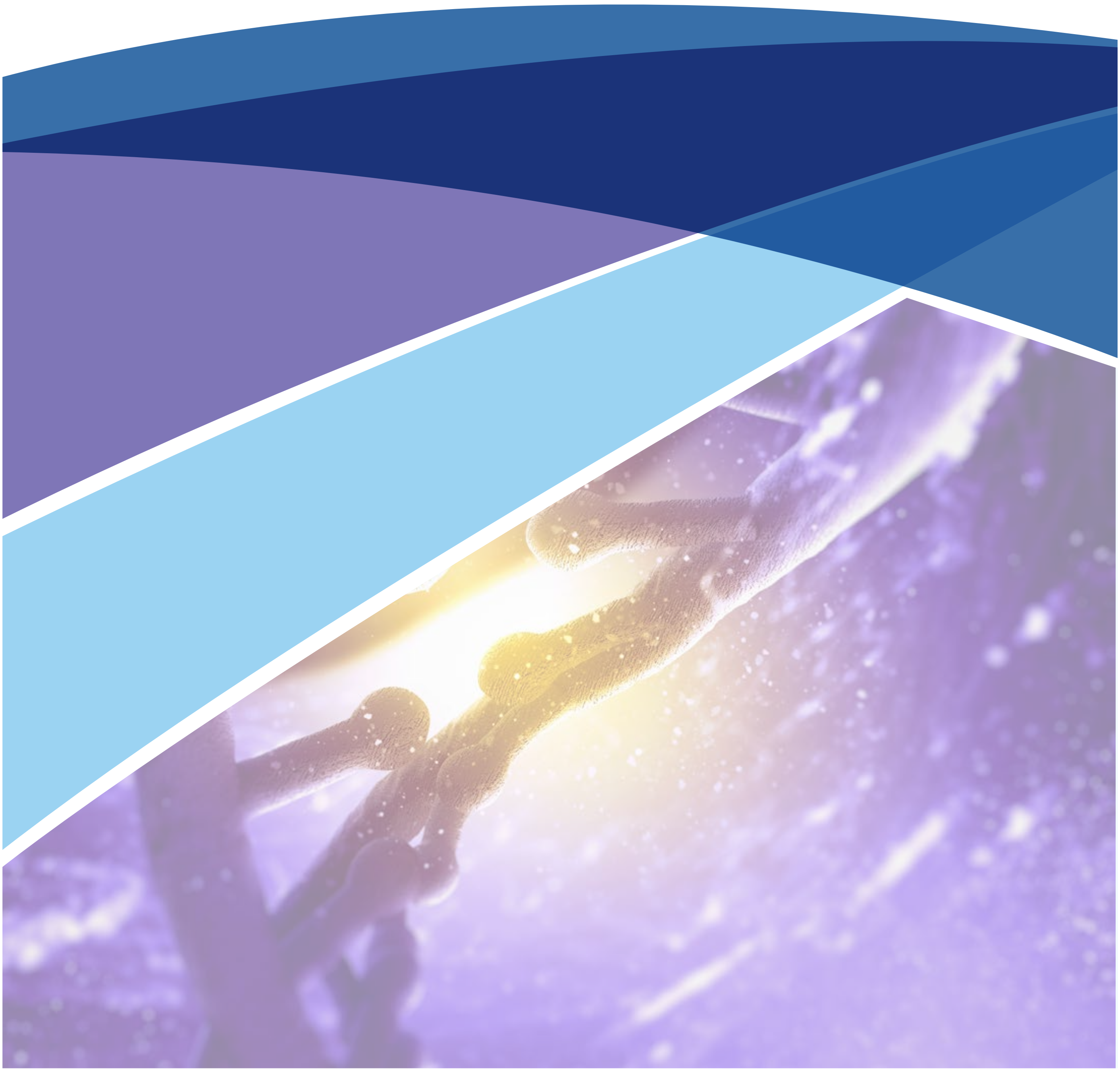
Intellectual Property Office

# Eight Great Technologies

## Life Sciences, Genomics and Synthetic Biology

### A patent overview



#8Great

This report was prepared by the
UK Intellectual Property Office Informatics Team
July 2014

e-mail: informatics@ipo.gov.uk



Intellectual Property Office
Concept House
Cardiff Road
Newport
NP10 8QQ
United Kingdom

www.ipo.gov.uk/informatics

# Contents

# 1 Introduction

The UK Government has identified ‘eight great technologies’ plus a further two which will propel the UK to future growth. These are:

- the big data revolution and energy-efficient computing;
- satellites and commercial applications of space;
- robotics and autonomous systems;
- life sciences, genomics and synthetic biology;
- regenerative medicine;
- agri-science;
- advanced materials and nanotechnology;
- energy and its storage;
- quantum technologies;
- the internet of things.

Patent data can give a valuable insight into innovative activity, to the extent that it has been codified in patent applications, and the IPO Informatics team is producing a series of patent landscape reports looking at each of these technology spaces and the current level of UK patenting on the world stage. As an aid to help people understand the eight great technologies and to consider the direction of future funding, the IPO is offering a comprehensive overview of what is already patented in the each of these technologies and in which direction the technology is developing. This information should not be taken as a direct measure of the level of innovation in the UK; it should be considered in conjunction with other sources of information to form a fuller picture.

This report gives an analysis of the worldwide patent landscape for life sciences, genomics, and synthetic biology. The scope of the search strategy included preparation of mutants and associated screening processes, hybrid cells, and all aspects of recombinant DNA technology. The dataset used for analysis was extracted from worldwide patent databases following detailed discussion and consultation with patent examiners from the Intellectual Property Office who are experts in the field and who, on a day-to-day basis, search, examine and grant patent applications relating to life sciences.

This report is based on analysis of published patent application data and not granted patent data. Data for published patent applications gives more information about technological activity than the figures for granted patents because a number of factors determine whether an application ever proceeds to grant. These include the inherent lag in patent processing at national IP offices worldwide and the patenting strategies of applicants who may file more applications than they ever intend to pursue.

# 2 Worldwide patent analysis

## 2.1 Overview

Table 1 gives a summary of the worldwide dataset used for this analysis of life sciences, genomics, and synthetic biology. The worldwide dataset was limited to applicants having a portfolio of at least ten patent families, with a publication date range of 2004 to 2013. This dataset contains over 500,000 published patents comprising more than 80,000 patent families. Publications may be at the application or grant stage, so are not necessarily granted patents. A patent family is one or more published patents originating from a single original (priority) application. Analysis by patent family more accurately reflects the number of inventions present because generally there is one invention per patent family, whereas analysis by raw number of patent publications inevitably involves double counting because one patent family may contain dozens of patent publications if the applicant files for the same invention in more than one country. Hence analysis by patent family gives more accurate results regarding the level of inventive activity taking place.

**Table 1: Summary of worldwide patent dataset for life sciences, genomics, and synthetic biology**

| | |
|---|---|
| **Number of patent families** | 81,379 |
| **Number of patent publications** | 501,716 |
| **Publication year range** | 2004-2013 |
| **Peak publication year** | 2012 |
| **Top applicant** | Pioneer Hi-Bred (US) |
| **Number of patent assignees** | 47,591 |
| **Number of inventors** | 144,827 |
| **Priority countries** | 40 |
| **IPC sub-group** | 9,905 |

Figure 1 shows the total number of published patents by publication year (top) and the total number of patent families by priority year (bottom – considered to be the best indication of when the original invention took place). The patent family chart in red does not show any patents filed after 2011 because a patent is normally published eighteen months after the priority date or the filing (application) date, whichever is earlier. Hence, the 2012 and 2013 data is incomplete and has been ignored.

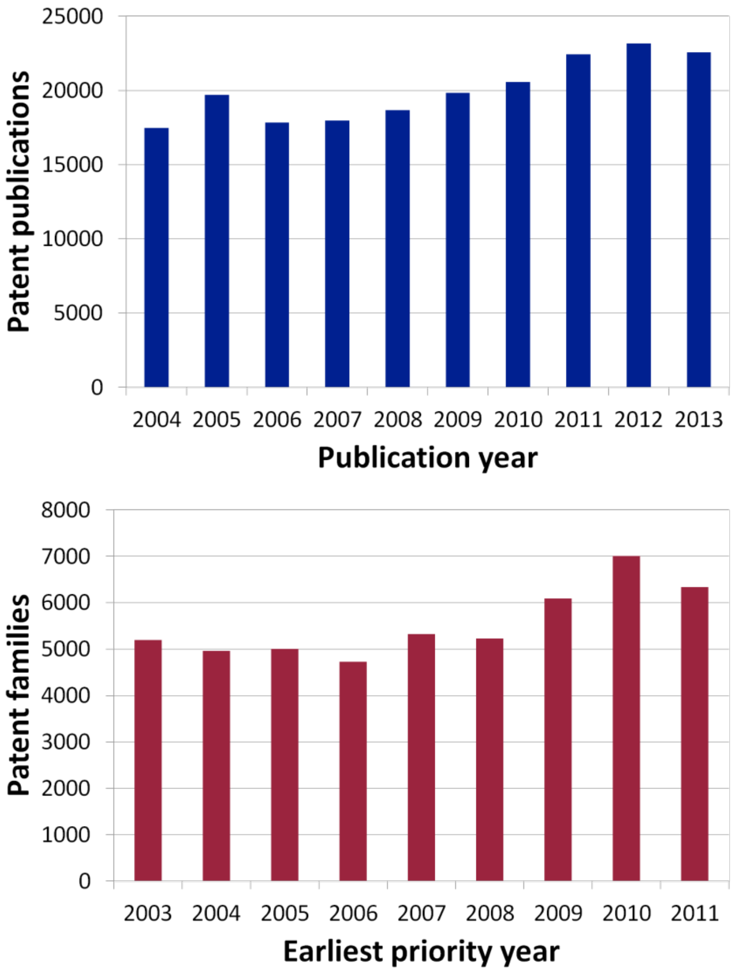



**Figure 1: Patent publications by publication year (top) and patent families by priority year (bottom)**

The number of publications dropped briefly from the local peak of 2005 and has grown throughout the last decade, but has levelled off and is possibly falling again most recently. When compared to the overall levels of patenting globally for all subject matter (Figure 2), this fall is more pronounced against a generally increasing trend (3.2% and -2.6% for the last two years compared to 12.7% and 8.6% respectively).

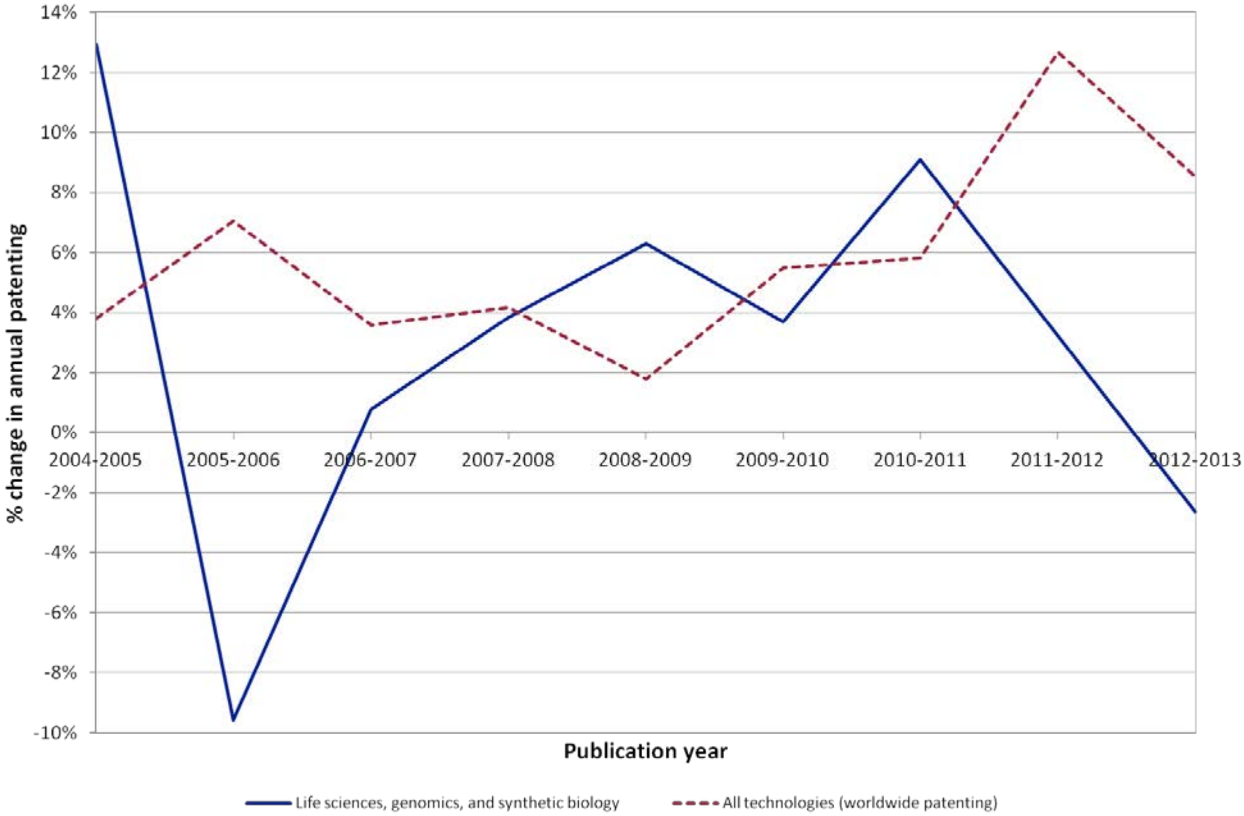



**Figure 2: Year-on-year change in life sciences, genomics, and synthetic biology patenting compared to worldwide patenting across all technologies**

Figure 3 shows the priority country distribution across the dataset. 44% of patent families have a patent first filed in the US. The US, Japan, and China together account for more than three quarters of first filings. Patent applicants frequently file an application first in their country of residence[1], but there may be other strategic reasons for an applicant choosing a different country of first filing (*e.g*. tax treatment).

