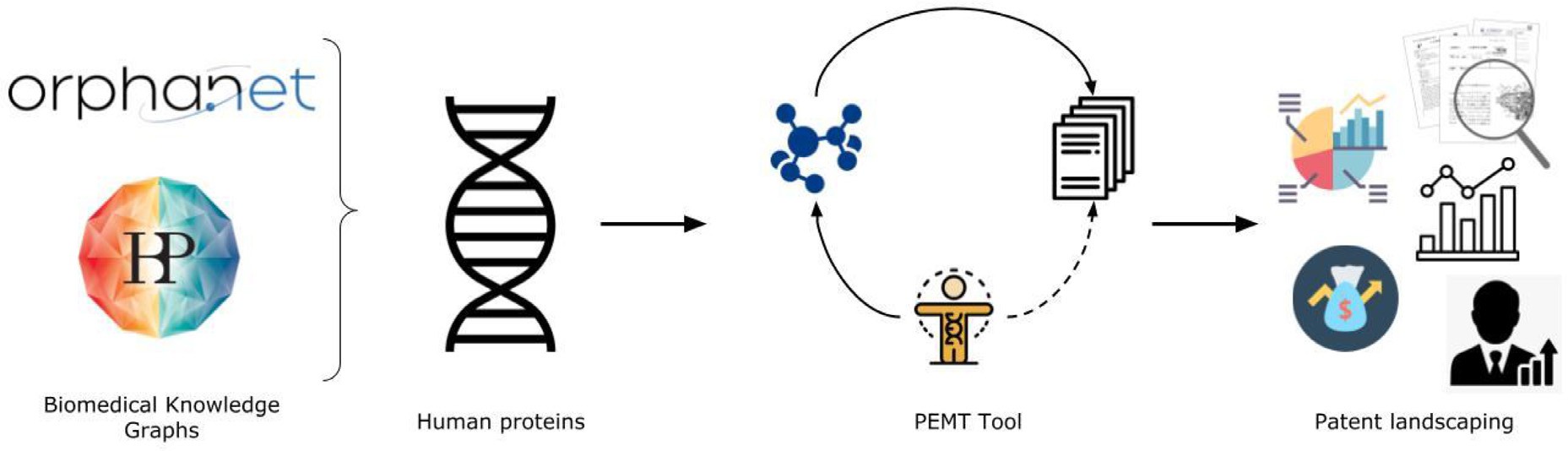
To expand the data resource with patent data, we used the PEMT tool [[13]](#_bookmark28), which systematically retrieves patent documents using a list of genes or proteins. To do so, it extracts manually annotated and veri- fied chemical and biological modulators that have binding or functional effects on proteins from ChEMBL and subsequently captures all patent documents related to the modulators using SureChEMBL. The patent documents collected are then filtered based on their pharmaceutical rel- evance and status (i.e. whether the patents are active or expired). Thus, the tool removed patent documents older than 20 years and only in- cluded those documents that were tagged with specific IPC code classes referenced in the publication [[13]](#_bookmark28). These codes are alphanumeric codes assigned by oﬃcials to patent documents and help in understanding the scope of the patent.

* 1. *Patent data harmonisation pipeline*

To aid systematic analysis, we developed data harmonisation and normalisation pipelines. In this pipeline, patent owner names were clas- sified into three categories: “Organisation”, “Acquired”, and “Individ- uals”. Organisations include patent owners that were part of compa- nies and universities, while Acquired owners are those that were within legacy entities which have been acquired or merged into larger Organ- isations. An example would be Dupont Pharmaceuticals, which Bris- tol Myers Squibb (BMS) acquired. Individuals are individuals who own a patent. Additionally, similar organisations were grouped for consis- tency. For example, all variations of Bristol Myers Squibb were mapped to Bristol Myers Squibb. This normalisation step allows for identifying all relevant patents from the same owner and has not been performed previously.

* 1. *Compound clinical stage annotation*

To annotate compounds based on their clinical trial stage, we utilised the information captured within ChEMBL. For each compound regis- tered in ChEMBL, an additional annotation is made based on whether the compound has been a clinical candidate and if so, at which stage. As a result, ChEMBL assigns phase numbers 0 - 4 where 0 indicates a compound being researched in the preclinical stage, 1 to 3 indicate the compound being tested in clinical trial phases I - III, and 4 indicates the compound has been approved by FDA and/or EMA. In our analysis, we make use of this ChEMBL data to identify compounds, with associated patents, that have been moved from the preclinical to the clinical setup within the drug discovery pipeline.



**Fig. 1. Synopsis of the patenting landscape analysis**. We collect all human proteins from the two biological networks relevant to diseases and symptoms. Using these proteins as a starting point, we run our PEMT tool and retrieve all associated modulators and patent documents based on predefined conditions. Finally, using the patent document dataset, also referred to as patent corpora, we conduct deductive and exploratory analysis to uncover patterns within the dataset.

* 1. *Patent data clustering*

To cluster patent owners, we made use of a hierarchical approach where two or more owners are clustered together based on a distance metric from singletons to cluster groups in a bottom-to-top manner. This type of clustering allows creation of dendrograms that help es- tablish patent owner groups that belong to the same cluster. The dis- tance metric used in the clustering approach was the distance correlation which enabled in measuring the dependency between the two owners ([Eq. (1)](#_bookmark7)). The clustering was achieved using Seaborn ‘clustermap‘ pack- age (<https://seaborn.pydata.org/generated/seaborn.clustermap.html>).

(*𝑢* − *𝑢*′) *.* (*𝑣* − *𝑣*′)

the study. [Fig. 1](#_bookmark6) provides an overview of the process involved in col- lecting and analysing the patent dataset, referred to as the patent cor- pora, in the following sections. The data and the scripts used during this analysis can be found on GitHub at https://github.com/Fraunhofer- ITMP/Pharmaceutical-patent-landscaping.

# Results

We analysed the patent extracted from 4314 to 10,237 proteins from OrphaNet and HBP respectively. In this analysis, we define “rare dis- eases” as diseases mentioned within the Orphanet database. In the fol- lowing section, details four subsections through which we explain how

*𝑥* = 1 − ′ ′

(1)

patent literature can be utilised in drug discovery. First, we provide an

||(*𝑢* − *𝑢* )||⋅||(*𝑣* − *𝑣* ) ||

[Eq. (1)](#_bookmark7). **Distance correlation.** The distance correlation is a metric used to clusters that have minimal distance together. In this equation, *u’* is the mean of vector *u* and *x . y* is the dot product of *x* and *y*.

*2.6. Target based publication data collection*

To extract publications relevant to targets of interest, we made use of open source platforms within NCBI Gene Browser (<https://www.ncbi.nlm.nih.gov/gene/>) to gather information on the hu- man targets. Through the link between the gene platform and PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), we collected all relevant publica- tions related to a target. For our purposes, we clustered all the publi- cation found in a year together and used this statistics as the basis for understanding the research trend on targets. Moreover, we rescaled the publication counts to be between 0 and 1 using the min-max normaliser ([Eq. (2)](#_bookmark8)).

overview of the information collected from the patent corpora for each disease. Next, we retrospectively analyse the patent corpora unveiling leading organisations in drug research and discovery (R&D) for individ- ual diseases. Following this, we investigated case studies for selected tar- gets based on biological networks and examined their patenting trend. Lastly, we illustrate the usefulness of patent documents in identifying drug repurposing opportunities.

* 1. *Overview of the patent corpora*

In this section, we investigate the patent corpora of each disease and provide a statistical overview of the information collected within the patent corpora. This summary will provide insights into the scope of patenting activity within the disease area, denoting whether the patents are filed for pharmacological relevance or for formulation purposes. We will also examine and compare the number of patent documents that have been approved versus those still in the application pipeline.

We retrieved 17,506 compound-patent pairs from the dataset, which

*𝑥*′ = *𝑥* − *𝑚𝑖𝑛*(*𝑥*)

*𝑚𝑎𝑥*(*𝑥*) − *𝑚𝑖𝑛*(*𝑥*)

(2)

included 585 compounds and 502 unique patent documents with the rare disease patent corpora. Out of the patents retrieved, 180 were

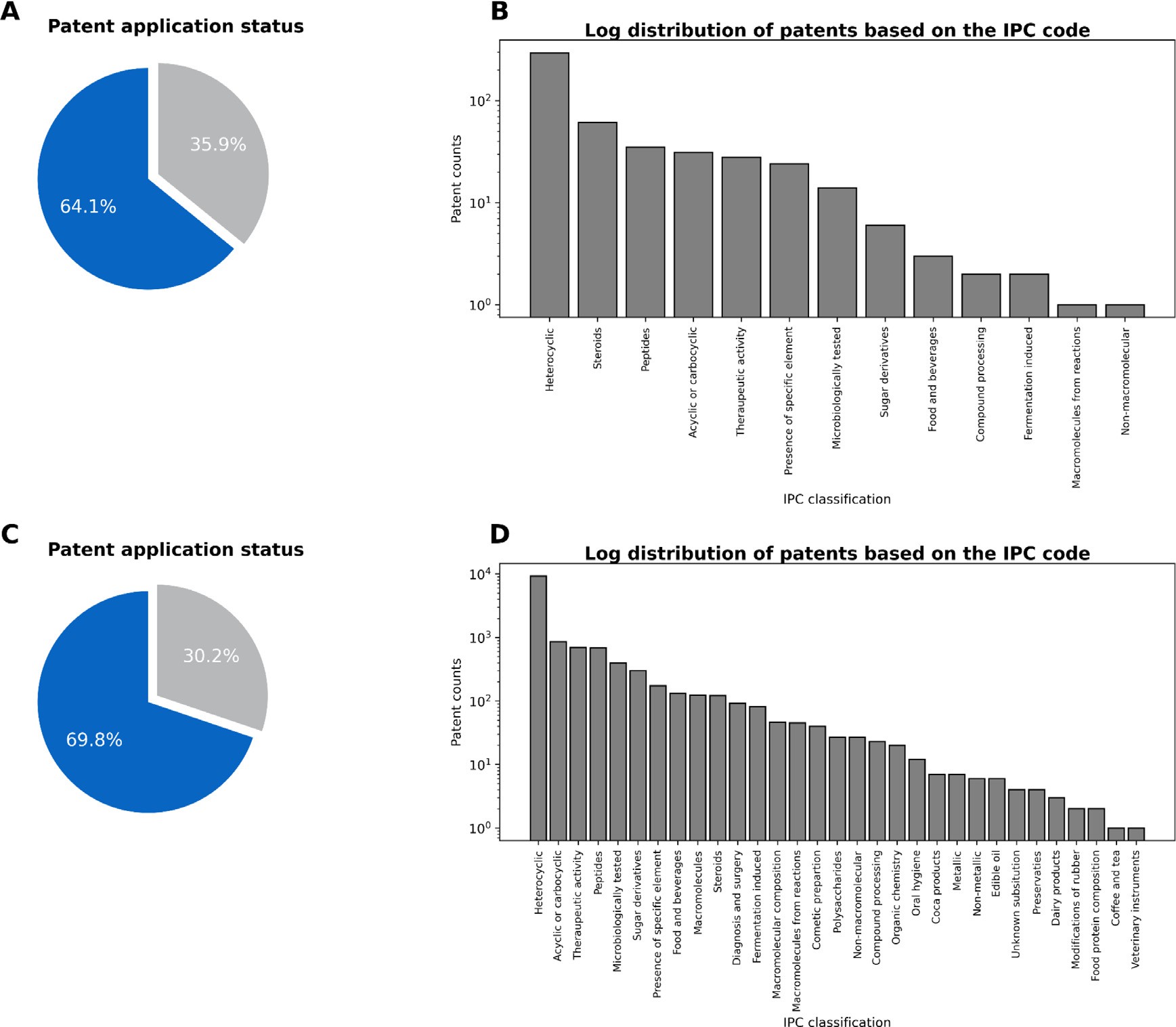
[Eq. (2)](#_bookmark8). **Min-max normalizer.** The min-max normalise allows us to scale the values to a define range. In this equation, *x* is an original value and *x’* is the normalized value.

