



Deliverable 2.5: Update of D2.1 “Mapping the scientific and Policy Landscape of PerMed in China, including the identification of stakeholders”.





Project Acronym	SINO-EU PerMed
Project Title	Widening Sino-EU policy and research cooperation in Personalised Medicine.
Grant Agreement no.	874556
Start date of the project	01/01/2020
End date of the project	31/12/2023
Name of the deliverable	Mapping the Scientific and Policy Landscape of PerMed in China, including the identification of stakeholders
Number of the deliverable	D2.5
Related WP number and name	WP2, Mapping of Personalised Medicine Initiatives in China and Identification of gaps and needs for better fostering integration of China in ICPeMed
Related task number and name	WP2 Task 2.1: Systematic Mapping of the Research and Innovation landscape of PerMed in China (M1#M6)
Deliverable dissemination level	Confidential
Deliverable due date	31/05/2023
Deliverable submission date	14/06/2023
Lead Beneficiary	TLS
Nature	Update of 2010-2020 mapping
DISCLAIMER	
<p>This document is provided with no warranties whatsoever, including any warranty of merchantability, non-infringement, fitness for any particular purpose, or any other warranty with respect to any information, result, proposal, specification or sample contained or referred to herein. Any liability, including liability for infringement of any proprietary rights, regarding the use of this document or any information contained herein is disclaimed. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted by or in connection with this document. This document is subject to change without notice. Sino-EU PerMed has been financed with support from the European Commission. This document reflects only the view of the author(s) and the European Commission cannot be held responsible for any use which may be made of the information contained herein.</p>	

Summary

Task 2 of the project includes a systematic mapping of available literature (scientific papers, patents) has been carried out to describe, in a comprehensive, transparent and objective manner the R&I and funding landscape of PerMed in China, the main areas in which research is happening, innovation activities and cooperation with European researchers being performed and how PM is being supported through Research Program.

The research was conducted with the effort of ErreQuadro S.r.l., an IT company specialized in Artificial Intelligence for business intelligence whose solutions combine cutting-edge software of computational linguistics with a sophisticated representation of technologies. ErreQuadro set up the tools to analyze cooperativity on patents and papers on Personalised Medicine, performing a data analysis which provided Sino-EU PerMed with strategic information for the activities to come.

The mapping was updated adopting the same workflow and the integrated data will be disseminated using an interactive web platform that is achievable through the project website.

Keywords

Patent Mapping, Scientific Mapping, Personalised Medicine, Precision Medicine, Preventive Medicine, Systems Biology, Stratified Medicine, Targeted Therapy, Tailored Treatment/Therapy, Deep Phenotyping, Omics Sciences, Big Data, Machine Learning Techniques, Traditional Chinese Medicine

Revisions			
Version	Submission date	Comments	Author
_v1	30/05/2023	Sent to Consortium	TLS
_v2	01/06/2023	Revised by Consortium	TLS
_v3	14/06/2023	Revised and Finalised	TLS

Table of contents

The Sino-EU-PerMed Project.....	5
List of Abbreviations.....	6
1.Introduction.....	7
2.Patents mapping	7
3. Papers Mapping.....	18
4. Preprints Mapping.....	23

The Sino-EU-PerMed Project

Personalised Medicine (PM) approaches bring along immense potentials to improve diagnosis and treatment of diseases and prevention strategies by considering individuals' geno- and phenotypes, biomedical, lifestyle or environmental data. The International Consortium for Personalised Medicine (ICPerMed) is a member state-driven initiative of over 40 international ministries and funding agencies. The European Commission (EC) is an active observer and supports by funding several related projects. ICPerMed aims at aligning efforts in all areas of PerMed by fostering international coordination of research and innovation funding.

To support these efforts “Widening Sino-EU policy and research cooperation in Personalised Medicine” (Sino-EU-PerMed) aims at connecting ICPerMed strategies and activities with relevant Chinese stakeholders. To achieve this goal the Sino-EU-PerMed will disseminate and coordinate the following activities. Inventory and analysis of the PerMed landscape in China through a mapping. In addition the Sino-EU-PerMed consortium will exchange expertise and experts through workshops and delegation trips to China and in Europe. Thereby we will gain a better mutual understanding of PerMed activities and related cultural aspects as well as health economics issues. By establishing technical taskforces

and running a series of targeted workshops, Sino-EU-PerMed will promote international standards and pave the way for future communication, collaboration, and undertakings. Therefore the implementation of a dissemination plan is an essential part of the project.

Our consortium and activities aim at maximising the unique and long-term impact PM approaches have on cooperation and are aligned with ICPerMed. The cooperation with China set up and planned in Sino-EU-PerMed will allow ICPerMed to further strengthen Europe's leading role in this area and contribute to a successful implementation of PerMed in the global context and foster joint PerMed projects between Europe and China.

List of Abbreviations

Abbreviation	Explanation
EC	European Commission
EPO	European Patent Office
I.I.	Innovation Index
IPC	International Patent Classification
MeSH	Medical Subject Headings
UNO	United Nations Organisation
PM	Personalised Medicine
PubMed	Free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics
R&I	Research and Innovation
Sino-EU PerMed	Acronym of the Project “Widening Sino-EU policy and research cooperation in Personalised Medicine”
WPOs	World Intellectual Properties

1. Introduction

This deliverable has been proposed to obtain an updated analysis of the mapping on the collaborations between Europe and China in PM, reported in D2.1. The analysis was also published in the form of a research paper titled “*Sino-European science and technology collaboration on personalized medicine: overview, trends and future perspectives*” in Personalised Medicine, Future Medicine journal. The latter can be consulted at the following [link](#).

The aim of this update is to integrate the previous outcomes with data from 2020 to 2022, thereby grasping an important insight into the changes in the Sino-EU Cooperation in PM over the past years, even in the light of the Covid- 19 outbreak. With the respect to the initial mapping, and in respect to the scientific publications, preprints (the version of scientific papers before they have been submitted it to a journal for peer review) were included as a novel category while the workflow was, overall, maintained. Preprints article had a very important role in disseminating knowledge during the pandemic months, and therefore we decided to include them in our mapping.

Thus, the global datasets cover a timeframe starting from January 2010 to January 2023 (included).

In addition, the SINO EU-PerMed database was upgraded and implemented with interactive tools and published online freely accessible without the necessity of credentials to login-in. A pilot version is yet available, and the final version will be published by July 2023.

2. Patents mapping

2.1 Operative Flow

The operative workflow designed for the initial mapping was maintained, the scheme is depicted in Figure 1.