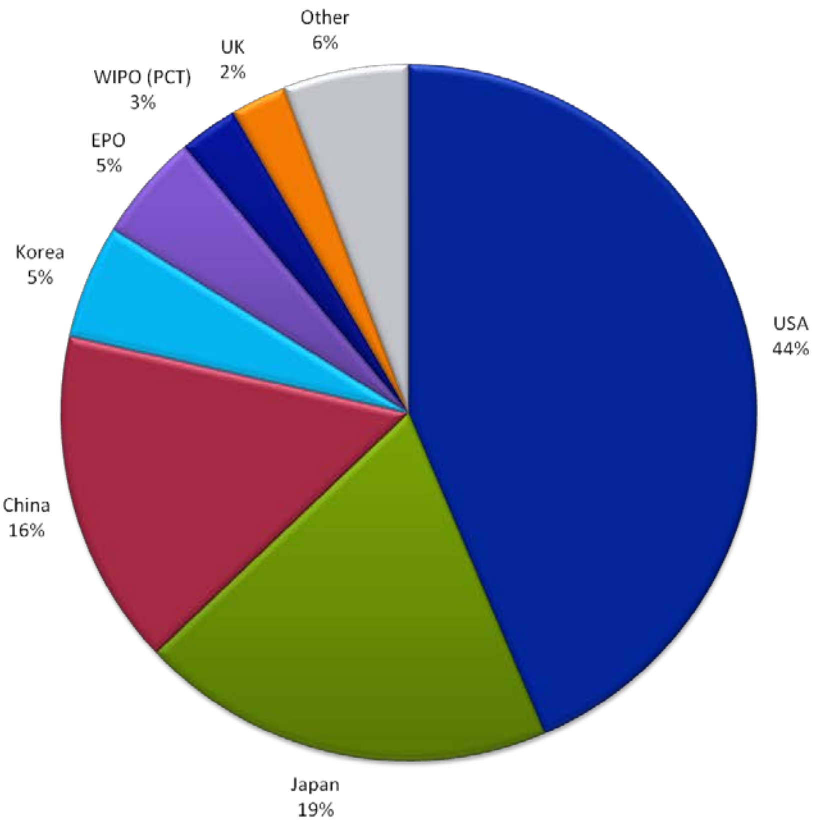



**Figure 3: Priority country distribution**

[1]In some countries this is/was a requirement (e.g. in the UK this was a requirement until 2005).

The applicant country distribution shown in Figure 4 demonstrate a similar distribution, this time with the US, China, and Japan accounting for just under three-quarters of all patent applicants. However, there is more diversity since applicants may have an address anywhere in the world but are likely to make applications in the countries with the most important markets. Note that EPO and WIPO[2] appear as "priority countries" but cannot appear as applicant countries since that information is derived from the applicant address.

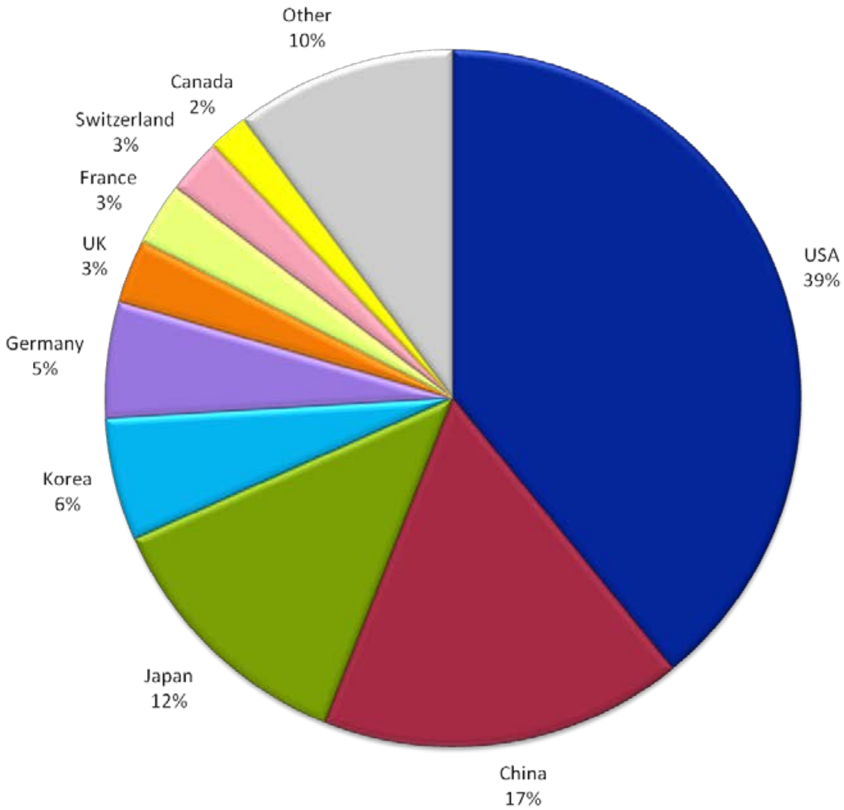



**Figure 4: Applicant country distribution**

There is a generally higher tendency to make patent applications in certain countries. A Relative Specialisation Index (RSI)[3] for each applicant country has been calculated to give an indication of the level of invention in life sciences, genomics, and synthetic biology for each country compared to the overall level of invention in that country, and is shown in Figure 5.

---

[2] Alternative filing routes, as outlined in Appendix A.3.
[3] See Appendix B for full details of how the Relative Specialisation Index is calculated.

Figure 5 indicates that Belgium, Israel, Australia, and Singapore have the most positive specialisation in this field. Taiwan has a very low specialisation. The UK has a small positive specialisation, indicating that patent applicants based in the UK have a slightly greater tendency to patent life sciences, genomics, and synthetic biology related inventions compared to other types of inventions in the UK.

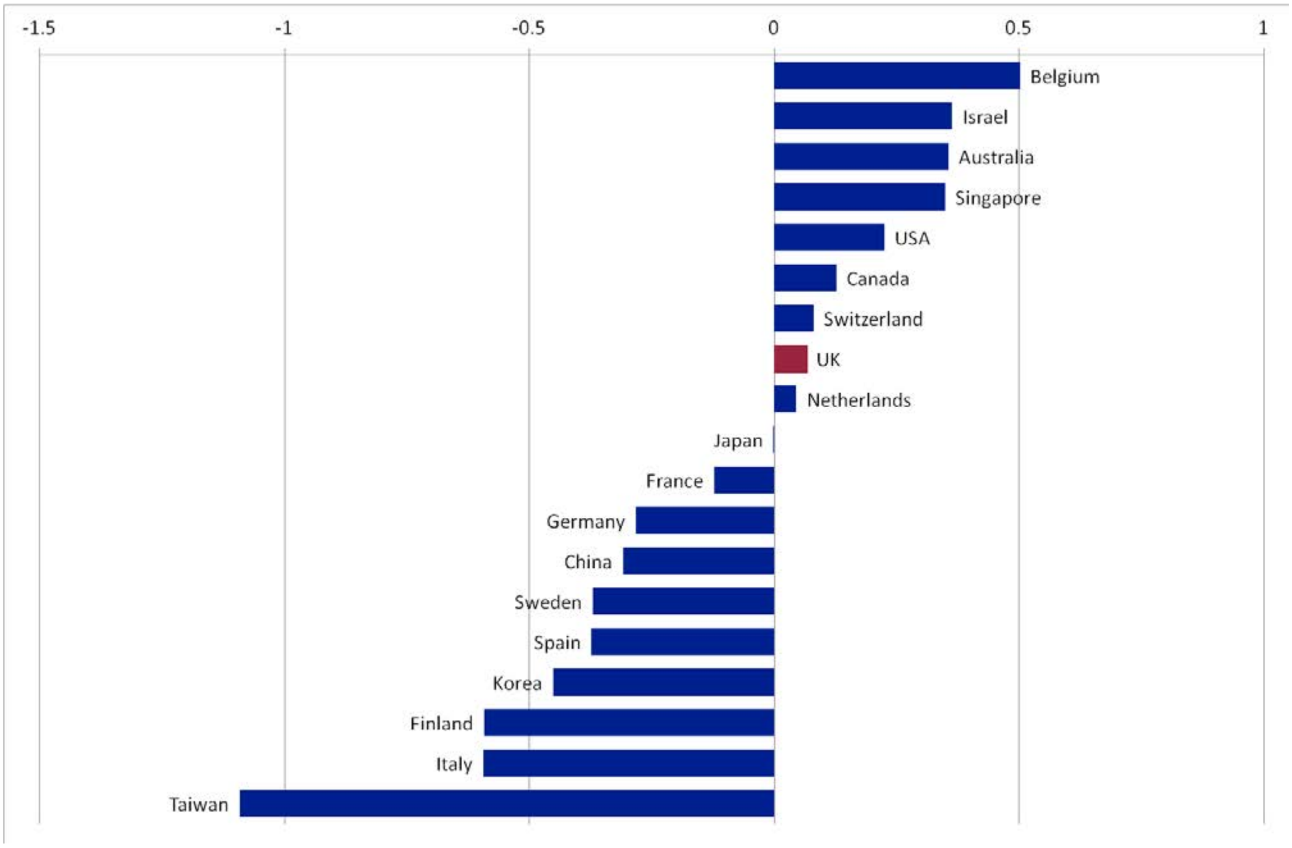



**Figure 5: Relative Specialisation Index (RSI) by applicant country**

Figure 6 shows the countries in which applicants in the field of life sciences, genomics, and synthetic biology are interested in seeking patent protection, with the strength of colour reflecting the quantity of publications in each jurisdiction. Patents filed via the EPO [] and WIPO (PCT) [] routes are also shown. The strongest coverage is through WIPO, indicating that inventions in life sciences, genomics, and synthetic biology are frequently patented across the world. Note, however, that the dataset excludes smaller organisations that may have a less international approach. Coverage in European countries directly is rather low, but these are covered indirectly by the EPO route.

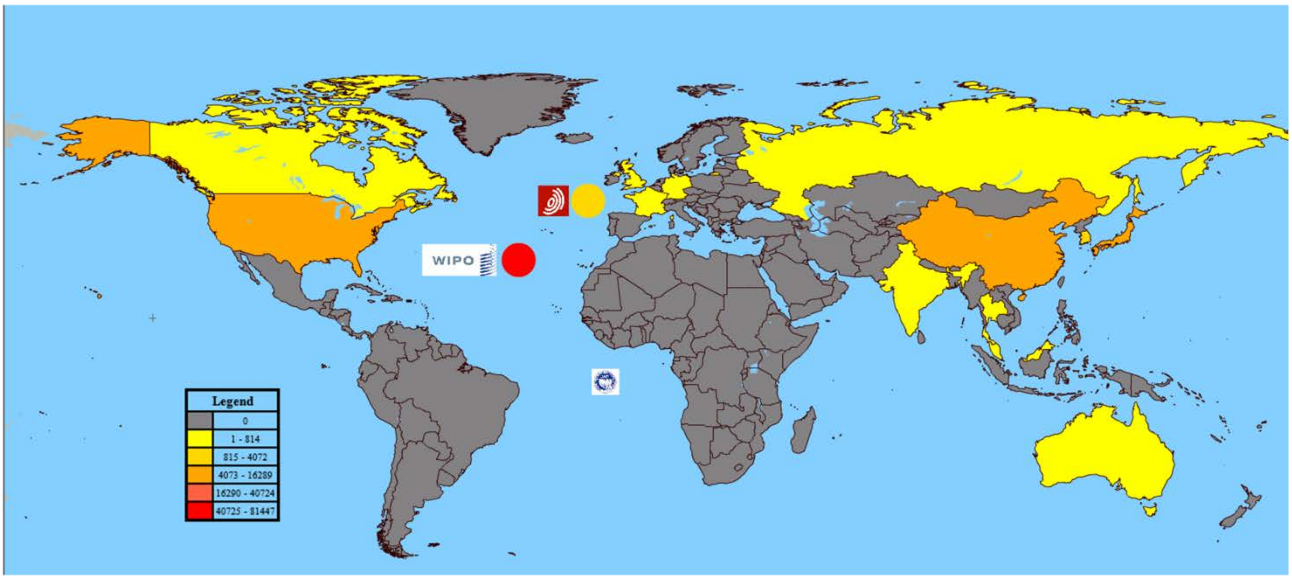



**Figure 6: Patent coverage (publication country coverage)**

## 2.2 Top applicants

Patent applicant names within the dataset were cleaned to remove duplicate entries arising from spelling errors, initialisation, international variation and equivalence[4]. Figure 7 shows the top twenty applicants in the life sciences, genomics, and synthetic biology dataset.