*2.7. Data availability and implementation details*

To run our PEMT tool, we extracted human proteins that mod- ulate a disease or symptom from both graphs. The PEMT tool was run on a Windows operating system and was followed by the har- monisation pipeline mentioned in [Section 2.3](#_bookmark5). Additionally, within the collected patent documents, we found 131 patents in rare diseases and 705 patents in Alzheimer’s diseases with no associated owner. This is likely due to the quarterly updation of public databases like SureChEMBL, causing a lag in the time for updating. As a result, ad- ditional filtering was done to exclude patents with no owners from

granted patents, represented by documents with kind codes starting with B or E, while the rest were still in the application process, represented by documents with kind codes starting with A ([Fig. 2](#_bookmark9)A). Additionally, based on the IPC classification, it was revealed that more than half of the [patents were related to C07D (https://www.wipo.int/classifications/ ipc/en/ITsupport/Version20170101/transformations/ipc/20170101/ en/htm/C07D.htm#C07D), which includes heterocyclic compounds.](https://www.wipo.int/classifications/ipc/en/ITsupport/Version20170101/transformations/ipc/20170101/en/htm/C07D.htm\043C07D) Following this, 12% of patent documents were in the IPC code C07J (steroids), 7% were in code C07K (peptides), and 6% were in code A61P (specific therapeutic activity of chemical compounds or medicinal preparations) ([Fig. 2](#_bookmark9)B).

Conversely, in the case of the Alzheimer’s disease patent corpora, we retrieved 76,321 compound-patent pairs that included 22,930 com- pounds and 13,181 unique patent applications. From these patents, 3980 were granted (indicated by kind codes starting with either B or



**Fig. 2. Patenting Landscape overview for Rare and Alzheimer’s Diseases**. (**A)** Rare disease patent application ratio denoting whether the patents have been granted (in grey) or are still pending (in blue). (**B)** Logarithmic distribution of the rare disease patents based on the IPC classification category. (**C)** Alzheimer’s disease patent application ratio denoting whether the patents have been granted (in grey) or are still pending (in blue). **D)** Logarithmic distribution of the Alzheimer’s disease patents based on the IPC classification category.

E), while the remaining were still under the application process (in- dicated by kind codes starting with A) ([Fig. 2](#_bookmark9)C). Upon reviewing the IPC classification of the patents, it was found that 70% of the docu- ments belonged to class C07D, representing patent applications men- tioning heterocyclic compounds. Alongside this class, other notable IPCs include C07C (representing acyclic and carboxylic compounds) with 7% of patent documents, A61P (representing a specific therapeutic activity of chemical compounds or medicinal preparations) with 5% of patent applications, and C07K (representing peptides) with 5% of patent appli- cations ([Fig. 2](#_bookmark9)D).

Thus, RD and AD patent cohorts revealed two recurring patterns:

i) a lower number of granted patents and ii) a dominance of “C07D coded” patent applications. The observation that granted patents are approximately half of all applications filed is consistent with the statis-

[epo.org/en/statistics-centre#/technologyfields?code=16). This filed- tics within the pharmaceutical domain generated by EPO (https://new.](https://new.epo.org/en/statistics-centre\043/technologyfields?code=16)

to-granted patents ratio can be attributed to several reasons. One of the reasons is the diﬃculty and expenditure involved in the research and development around patents. With the limited number of individuals re- viewing the documents, the search for potential competitiveness within [pre-existing inventions is time-consuming (https://www.epo.org/ learning/materials/inventors-handbook/protection/patents.html). Ad-](https://www.epo.org/learning/materials/inventors-handbook/protection/patents.html) ditionally, the prevalence of C07D patents is likely due to the diver- sity provided by this class of compounds, which makes them a desirable

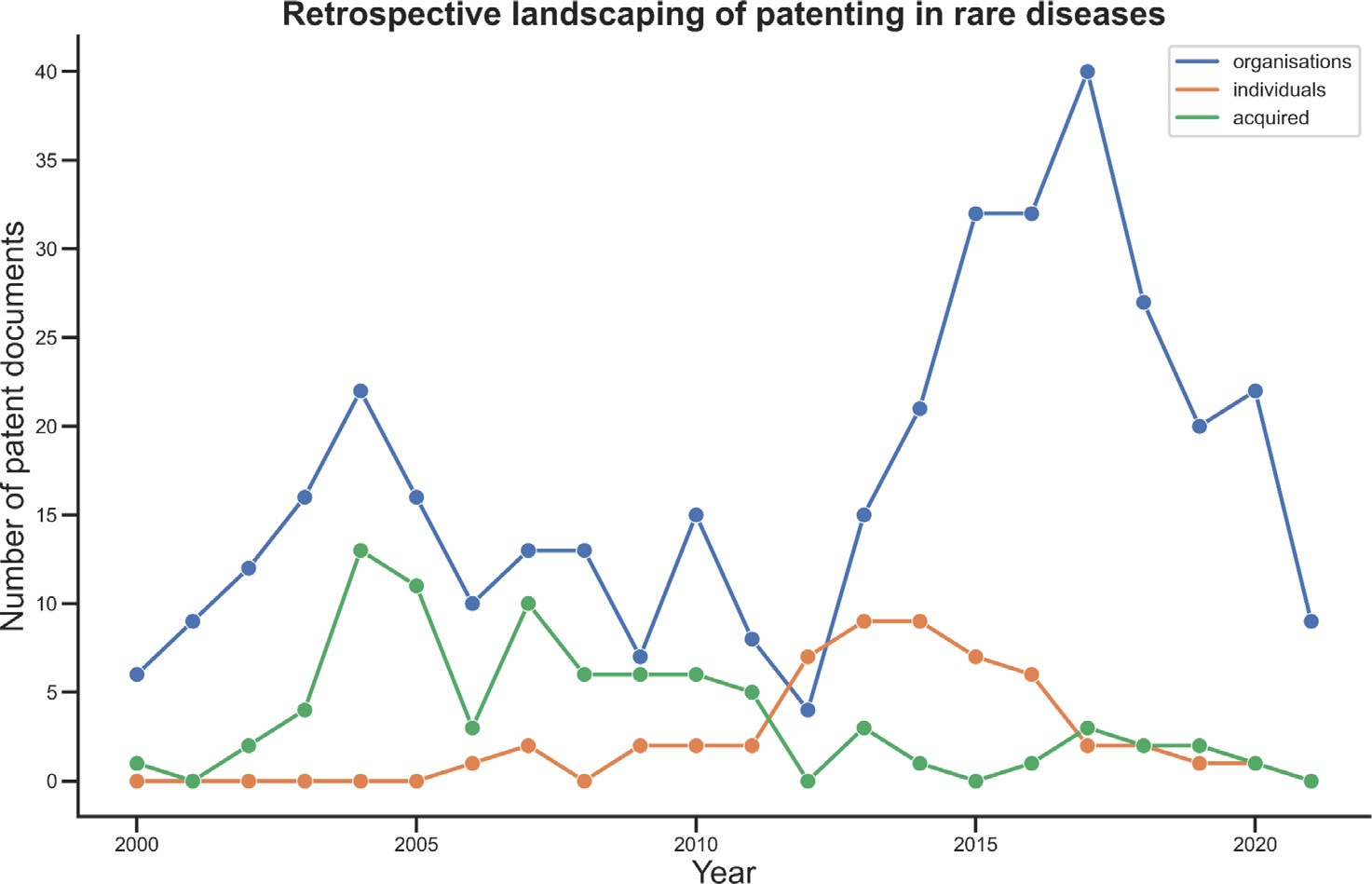
candidate for drug discovery patent innovation compared to other com- pound classes [[17]](#_bookmark35).

* 1. *Trends in the patenting landscape from the patent owners’ perspective*

Next, we analysed the patenting activity for each disease to deter- mine the major pharmaceutical companies and academic institutions that have made significant contributions to the R&D of potential drugs in the field. To do so, we leveraged the high-level patent owner classifi- cation (i.e., organisation, acquired, and individuals) and grouped patent owners based on the number of applications they have filed. Then, we reviewed the top patent owners and deduced their historic patenting activity in the field. We examined RD and AD separately to understand better the patent landscaping related to each disease.

* + 1. *Identifying key players in rare disease patenting over the past 20 years*

For RD’s, from 2000 to 2021, 369 organisations, 53 individuals, and 80 acquired organisations have played a role in the patent landscape. As illustrated in [Fig. 3](#_bookmark10), patenting activity has grown over the years, with or- ganisations being the most active, followed by individuals. The boost in the patenting of large organisations, mainly pharmaceutical companies, is ascribed to legislations like the Orphan Drug Act (ODA), introduced in 1983, which gave financial incentives for developing therapies for

**Fig. 3. The retrospective landscape of the rare diseases patent corpora in the past two decades**. The line plots indicate the growth of the patent documents years based on assignee type: organisations (blue), individuals (orange), and acquired (green). As expected, a large contri- bution is evident for organisations compared to others. A large boost in the Individual owners is seen around 2012, after which there has been a decline due to the legislatures.

rare diseases [[18]](#_bookmark37). On top of that, the surging patent activity within the field after 2012 was correlated to the establishment of the Patient Protection and Affordable Care Act (ACA) which catalysed engagement of investors in financing R&D for orphan drugs and diseases [[19]](#_bookmark38). On the contrary, a decrease in independent patent applications by individ- uals was observed since its peak in 2013. This effect could be due to two reasons: firstly, changes made in US and EU legislatures during the early 2000s, moving the “professor’s privileges” at IP generation to the universities, and secondly, professors and professionals, due to their leg- islation, were either hired by commercial companies or became part of larger academic groups [[20]](#_bookmark40).

