4 Knowledge Network Influences on the Development of Drug Discovery Technologies: A Longitudinal Case Study

Abstract
Mechanisms of knowledge transfer from academia to industry have long been debated. The knowledge inputs required may stem from research conducted many years prior to a technology being adopted and adapted by industry. A supporting knowledge base is required to facilitate knowledge transfer. In this paper we utilise the publishing and patenting history of an individual scientist to demonstrate in detail how new technologies emerge from the work of academic inventors and from their cognitive environment. It provides a detailed description of the knowledge within these technologies. In particular, we will address the role of absorptive capacity in priming their development. We find clear linkages between the technologies in their present form and the long-past specific outputs authored by the individual, and that the knowledge contained therein has undergone varied degrees of transformation. The individual scientist demonstrates a high level of absorptive capacity, incorporating exogenous knowledge into their own knowledge base.

4.1 Introduction
In Nelson’s seminal essay entitled “The market economy, and the scientific commons” (R. Nelson, 2004), technologies need to be understood as: “[I]nvolving both a body of practice, manifest in the artefacts and techniques that are produced and used, and a body of understanding, which supports, surrounds, and rationalizes the former.” (p.457)

In knowledge and technology transfer, analyses address a multitude of aspects and levels of science-technology interaction (Bozeman, 2000; Ponomariov & Boardman, 2012). A minute analysis of knowledge transfer mechanisms and mediums - such as using tacit and codified knowledge, R&D networks, formal and informal collaborations - runs into many difficulties. These difficulties stem from the enormous complexity of the knowledge that is transferred. In most cases, the final technological product is the result of heterogeneous knowledge inputs and its accretion over time into a coherent system (Nelson, 2004).

In this paper, we focus on the knowledge inputs of a firm and on the nature of the quanta of knowledge and information themselves. We will add to the current literature on knowledge transfer by examining in detail the specific knowledge and technologies involved, by means of an analysis of the actual uptake and implementation of the associated knowledge concepts into the technologies developed.

1 This chapter has been published as Gurney, T. et al. (2013). Knowledge capture mechanisms in bioventure corporations: a case study. Proceedings of 14th International Society of Scientometrics and Informetrics Conference, Vienna. It has been submitted for publication in Research Policy.
We have developed a method to discern the knowledge contributions of inventors and scientists to a corpus of patents and the technologies they represent (Chapter 3). The method is applied to the two main output indicators typically used in other studies, those of patents and publications. Publications are the typical output of scientific endeavours, and patents are the technological result of the application of the results of those endeavours. The concepts and practices embodied and codified in the publications and patents were linked to each other through the citations to literature found in the patent documents. By linking the two corpora of scientific and technological knowledge, we are able to address our research question: To what degree does an existing knowledge base contribute to the development of novel technologies and how can we effectively measure these contributions?

As such, we aim to demonstrate the origins of the knowledge contributions to the development of an idea over time, from its inception, through its transformation and finally to its application in a technological product. The initial sections of this paper discuss the multiple aspects of absorptive capacity, knowledge transfer and transformation, including how scientific knowledge is incorporated into practices, skill sets and artefacts. We then discuss the national policy context of our test case study, and provide a brief history of our test case study. Following this, we briefly summarise our previous methodology, along with descriptions of the indicators we use, followed by the visualisation and clustering techniques employed in our analysis. Our results and conclusions follow, ending with our discussion and implications for further analyses and policy.

4.2 Conceptual Framework

The most common and widely cited knowledge transfers, capture mechanisms and inputs are patents, publications, informal and formal interactions, personnel hiring, licensing, R&D collaborations, contract R&D and consulting (Cohen et al., 2002). In each of these mechanisms, the medium of knowledge transfer is either codified (such as, for example, patents and publications) or tacit (such as, for example, R&D collaborations and personnel hiring). A third medium, that of embedded knowledge, resides in the material aspects, such as new equipment (Gorman, 2002). Key to the reception and implementation of these mediums is the absorptive capacity of the unit under study.

4.2.1 Absorptive capacity

The organisational infrastructure required to facilitate the development, transfer and capture of knowledge depends heavily on its recipient. The recipient is understood to demonstrate a need for ‘absorptive capacity’ (Cohen & Levinthal, 1990) which is described as “[t]he ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends [and this] is critical to its innovative capabilities,” (p.128).