The three top positions are occupied by US organisations: Pioneer Hi-Bred with 1667 patent families (now owned by Du Pont, who themselves occupy eighth position with 633 patent families), closely followed by Monsanto with 1610 patent families, and the University of California with 1059 patent families. Although many US organisations appear in this list, and occupy four of the top five positions, the nationalities in the list are quite varied, additionally including Japan, Germany, China, France, Switzerland, and Korea. The leading UK applicant, Glaxo Group, holds 229 patent families. Eight of these twenty leading applicants are universities, research institutes or agencies, reflecting the science-intensive nature of the technology.

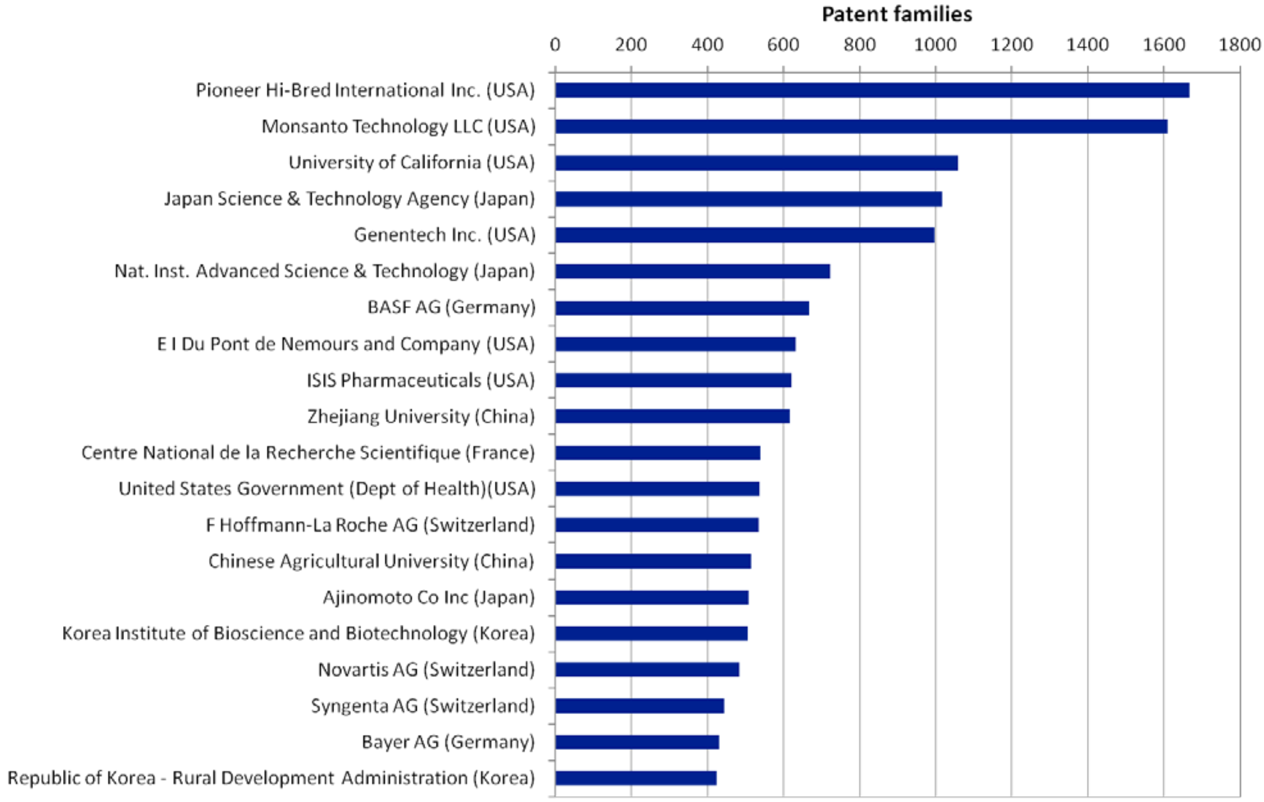



**Figure 7: Top applicants**

---

[4] See Appendix A.4 for further details

Figure 8 is a bubble map showing a timeline for the top twenty applicants and shows the filing activity of these applicants in the last ten years. It shows that most of the top applicants have been involved in life sciences, genomics, and synthetic biology patenting throughout the period, although some applicants have grown and others have not. The Chinese Agricultural University is absent from the chart until 2008 and has grown strongly since then. Monsanto and Pioneer Hi-Bred have also seen substantially increased publications since 2004, whereas the Japan Science and Technology Agency has seen reduced publications.

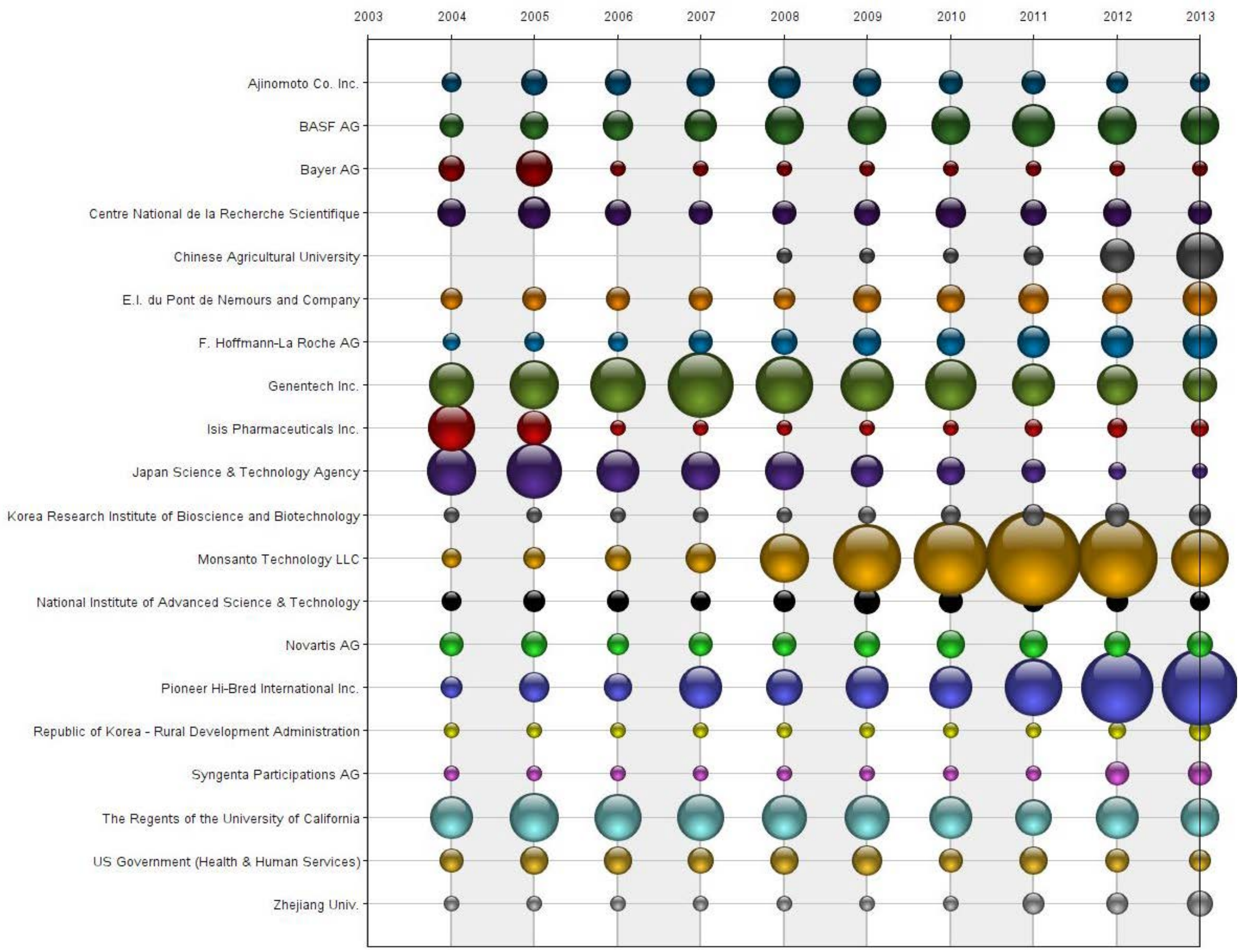



**Figure 8: Applicant timeline of publications by publication year**

## 2.3 Collaboration

Figure 9 is a collaboration map showing main collaborations between the top five applicants in the dataset (the top five shown in Figure 7) and their collaborators. The collaborations form two major clusters, one of which is geographically centred mostly in the US and the other in Japan, with minimal connection between the two. In the US, Monsanto has a large shared portfolio with Stine Seed Farm, as well as a further, larger portfolio of its own. Du Pont and Pioneer Hi-Bred also show lots of collaboration with each other. La Roche, a Swiss firm, is the only non-US organisation in this cluster. Most organisations in this cluster are companies. The University of California forms the sole link to the second cluster, based in Japan, having collaborations with the Institute of Physical and Chemical Research, the Japan National Institute for Infectious Diseases, and Riken. In contrast with the first cluster, this cluster includes mostly non-commercial organisations, and contains universities, national institutes, and agencies.