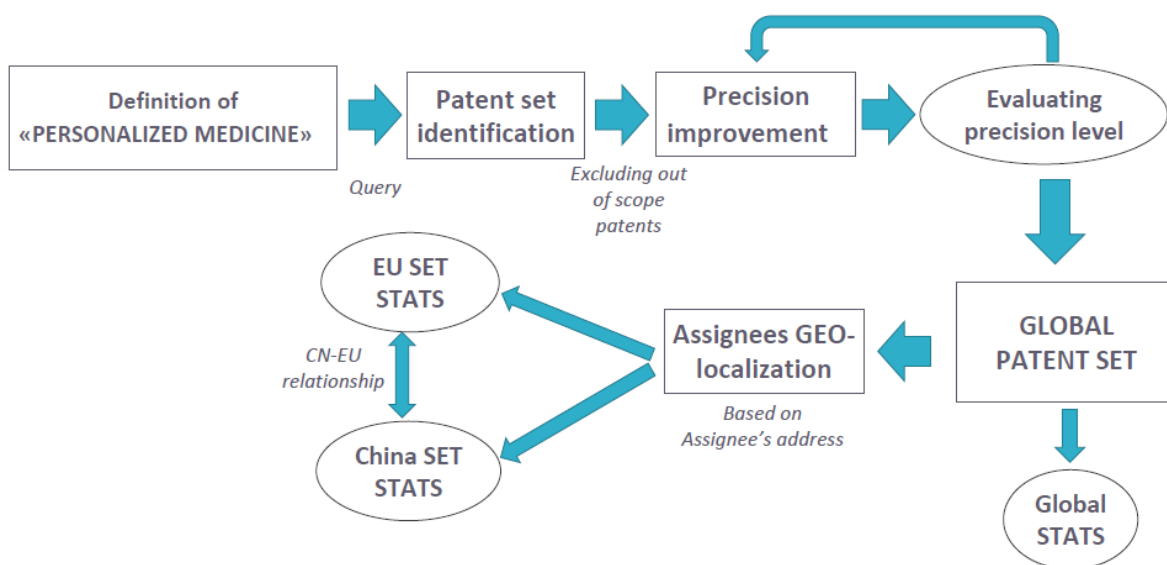


Figure 1. Mapping Workflow: It depicts how we have developed the global, European and Chinese patents datasets.

As previously performed, the mapping needs of appropriate keywords and definitions of PM, necessary to shape the query for the mining of the European Patent Office Database (Espacenet).

The **definitions** are:

1. **Personalised medicine:** a medical model using characterization of individuals' phenotypes and genotypes (e.g. molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention, which makes the 4P medicine concept (European Council Conclusion 2015/C 421/03).
2. **Precision medicine:** treatments targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations" (Precision medicine--personalized, problematic, and promising. Jameson JL, Longo DL N Engl J Med. 2015 Jun 4; 372(23):2229-34.)
Focus on process and used data: precision medicine as "a model that integrates clinical and other data to stratify patients into novel subgroups.
3. **Personalised health care & Precision public health:** the application of clinical know-how, concepts of systems medicine, and personalised medicine technologies to improve health and minimize disease.

The **keywords** to be used within the searches are personalised medicine, personalized medicine, precision medicine, preventive medicine, predictive medicine, systems biology, systems medicine, stratified medicine, targeted therapy, tailored treatment/therapy, deep phenotyping, omics sciences, big data, machine learning techniques.

The **established goal of PM** can be:

- Diagnostics
- Design of patient-specific therapy/treatment strategy
- Forecast of success of therapy/treatment strategy

In addition, it was assumed that the **tools** of PM are genetic analysis, statistical analysis, patient-specific treatment, patient-specific drug.

To eliminate non-coherent results with the definitions and approaches of PM, the search was refined and additionally filtered starting by assumptions related to what PM is and what it is not:

PM is:

- Drugs, devices, methods for patient-tailored treatment/therapy
- Procedures using an individualized approach based on genetic evaluation and treatment
- Treatment strategies based on individual data of patient genotype, phenotype, lifestyle
- Methods for synthesizing or designing customized and patient-specific drugs

- Methods for predictive medicine (e.g. statistical method for predicting life expectation, therapy response based on patient genotype, phenotype, lifestyle)

PM is not:

- Customized prosthesis/implants
- Medical instrument designed for specific group of patients (e.g. obese patients)
- Diagnostic, therapeutic and prophylactic approaches for a specific disease (and not for a specific patient)
- Genetic analysis for general purposes

Other fields excluded:

- Veterinary
- Personalised nutrition plans

The indications related to the national validations of the patents contained in the dataset allowed to divide the set in two parts, Europe and China. Differential data were presented together with the global set data.

2.2 Results of the update – Global patent set

The presented analysis has been carried out using documents from the ErreQuadro proprietary database, based on the data provided by the PATSTAT service based on European Patent Office (EPO) database (<https://www.epo.org/searching-for-patents/business/patstat.html>).

In some cases, data supplied by Chinese Patent&Trademark Office to EPO are partially omitted. For those documents, the geographic location of the address is retrieved using other proprietary tools developed by ErreQuadro. However, for a restricted number of documents, it is not possible to obtain a reliable geographical location. For such reasons, all the information based on geographical localization retrieved in this analysis can be underestimated. The presented outcomes overcome part of the blind period from the previous mapping, giving information on patents deposited by 2021. However, they are limited to 18 months before 31 December 2022.

The following statistics on patents in PM worldwide were obtained and compared with previous data:

Previous Mapping (2010-2020)	Parameter	Updated (2010-2022)
~73000	Patents	~119000
~12000	Patent families	~20000
94,5%	Precision	94,5%
High	Level of recall	High
6,1	Innovation Index	6,0

Table 1 - Comparison of the updated outcomes (global set) with respect to the results reported in D2.1 from 2010-2020.

Where:

- Precision (or positive predictive value) is intended as the fraction of relevant patents among the global patent set. This value should be high enough to grant the proper significance to statistical data retrieved from the set. Precision is evaluated by manually analysing a sample subset of patents;
- Patent families is the number used for classification;
- Level of recall define the fraction of relevant patents that were included in the set (total number of documents retrieved that are relevant/total number of relevant documents in the database).
- The Innovation Index (I.I.) represents a qualitative information on patenting strategies.

$$I.I. = \frac{\# \text{ patent}}{\# \text{ patent families}}$$

The number is obtained from the ratio between the global set of patents and the bibliographic data (DOCDB [document database]) families of a patent set, as shown in the formula below; the higher the level of extended patents, the higher the value generated by those patents. To give an idea of the indexing of various category, Machine Learning I.I. is 2,7 and turbines is 3,9. The number of patent families is more related to inventive activity. Compared to the previous analysis, the I.I. confirms the high tendency of PM to extend the patent beyond the region where the invention was generated. Nevertheless, while patents families maintain the tendency to slowly grow, there is a drastic reduction of the total patents number (data not shown) that led to a slightly minor decrease of the I.I PM from 6,1 to 6,0. The global trend of filed families is reported in Figure 2. The drop in the number during 2021-2022 is due to the 18 months blind period of the patents before publications.