There are two key aspects involved in absorptive capacity, firstly on an individual level, in that the absorptive capacity of a firm is tied to its constituent individuals’ absorptive capacity, i.e. the right personnel are in place to take advantage of incoming information. In this aspect, the communication infrastructure between a firm and its external environment is key, with select individuals acting as gatekeepers, much like the star scientists of Zucker and Darby (1996) or the core scientists of Furukawa and Goto (2006). In the case of a dedicated university-industry spanning role, as might be fulfilled by a bridge scientist (a scientist who is active within both academia and
industry), the absorptive capacity could be bolstered by a more active search process. As Cohen and Levinthal (1990) state, “Beyond diverse knowledge structures, the sort of knowledge that individuals should possess to enhance organizational absorptive capacity is also important. Critical knowledge does not simply include substantive, technical knowledge; it also includes awareness of where useful complementary expertise resides within and outside the organization” (p.133).

Secondly, on an organisational level, the requirements for absorptive capacity are related to the network-wide communication structure, such as “the relationships between corporate and divisional R&D labs or, more generally, the relationships among the R&D, design, manufacturing, and marketing functions” (Cohen & Levinthal, 1990) (p.134). In essence, this form of absorptive capacity links the interactive, exchange aspects of innovation to the communication aspects. This form relies on the positive and/or negative feedback between departmental/divisional entities to further develop an idea or product. In this paper, we do not focus on this level of absorptive capacity - which we do elsewhere (Chapter 3 and (Lanciano-Morandat et al., 2009)) - but on the knowledge capture mechanisms carried out by the star and bridge scientists discussed above.

The concept of absorptive capacity was expanded on significantly by Zahra & George (2002) who include potential and realised absorptive capacity. This expansion has clarified what absorptive capacity encompasses, by defining its various dimensions. We utilise Zahra & George’s dimensions in our case study. These dimensions include, for potential absorptive capacity: Acquisition - details the role of prior knowledge or capabilities and the infrastructure already in place; and Assimilation - exogenously generated knowledge (by others outside the firm) needs to be understood prior to incorporation within a firm. For realised absorptive capacity: Transformation - the ability to combine knowledge generated exogenously and endogenously (within the firm) to create novel fundamental or applied knowledge; and Exploitation - the usage of novel knowledge generated during transformation. The outputs of exploitation can be patents, publications, products or new processes.

4.2.2 Proxies of knowledge input and output
All four dimensions of absorptive capacity introduced by Zahra & George (2002) include patents and publications as typical knowledge mediums (as well as outputs). We use these codified knowledge mediums as proxy indicators of exogenous and endogenous knowledge. These proxies embody and detail the knowledge units acquired, assimilated, transformed and finally exploited. The use of patents and publications as such proxies is well developed and details of which are given below.

The use of patents as indicators was pioneered by Schmookler (1966) with many applications following (e.g. Griliches (1998), Schmoch (1993) and Fleming (2001)). Patents have been used as indicators for multiple purposes as they are considered detailed evidence of technological progress (Tijssen, 2002). They are inherently scalable (Narin, 1994), contain multiple metadata useful for analysis (Nelson, A.J., 2009), and are commonly regarded as substantive links to R&D activity (Ahuja & Katila, 2001). There are drawbacks in that not all innovations are patented (Arundel, 2001; Arundel & Kabla, 1998; Pavitt, 1988). However, they arguably remain the best indicators of R&D input (Narin, 1994).
When patents are used as indicators, the analysis is typically based on the metadata found in patents. Title words, abstract words and keywords (Courtial et al., 1993; Engelsman & van Raan, 1994), patent classifications (Leydesdorff, 2008; Tijssen & Van Raan, 1994), and patent/non-patent citations (Karki, 1997; Meyer, M.S., 2001) have all been used extensively. We use the PatSTAT database (April 2011 version), prepared by the EPO, because it is widely available, has global coverage and is data-rich and comprehensive.

Publications serve as the primary indicators for the defining characteristics and development of science. They are the most visible outcome of scientific endeavours, and an extensive range of indicators and methodologies have been developed. The analysis of publications shares a number of analytical approaches with patent analyses, such as word mapping (Callon et al., 1991) and citation analysis (Garfield & Welljams-Dorof, 1992; White & McCain, 1998). Sequential usage (Braam et al., 1991) and combinations (van den Besselaar & Heimeriks, 2006) of title words and cited references of publications has also become a common analytical technique.