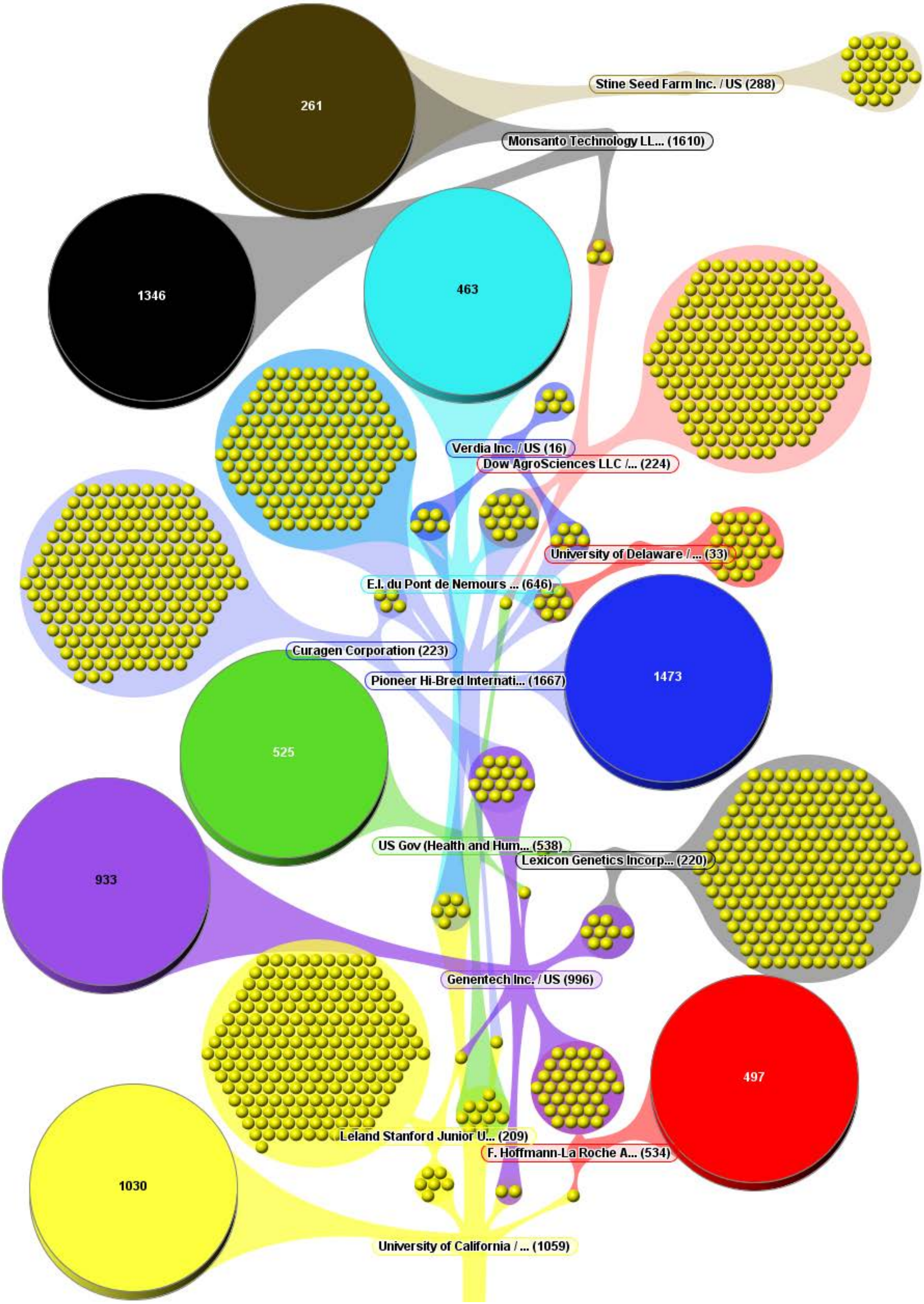



**Figure 9: Collaboration map showing collaborations of the top ten applicants and their collaborators**

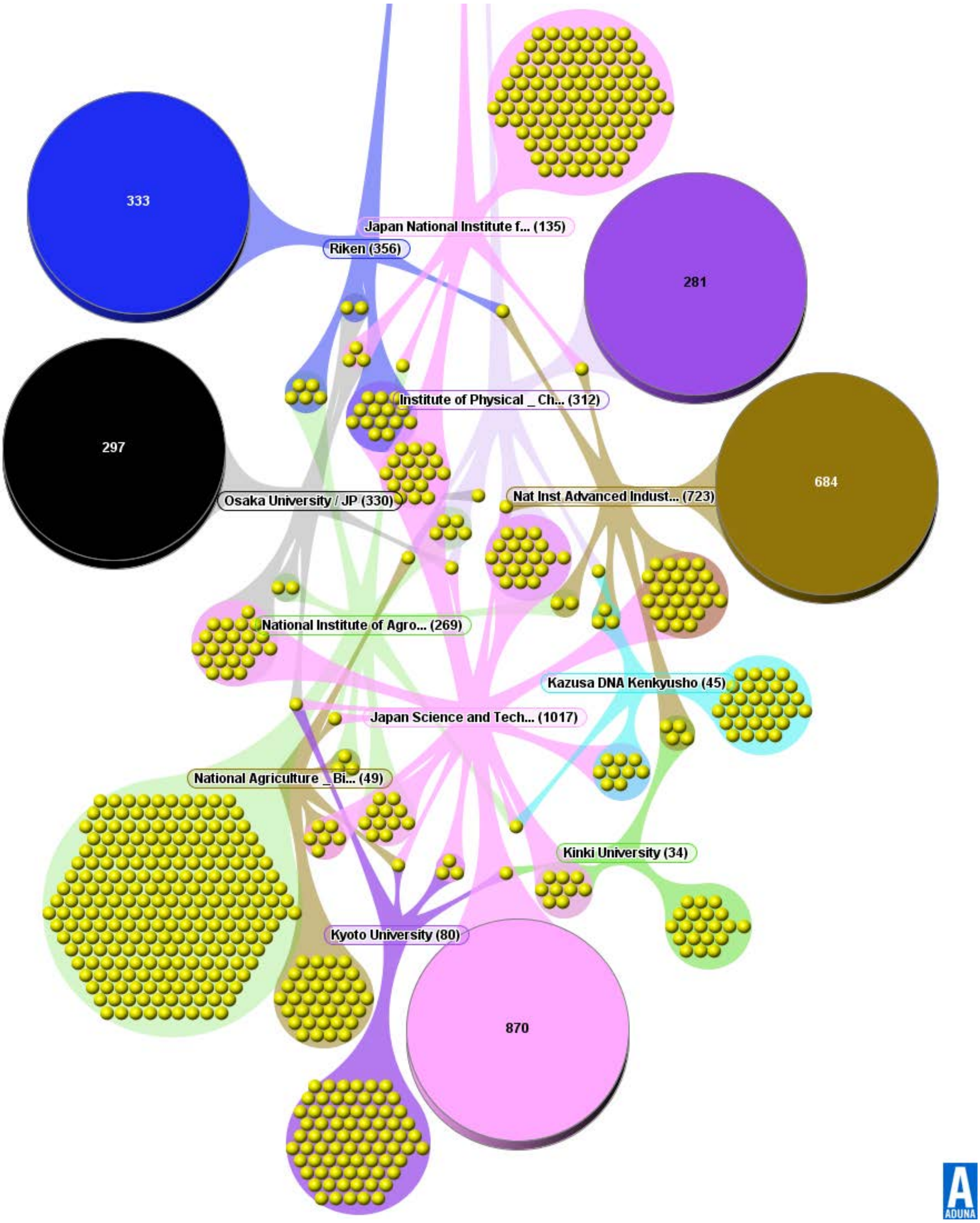



**Figure 9 (cont.): Collaboration map showing collaborations of the top ten applicants and their collaborators**

Figure 10 shows collaborations for UK leading applicants and their collaborators. This dataset shows a significant degree of collaboration, which forms a complex web of interactions between UK universities and companies, and overseas universities and companies. There are a few primary nodes immediately apparent within the web: the Medical Research Council, Glaxo Group, Isis Innovation, Astrazeneca, and Imperial

College. Nevertheless, these are all connected to each other directly or indirectly (via a smaller third party). It is clear that the Medical Research Council in particular collaborates widely and successfully, although notably many of these collaborations appear to be on a one-off basis. The Medical Research Council also keeps a large portfolio solely in its own name. All of the UK's leading applicants show some collaboration.