To get a historical perspective of these patents, we selected 10 top patent owners and organised their patent applications based upon the timestamp of each patent application date, also known as the prior- ity date. This subset of owners consisted only of global pharmaceutical companies such as Pfizer, GlaxoSmithKline, etc. ([Fig. 4](#_bookmark11)). It is essential to note that many start-up and medium-sized companies have emerged and subsequently been absorbed by larger organisations providing them with the leverage of patent acquisition filed by the smaller companies. One example is Sterix Ltd., which had 33 patents and was acquired by Ipsen (in 2004), adding to the company’s patent portfolio. Furthermore, most top assignees had a decreasing patenting activity in the past years. For instance, global giant Takeda, a Japanese multinational pharmaceu- tical company, just had one compound, Luvadaxistat, out of 122 from the patent compound space between 2000 and 2021 that went into the clinical trial phase II for Friedreich’s ataxia [[21]](#_bookmark43). Thus, indicating a rea- son for decreasing interest of Takeda in rare diseases patenting ([Fig. 4](#_bookmark11)). Regardless, the top 10 assignees accounted for 44% of the 502 patent applications analysed, and the top 20 assignees accounted for 61%. Ad- ditionally, we observed the cumulative positive patenting trend for the top patent inventors in rare disease. The trend between 2013 and 2021, however, can be observed with a peak around 2016 and a gradual de- cline in the later years (**Supplementary Figure 1**).

Additionally, we looked into the correlation between the patent-

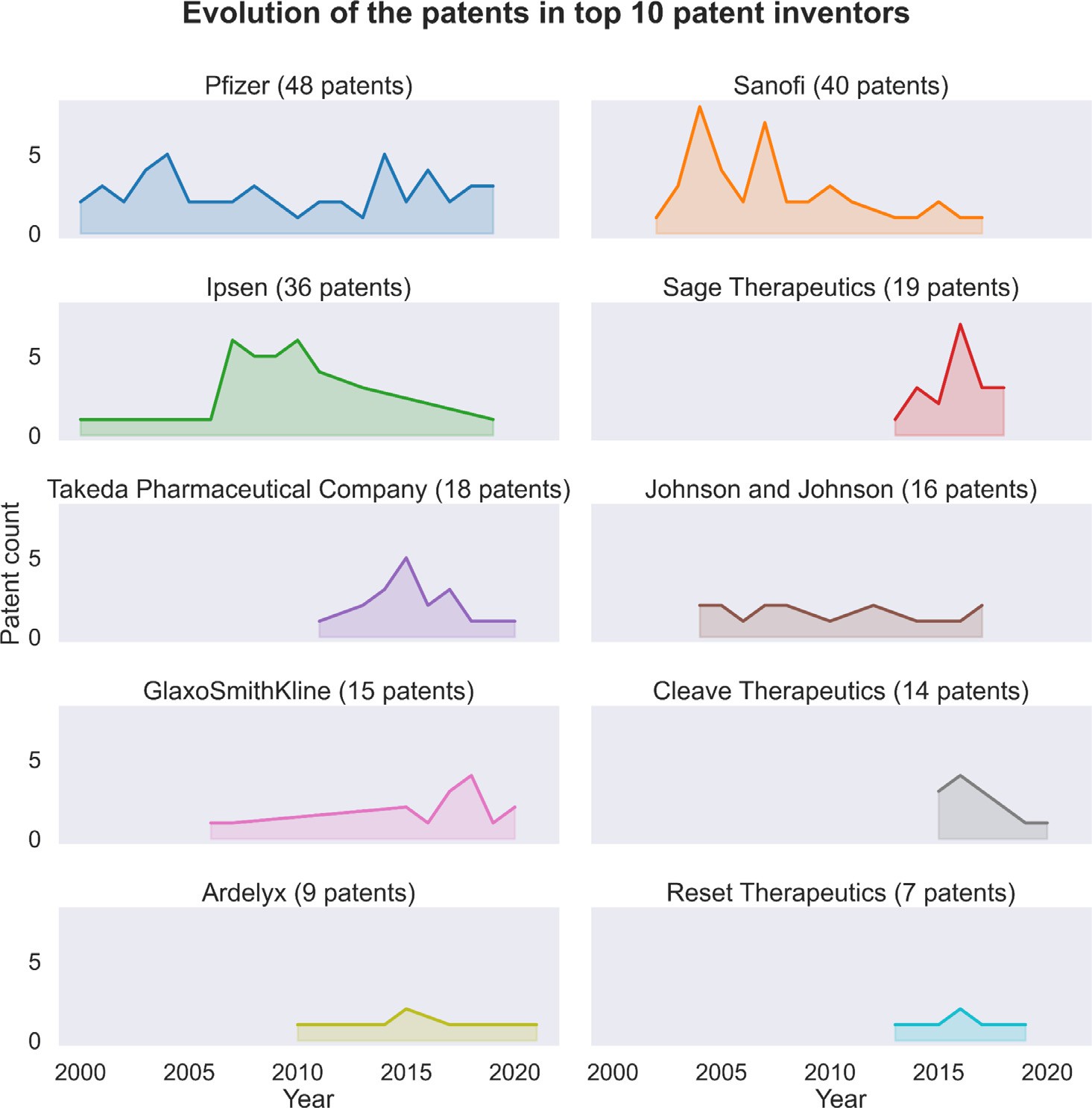
ing activity for the top owners based on the target portfolio to iden- tify pharmaceutical companies that show similar target portfolios. We identified two clusters when correlating the target-based portfo- lio of the top 10 owners in rare diseases ([Fig. 5](#_bookmark12)). One of the clus- ters was between Sanofi and Ardelyx. We suspect this was due to their collaboration on sodium and potassium channel inhibitors (which include KCNK3, KCNK9 and SLC9A3), allowing them to apply for

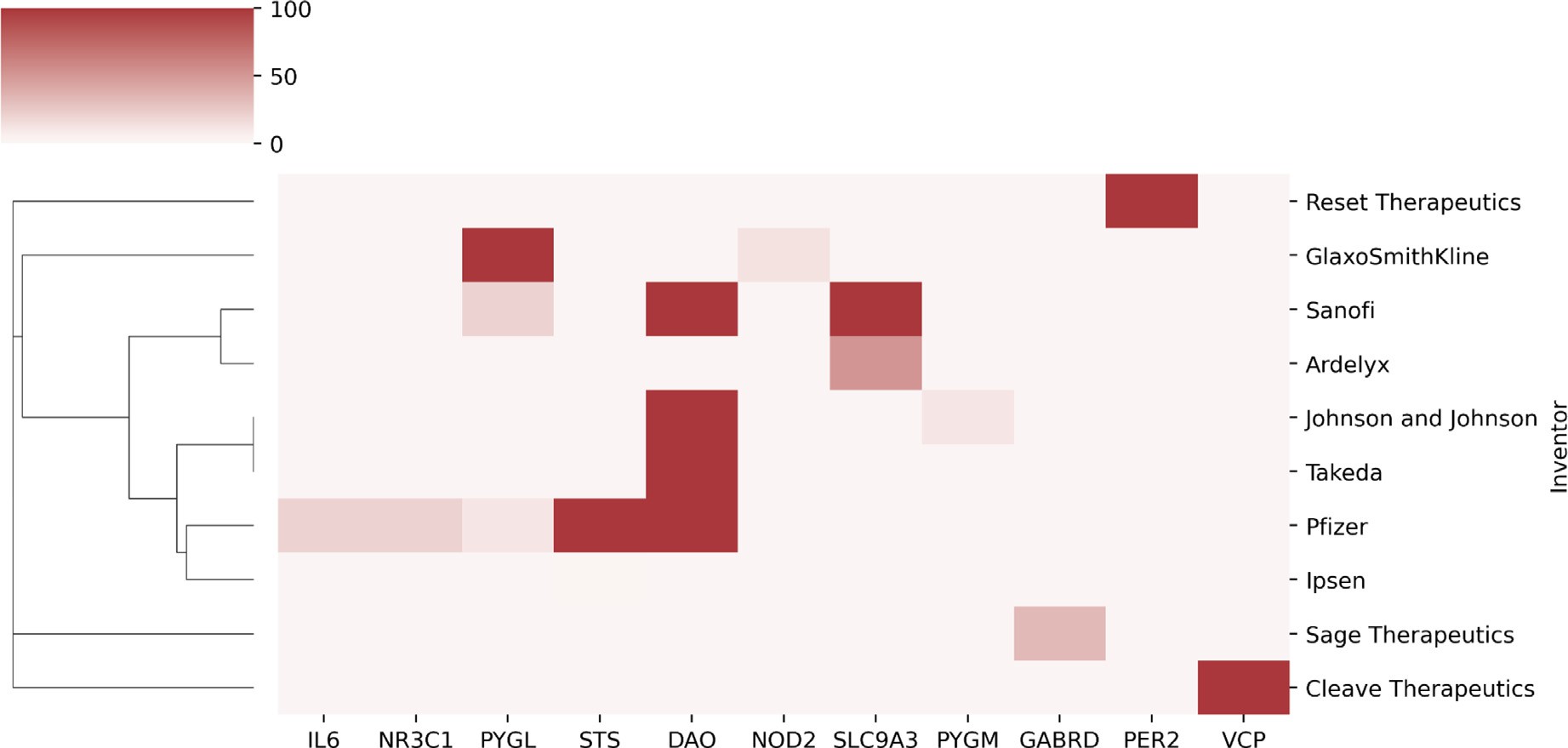
[over 50 patents (https://ir.ardelyx.com/news-releases/news-release- details/ardelyx-licenses-nap2b-phosphate-inhibitor-program-kidney). A](https://ir.ardelyx.com/news-releases/news-release-details/ardelyx-licenses-nap2b-phosphate-inhibitor-program-kidney) prominent target in the cluster space is DAO, with competitive patenting between large pharmaceutical companies like Pfizer, Takeda, and John and Johnson. A number of singletons are also seen, indicating the devel- opment of candidates targeting isolated targets (e.g. GSK, Cleave). To- gether this target portfolio approach allows for identifying top patented targets within the indication areas and comparing patent owners’ target portfolios.

As mentioned, no patent assignees in the top 10 were aﬃliated with an academic institute. To depict the difference, we divided organisa- tions into two categories, industry and academia. Interestingly, a boost in patent applications from the academic sector was seen after 2012 has been tracked. (**Supplementary Figure 2A**). However, a general pos- itive trend in patent applications by the industry is observed denoting increasing interest within pharmaceutical drug discovery for novel treat- ments for this disease. During the same time, patent law amendments took place in the United States of America, commonly referred to as the 2011 Patent Reform Act [[22]](#_bookmark44). The reformed law pointed out the change in the prosecution of patent applications within the U.S. Patent and Trademark Oﬃce (USPTO) by moving from a “first-to-invent” to a “first-to-file” system. This law thereby redefined what was required in terms of prior art, modifying the application process in a significant way.