4.2.3 Non-patent literature references
As part of the patenting process, listing prior art is a requirement across all patenting offices (although the completeness of submitted prior art varies between offices (Criscuolo & Verspagen, 2008)). Citation studies using patent-to-literature citations have progressed immensely (Meyer, M., 2000, 2002; Meyer, M.S., 2001; Narin, 1976, 1994). These studies examined in detail the literature citing proclivity and distribution across fields. Key to these studies was the recognition that references to literature in patents do provide a substantive link between the technologies and the sciences.

Non-patent literature references (NPLRs) exhibit different characteristics based on their source and who includes the reference. The scientific nature of NPLRs has been examined by Callaert et al. (2006). They found that there may be occurrences of citations to non-scientific literature but, overall, most citations are to peer-reviewed journals. NPLRs may come from applicants or examiners and have typically been treated as being of differing importance (Karki, 1997). The complex negotiation processes between applying for a patent and the granting of that patent result in the strategic inclusion (or exclusion) of specific references by the applicant. Examiners must rely on references supplied by the applicant and their own examination process to ensure a complete check. Following this, studies have typically chosen the front-page references (on the patent document), i.e. the examiner references, over the applicant references. In our study we choose to utilise both sources of references, bibliographic-NPLRs (examiner) and in-text NPLRs (applicant). We do so as it is generally understood and accepted that the presence of citations to literature in patent documents indicates a cognitive link to, or awareness of, the related scientific concepts (Tijssen, 2001), regardless of the source of the NPLRs. By combining the two sources of NPLRs we aim to provide a more comprehensive view and call upon the judgement of not only the inventor but also the examiner as to what is relevant.

4.2.4 Publishing and patenting in academia and industry
In studies such as ours, which examines patents or publications across the academic-industrial divide, it is important to note that each sphere maintains different approaches to each aspect of knowledge production. When comparing university-based and firm-based scientists and their
propensity for the different types of knowledge production, it is important to consider the underlying motivations. University-based scientists publish primarily to extend their professional and intellectual prowess (through which resources are allocated for future projects) and regular publishing is considered a requirement. With patenting, there has been a recent explosion of sorts in the rate of university patenting. This has been linked to institutional and national level changes (Owen-Smith & Powell, 2003; Zucker, L.G. & Darby, 1996), and the increased interest in spin-offs and IP spin-outs generated in academia (Owen-Smith & Powell, 2003; Zucker, L.G. & Darby, 1996; Zucker, L.G. et al., 1999).

Firms benefit from carrying out their own basic research, as they become intimately involved in the fundamental aspects of their own applied research (Rosenberg, 1990) and the output of this basic research typically results in publications. With firm-based publishing efforts, the underlying motivation is generally a directed and concerted effort from within the corporate infrastructure, as the firm stands to gain (or lose) more from the publication process than the author. For example, higher rates of approval of patents (McMillan et al., 2003), a window and source into various fields (Schartinger et al., 2002) and stronger ties with future progenitors of knowledge (Zucker, L.G. & Darby, 1996). Publishing in firms has also been shown to be strongly linked to future patenting areas and recruitment efforts (Hicks, 1995).

Industry has particularly realised the benefits of collaborative efforts with academia, as multiple studies have demonstrated positive results in terms of innovation output, ranging from the life sciences sector to the nanosciences (Baba et al., 2009; Meyer, M., 2007).

4.3 Knowledge utilisation

In this case study we aim to comprehensively describe the knowledge utilisation processes, including absorptive capacity, transfer and capture within a company, using a methodology developed previously in Chapter 3 of this thesis. We use this methodology to investigate the research question: To what degree does an existing knowledge base contribute to the development of novel technologies and how can we effectively measure these contributions?