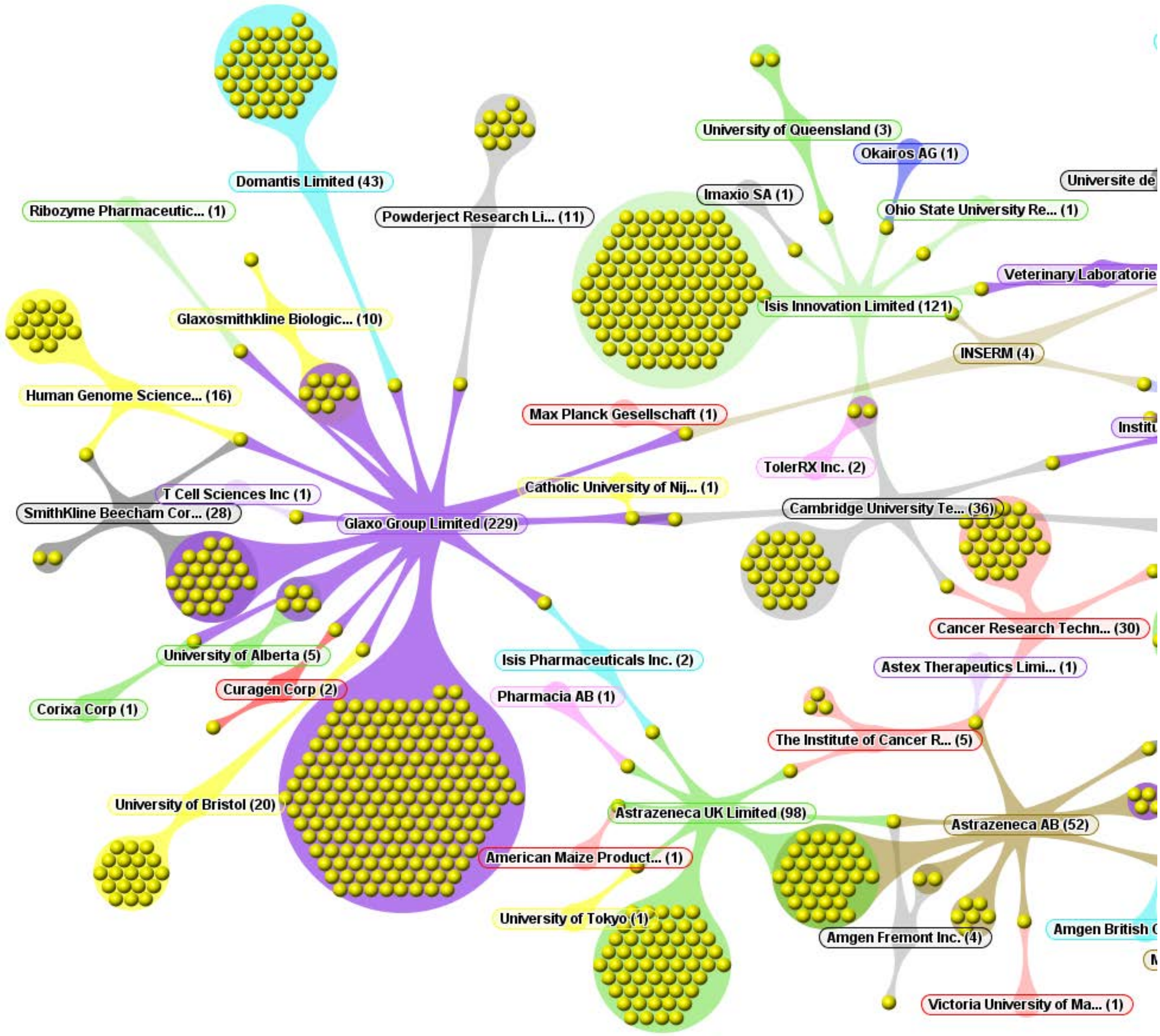



**Figure 10: Collaboration map showing collaborations of UK applicants and their collaborators**

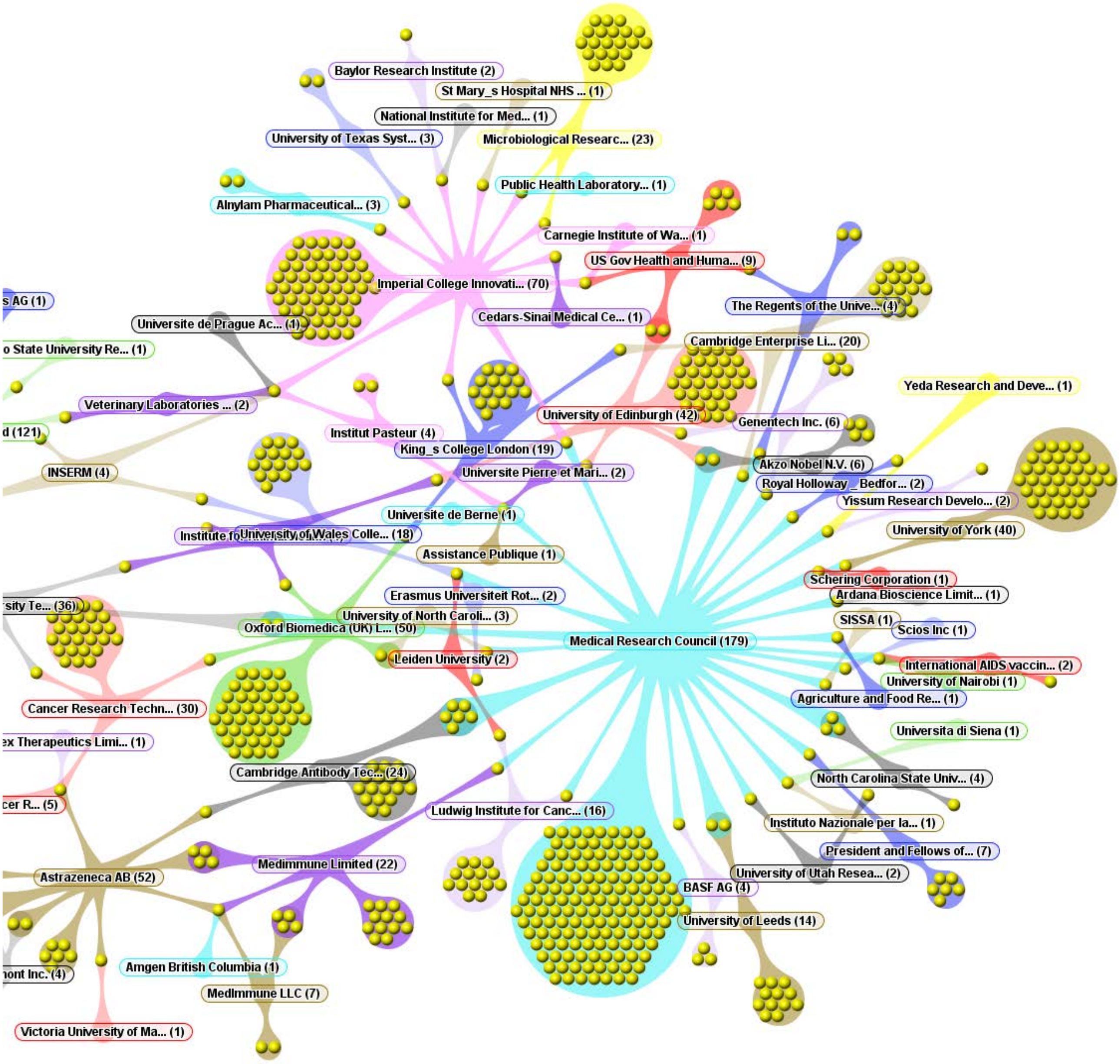



**Figure 10 (cont.): Collaboration map showing collaborations of UK applicants and their collaborators**

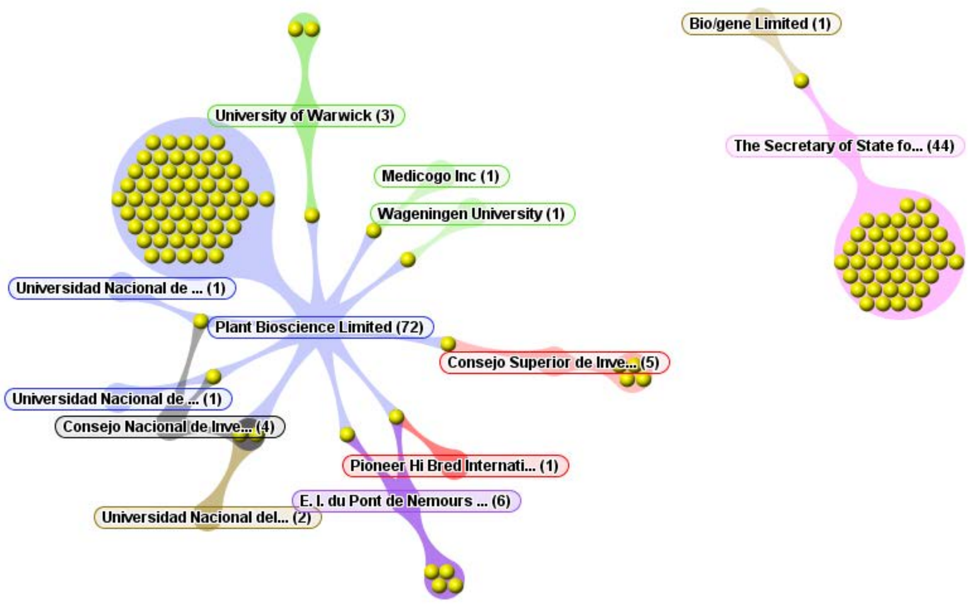



**Figure 10 (cont.): Collaboration map showing collaborations of UK applicants and their collaborators**

## 2.4 Technology breakdown

Figure 11 shows the top IPC subgroups, and Table 2 lists the description of each of these subgroups. The most frequently applied subgroup relates to recombinant DNA technology. Some subgroups relate to medicine, others to chemistry, and yet others to assay (measurement and detection). The classifications are not mutually exclusive and each patent family will have many of these classifications applied.

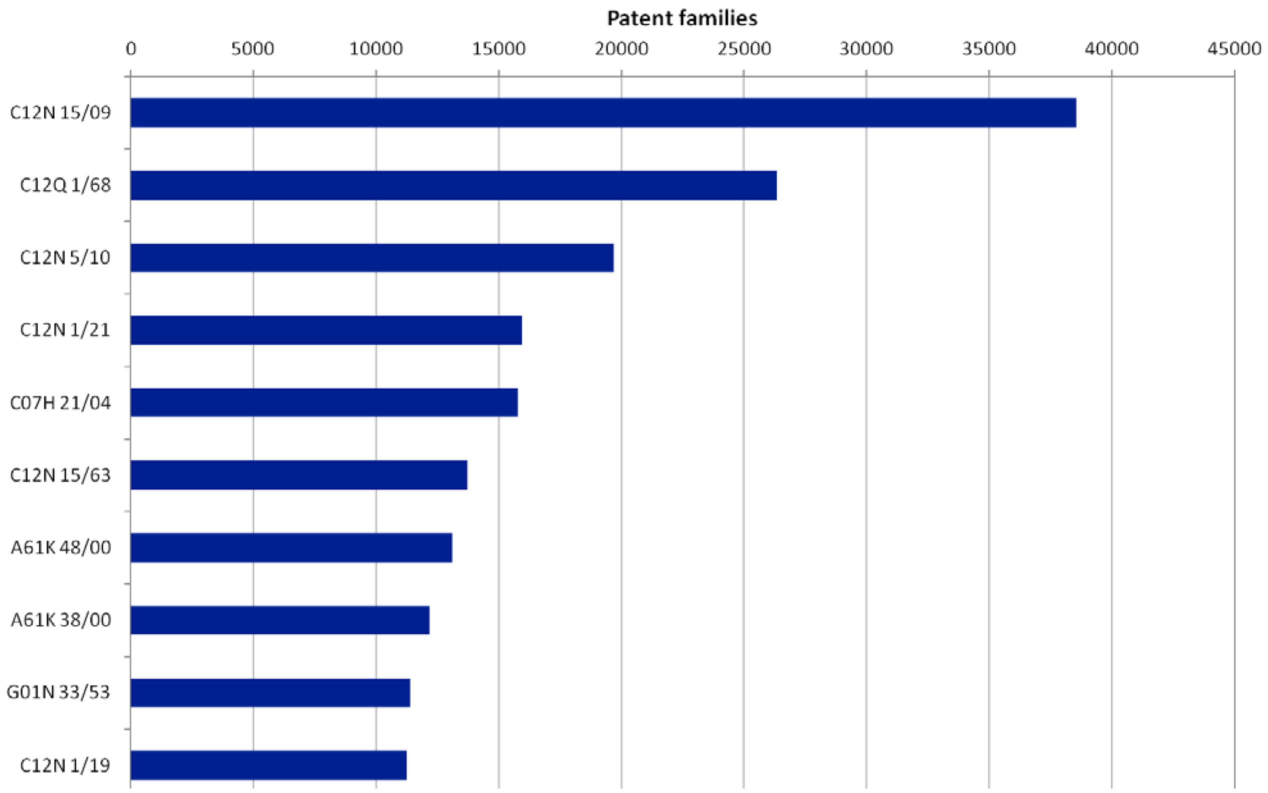



**Figure 11: Top IPC sub-groups**

**Table 2: Key to IPC subgroups referred to in Figure 11**

| | |
|---|---|
| C12N 15/09 | Mutation or genetic engineering; recombinant DNA-technology |
| C12Q 1/68 | Measuring or testing processes involving enzymes or micro-organisms; involving nucleic acids |
| C12N 5/10 | Undifferentiated human, animal, or plant cells; modified by introduction of foreign genetic material |
| C12N 1/21 | Bacteria; modified by introduction of foreign genetic material |
| C07H 21/04 | Compounds containing two or more mononucleotide units, having separate phosphate or polyphosphate groups linked by saccharide radicals of nucleoside groups; with deoxyribosyl as saccharide radical |
| C12N 15/63 | Mutation or genetic engineering; introduction of foreign genetic material using vectors; regulation of expression |
| A61K 48/00 | Medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases; Gene therapy |
| A61K 38/00 | Medicinal preparations containing peptides |
| G01N 33/53 | Immunoassay; Biospecific binding assay |
| C12N 1/19 | Fungi; yeast; modified by introduction of foreign genetic material |

# 3 The UK landscape

## 3.1 Top UK applicants

Figure 12 shows the top UK-based applicants within the life sciences, genomics, and synthetic biology dataset. Several major pharmaceutical companies appear in the top ten of patent applicants. Glaxo Group Limited leads, and the UK based divisions of Astrazeneca and Unilever follow. Pfizer is found beyond this list, in eighteenth position. Non-corporate entities are prominent in this list, namely the UK's Medical Research Council and the Ministry of Defence (notionally represented by its Secretary of State). Universities and their associated spin-out companies or technology transfer companies are also prominent in the list: Isis Innovation and Oxford Biomedica of Oxford, and Imperial College appear in the top ten. The universities of Edinburgh, York, Cambridge, Sheffield, Glasgow, and Bristol are also found in the top thirty.

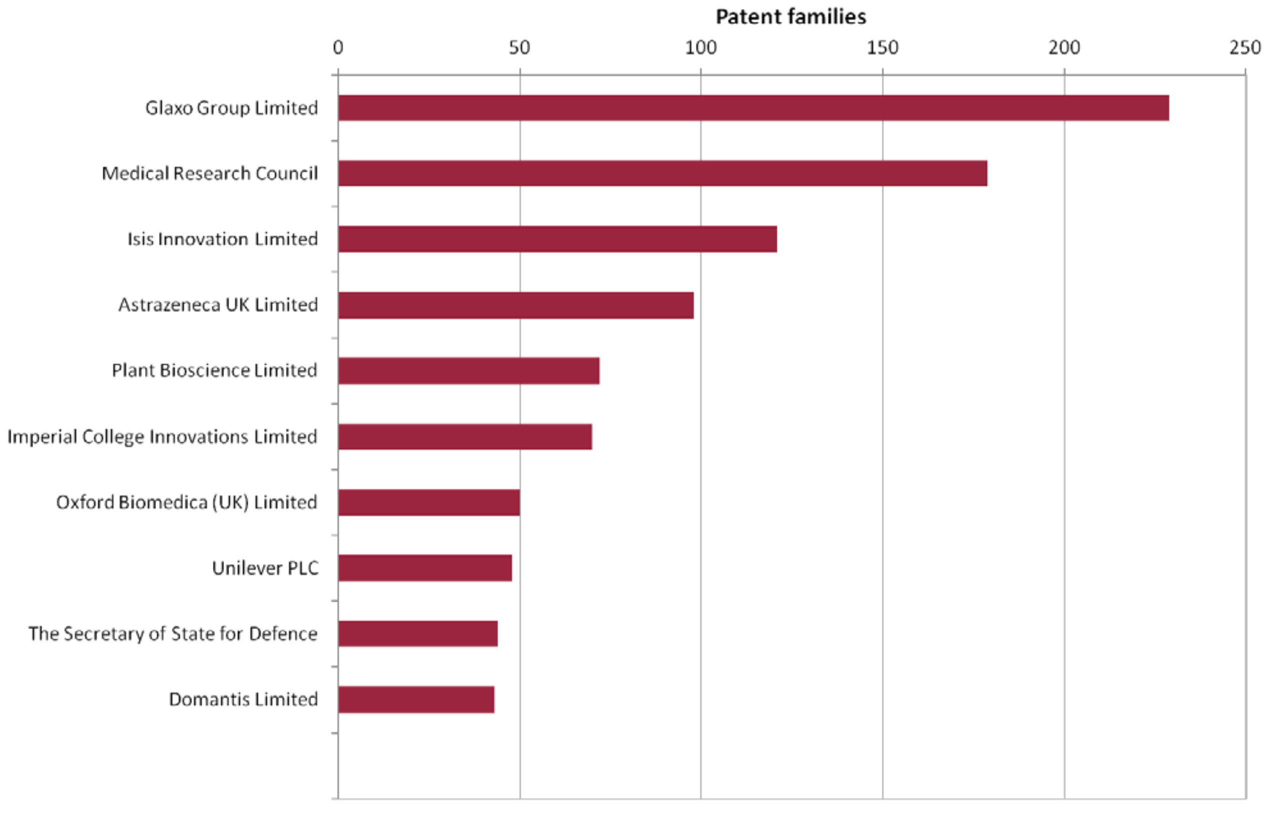



Figure 12: Top UK applicants

## 3.2 UK inventor mobility

Figure 13 shows applicants with the top UK inventors named on their publications. It is clear from this chart that the leading UK-based inventors are associated with the leading UK-based organisations. For this reason, the list largely mirrors that of Figure 12.