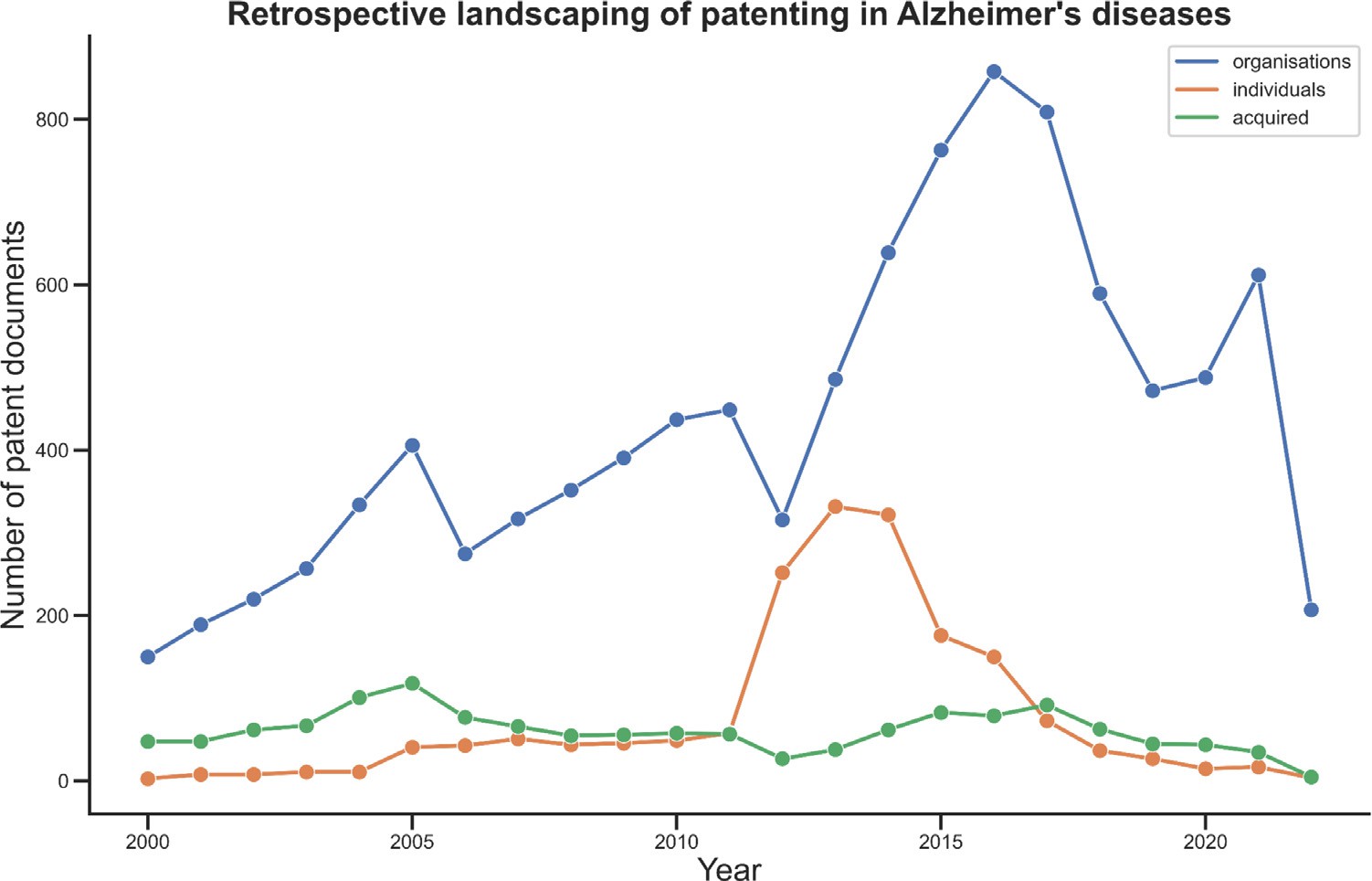
* + 1. *Identifying key players in Alzheimer’s disease patenting over the past 20 years*

In Alzheimer’s disease, 10,017 organisations, 1778 individuals, and 1386 acquired organisations have contributed towards the patenting ac- tivity. Similar to the rare disease patent landscape, organisations are in the lead compared to individuals and acquired organisations (as shown in [Fig. 6](#_bookmark13)). This is attributed to the exceptionally high cost involved in the research and development of clinical trials, which cannot be afforded by publicly funded universities or small-scale organisations unless in partnerships with pharmaceuticals [[23]](#_bookmark46). On the other hand, individuals showed an increase in patenting activity in 2011, with a peak in 2013 and then a decline afterwards, likely due to changes in patenting laws made by US and EU government bodies that allowed for commercial in- novation rights to universities as well. Thus, individuals aﬃliated with universities lost their right to file patents to universities, while other in-

**Fig. 4. Top owners-based patent analysis in the field of rare diseases.** Patent landscaping of the top 10 owners in the rare disease domain in descending order. The top owners were deduced from the In the figure, it can be seen that large pharmaceutical companies historically active in rare diseases, like Sanofi or Takeda, declined their patenting activity in the past few years, allowing other biopharmaceuticals such as GlaxoSmithK- line, Pfizer, and Sage therapeutics to enter the field.



**Fig. 5. Target portfolio of top 10 owners in rare diseases**. The hierarchically-clustered heatmap denotes the correlation distance between the top 10 rare disease owners based on their respective target patent landscape. The colour scale in the heatmap represents the number of patent documents associated with these targets. Together this clustering approach allows for the identification of clusters, here between Sanofi and Ardelyx and Pfizer and Ipsen, thereby indicating overlap between the targeting portfolio of these pharmaceuticals. From the plot, we can also notice the focus on DAO inhibitors within the rare disease industry.

**Fig. 6. Patent activity by owner type in the Alzheimer’s field**. The line plots indicate the growth of the patent documents years based on assignee type: organisations (blue), individ- uals (orange), and acquired (green). Pharma- ceutical companies and universities are leading in patenting activity compared to others. Com- pared to the rare disease field, the contribution of the acquired owners is greater. A decline in Individual owner activity was observed post- 2012 due to the legislature changes in US and EU.

dividuals attracted pharmaceutical companies towards them and soon were hired by these large companies.

We generated a retrospective perspective of the top 10 organisations found in Alzheimer’s patent corpora ([Fig. 7](#_bookmark14)). Even though the total num- ber of patent applications was higher for Alzheimer’s disease compared to rare diseases, a similar trend was observed. The top patent owners include Roche, Bristol Myers Squibb, Pfizer, etc., which are global phar- maceutical companies. There were “patent cliffs”, where a sudden loss of patent rights occurs due to concomitant expirations of several patents at the same time [[24]](#_bookmark48), within the underlying cohorts with decrements and increments in the patenting activity during the two decades. The increment denoted the increased interest of researchers in understand- ing the complex mechanism involved in Alzheimer’s disease. Despite the high investment in basic and clinical research, the mechanisms involved in the modulation of Alzheimer’s disease remained unclear, with a high attrition rate seen in clinical trials [[25](#_bookmark50),[26](#_bookmark52)]. We would like to point out that the trend mentioned above for Alzheimer’s disease would not be fully concordant with the current market trends as we did not consider any “biologicals” compounds, which have played an enormous role in drug discovery against Alzheimer’s disease lately [[27]](#_bookmark54). A similar trend as that of rare diseases was observed in the cumulative patenting of the top patent inventors in Alzheimer’s disease. A gradual increase in activity was observed from 2013 with a peak around 2016 followed by a decline in the activity as the years came closer to the COVID-19 pandemic age (**Supplementary Figure 1**).

Next, we looked at the top patent owners with Alzheimer’s dis-

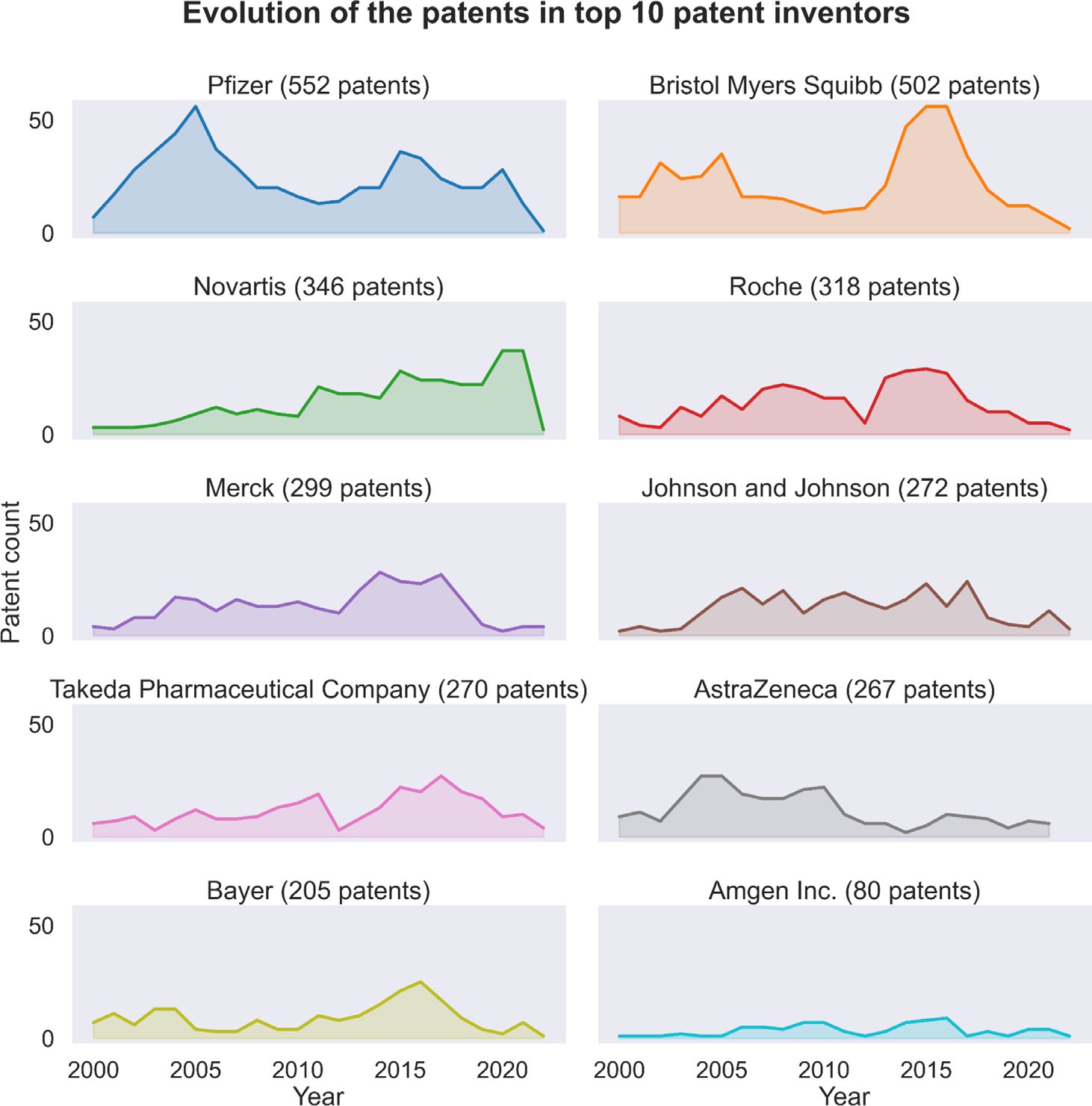
ease and plotted their target portfolio to examine whether two or more owners follow a similar target portfolio. Prominent col- laboration partners such as Novartis and Amgen have coordi- nated neuroscience efforts and hence can be seen within a clus- [ter in](https://www.genengnews.com/news/amgen-novartis-launch-neuroscience-drug-collaboration/) [Fig.](#_bookmark15) [8 (https://www.genengnews.com/news/amgen-novartis- launch-neuroscience-drug-collaboration/). Moreover, individual com-](https://www.genengnews.com/news/amgen-novartis-launch-neuroscience-drug-collaboration/) panies such as Takeda are observed due to their dominance, in particu- lar, targeting mechanisms such as LPAR5 and SLC6A9. It is noteworthy that despite having common target spaces with leading pharmaceuti- cals, each top owner has at least one other neurodegenerative-related target space where it is leading with respect to the patenting landscape. ROCK2 for BMS, TTK for Bayer, and GCG for Pfizer are examples of this case. Overall, this approach enables identification of patent owners with comparable target focus and uncovers patterns in neurodegenera- tive drug target prioritisation.