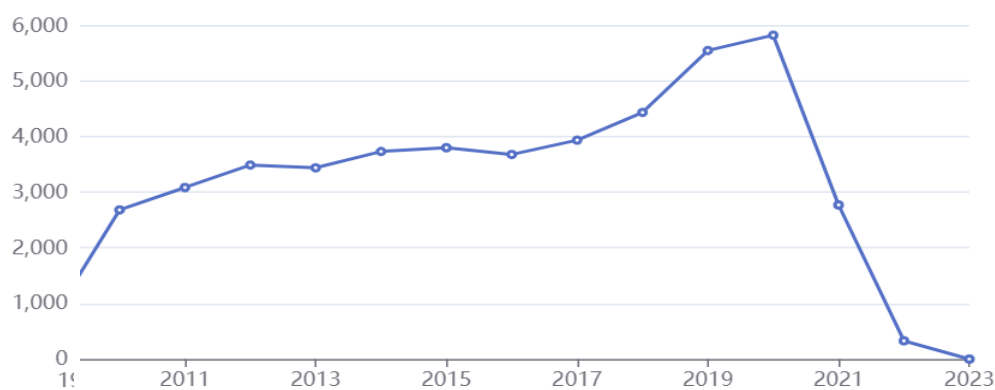


Figure 2. On the top, the global annual trend of filed patent families related to PM. The plot shows an increasing of deposited documents, the drastic drop depends on the 18months blind period that limits the information from 2021-2022.

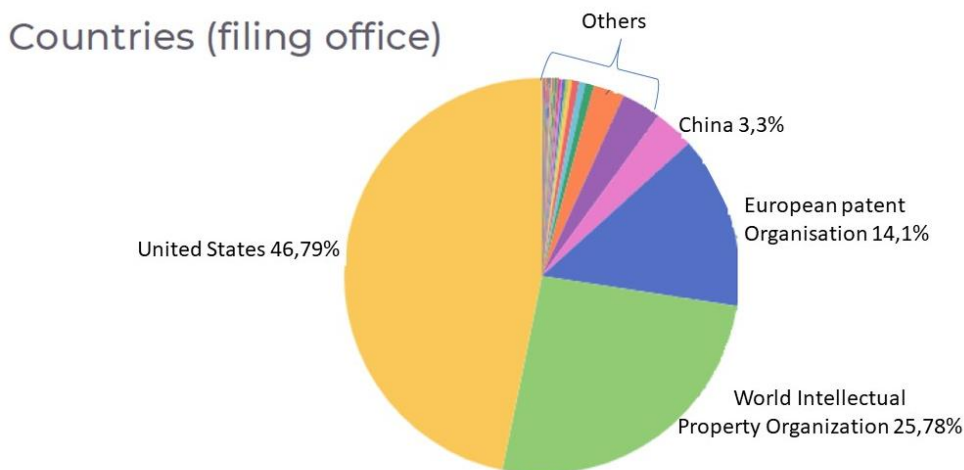


Figure 3. A pie plot showing the percentage of the global filed families for filing office. United States are depicted in yellow, European Patent Organisation in blue, China in pink while green represents World Intellectual Property Organization, a specialised agency of United Nation (UNO). Others are also defined.

Private held companies vs Public research institutes

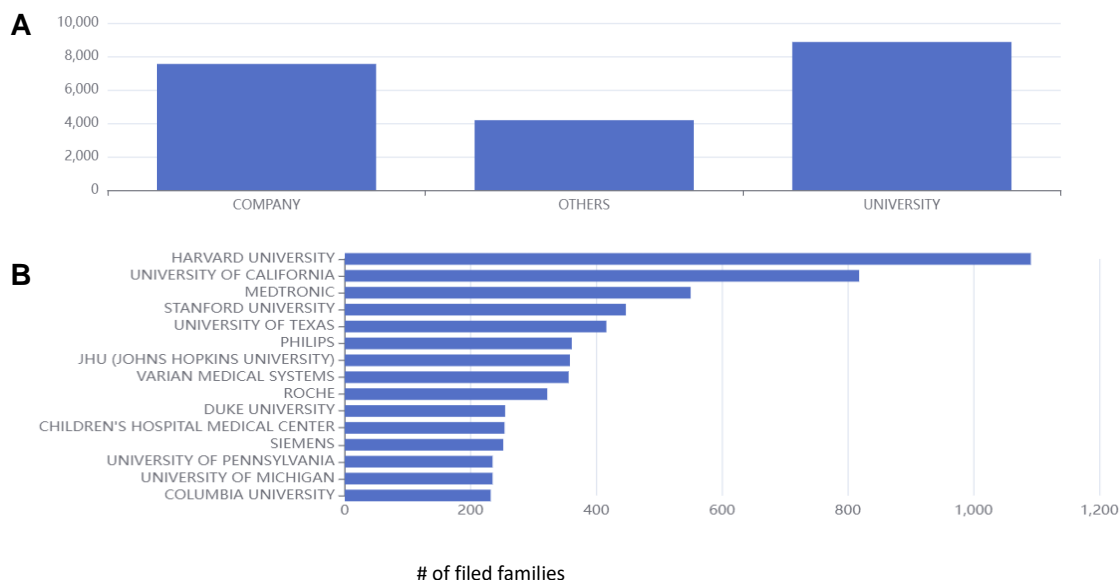


Figure 4 – The top plot (A) represents the distribution of patent activity (global set) among private and public institutions. The plot on the bottom (B) details the top #15 global assignees emerged from the analysis.

First of all, it should be pointed on the presence of the World intellectual property organisation (WPO or WIPO, www.wipo.int/portal/en/index.html) and the European patents organisation (EPO, www.epo.org/). WPO is a self-funded agency of the United Nations acting as global forum for intellectual property (IP) services, policy, information and cooperation. EPO is an international public organisation created by the European Patent Convention that works on national and international patent grants.

Among all the countries, United States retains top spot in global patent activity (46.79%), followed by Europe (14,1%) and China (3,3%). Interestingly, the patenting activity (which can be a primary indicator for innovation in PM) presents a turnaround and it is now lead by universities followed by industries and other categories (i.e., non-profit, hospitals, government, etc) as shown in Figure 4. This also emerges by the assignees analysis wherein the prevalence is pointed out. The top 5 rank shows only the main assignees: the Harvard University (US) followed by University of California (US), Medtronic (Ireland, EU), Stanford University (US), University of Texas (US) and Philips (NL) among others.

IP Classes

The International Patent Classification (IPC) was used to categorise and list the main classes for which patents are applied, in decreasing order, as shown in Table 2. IPC provides for a hierarchical system of language independent symbols for the classification of patents according to the different areas of technology to which they pertain [[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:International_patent_classification_\(IPC\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:International_patent_classification_(IPC))].