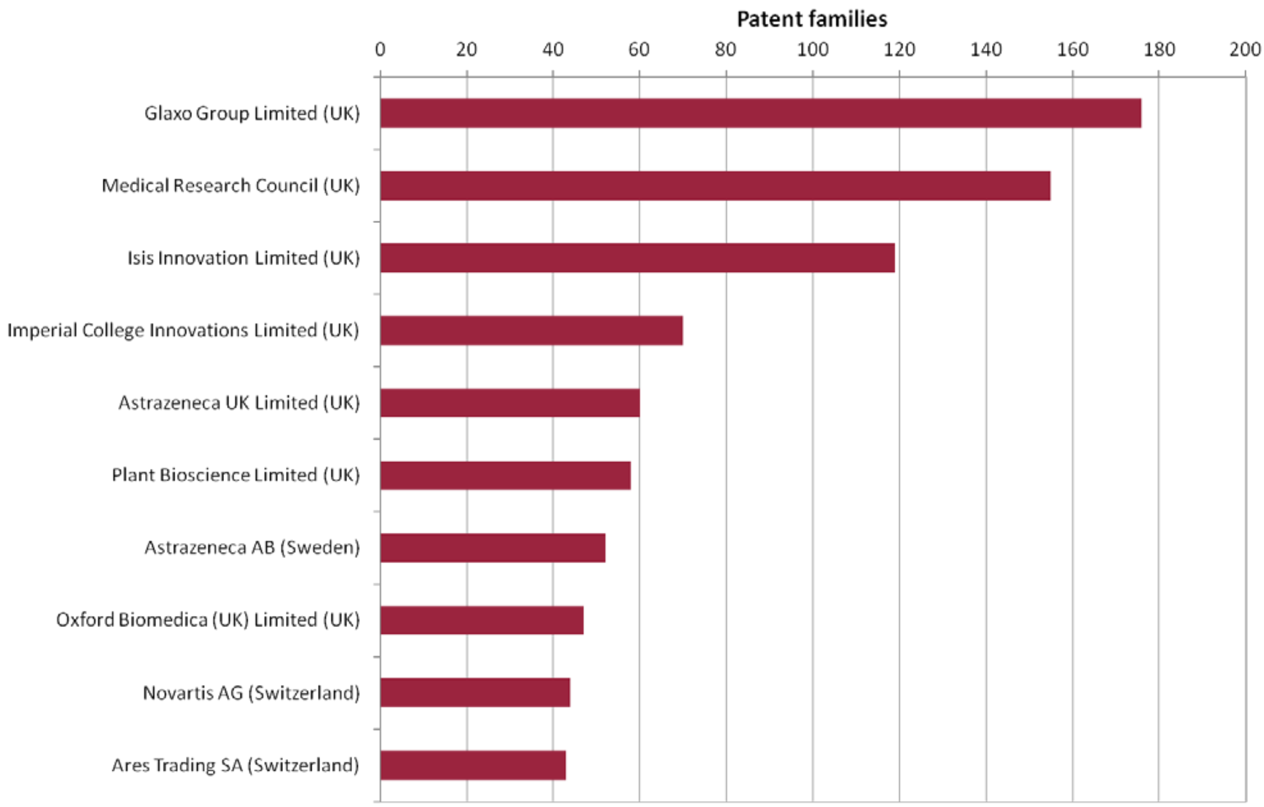



Figure 13: Top worldwide applicants with named UK-based inventors

## 3.3 How active is the UK?

A subset of the main worldwide dataset designed to reflect all UK patenting activity was selected. Figure 14 shows the year-on-year change in UK patenting activity against the worldwide year-on-year change in life sciences, genomics, and synthetic biology patenting shown in Figure 2; this shows a significant and sustained drop in the UK over the period. Publication numbers have fallen from 2962 in 2005 to 1609 to 2013

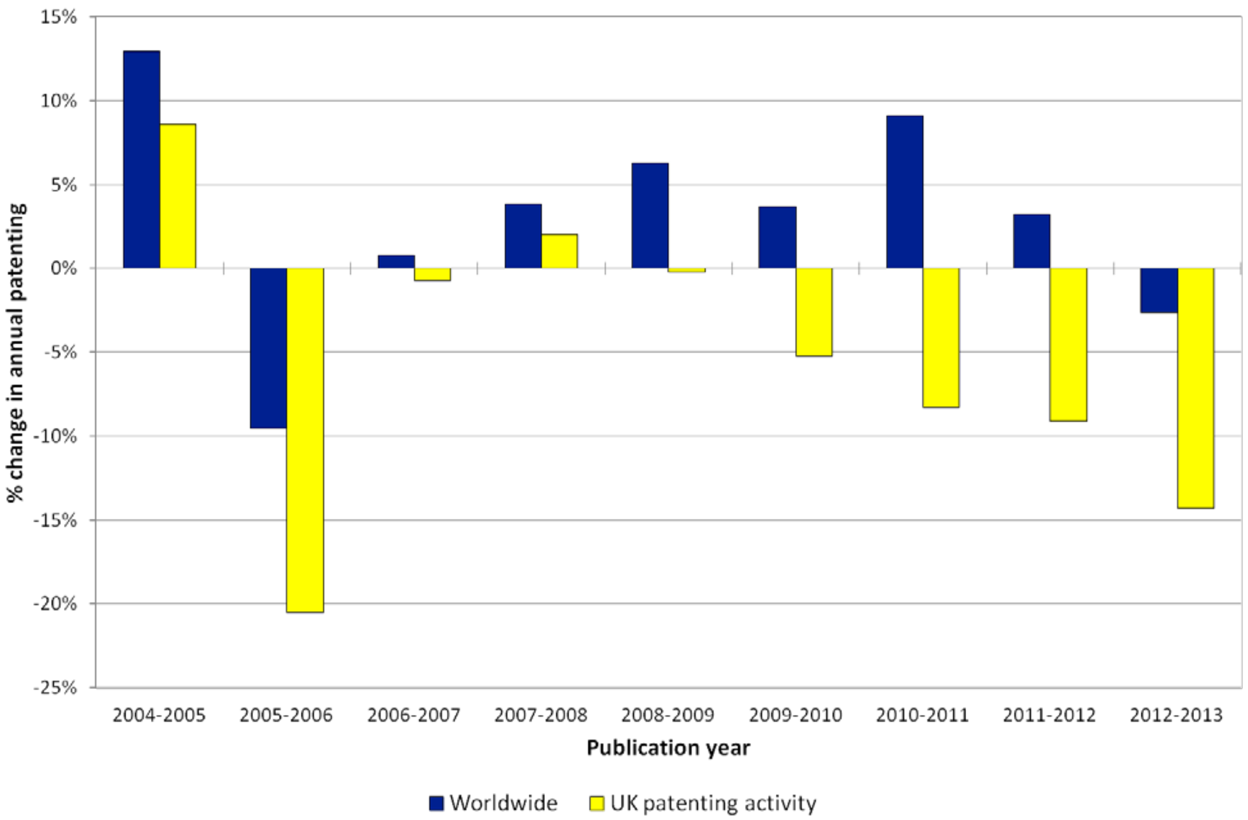



**Figure 14: Year-on-year change in UK and worldwide for life sciences, genomics, and synthetic biology patenting**

Similar patent subsets were created to reflect patenting activity taking place in several comparator countries (France, Germany, USA, Japan, China, and Korea) to produce the comparison chart shown in Figure 15. It is clear that the pattern in the UK is repeated to some extent in other comparator countries, with the notable exception of Korea and China, which have seen substantial, consistent growth from 2004 up to the present. The growth for China is such that, by 2013, patenting activity has overtaken all countries except the US, and for Korea patenting activity has overtaken Germany, France, and the UK. This growth appears, in both cases, to be driven by research institutes and agencies, rather than by corporate entities.

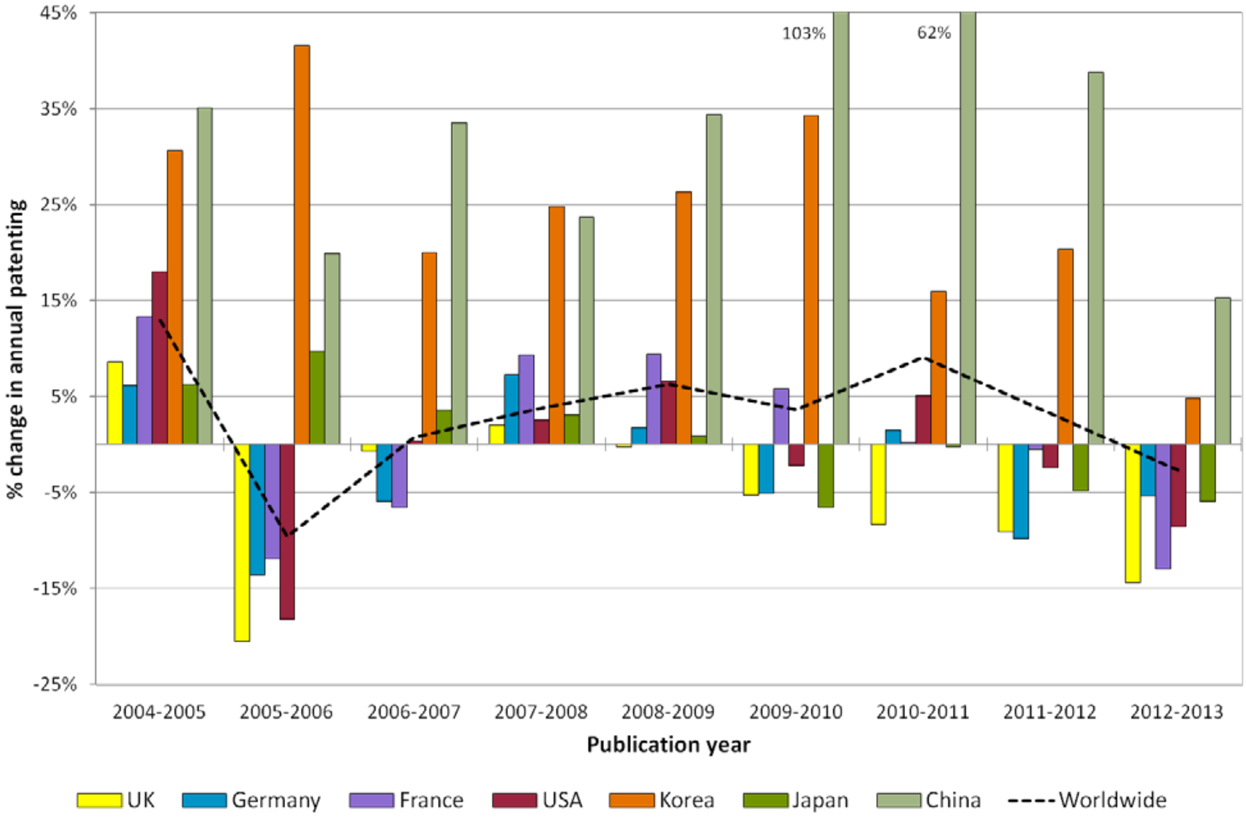



**Figure 15: Year-on-year change in UK life sciences, genomics, and synthetic biology patenting activity against comparator countries**

# 4 Patent landscape map analysis

In order to give a snapshot as to what the patent landscape looks like for this technology space, a patent map provides a visual representation of the dataset. Published patents (not patent families) are represented on a patent map by dots and the more intense the concentration of patents (*i.e.* the more closely related they are) the higher the topography as shown by contour lines. The patents are grouped according to the occurrence of keywords in the abstract and major topics appear on the patent map[5].

The landscape map for life sciences, genomics, and synthetic biology is shown in Figure 16. This map was produced using only the uses for the invention as described in the abstract (rather than the details of the processes and chemistry involved), in order to provide a clearer picture of the dataset.

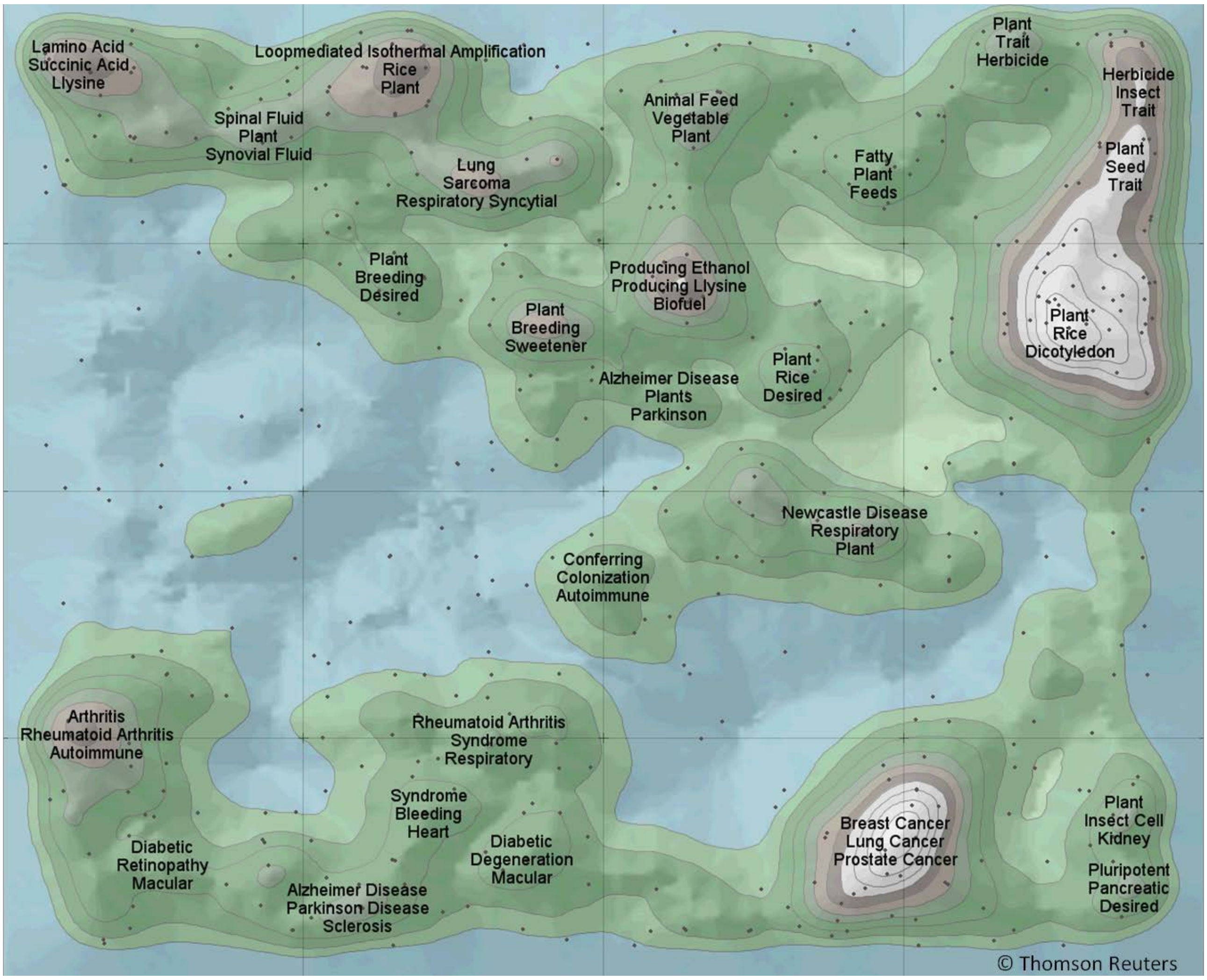



**Figure 16: Patent landscape map of life sciences, genomics, and synthetic biology**

[5] Further details regarding how patent landscape maps are produced is given in Appendix C.