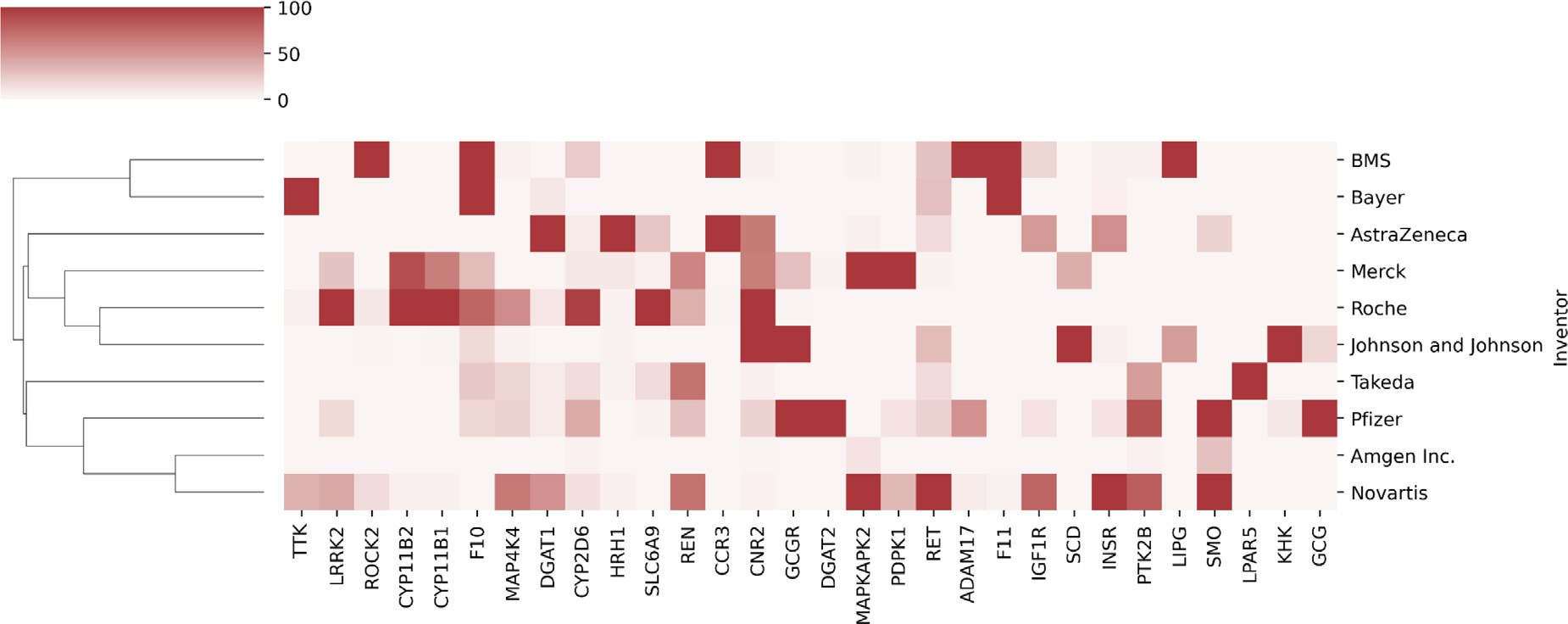
Similarly to the rare disease case, none of the top patent owners were from the academic sector, which is not surprising given the R&D cost involved in the case of Alzheimer’s disease [[23]](#_bookmark46). Thus, we looked into the patenting activity of the academic sector and compared their growth with the industry (**Supplementary Figure 2B**). Similar to rare diseases, a general positive trend in patent applications by the indus- try is observed. An incremental rise in academic patenting activity was observed due to the amendment of the 2011 Patent Reform Act that was highlighted previously. During the same period, there was es- tablishment of an act in the United States, the National Alzheimer’s Project Act (NAPA), that promoted research and development in the [field of Alzheimer’s (https://www.nia.nih.gov/about/nia-and-national- plan-address-alzheimers-disease). In Europe, collaborative projects in-](https://www.nia.nih.gov/about/nia-and-national-plan-address-alzheimers-disease) volving public-private partnerships with key stakeholders from the phar- maceutical industry were developed. One such project was the Euro- [pean Prevention of Alzheimer’s Dementia Consortium EPAD (https://ep- ad.org/) which involved data collection and analysis to speed up the](https://ep-ad.org/) drug discovery process for the disease, and support for improving clini- cal trial practice and design.

* 1. *Elucidation of the research trends of targets using patent documents*

Many factors contribute to the successful discovery of small molecule drugs for specific therapies. One of the essential factors is the mode of action (MoA) of the drug in the human body, which refers to the ef- fect a drug has on a biological protein (also known as a target) that is relevant to mechanisms involved within diseased conditions. With the help of patent documents, we can investigate the target-based MoA claims within patent applications based on their effect in certain indi- cation areas or on their importance over time from success achieved in clinical trials. Additionally, identifying potential disease targets, their prevalence, and their role in affected patient populations can be useful in developing new intellectual property. In this section, we will focus only on trends in the prioritisation of targets in patent literature over the years, while the therapeutic usage of targets has been previously described [[28]](#_bookmark56).

To investigate the research trend for targets, we first linked the patent information to targets. This was done by retrospectively connect- ing the patents to the targets through small molecule modulators. We would like to highlight that no prospective mapping of patents to targets based on their presence in the patent document was performed to con-

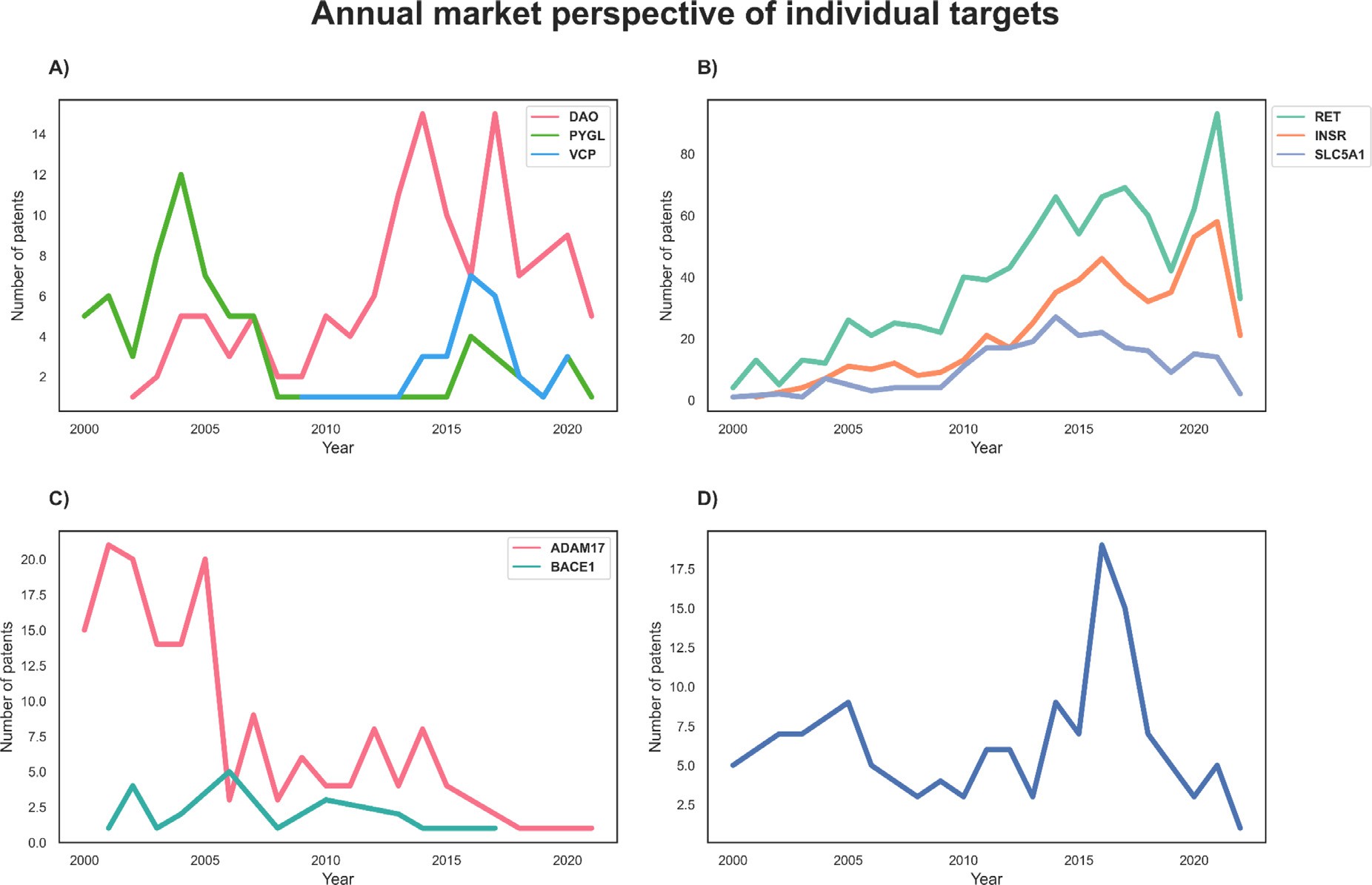
**Fig. 7. Patenting activity from the top owners in Alzheimer’s disease.** In descending order, the plots showcase the patenting activity of the past two decades for the 10 top owners. The leaders in the field are global pharmaceutical companies such as Pfizer and BMS. All the top owners have been active within the field of Alzheimer’s with consistent patenting activity.



**Fig. 8. Target portfolio of top 10 patent owners in Alzheimer’s diseases**. The hierarchically-clustered heatmap denotes the correlation distance between the top 10 Alzheimer’s patent owners based on their respective target patent landscape. Each coloured box in the plot indicates the number of patents from the owner linked to the specific target. We can observe three main clusters in this plot: the first is that of BMS and Bayer, the second is that of Roche and Johnson & Johnson, and the third is composed of Amgen and Novartis.

firm the link. Similar to the patent assignee analysis, we looked into the prevalence of targets over time **(Supplementary Tables 1 and 2)**. Fur- thermore, we will clearly distinguish the target priorities between RD and AD to understand the involved targets’ evolution. For both RD’s and AD, only the top targets, represented by the number of corresponding patent documents, were used to understand the target portfolio focus within the pharmaceutical companies.

We first started to evaluate the variations of the targets that are prioritised each year, to form the basis for target-centric landscaping, enabling us to identify targets that have been continued, diverted (to other targets), or halted, with respect to research, during their develop- ment. During this analysis, it was observed that the target perspective of the patent landscape was characterised by waves, which meant that not all targets were prioritised simultaneously **(Supplementary Figure 3)**.