#	Class	# of families	# of patents	Class description	# rank in previous mapping
1	A61K	8491	50115	Human Necessities Preparations For Medical, Dental, Or Toilet Purposes	2
2	G01N	7416	38291	Physics Investigating Or Analysing Materials By Determining Their Chemical Or Physical Properties	1
3	C12Q	6750	33377	Chemistry; Metallurgy Measuring Or Testing Processes Involving Enzymes, Nucleic Acids Or Microorganisms	3
4	A61P	6295	29704	Human Necessities Specific Therapeutic Activity Of Chemical Compounds Or Medicinal Preparations	4
5	C12N	3877	17004	Chemistry; Metallurgy Nucleic acids, Microorganisms Or Enzymes ; Compositions Thereof	5
6	C07K	3538	21892	Chemistry; Metallurgy Peptides	6
7	G06F	2806	10124	Physics Electric Digital Data Processing	7
8	G16H	2722	8195	Physics Healthcare Informatics , I.E. Information And Communication Technology [Ict] Specially Adapted For The Handling Or Processing Of Medical Or Healthcare Data	-
9	A61B	2456	9369	Human Necessities Diagnosis; Surgery ; Identification	8
10	A61N	1225	5421	Human Necessities Electrotherapy; Magnetotherapy; Radiation Therapy; Ultrasound Therapy	-

Table 2. Global IPC Classes of Patent Filing from 2010-2022 global set

The last column of the table presents the 2010-2020 ranking. Classes **G16H** and **A61N** are new in the list replacing C40B (Combinatorial chemistry libraries, e g chemical libraries, in silico libraries) and G06Q

(Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory, or forecasting purposes systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes).

The **EU-China compared statistics** were obtained by filtering the global analysis; the EU/China attribution is based on the geolocation of the patent family assignee. In the previous work, third countries such as Canada and Israel were included in view of the important links and collaborations already established within the European PM initiatives. These Countries have been maintained in the update analysis.

For this reason, countries that are considered EU are:

- EU members ✓
- European Free Trade Association ✓
- European Economic Area ✓
- Canada ✓
- Israel ✓
- Balkan countries ✗ (other than Croatia, Greece ✓)

Included in STATS ✓

Austria, Belgium, Canada Cyprus, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Liechtenstein Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Romania, Slovenia, Sweden, Switzerland, UK

Excluded from STATS ✗

Albania, Andorra, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Georgia, Kosovo, Macedonia, Moldova, Monaco, Montenegro, Russia, San Marino, Serbia, Turkey, Ukraine, Vatican Cit.

An overview of the statistic is depicted in Figure 5. The green line represents European filed patent families while the Chinese are presented in blue.

Area	# of Families
Global SET	20123
EU Assignees	5860 (29,1%)
Chinese Assignees	906 (4,5%)

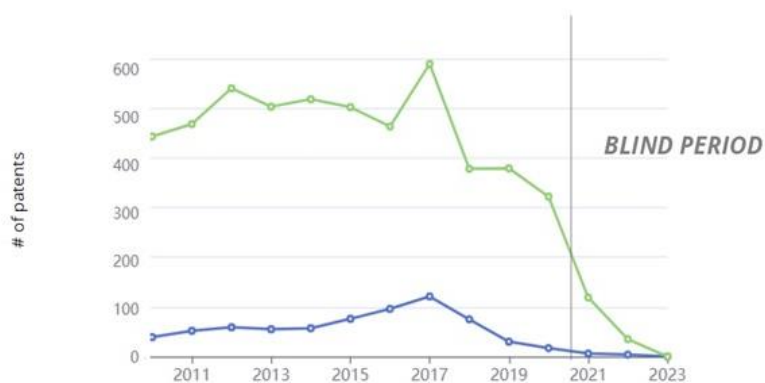


Figure 5 – Europe-China statistics obtained from the global updated set. The trend shows the European filed patent families in green and the Chinese trend in blue. The table presents the numbers of deposited families with the respect to the global set.

Patents can be filed by the European or Chinese headquarters of a non-EU/non-Chinese Company or can be coassignee to a non-EU company. Analysis of “filing” is presented in Figures from 6 to 9. Details are provided in the respective captions.

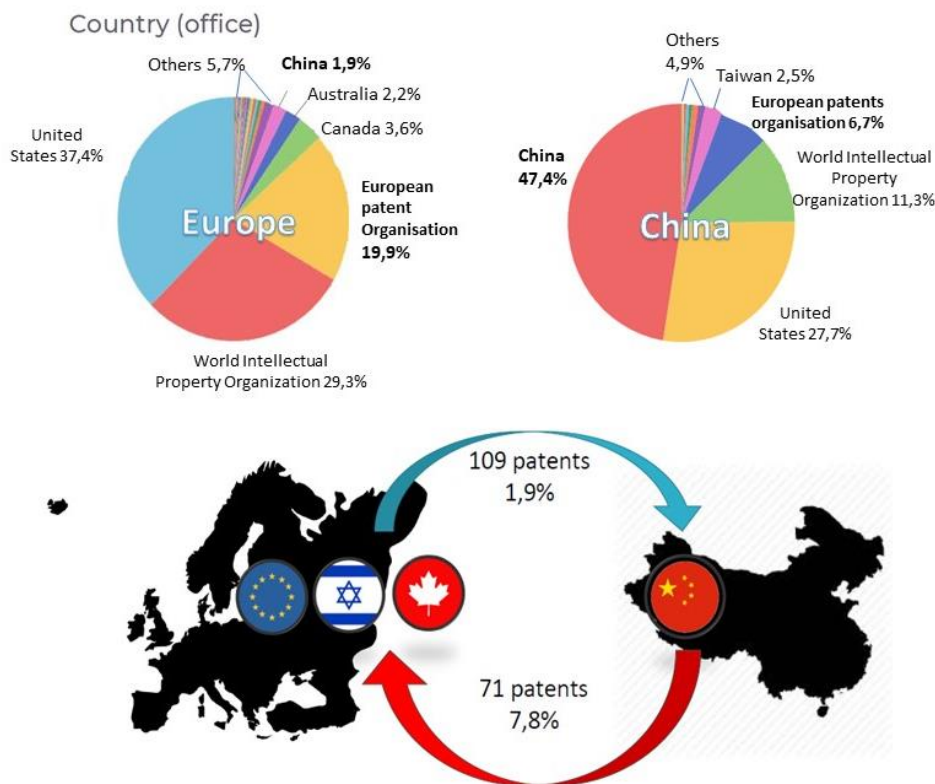


Figure 6. The pie plots representing the main filing offices for Europe (left) and China (right). Most of the EU patents are filed in US while almost half of the Chinese ones are deposited in China. Below, the percentage of filing of European patent families to China and vice versa.

Applications vs Granted Ratio - EU vs CN

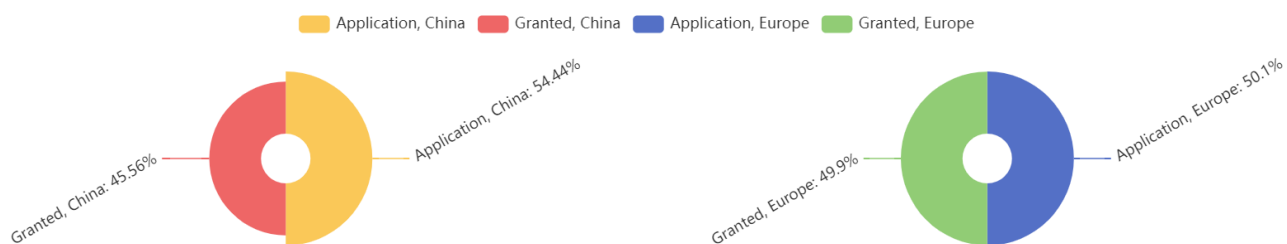


Figure 7. Percentage of granted patents vs applications for both China (left plot) and Europe (right plot)