Figure 16 shows that the major uses of the technology are in plants and crops, treatment of diseases, and in industrial processes. A large number of patents relate to the production of new varieties of staple crops such as rice, wheat, maize, and soybeans, and also cotton, generally seen in the top right part of the landscape. Also found here are patents relating to animal feed and varieties with enhanced nutritional content. Patents relating to the treatment of disease include many types of cancer, such as breast, lung and prostate as seen in the peak towards the bottom right of the landscape, but also in the landscape are colon, ovarian, pancreatic, and many other types including carcinoma and sarcoma. Degenerative and autoimmune disorders such as rheumatoid arthritis, Alzheimer's disease and Parkinson's disease are found in the bottom left of the landscape, together with complications of diabetes. Other patents in the landscape are for varieties of microorganisms which are used in industrial processes. These are exemplified by the production of ethanol as a biofuel, but also found are production of butanol and of chemicals for use in industrial processes such as the manufacture of carboxylic acid. Deeper within the landscape are varieties of tobacco alleged to have lower than usual nicotine content, reduced carcinogenic potential, and reduced heavy metal uptakes.

The distribution of leading applicants throughout the map can be shown by colouring the patents of each applicant, as shown in Figure 17–Figure 19. Of the leading applicants, those which are non-corporate entities (universities and agencies) have much more varied portfolios covering most parts of the patent landscape. These are Zhejiang University, Japan Science and Technology Agency, and University of California (Figure 17). Companies appear to be much more focussed in what they do, possibly as a consequence of trying to create clusters of patents which protect products in the marketplace. Two groups fall out of the top ten applicants: those having patents in the plants and crops part of the landscape, namely BASF, Syngenta, Dupont, and Pioneer (Figure 18); and those having patents in the disease part of the landscape, namely Genentech and Isis Pharmaceutical (Figure 19).

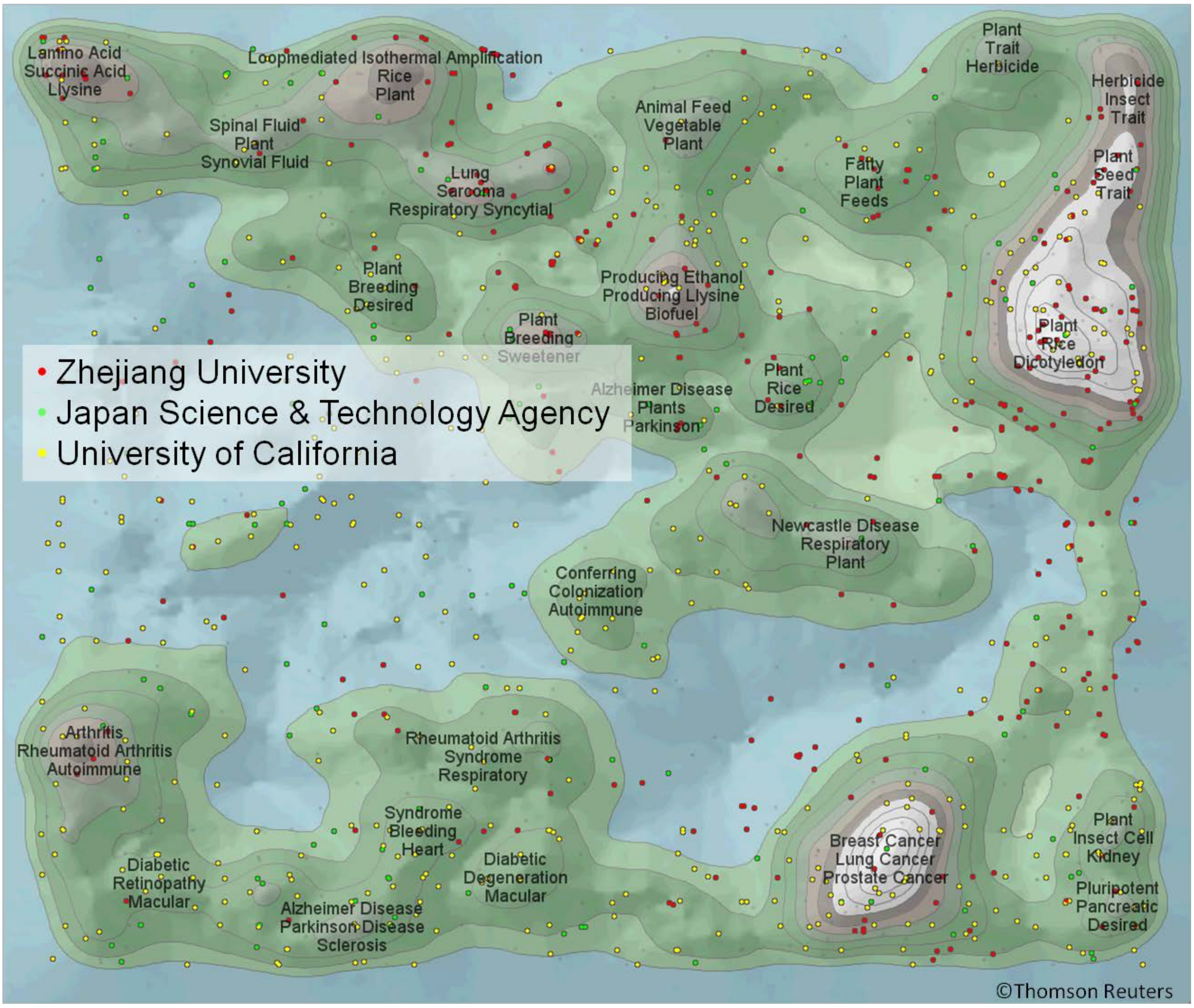



**Figure 17: Patent landscape map of life sciences, genomics, and synthetic biology with leading non-corporate applicants' patents coloured**

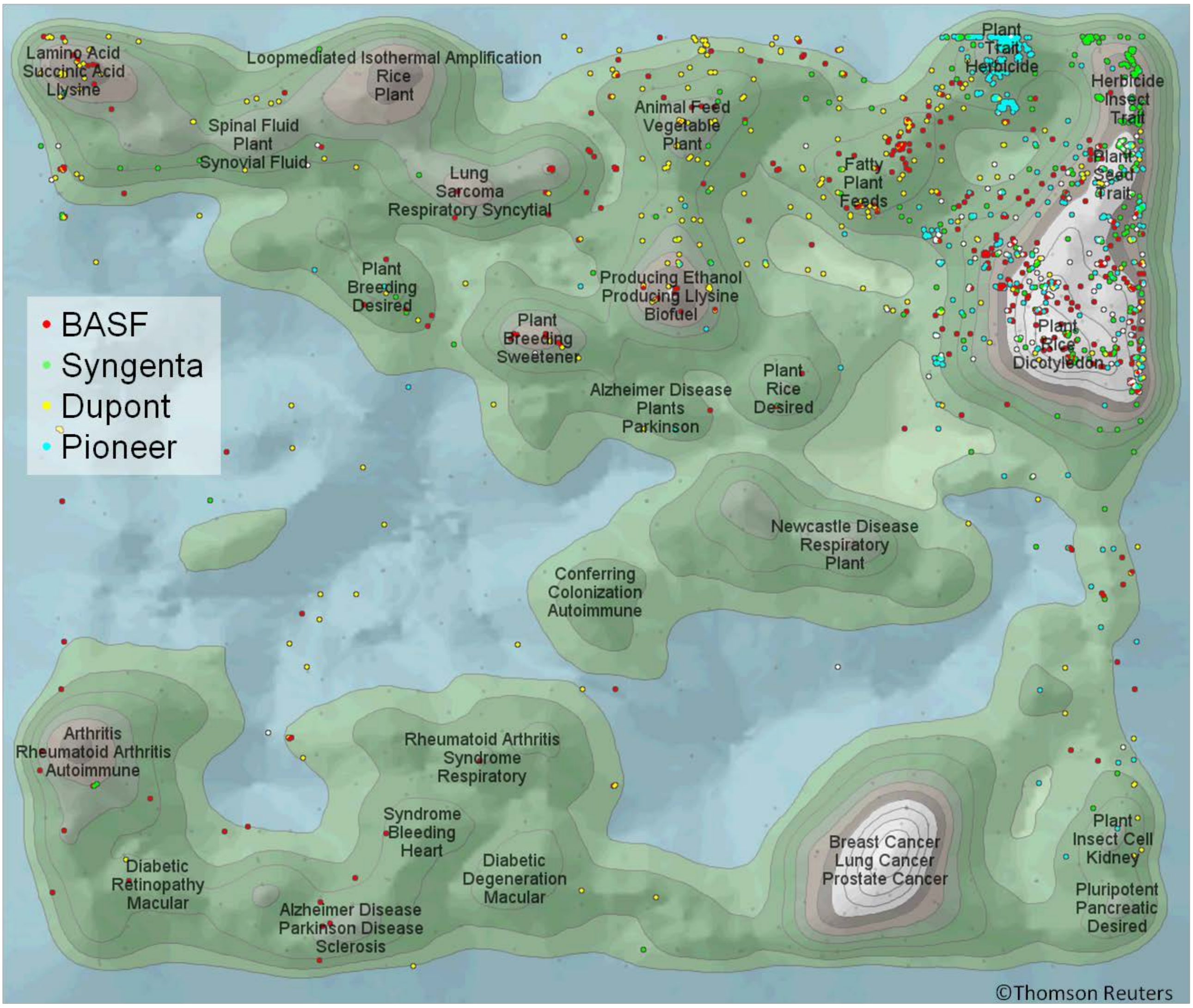



**Figure 18 : Patent landscape map of life sciences, genomics, and synthetic biology with leading plant and crop applicants' patents coloured**