**Fig. 9. The patenting frequency of selected targets within each disease domain. (A) Targets for rare diseases.** In the case of rare diseases, we looked into three targets, namely DAO (red), PYGL (green), and VCP (blue), to understand the underlying market perception concerning patents. (**B) Targets for insulin metabolism- related neurodegenerative diseases.** In the case of Alzheimer’s disease, we looked into a set of targets that play a role in insulin and sugar metabolism, namely RET (turquoise), INSR (orange), and SLC5A1 (violet). Each of the three targets follows a similar patenting profile. (**C) Secretase-related targets in Alzheimer’s diseases.** The major secretase enzymes related to Alzheimer’s disease include ADAM17 (pink) and BACE1 (teal), and both of them follow a different patenting profile despite being in the same class. (**D) Patenting profile for ELANE, a target within Alzheimer’s disease.**

Thus, we looked at selected high profile targets for each disease and compared the patent activity with the corresponding therapeutical trial activities to understand the correlation between the agglomerated tar- gets in a specific year and the preclinical research in drug discovery.

Glycogen phosphorylase L (PYGL) was one of the high incidence tar- gets with 71 patent references. The mutation in the PYGL gene causes inhibition in the conversion of glycogen to glucose, thus, associating it with an autosomal recessive rare disease called glycogen storage dis- ease type VI (GSDVI) [[29]](#_bookmark57). This key link between the mutant PYGL and GSDVI was the reason for its early research and patenting activity. How- ever, over the past years, interest in this target has gradually decreased, with almost no patent documents between 2014 and 2019 ([Fig. 9](#_bookmark16)A). In early 2019, small molecules called glucokinase activators (GKAs) were thought to be a potential solution for type II diabetes due to their ability to regulate glucose-6-phosphate and PYGL [[30]](#_bookmark59). As research progressed on GKAs, it was found that they induced hypoglycemia, a condition char- acterised by decreased blood glucose levels, causing all clinical trials during that time to be terminated [[31]](#_bookmark61). As a result, researchers refo- cused on the PYGL gene post-2014.

Another illustrative target in the ranked list of targets was ATPase valosin-containing protein (VCP/p97). This protein plays a role in in- tracellular homoeostasis by regulating protein metabolism and is asso- ciated with multiple diseases such as Paget disease and Amyotrophic Lateral Sclerosis and its types [[32](#_bookmark62),[33](#_bookmark64)]. However, due to the complex na- ture of the gene’s function in the body, research took an extended period, causing a dormant period in the patenting world. As soon as it gained interest, there was a steep increase in the number of patent documents, but this number soon fell ([Fig. 9](#_bookmark16)A). The key reason for this decline could be realistically attributed to VCP inhibitor CB-5083, a molecule devel- oped by Cleave Bioscience, that failed at phase-1 due to off-target effects

[[34]](#_bookmark65). Subsequently, it was deduced that the covalent binding towards the target caused mutations that modified the morphology of protein, thus, making cells resistant to the drug [[35]](#_bookmark66). Moreover, since the target is involved in cellular homoeostasis [[36]](#_bookmark68), a deeper understanding of the MoA of drugs is needed for such targets.

Another target that surfaced in the rare disease domain with the highest number of associated patent documents was d-amino acid ox- idase (DAO). It is a neuromodulator that was found to be involved in psychotic disorders [[37]](#_bookmark30). This target has been active over the years from both the patenting ([Fig. 9](#_bookmark16)A) and scientific literature, with several clini- cal trials. Additionally, the discovery of the target’s role in schizophrenia and cognitive dysfunction, and the need for its inhibition in these condi- tions, is one of the reasons for its active patenting landscape [[38]](#_bookmark33). How- ever, in the past two years, there has been a decrease in patents. This decline in patent activity could be ascribed to the COVID-19 pandemic, as global attention and resources have been focused on addressing the outbreak [[39]](#_bookmark34).

Interestingly, in the context of neurodegenerative diseases, some of the most well-known targets for Alzheimer’s, such as amyloid beta pre- cursor protein (APP) and tau (MAPT) [[40]](#_bookmark36), were not found in the top target list **(Supplementary Table 2).** This was likely due to an ab- sence of publicly available bioactivity data within resources such as ChEMBL on these targets, as antibodies and other biologicals are not considered. Contrary to these, coagulation factors, such as coagulation factor X (F10), which have a direct link to neuro-related diseases, were found in the top list [[41]](#_bookmark39) due to several novel small molecule modu- lators discovered. At the top of the target list was ret proto-oncogene (RET), with 886 patent application documents, demonstrating an active patenting profile ([Fig. 9](#_bookmark16)B). This activity profile can be attributed to the growing interest and research regarding insulin and sugar metabolism

in the brain, where the receptor RET contributes towards metabolic ho- moeostasis [[42]](#_bookmark41). Other targets that contribute to the same mechanisms include insulin receptors (INSR) and glucose transporters solute carrier family 5 member 1 (SLC5A1), which were also found in the list later with an active patenting profile ([Fig. 9](#_bookmark16)B).

Secretases such as alpha-secretase (ADAM17) have gained interest within the pharmaceutical industry since the early 2000s due to their role in inhibiting amyloid beta formation [[43]](#_bookmark42). Despite high patenting activity in the early years, there was a sharp decrease in 2005, and in- terest in such targets has been declining since then ([Fig. 9](#_bookmark16)C). Possible reasons for this decline include the failure of clinical trials for secretase inhibitors Lundbeck’s Flurizan and Lilly’s semagacestat in later years [[44](#_bookmark45),[45](#_bookmark47)]. The major drawback of these compounds was their inability to reach the desired concentration in the brain, as they could not pass the blood-brain barrier, thus, not reaching long residency time.

Other targets, such as neutrophil elastase (ELANE), also demonstrate patenting cliffs [[25]](#_bookmark50). As described previously, this is correlated with ei- ther research activities on the target or clinical trial outcomes of drugs. The same pattern was observed in ELANE ([Fig. 9](#_bookmark16)D), which involved the development of elastase inhibitors and their advanced clinical trials [[46]](#_bookmark49). The declining patenting activity was attributed to the identifica- tion of multiple mutations for the targets and the involvement of these mutants in multiple diseases. This target complexity thus required fur- ther research on both the biological and drug development sides.

Lastly, it is already known that there is a “lag” in the period between the filing of patent applications and the scientific research involved. We extrapolated this lag in the target space by looking into selected top tar- gets and correlating the patent applications and research publications over time (**Supplementary Figure 4**). It can be observed that, on aver- age, there is a lag of 5 years between the research involved on a target (with or without disease context) and the filing of patents relevant to the target. There are indeed certain outliers to this lag period such as INSR, which took at least 14 years to enter into the Alzheimer’s patent- ing world. One reason for this could be due to the importance of the target on other diseases, causing it to be extensively researched in the early 2000s. However, only after 2009 did the study of INSR in neu- rodegeneration start following the recognition of type 2 diabetes as a risk factor for Alzheimer’s [[47]](#_bookmark51).

* 1. *Identification of drug repurposing scenarios of diseases using patent documents*

As drug discovery is so resource intensive, a preliminary assessment of promising or under-exploited protein target and/or chemical scaf- fold combinations repurposing drugs could be a convenient alternative approach. We aimed to investigate whether drugs that have been repur- posed for different therapeutic areas are also present in the patent cor- pora for the two disease areas. To achieve this, we clustered the patent documents based on the compounds they mentioned and ranked them in descending order of their counts. Since we were interested in repurposed compounds, we restricted the patent search to granted documents only. It is important to note that a patent application does not exclusively cover a single compound due to the presence of Markush structures [[48]](#_bookmark53). These structures are a scaffold representation of a compound with the addition of variable side chains, thus making it more diﬃcult for re- searchers to identify a specific compound via a direct search match, al- though commercial software tools (like SciFinder) have been developed to solve this problem. A single compound can also be covered by mul- tiple patents, as the same compound can be used for varied innovative features (such as disease type or treatments, physical pharmaceutical forms, combinations with other therapeutic compounds, etc.), allow- ing for independent drafting for each area. Moreover, multiple patent drafting does not necessarily mean the compound is used for multiple purposes, as they may be patents filed for the compound’s formulations or non-obvious structural modifications. Thus, a systematic case-by-case study needs to be done to identify such compounds and ranking of com-

pounds based on patent documents can serve as a starting point for such analysis.

Aligning our definition of repurposed compounds to a patent per- spective, we categorised a compound as repurposed if it was found in more than two granted patent documents (belonging to B or E kind code). This resulted in a reduced dataset of 145 repurposed compounds in rare diseases and 1928 in Alzheimer’s diseases, out of 585 and 22,682 compounds, respectively. Moreover, from the entire compound dataset, only four compounds for rare disease compounds and 215 com- pounds for Alzheimer’s diseases have been tested in clinical trials (as of 1st of Jan 2023 [http://clinicaltrial.gov](http://clinicaltrial.gov/)) (as shown in **Supplementary Figure 5)** with minimal overlap with the dataset of repurposed com- pounds. This indicates that most of the compounds within the granted patent space under study are still in the research or pre-clinical phase, likely due to the limited availability of bioassay data on clinically vali- dated compounds in public data repositories such as ChEMBL.

Cleave Biosciences’s CB-5083 is one of the RD repurposed drugs forming an interesting case study. The company had patented the com- pound in Europe (EP-2,875,018-B1) and the United States of Amer- ica (US-10,010,554-B2) as an inhibitor of VCP for cancer treatment. Following a phase I clinical trial, which revealed off-target activities, clinical development was halted [[49]](#_bookmark55). Despite the unsuccessful clinical trial, it has made its way to the RD community regarding its poten- tial use in treating Paget disease, a VPC-dependant disease [[32]](#_bookmark62). An- other example is Fosdagrocorat, a compound patented by Pfizer as a glucocorticoid receptor modulator for osteoporosis (EP-2,114,888-B1). Research has linked the same receptor to rheumatoid arthritis [[50]](#_bookmark58), thus making it possible to repurpose the compound for this rare dis- ease [[51]](#_bookmark60). Upadacitinib, commonly known by the trade name Rinvoq, is an Abbvie-developed kinase inhibitor that has been approved by the FDA for the treatment of multiple diseases, such as rheumatoid arthritis

[[52]](#_bookmark63) and atopic dermatitis. It inhibits the Spleen tyrosine kinase (Syk) (US-10,072,034-B2), which has recently been identified as a key modu- lator of tauopathy, a disorder involving the accumulation of Tau [[53]](#_bookmark64). This discovery opens up the potential for the drug to be used to treat neurodegenerative diseases. Similar to rare diseases, we found a few compounds derived from the cancer domain. Sotorasib is a compound that inhibits a specific KRAS mutant, KRAS G12C and demonstrates anti- cancerous properties (US-10,519,146-B2) [[54]](#_bookmark65). Several researchers have identified several KRAS mutants that play a role in brain and spinal- related haemorrhage [[55]](#_bookmark67). This presents an opportunity to test the com- pound on multiple KRAS mutations and optimise it to create a mutant- specific or general-purpose compound. In summary, it is evident through the above examples that the mining of patent documents from a com- pound perspective can provide insights into how compounds that have been repurposed for multiple disorders beyond the scope of their origi- nal patents.