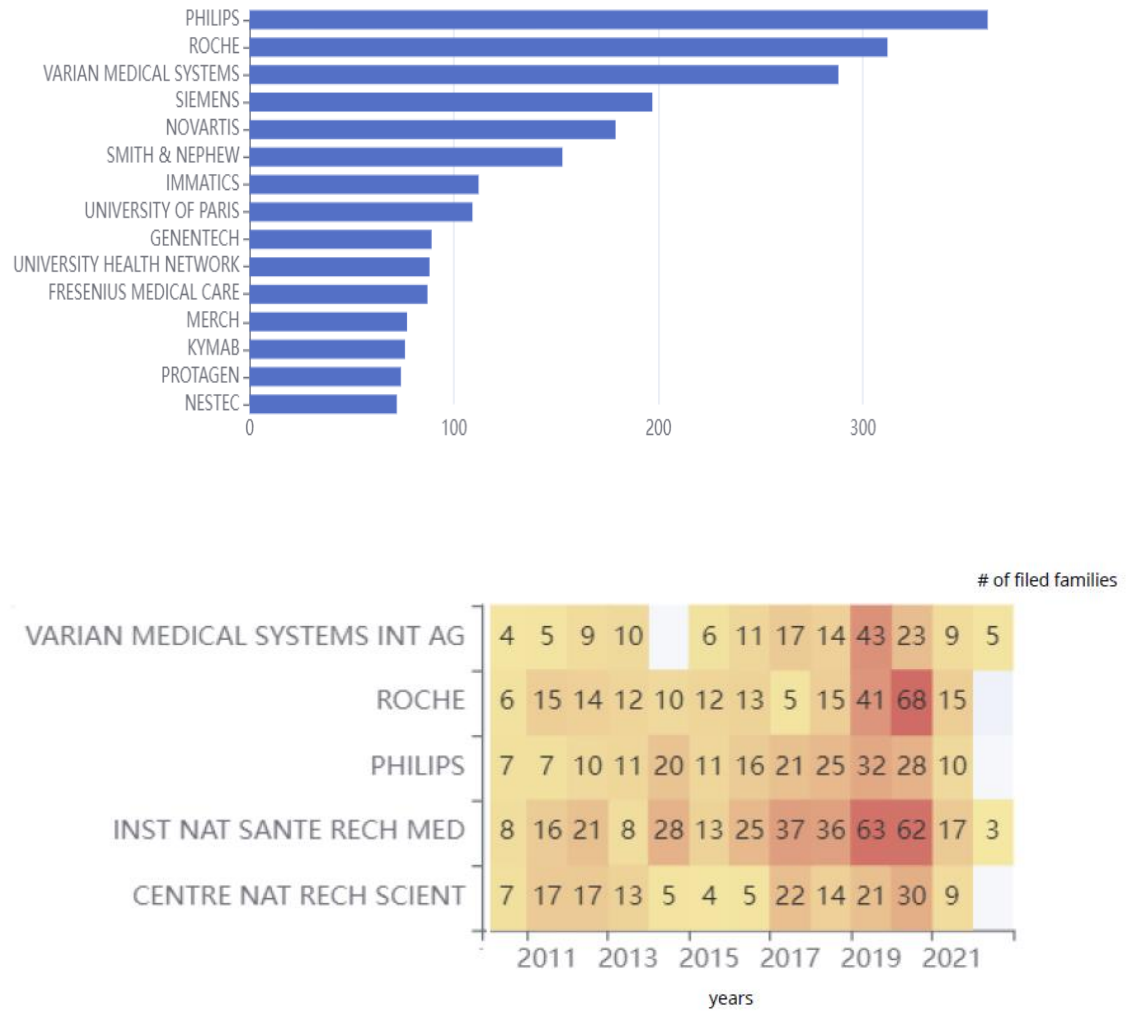


Figure 8. Main European Assignees (top) and filing per year (bottom). In the second graph, for each year, the number of filed patent families is shown.

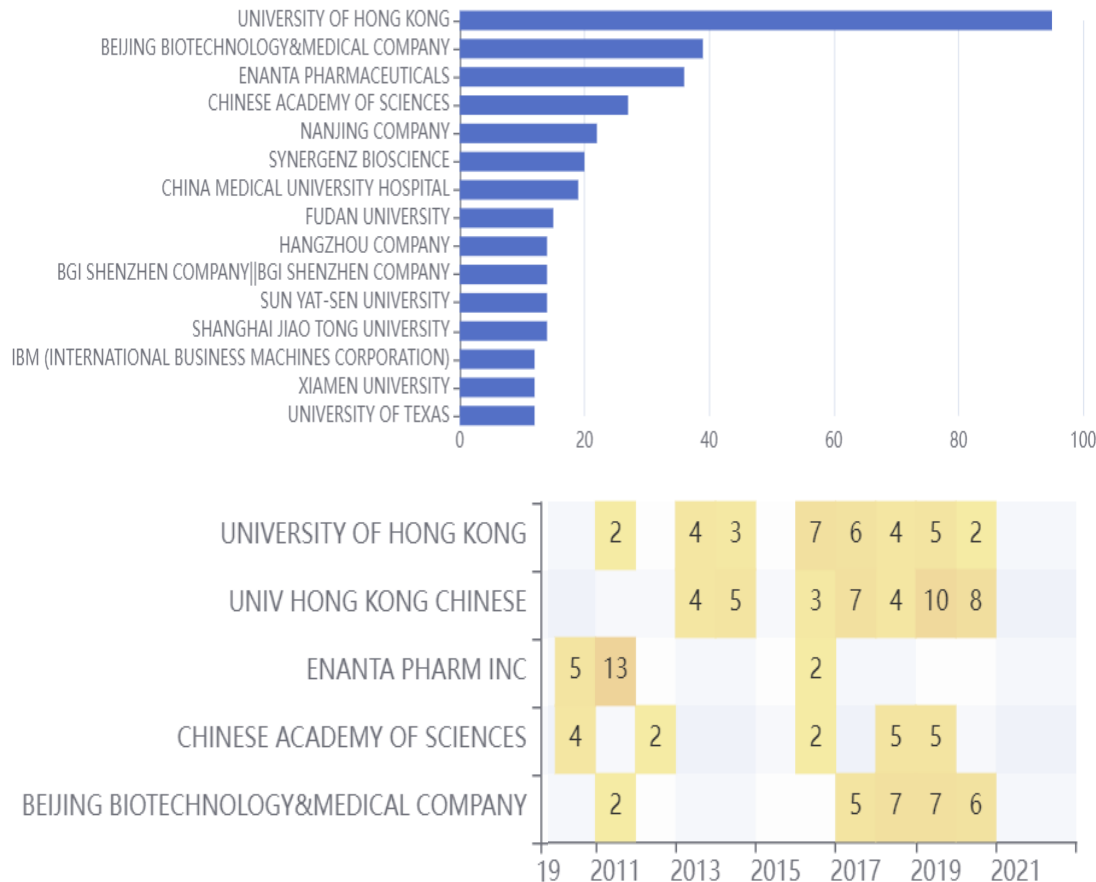


Figure 9. Main Chinese assignees (top) and filing per year (bottom). In the second graph, for each year, the number of filed patent families is shown.