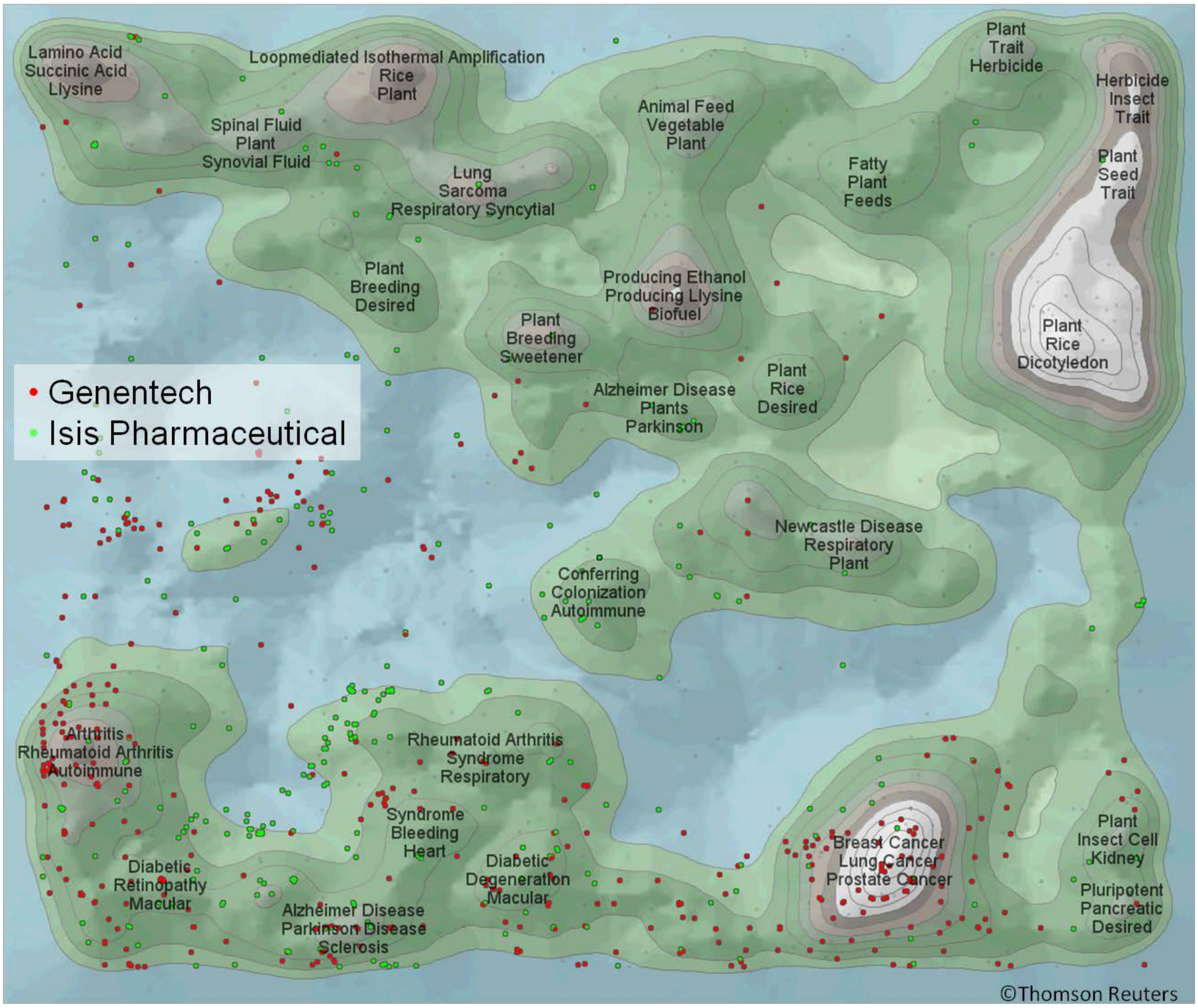



**Figure 19 : Patent landscape map of life sciences, genomics, and synthetic biology with leading disease applicants' patents coloured**

# 5 Conclusions

Life sciences, genomics, and synthetic biology patenting has grown only slightly over the ten year period that has been studied, and has not kept up with general patenting trends for all technologies worldwide: publications appear to have peaked in 2012 with a 2.6% fall into 2013, whilst other technologies continue to grow at 8.6% in 2013. Analysis by country reveals diverging trends between China and Korea, which have grown substantially and consistently since 2004 at year-on-year of up to 103%, and other countries, including the UK, which are beginning to show a general decline in patenting in the life sciences, genomics, and synthetic biology field.

The US has a large lead over other countries as a place of first making a patent application and as the country of the applicant (44% and 39% respectively). The US, together with China and Japan, are the countries where over three-quarters of patent applications are first filed, and where nearly three quarters of patent applicants are resident.

According to relative specialisation, Belgium, Israel, Australia, and Singapore show a positive specialisation in the life sciences, genomics, and synthetic biology (indicating a greater than expected degree of patenting for that country). The UK shows a small positive degree of specialisation. Despite their substantial growth in recent years, China and Korea have a negative specialisation, which is a reflection of the very low baseline from which the growth has developed at the beginning of the period, and also of the substantial level of patenting in other fields in those countries. Finland, Italy, and Taiwan show the lowest degrees of specialisation in this field.

Applicants in life sciences, genomics, and synthetic biology appear to be heavy users of the worldwide patent application system administered by the World Intellectual Property Organization, and of the European Patent Office, in preference to direct applications to national patent offices.

Many of the leading applicants are US based, with four of the top five positions being occupied by Pioneer Hi-Bred International, Monsanto Technology, University of California, and Genentech. The Japan Science & Technology Agency also appears in the top five. Other chemical and pharmaceutical companies from Germany (BASF, Bayer), USA (Du Pont, ISIS Pharmaceuticals), Switzerland (La Roche, Novartis, Syngenta), and Japan (Ajinomoto) are also among the leaders. However, several of the leaders are non-commercial entities such as universities (Zhejiang University, Chinese Agricultural University) or research institutes (Centre National de la Recherche Scientifique, Korea Institute of Bioscience and Biotechnology), and government agencies are also present.

The Chinese Agricultural University, Monsanto Technology, and Pioneer Hi-Bred International have grown significantly since 2007, having overtaken the Japan Science and Technology Agency and Genentech, whose publication numbers have dropped since 2006 and 2007 respectively.

A significant amount of collaboration appears to be going on between leading applicants in the life sciences, genomics, and synthetic biology field, as measured by applicants who co-occur in patent publications. Much of this collaboration is limited to within the US and within Japan, as two mostly separate clusters. However, the University of California bridges the gap with collaborations with several Japanese research institutes. All leading UK applicants also show significant collaboration, with Glaxo Group, Isis Innovation, Astrazeneca, Imperial College, but especially the Medical Research Council being found at the centres of clusters, which include many collaborations with universities reflecting the science-intensive nature of the field. Other leading UK applicants are Plant Bioscience, Oxford Biomedica, Unilever, the Ministry of Defence, and Domantis. Leading UK inventors are generally associated with the leading UK applicants, suggesting that life sciences, genomics, and synthetic biology is a field where inventors stay within the UK.

Patent landscape analysis shows that major applications of the technology being patented are plants and crops (such as new varieties of staple cereal crops and soybeans, cotton, animal feed, and the enhancement of nutritional content), treatment of diseases (including many types of cancer, degenerative diseases, autoimmune disorders, and complications of diabetes), and microorganisms used in industrial processes (such as production of biofuels and chemicals used in further industrial processes). Universities and research organisations are found to have patents distributed more widely around the landscape when compared to companies, which tend to be focussed on particular areas in the landscape. This is possibly a reflection of research activity being more fundamental and of general application, whereas companies are concerned with the protection of specific products in the marketplace.

# Appendix A Interpretation notes

## A.1 Patent databases used

The *Thomson Reuters* World Patent Index (WPI) was interrogated using *Thomson Innovation*[6], a web-based patent analytics tool produced by *Thomson Reuters*. This database holds bibliographic and abstract data of published patents and patent applications derived from the majority of leading industrialised countries and patent organisations, *e.g.* the World Intellectual Property Organisation (WIPO), European Patent Office (EPO) and the African Regional Industry Property Organisation (ARIPO). It should be noted that patents are generally classified and published 18 months after the priority date. This should be borne in mind when considering recent patent trends (within the last 18 months).

The WPI database contains one record for each patent family. A patent family is defined as all documents directly or indirectly linked via a priority document. This provides an indication of the number of inventions an applicant may hold, as opposed to how many individual patent applications they might have filed in different countries for the same invention.

## A.2 Priority date, application date and publication date

**Priority date**: The earliest date of an associated patent application containing information about the invention.

**Publication date**: The date when the patent application is published (normally 18 months after the priority date or the application date, whichever is earlier).

Analysis by priority year gives the earliest indication of invention.

## A.3 WO and EP patent applications

International patent applications (WO) and European patent applications (EP) may be made through the World Intellectual Property Organization (WIPO) and the European Patent Office (EPO) respectively.

International patent applications may designate any signatory states or regions to the Patent Cooperation Treaty (PCT) and will have the same effect as national or regional patent applications in each designated state or region, leading to a granted patent in each state or region.

European patent applications are regional patent applications which may designate any signatory state to the European Patent Convention (EPC), and lead to granted patents having the same effect as a bundle of national patents for the designated states.

[6] *http://info.thomsoninnovation.com*

Figures for patent families with WO and EP as priority country have been included for completeness although no single attributable country is immediately apparent.

## A.4 Patent documents analysed

The life sciences, genomics, and synthetic biology patent dataset for analysis was identified in conjunction with patent examiner technology-specific expertise. A search strategy was developed and the resulting dataset was extracted in April 2014 using International Patent Classification (IPC) codes, Co-operative Patent Classification (CPC) codes and keyword searching of titles and abstracts in the *Thomson Reuters* World Patent Index (WPI) and limited to patent families with publications from 2004 to 2013.

The applicant and inventor data was cleaned to remove duplicate entries arising from spelling errors, initialisation, international variation (Ltd, Pty, GmbH *etc*.), or equivalence (Ltd., Limited, *etc*.).

## A.5 Analytics software used

The main computer software used for this report is a text mining and analytics package called *VantagePoint*[7] produced by *Search Technology* in the USA. The patent records exported from *Thomson Innovation* were imported into *VantagePoint* where the data is cleaned and analysed. The patent landscape maps used in this report were produced using *Thomson Innovation*.

[7] http://www.thevantagepoint.com

# Appendix B Relative Specialisation Index

Relative Specialisation Index (RSI) was calculated as a correction to absolute numbers of patent families in order to account for the fact that some countries file more patent applications than others in all fields of technology. In particular, US and Japanese inventors are prolific patentees. RSI compares the fraction of life sciences, genomics, and synthetic biology patents found in each country to the fraction of patents found in that country overall. A logarithm is applied to scale the fractions more suitably. The formula is given below:

$$\log_{10}\left(\frac{n_i/n_{total}}{N_i/N_{total}}\right)$$

where

ni = number of life sciences, genomics, and synthetic biology patents in country i
ntotal = total number of life sciences, genomics, and synthetic biology patents in dataset
Ni = total number of patents in country i
Ntotal = total number of patents in dataset

The effect of this is to highlight countries (such as Belgium, as shown in Figure 5) which have a greater level of patenting in life sciences, genomics, and synthetic biology than expected from their overall level of patenting, and which would otherwise languish much further down in the lists, unnoticed. Please not that India is not included in the RSI measure because the worldwide patent databases have poor coverage of Indian applicant address (applicant country) data.

# Appendix C Patent landscape maps

A patent landscape map is a visual representation of a dataset and is generated by applying a complex algorithm with four stages:

*i)* **Harvesting documents** – When the software harvests the documents it reads the text from each document (ranging from titles through to the full text). Non-relevant words, known as stopwords, (e.g. “a”, “an”, “able”, “about” etc) are then discounted and words with common stems are then associated together (e.g. “measure”, “measures”, “measuring”, “measurement” etc).

*ii)* **Analysing documents** – Words are then analysed to see how many times they appear in each document in comparison with the words’ frequency in the overall dataset. During analysis, very frequently and very infrequently used words (i.e. words above and below a threshold) are eliminated from consideration. A topic list of statistically significant words is then created.

*iii)* **Clustering documents** – A Naive Bayes classifier is used to assign document vectors and Vector Space Modelling is applied to plot documents in n-dimensional space (i.e. documents with similar topics are clustered around a central coordinate). The application of different vectors (i.e. topics) enables the relative positions of documents in n-dimensional space to be varied.

*iv)* **Creating the patent map** – The final n-dimensional model is then rendered into a two-dimensional map using a self-organising mapping algorithm. Contours are created to simulate a depth dimension. The final map can sometimes be misleading because it is important to interpret the map as if it were formed on a three-dimensional sphere.

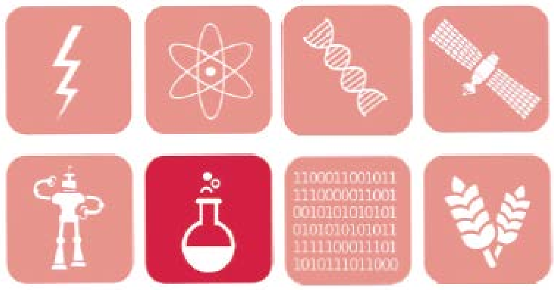
Thus, in summary, patents are represented on the patent map by dots and the more intense the concentration of patents (*i.e.* the more closely related they are) the higher the topography as shown by contour lines. The patents are grouped according to the occurrence of keywords in the title and abstract and examples of the reoccurring keywords appear on the patent map. Please remember there is no relationship between the patent landscape maps and any geographical map.

Please note that the patent maps shown in this report are snapshots of the patent landscape, and that patent maps are best used an interactive tool where analysis of specific areas, patents, applicants, inventors *etc* can be undertaken ‘on-the-fly’.

Intellectual Property Office

Concept House
Cardiff Road
Newport
NP10 8QQ
United Kingdom

www.ipo.gov.uk



#8Great