* 1. *Limitations and future perspective of the approach*

The current study illuminates the historical trends observed in neu- rodegenerative and rare disease drug discovery and research through the analysis of patents. However, it is important to note that the method- ology employed has limitations due to the lack of inclusion of “biolog- icals” compounds such as vaccines or long peptides. One such limita- tion is the indirect relationship between proteins and patents, as the connections were made based on chemical modulators, and the initial proteins may not necessarily be mentioned in the patent. To address this, further analyses using natural language processing (NLP) methods to extract gene annotations in patent documents are the best alterna- tive. Additionally, the modulators linked to the proteins were sourced from open-source data repositories, such as ChEMBL. As the deposition of clinical candidate structures is infrequent and often only occurs after early clinical trials, the patent analysis may miss important compounds currently in development. To mitigate this, we focused on highly ac- tive compounds with submicromolar activity to increase the chances of

identifying "champion" compounds in a patent. Lastly, the time range chosen for the analysis (2000–2021) may have skewed the overall dis- tribution of results, as certain compounds approved for treatment before 2000 were not included in the analysis. For instance, Jeffrey Cummings (2018) mentioned in his article that “Agents producing cognitive en- hancement may have mechanisms independent of AD-specific pathol- ogy (e.g., 5-HT6 antagonists)” [[56]](#_bookmark69). 5-hydroxytryptamine receptor 6 (5-HT6) antagonists like memantine and acetylcholinesterase inhibitors have been approved for AD treatment and only alleviate some AD symp- toms by enhancing cholinergic signalling. These chemicals are not cu- rative and do not surface in our analysis, as they were patented before 2000. Despite these limitations, the study aims to provide insight into the intellectual effort required within the pharmaceutical industry to identify competitors’ patents containing "clinical candidates" and estab- lish a successful "me-too" strategy.

# Conclusion

Finding safe and effective therapies for diseases with unmet patient needs is a crucial goal of pharmaceutical research. Traditionally, the discovery process was initiated by systematically reviewing the scien- tific literature for the most promising disease-linked therapeutic targets or pathways. Once a targeting mechanism has been selected, it would either be subsequently validated or eventually discarded and an alterna- tive strategy adopted based on new experimental insights or literature analysis. Additionally, assessment of target opportunities and estimating their potential for successful translation in development candidates is a core strategic activity within the pharmaceutical industry. Such analy- sis is driven by machine learning models trained on large multimodu- lar data, including toxicological effects and druggability characteristics. Moreover, the novelty associated with the chemical structure of candi- date molecules is a critical aspect of patent assessment, as companies need to be aware of pre-existing successful scaffolds. This chemoinfor- matic integration of structural features is often a key factor in defining medicinal chemistry strategies when optimising leads and further pro- gressing compounds to development stages.

This study uses the PEMT tool to extract, analyse, and explore the patent landscape for rare and Alzheimer’s diseases. The results revealed historical trends in the past 20 years and identified key patent hold- ers and application dates. Such a retrospective trend paved the way for generating a longitudinal visualisation of the importance of the targets in both diseases. Additionally, the analysis showed the potential for us- ing patent literature in drug repurposing and annotating the clinical significance of targets. Last, but not least, the study was able to iden- tify and reflect the shift between the research and patenting period in- volved in drug discovery. Certainly, this effort cannot be assumed to reach the same proficiency as a professional patent analysis as it takes advantage of public databases, which themselves are limited, as dis- cussed in the previous sections. This paper is focused on the integration of public bioactivity datasets with patent applications and their analy- sis. We will inspect through future analyses how the patent information can optimise drug search within disease mechanisms networks, which is out of the scope of the current paper. This is just the tip of the ice- berg as huge effort and resources need to be dedicated to standardis- ing the data and making them a potential source for future ML and LM models.

Our work demonstrates that the PEMT tool provides a comprehen-

sive and informative patent cohort to understand the research in the field of small molecule drugs. This landscape not only aligns with sci- entific literature but also allows for competitive intelligence analysis in pharmaceutical R&D. Overall, this patent landscaping approach from a drug discovery perspective highlights the importance of analysing patents and their usefulness in making decisions for future drug devel- opment efforts. The possibility of tracking patenting activities from pub- licly available resources might help define strategies for target ranking within a therapeutic focus. Our methodology will contribute towards

simplifying and streamlining the process of target prioritisation in drug discovery campaigns at a pharmaceutical company.

# Funding

This work was supported by the German Federal Ministry of Educa- tion and Research (BMBF) [[01GM2003A](#_bookmark17)] (project ‘TreatKCNQ’).

# Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

# CRediT authorship contribution statement

**Yojana Gadiya:** Project administration, Formal analysis, Writing – original draft. **Philip Gribbon:** Writing – review & editing. **Martin Hofmann-Apitius:** Writing – review & editing. **Andrea Zaliani:** Project administration, Formal analysis, Writing – review & editing, Writing – original draft.

# Data availability

The [data and code is available on GitHub at https://github.com/ Fraunhofer-ITMP/Pharmaceutical-patent-landscaping.](https://github.com/Fraunhofer-ITMP/Pharmaceutical-patent-landscaping)

# Acknowledgements

The authors thank Daniel Domingo-Fernández, Bruce Schultz, and the anonymous reviewers for their comments and suggestions on the manuscript. We would also like to thank Bruce Schultz for providing the protein data for the HBP network, hosted by Fraunhofer SCAI.

# Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ailsci.2023.100069](https://doi.org/10.1016/j.ailsci.2023.100069).