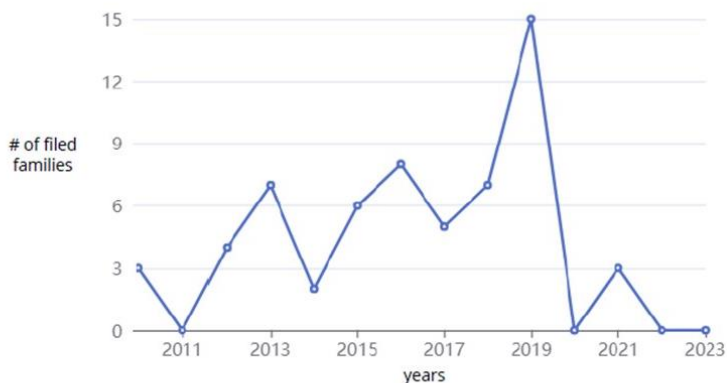
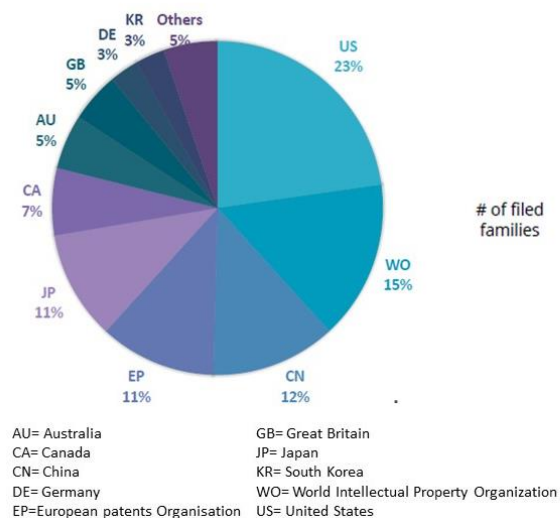
Companies are predominant in patent filing both for China and Europe even if the presence of academia is higher in China, confirming the information gathered from the previous mapping.

Key player are Philips, Roche, Varian Medical Systems and Siemens for Europe. In China, the strong presence of academia is manifested with the presence of University of Honk as the main assignee, followed by Beijing Biothec&Medical Company (Beijing), Enanta Pharmaceutial (US) and the Chinese Academy of Sciences (Beijing). It should be pointed out that many companies in China are owned by the state.

Collaborations are also analysed. In this regard it is necessary to discriminate the two definitions as follow:

- *CO-ASSIGNED*: patent having at least 1 Chinese assignee and at least 1 EU assignee
- *COLLABORATION*: patent of a family having at least 1 Chinese assignee and 1 EU assignee

in this framework, we found a total of **256** collaboration patents from **35** collaboration families, **116** additional documents to the previous dataset. Data trend is presented in Figure 10.



	Assignee	Families	TOT
1	IBM	11	64
2	Bio Merieux	3	23
3	Inserm	3	16
4	Pathway Pharmaceuticals	3	5
5	Lifshitz Susan Eve	3	4
6	Novartis	2	27
7	Adagene	2	16
8	Basf	2	15
9	CNRS	2	2
10	Hospices Civils De Lyon	2	2
11	Universite De Nice Sophia Antipolis	2	2
12	Shanghai Jiao Tong University	2	2
13	Universite Claude Bernard Lyon 1	2	2
14	Centre Hopitalier Universitaire De Nice	2	2

Figure 10. Statistics emerged from Collaboration patent documents analysis. The annual trend is shown in the upper part together with a pie plot which presents the filing offices (acronyms are defined at the bottom). The principal assignees are listed in the table

2.3 Differential data summary 2020-2022

- **Global set of families:** 7946
- **EU Assignees:** 2148
- **Chinese assignees:** 339
- **Collaboration families:** 9
- **Collaboration patents:** 77
- **Patent filing from EU to China:** 3.6%
- **Patent filing from China to EU:** 9.6%

3. Papers Mapping

3.1 Operative Flow

The same strategy used for the previous literature mapping was adopted for the update and schematically summarize below.

DATABASE: PubMed

QUERY STRATEGY: Hybrid approach, carried out merging two different set

Set 1) “precision medicine” and “personalized medicine” as MeSH terms (Untagged terms that are entered in the PubMed search box are automatically mapped to the MeSH vocabulary when a match is found.

Set 2) customized query, using keywords (Note: this set includes non-English written article only if at least one of title, abstract or keyword are available in PubMed in English)

QUERY FILTERS

- Human
- Years from Gen 2010 to Gen 2023 (included)

RESULTS

- **66 950** scientific papers (updated dataset)

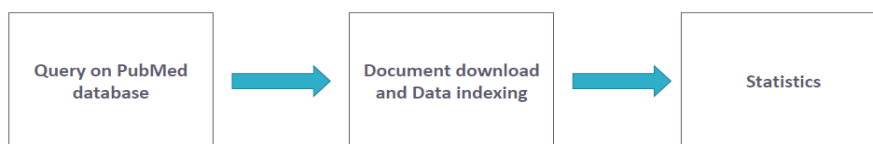


Figure 9. Workflow for the mapping of scientific publications

PubMed database provides different name for a single affiliation. Moreover, the name can be expressed in different formats. A database “normalisation” was then performed, to assign the number of submitted papers to a single affiliation. For instance, symbols and punctuation are deleted and some words as “centre” and “center” were standardized. In such way it is possible to assign a paper to affiliation even if the name of the affiliation itself is provided in a different format. However, the global number of papers assigned to each affiliation might be underestimated because of errors in database or complex name formats. January 2023 was included because of some submitted document which indicated a postponed publication date.

PubMed database could be partially updated with 2022 and 2023 publications. Additionally, some countries or some journals could provide articles to PubMed with a bigger delay.

The results that emerged from the query have been downloaded in as .xls format including information about authors, affiliation, year of publication, journal, volume, etc. More precisely, the global paper set has been downloaded in a Medline format and through EndNote (a commercial reference management software used to manage bibliographies and references when writing essays and articles) This programme outputs the information in Excel format. The information was then clustered in Global and EU/CN collaborations.

3.2 Results of the update - Global Publication set

The plotted global analysis data shows a constant growth with a peak in publication around 2019. This depends to the Pubmed's updating practice on which we have little information. Some documents deposited in 2019 were only retroactively uploaded after 2020 and this created an ambiguity that produces the scattered data shown in Figure 11. For this reason, it is necessary to recall the data from latter two years could be underestimated.