# References

1. Walker RD. Patents as a source of scientific and technical information in developing nations. World Pat Inf 1988;10(1):5–10. doi:[10.1016/0172-2190(88)90210-4](https://doi.org/10.1016/0172-2190(88)90210-4).
2. Morgan MR, Roberts OG, Edwards AM. Ideation and implementation of an open science drug discovery business model–M4K Pharma. Wellcome Open Res 2018;3. doi:[10.12688/wellcomeopenres.14947.1](https://doi.org/10.12688/wellcomeopenres.14947.1).
3. Scianna M, Munaron L. Computational approaches for translational oncology: concepts and patents. Recent Pat Anticancer Drug Discov 2016;11(4):384–92. doi:[10.2174/1574892811666161003111543](https://doi.org/10.2174/1574892811666161003111543).
4. Pan CL, Chen FC. Patent trend and competitive analysis of cancer immunother- apy in the United States. Hum Vaccin Immunother 2017;13(11):2583–93. doi:[10.1080/21645515.2017.1361074](https://doi.org/10.1080/21645515.2017.1361074).
5. Zhang T, Chen J, Jia X. Identification of the key fields and their key technical points [of oncology by patent analysis. PLoS ONE 2015;10(11):e0143573. doi:10.1371/jour- nal.pone.0143573.](https://doi.org/10.1371/journal.pone.0143573)
6. Xiong W, Cao J, Zhou X, Du J, Nie B, Zeng Z, Li T. Design and evaluation of a pre- scription drug monitoring program for Chinese patent medicine based on knowledge graph. EvidBased Complement Alter Med 2021;2021. doi:[10.1155/2021/9970063](https://doi.org/10.1155/2021/9970063).
7. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. Drug discovery today 2010;15(23–24):1052–7. doi:[10.1016/j.drudis.2010.10.003](https://doi.org/10.1016/j.drudis.2010.10.003).
8. Papadatos G, Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, Overington JP. SureChEMBL: a large-scale, chemically annotated patent document database. Nu- cleic Acids Res. 2016;44(D1):D1220–8. doi:[10.1093/nar/gkv1253](https://doi.org/10.1093/nar/gkv1253).
9. Mucke HA. What patents tell us about drug repurposing for cancer: a land- scape analysis. In: Seminars in cancer biology, 68. Academic Press; 2021. p. 3–7. doi:[10.1016/j.semcancer.2019.09.010](https://doi.org/10.1016/j.semcancer.2019.09.010).
10. Falaguera MJ, Mestres J. Congenericity of claimed compounds in patent applica- tions. Molecules 2021;26(17):5253. doi:[10.3390/molecules26175253](https://doi.org/10.3390/molecules26175253).
11. Choi Y, Park S, Lee S. Identifying emerging technologies to envision a future in- novation ecosystem: a machine learning approach to patent data. Scientometrics 2021;126:5431–76. doi:[10.1007/s11192-021-04001-1](https://doi.org/10.1007/s11192-021-04001-1).
12. Leach AR, Magarinos MP, Gaulton A, Felix E, Kiziloren T, Arcila R, Oprea TI. Illu- minating the Druggable Genome through Patent Bioactivity Data. *bioRxiv*, 2022-07 2022. doi:[10.1101/2022.07.15.500187](https://doi.org/10.1101/2022.07.15.500187).
13. Gadiya, Y., Zaliani, A., Gribbon, P., & Hofmann-Apitius, M. PEMT: a patent enrich- ment tool for drug discovery. *Bioinformatics*. doi:[10.1093/bioinformatics/btac716](https://doi.org/10.1093/bioinformatics/btac716).
14. [Weinreich SS, Mangon R, Sikkens JJ, Teeuw ME, Cornel MC. Orphanet: a European database for rare diseases. Ned Tijdschr Geneeskd 2008;152(9):518–19.](http://refhub.elsevier.com/S2667-3185(23)00013-2/sbref0014)
15. Lage-Rupprecht V, Schultz B, Dick J, Namysl M, Zaliani A, Gebel S, Hofmann- Apitius M. A hybrid approach unveils drug repurposing candidates target- ing an Alzheimer pathophysiology mechanism. Patterns 2022;3(3):100433. doi:[10.1016/j.patter.2021.100433](https://doi.org/10.1016/j.patter.2021.100433).
16. Hoyt CT, Konotopez A, Ebeling C. PyBEL: a computational framework for Biological [Expression Language. Bioinformatics 2018;34(4):703–4. doi:10.1093/bioinformat- ics/btx660.](https://doi.org/10.1093/bioinformatics/btx660)
17. Kabir E, Uzzaman M. A review on biological and medicinal impact of heterocyclic compounds. Res Chem 2022:100606. doi:[10.1016/j.rechem.2022.100606](https://doi.org/10.1016/j.rechem.2022.100606).
18. Bagley N, Berger B, Chandra A, Garthwaite C, Stern AD. The Orphan Drug Act at 35: observations and an outlook for the twenty-first century. Innov. Policy Econ. 2019;19(1):97–137. doi:[10.1086/699934](https://doi.org/10.1086/699934).
19. Stella P, Gold-von Simson G. Pharmaceutical pricing, cost containment and new treatments for rare diseases in children. Orphanet J Rare Dis 2014;9(1):1–4. doi:[10.1186/s13023-014-0152-2](https://doi.org/10.1186/s13023-014-0152-2).
20. Hvide HK, Jones BF. University innovation and the professor’s privilege. Am Eco Rev 2018;108(7):1860–98. doi:[10.1257/aer.20160284](https://doi.org/10.1257/aer.20160284).
21. Wang H, Norton J, Xu L, DeMartinis N, Sen R, Shah A, Lynch D. Results of a ran- domized double-blind study evaluating luvadaxistat in adults with Friedreich ataxia. Ann Clin Transl Neurol 2021;8(6):1343–52. doi:[10.1002/acn3.51373](https://doi.org/10.1002/acn3.51373).
22. [Ashworth AL. Race You to the patent oﬃce: how the new patent reform act will affect technology transfer at universities. Alb LJ Sci Tech 2012;23:383.](http://refhub.elsevier.com/S2667-3185(23)00013-2/sbref0022)
23. Cummings JL, Goldman DP, Simmons-Stern NR, Ponton E. The costs of developing treatments for Alzheimer’s disease: a retrospective exploration. Alzheimer’s Dement 2022;18(3):469–77. doi:[10.1002/alz.12450](https://doi.org/10.1002/alz.12450).
24. Song CH, Han JW. Patent cliff and strategic switch: exploring strategic de- sign possibilities in the pharmaceutical industry. Springerplus 2016;5(1):1–14. doi:[10.1186/s40064-016-2323-1](https://doi.org/10.1186/s40064-016-2323-1).
25. Anderson, R.M., Hadjichrysanthou, C., Evans, S., & Wong, M.M. (2017). Why do so many clinical trials of therapies for Alzheimer’s disease fail?. *The Lancet*, 390(10110), 2327–9. [10.1016/S0140-6736(17)32399-1](https://doi.org/10.1016/S0140-6736\05017\05132399-1).
26. Koynova R, Tenchov B. Natural product formulations for the prevention and treat- ment of Alzheimer’s disease: a patent review. Recent Pat Drug Deliv Formul 2018;12(1):23–39. doi:[10.2174/1872211312666171207152326](https://doi.org/10.2174/1872211312666171207152326).
27. Senior M. Fresh from the biotech pipeline: fewer approvals, but biologics gain share. Nat. Biotechnol. 2023;1. doi:[10.1038/s41587-022-01630-6](https://doi.org/10.1038/s41587-022-01630-6).
28. Zdrazil, B., Richter, L., Brown, N., & Guha, R. (2020). Moving targets in drug dis- covery. Sci Rep, 10(1), 1–15. [10.1038/s41598-020-77033-x](https://doi.org/10.1038/s41598-020-77033-x).
29. Burwinkel B, Bakker HD, Herschkovitz E, Moses SW, Shin YS, Kilimann MW. Muta- tions in the liver glycogen phosphorylase gene (PYGL) underlying glycogenosis type VI (Hers disease). Am J Hum Gene 1998;62(4):785–91. doi:[10.1086/301790](https://doi.org/10.1086/301790).
30. Matschinsky FM. Assessing the potential of glucokinase activators in diabetes ther- apy. Nat Rev Drug disc 2009;8(5):399–416. doi:[10.1038/nrd2850](https://doi.org/10.1038/nrd2850).
31. Nakamura A, Terauchi Y. Present status of clinical deployment of glucokinase acti- vators. J Diabetes Investig 2015;6(2):124–32. doi:[10.1111/jdi.12294](https://doi.org/10.1111/jdi.12294).
32. Costantini S, Capone F, Polo A, Bagnara P, Budillon A. Valosin-Containing Protein (VCP)/p97: a Prognostic Biomarker and Therapeutic Target in Cancer. Int J Mol Sci 2021;22(18):10177. doi:[10.3390/ijms221810177](https://doi.org/10.3390/ijms221810177).
33. Scarian E, Fiamingo G, Diamanti L, Palmieri I, Gagliardi S, Pansarasa O. The role of VCP mutations in the spectrum of amyotrophic lateral sclerosis-frontotemporal dementia. Front Neurol 2022;271. doi:[10.3389/fneur.2022.841394](https://doi.org/10.3389/fneur.2022.841394).
34. Le Moigne R, Aftab BT, Djakovic S, Dhimolea E, Valle E, Murnane M, Rolfe M. The p97 inhibitor CB-5083 is a unique disrupter of protein home- ostasis in models of multiple myeloma. Mol. Cancer Ther. 2017;16(11):2375–86. doi:[10.1158/1535-7163.MCT-17-0233](https://doi.org/10.1158/1535-7163.MCT-17-0233).
35. Bastola P, Wang F, Schaich MA, Gan T, Freudenthal BD, Chou TF, Chien J. Specific mutations in the D1–D2 linker region of VCP/p97 enhance ATPase activity and con- [fer resistance to VCP inhibitors. Cell Death Discov 2017;3(1):1–9. doi:10.1038/cd- discovery.2017.65.](https://doi.org/10.1038/cddiscovery.2017.65)
36. Ahlstedt BA, Ganji R, Raman M. The functional importance of VCP to main- taining cellular protein homeostasis. Biochem. Soc. Trans. 2022;50(5):1457–69. doi:[10.1042/BST20220648](https://doi.org/10.1042/BST20220648).
37. Mitchell J, Paul P, Chen HJ, Morris A, Payling M, Falchi M, de Belleroche J. Familial amyotrophic lateral sclerosis is associated with a mutation in d-amino acid oxidase. Proc Natl Acad Sci 2010;107(16):7556–61. doi:[10.1073/pnas.0914128107](https://doi.org/10.1073/pnas.0914128107).
38. Sacchi S, Rosini E, Pollegioni L, Molla G. d-amino acid oxidase inhibitors as a novel class of drugs for schizophrenia therapy. Curr pharma des 2013;19(14):2499–511. doi:[10.2174/1381612811319140002](https://doi.org/10.2174/1381612811319140002).
39. Riccaboni M, Verginer L. The impact of the COVID-19 pandemic on scientific [research in the life sciences. PLoS ONE 2022;17(2):e0263001. doi:10.1371/jour- nal.pone.0263001.](https://doi.org/10.1371/journal.pone.0263001)
40. Kodamullil AT, Zekri F, Sood M, Hengerer B, Canard L, McHale D, Hofmann- Apitius M. Trial watch: tracing investment in drug development for Alzheimer dis- ease. Nature reviews. Drug discovery 2017;16(12):819. doi:[10.1038/nrd.2017.169](https://doi.org/10.1038/nrd.2017.169).
41. De Luca C, Virtuoso A, Maggio N, Papa M. Neuro-coagulopathy: blood coagula- tion factors in central nervous system diseases. Int J Mol Sci 2017;18(10):2128. doi:[10.3390/ijms18102128](https://doi.org/10.3390/ijms18102128).
42. Zhao M, Jung Y, Jiang Z, Svensson KJ. Regulation of energy metabolism by receptor tyrosine kinase ligands. Front Physiol 2020;11:354. doi:[10.3389/fphys.2020.00354](https://doi.org/10.3389/fphys.2020.00354).
43. Qian M, Shen X, Wang H. The distinct role of ADAM17 in APP proteolysis and microglial activation related to Alzheimer’s disease. Cell. Mol. Neurobiol. 2016;36(4):471–82. doi:[10.1007/s10571-015-0232-4](https://doi.org/10.1007/s10571-015-0232-4).
44. Wan HI, Jacobsen JS, Rutkowski JL, Feuerstein GZ. Translational medicine lessons from flurizan’s failure in Alzheimer’s disease (AD) trial: implication for future drug discovery and development for AD. Clin Transl Sci 2009;2(3):242. doi:[10.1111/j.1752-8062.2009.00121.x](https://doi.org/10.1111/j.1752-8062.2009.00121.x).
45. Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, Joffe S, Mohs R. A phase

3 trial of semagacestat for treatment of Alzheimer’s disease. N Engl J Med 2013;369(4):341–50. doi:[10.1056/NEJMoa1210951](https://doi.org/10.1056/NEJMoa1210951).

1. Ahmad S, Saleem M, Riaz N, Lee YS, Diri R, Noor A, Elsebai MF. The natu- ral polypeptides as significant elastase inhibitors. Front Pharmacol 2020;11:688. doi:[10.3389/fphar.2020.00688](https://doi.org/10.3389/fphar.2020.00688).
2. Wang H, Wang R, Zhao Z, Ji Z, Xu S, Holscher C, Sheng S. Coexistences of in- sulin signaling-related proteins and choline acetyltransferase in neurons. Brain Res. 2009;1249:237–43. doi:[10.1016/j.brainres.2008.10.046](https://doi.org/10.1016/j.brainres.2008.10.046).
3. Geyer P. Markush structure searching by information professionals in the chem- ical industry–our views and expectations. World Pat Inf 2013;35(3):178–82. doi:[10.1016/j.wpi.2013.05.002](https://doi.org/10.1016/j.wpi.2013.05.002).
4. Leinonen H, Cheng C, Pitkänen M, Sander CL, Zhang J, Saeid S, Palczewski K. A p97/valosin-containing protein inhibitor drug CB-5083 has a potent but reversible off-target effect on phosphodiesterase-6. J Pharmacol Exp Ther 2021;378(1):31–41. doi:[10.1124/jpet.120.000486](https://doi.org/10.1124/jpet.120.000486).
5. Schlaghecke R, Kornely E, Wollenhaupt J, Specker C. Glucocorticoid receptors in rheumatoid arthritis. Arthritis Rheum Off J Am Coll Rheum 1992;35(7):740–4. doi:[10.1002/art.1780350704](https://doi.org/10.1002/art.1780350704).
6. Stock T, Fleishaker D, Wang X, Mukherjee A, Mebus C. Improved disease activity with fosdagrocorat (PF-04171327), a partial agonist of the glucocorticoid receptor, in patients with rheumatoid arthritis: a Phase 2 randomized study. Int J Rheum Dis 2017;20(8):960–70. doi:[10.1111/1756-185X.13053](https://doi.org/10.1111/1756-185X.13053).
7. Duggan S, Keam SJ. Upadacitinib: first approval. Drugs 2019;79(16):1819–28. doi:[10.1007/s40265-019-01211-z](https://doi.org/10.1007/s40265-019-01211-z).
8. Schweig JE, Yao H, Coppola K, Jin C, Crawford F, Mullan M, Paris D. Spleen tyrosine kinase (SYK) blocks autophagic Tau degradation in vitro and in vivo. J Biol Chem 2019;294(36):13378–95. doi:[10.1074/jbc.RA119.008033](https://doi.org/10.1074/jbc.RA119.008033).
9. Nakajima EC, Drezner N, Li X, Mishra-Kalyani PS, Liu Y, Zhao H, Singh H. FDA ap- proval summary: sotorasib for KRAS G12C-mutated metastatic NSCLC. Clin Cancer Res 2022;28(8):1482–6. doi:[10.1158/1078-0432.CCR-21-3074](https://doi.org/10.1158/1078-0432.CCR-21-3074).
10. Hong T, Yan Y, Li J, Radovanovic I, Ma X, Shao YW, Wang Y. High prevalence of KRAS/BRAF somatic mutations in brain and spinal cord arteriovenous malforma- tions. Brain 2019;142(1):23–34. doi:[10.1093/brain/awy307](https://doi.org/10.1093/brain/awy307).
11. Cummings J. Lessons learned from Alzheimer disease: clinical trials with negative outcomes. Clin Transl Sci 2018;11(2):147. doi:[10.1111/cts.12491](https://doi.org/10.1111/cts.12491).