Previous Mapping (2010 – 2020)	Parameter	Updated (2010 – 2022)
~42000	Scientific Manuscripts	~67000
434	Collaboration papers	865
USA	Main affiliation	USA
Pharmacogenomics	Main Journal	Pharmacogenomics

Table 3. Comparison of the updated outcomes from scientific mapping (global set) with respect to the results reported in D2.1 from 2010-2020

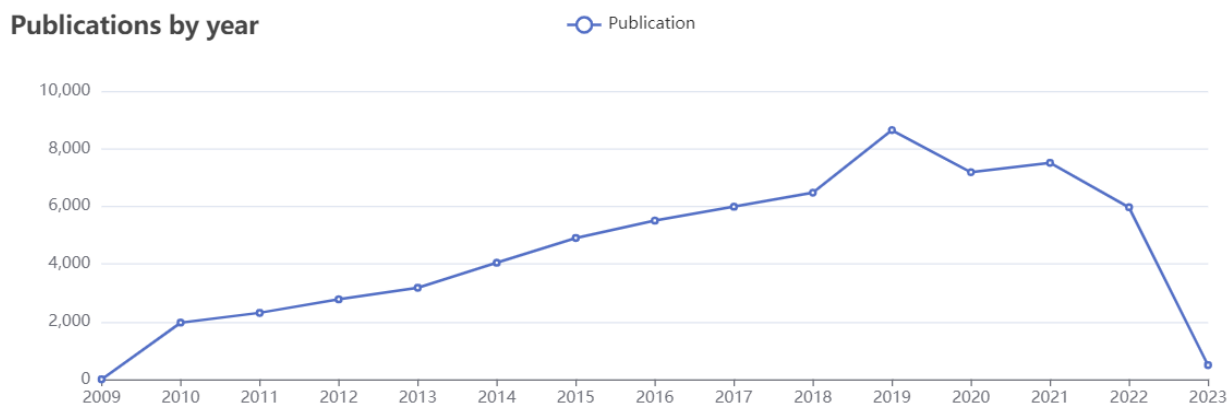


Figure 11. Global trend of Scientific publication on Personalised Medicine from 2010 to 2022

Analysis of assignees was conducted by considering a scientific paper is assigned to a country if at least an affiliation of that country is present. Now the global set presents a total of **66950** publications, **28289** by EU affiliations and **7306** by Chinese. The detailed list of countries is presented in the following tables 4A and 4B.

A		B		
Area	# of Articles	#	Country	# of papers
Global SET	66 950	1	USA	22689
European Affiliation	28 289	2	China	7306
Chinese Affiliation	7 306	3	United Kingdom	6120
		4	Italy	5007
		5	Germany	4627
		6	Canada	3384
		7	France	3313
		8	Netherlands	3264
		9	Japan	2579
		10	Spain	2575
		11	Australia	2381
		12	Switzerland	1682
		13	Sweden	1439
		14	Belgium	1174
		15	Denmark	1146

Tables 4A and 4B. Global updated dataset of Scientific mapping (A). rank for countries (B).

Compared to the recent mapping, while USA retains the top spot, China ranks second, Italy is in 3rd position, followed by UK and Germany. Overall, the scenario is not changed but slight variations could be appreciable while comparing the rank of top 20 global affiliations (together with the respective total number of deposited manuscripts), whose update is presented in table 5. At present, 50% of the affiliations are based in USA, which raises in the ranking to the first positions together with China, that is in now in 2nd position with the Department of Clinical Pharmacology from Xiangya Hospital, Central South University from Changsha. The latter leads the rank also having its Institute of Clinical Pharmacology in 4th position. This had a boost in terms of scientific publications on PM (from 151 publications [old mapping] to 405 papers), moving up the institution of three places.

Saudi Arabia drops down from 1st to 5th place while the Netherland and Japan dropped out the ranking. The new entries are Germany, Italy, Iran and Iceland.

#	Affiliation	# of papers	Country
1	Department of Pathology and Genomic Medicine, Houston Methodist Hospital, Houston, Texas.	452	USA
2	Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China.	414	China
3	Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.	406	USA
4	Institute of Clinical Pharmacology, Central South University	405	China
5	Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia.	390	Saudi Arabia
6	German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Germany.	373	Germany
7	Università Cattolica del Sacro Cuore, Rome, Italy.	355	Italia
8	Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.	330	Taiwan
9	Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.	275	USA
10	University of Medical Sciences, Tehran, Iran.	272	Iran
11	Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas.	271	USA
12	Faculty of Medicine, University of Iceland, Reykjavik, Iceland.	267	Iceland
13	Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas.	263	USA
14	Department of Molecular and Human Genetics, Baylor College of Medicine, Houston	257	USA
15	Department of Genetics, Stanford University, Stanford, California, USA.	222	USA
16	Department of Psychiatry, University of California San Diego, La Jolla, CA, United States	220	USA
17	Department of Psychiatry, University of Toronto, Toronto, ON, Canada.	216	Canada
18	Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.	213	USA
19	Department of Psychiatry, Harvard Medical School, Boston, MA, United States	209	USA
20	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.	201	UK

Table 5. Top #20 affiliations emerged from the global updated dataset.

The EU/CN compared statistics were obtained using the same group of EU countries used for the patent mapping and presented at page 13 of this report (Austria, Belgium, Canada Cyprus, Croatia, Czech

Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Liechtenstein Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Romania, Slovenia, Sweden, Switzerland, UK).

The differential global set shows a total of **25418** documents, 45,11% EU affiliations and 15,73% Chinese affiliations. The trend is also depicted in figure 12 and shows a slight but constant increase in the overall number of published papers. The raw data are filtered to highlight EU and CN affiliations.

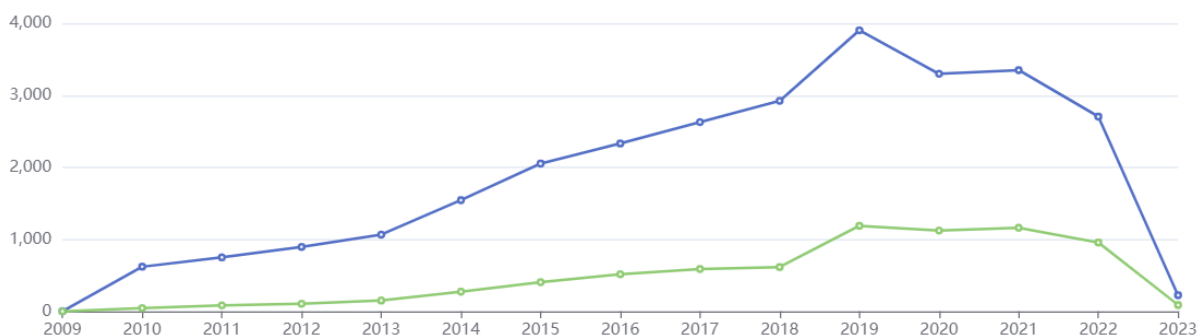


Figure 12. Global trend of EU (blue) and Chinese (green) scientific collaborations on PM

It is worth noting is not trivial to trace affiliations for scientific publications because of some issues related to the PubMed raw data. This allows to have a qualitative vision of the top affiliations. Numerosity has a degree of uncertain that must be considered.

The top 10 affiliations which are more active in scientific collaborations are listed in tables 6A and 6B.

Affiliation EU

A

#	Affiliation	# of papers	Country
1	German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Germany.	373	Germany
2	Università Cattolica del Sacro Cuore, Rome, Italy.	355	Italia
3	Faculty of Medicine, University of Iceland, Reykjavik, Iceland.	267	Iceland
4	Department of Psychiatry, University of Toronto, Toronto, ON, Canada.	216	Canada
5	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.	201	UK
6	Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands.	190	The Netherlands
7	Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.	180	Sweden
8	Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands.	178	The Netherlands
9	Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.	177	Finland
10	Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.	164	Denmark

B

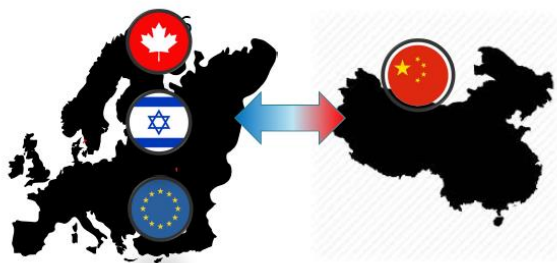
#	Affiliation	# of papers
1	Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China.	414
2	Institute of Clinical Pharmacology, Central South University	405
3	Institute of Clinical Pharmacology, Hunan Key Laboratory of Pharmacogenetics, Central South University, Changsha, China.	152
4	Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China.	108
5	The Key Laboratory of Carcinogenesis and Cancer Invasion of the Chinese Ministry	84
6	Department of Oncology, Shanghai Medical College, Fudan University, Shanghai,	78
7	State Key Laboratory of Oncology in South China, Collaborative Innovation Center	71
8	Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China.	51
9	Engineering Research Center of Applied Technology of Pharmacogenomics, China	36
10	College of Chemistry, Chemical Engineering and Biotechnology, Donghua University	27

Tables 6. (A) Main European affiliations collaborating with China on scientific publications. (B) Main Chinese affiliations collaborating with Europe on scientific publications.

The EU affiliations updated list has several differences to the old mapping. Regarding EU affiliations, Germany, Italy and Iceland are in the top three positions overtaking UK from 1st to 5th and Netherlands, now 6th and 8th. For Germany, main affiliations are German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) with 256 new papers. Università Cattolica del Sacro Cuore, Rome, Italy (total of 355 publications) and (total of 257 publications) are also new entries along with Denmark, which ranks 10th with the Faculty of Health and Medical Sciences, University of Copenhagen (total of 164 papers). Regarding Chinese affiliation, the predominance of the Central South University from Changsha appreciated from the global mapping, is well reflected also in this rank. The top 3 is led by Department and Institute of Clinical Pharmacology which, overall, duplicated its publications. The Wenzhou’s Institute of Genomic Medicine (Wenzhou Medical University) drops from 3rd to 4th position, but it presents just 2 new publications. A new entry that is interesting to highlight due to the total high number of publications (84) is The Key Laboratory of Carcinogenesis and Cancer Invasion of the Chinese Ministry. Collaborations in scientific collaborations between Europe and China were obtained by considering papers that have at least 1 co-author with a Chinese affiliation and at least 1 co-author with a European affiliation.

A total of **865** collaboration papers (1.3 % of global set) emerged:

- 11.8% of Chinese papers
- 3.0% of European papers



#	Country	# of collaboration
1	UK	265
2	Germany	203
3	Canada	167
4	France	123
5	Italy	112
6	Netherlands	114
7	Sweden	53
8	Denmark	59
9	Spain	78
10	Israel	51
11	Finland	37
12	Switzerland	57
13	Belgium	54

Figure 13. A table containing the main EU Countries who publish in collaboration with China.

Compared to previous data, even if the percentage of collaboration papers has increased significantly in number (from 434 to 865, with a total of 418 new joint publications during 2020-2022), the scenario is unchanged. EU/CN collaboration papers are the 1,3% of the global set (+0,2%); there is a small drop of Chinese papers percentage (-3,3%) together with a little increase of European papers (0,4%). However, considering the small entity of the variations, any comparison could be speculative.

About the Scientific journals, many relevant Journal are present, such as *Nature* and *PloS one*. The *Journal of Pharmacogenomics* (2436) and *PLoS One* (1035) maintain the first positions while *Scientific reports* has double the number of publications (+405), ranking to the third position with 785 papers. Table 7 presents the ranking of the top 10 journals. It would be worth noting the list does not present special journals dedicated to PM.

#	Journal	# of papers
1	Pharmacogenomics	2436
2	PloS one	1035
3	Scientific reports	785
4	International journal of molecular sciences	614
5	Nat Commun	401
6	Clin Pharmacol Ther	395
7	Oncotarget	315
8	Nature	298
9	Theranostics	278
10	BMC cancer	240

Table 7. The Rank of scientific journals emerged after the update.

3.3 Differential data summary from 2020-2022update

- **Global set of scientific publication:** 25418
- **EU Assignees:** 11467
- **Chinese assignees:** 3999
- **Collaboration papers:** 418 (1.6% of global set, 10.4% of EU set, 3.6% of CN set)

4. Preprints Mapping

Peer-review is a fundamental process designed to assess the integrity and the consistency of a scientific papers, a mandatory step for publication in scientific journals, regardless of their prestige. While unavoidable, it can often significantly slow down publication time due to various biases and problems that will not be addressed here because they are beyond the scope of this report. The consequences, however, have a strong impact on scientific communication, decreasing its efficiency and prompting many researchers to find alternative strategies to speed up the dissemination of their results. The need triggered the embracing of the use and spread of non-peer reviewed articles, the so-called **preprints**, as an accepted dissemination practice. Preprints were familiar in different field of research since early 90s, particularly physics, and indeed the first established server arXiv.org begun in 1991 and now counts more than 18000 submission per months, with a total number of 2,249,493 (as of May 4, 2023, source:

https://arxiv.org/stats/monthly_submissions). Since the last decade, life sciences researchers also had back of special dedicated databases which growth was boosted during the recent pandemic because of the urgency of sharing potentially benefitting results as quick as possible (*Clare Watson, "Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever", Nature Medicine, 2022*).

On these grounds and according to the same workflow strategy adopted for the scientific mapping and presented in paragraph 3.1, the mining of some specific databases, bioXriv and medXriv, was included in the update. The objective is to retrieve the possible slice/part of literature of interest that has not yet been published but is present in the form of preprints. It must be borne in mind that the lack of expert supervision may affect the integrity of the preprints, which must therefore be used critically.

A total of **1682** items on PM were collected. Unfortunately, due to the difficulties of handling this raw data, it was not possible to perform accurate statistics. Nevertheless, the results are included in the web platform and accessible for consultation.

4. SINO-EU PerMed web platform

The differential data obtained from the mapping update and shown in this report were integrated into the preliminary set presented in D2.1 and will be available on the online platform. Through this tool, which will be interactive, it is possible to access and consult the complete information or filter data according to chronological criteria, geolocation or medical topic addressed. It is also possible to view the plots of the results statistics and download the related images, including all the graphs presented in this document.

The platform will be accessible through a dedicated section of the project web page <https://www.sino-eu-permed.eu/>