We address this question by developing descriptors (indicators) of the degree of similarity between a researchers’ publication corpus and the patent applications of which he is an inventor. These descriptors, in general terms, describe the level of endogenous versus exogenous (to the individual or research group) knowledge influences on the patent output. More precisely, they relate to the receptivity of the researcher/inventor and the firm involved to the sources of knowledge, and the specific knowledge that was transferred or captured. These include (i) the researcher/inventor’s own knowledge that he or she brings in; (ii) knowledge brought in through research collaboration with others; (iii) knowledge produced by others, but which is similar enough to the companies’ knowledge base to absorb it directly; and (iv) new research conducted to be able to absorb knowledge which is rather dissimilar to the existing knowledge base. The descriptors are:

1. The reputational and applicability aspects of the scientific base work (Hullmann & Meyer, 2003) conducted by an individual;
   - The reputational aspect (as defined by the number of citations) and applicability aspect (as
The Intellectual Salmon Run: Knowledge Transfer and Dynamics between Academia and Industry

defined by the proportion of the researcher/inventor’s work cited by the patent applications versus their total corpus) determine the quality and relative importance of the researcher/inventor’s research to the technology.

2. If the individual’s overall research trajectory is, or is not, located in the field(s) of research necessary for the technologies;
   - This provides detail of the knowledge base cited by the patent applications and highlights the similarities (if any) between the publications cited by the patent applications and the overall research corpus of the researcher/inventor.

3. The markers for the other fields of science that are being utilised by the technologies (Karki, 1997; Schmoch, 1993);
   - This describes knowledge sourced from outside the researcher/inventor’s own expertise or network. This is defined by what research is cited by the patent applications and does not fall into the fields of research of the researcher/inventor.

4. The level of input by collaborators;
   - Externally sourced knowledge, from collaborators or co-inventors, may be required for the technologies. These contributions may be identified by the cited literature that is authored by collaborators/co-inventors without the researcher/inventor as an author.

5. The degree that the knowledge features (such as concepts, knowledge bases and, to a certain extent, skill sets) utilised by the technologies are shared amongst their sources (the inventor, his co-inventors, or other researchers);
   - The similarities or differences between the knowledge cited by the patent applications and the overall corpus of the researcher/inventor may indicate whether the required research and the associated skill sets are already present or need to be developed.

6. Whether the individual incorporated skill sets that were acquired during the development of the technologies and applied them to further his or her fundamental scientific research by knowledge creation feedback (Fischer, 2001; Tijssen, 1998).
   - If research conducted by the researcher/inventor after applying for a patent displays a degree of similarity to the publications cited by the patent applications - in which the researcher/inventor had no previous presence - this indicates that the knowledge and research skills obtained during the development of the technologies were applied in his or her further work.

We have chosen a case study in which an individual occupies a significant bridging role between academia and industry. This role allows us to effectively isolate his contributions to both the technologies and the underlying sciences involved, and allow us to fully investigate the descriptors mentioned above. The method can also be applied on more hybrid cases, but for reasons of clarity we selected this one.
4.4 Case study selection and history of Japanese biotechnology

4.4.1 Case study selection

Our case study involves a prominent Japanese biotechnology researcher, Professor Yusuke Nakamura, who is heavily involved in cancer therapeutics at the University of Tokyo, where he was head of the Human Genome Center. Nakamura founded OncoTherapy Science Inc. (OTS) in April 2001 to “[…] contribute in research and development of anti-cancer medicine, cancer therapy and cancer diagnosis based on oncogenes and proteins”.2 OTS’s business outline is: “…to provide innovative anti-cancer medicines with higher efficacy and a minimum risk of adverse events based on the comprehensive research on cancer genomics and biomedical analysis conducted by Nakamura of Human Genome Center, Institute of Medical Science, the University of Tokyo.”

Although Nakamura is not currently listed on the board of directors, he maintains direct links between his research at the University of Tokyo and research conducted at OTS. This direct link between academia and industry as manifested by Nakamura is the primary reason for choosing Nakamura and OTS. It enables us to draw upon his extensive publishing history as well as his numerous patenting activities, both at the University of Tokyo and OTS.

Below we provide some background information on the Japanese biotechnology sector from our own research into public documents to provide more detail on our selection of this case study.

4.4.2 Japanese firms in biotechnology research

Japan has a long tradition of biotechnology and can be regarded as a biotechnology-orientated country. In 1908, Dr Kikunae Ikeda at the University of Tokyo found that monosodium-glutamate is the true substance of the “Umami” taste. In the same period, Mr Saburosuke Suzuki was extracting iodine from konbu (seaweed), for pharmaceuticals. Collaboration between Dr Ikeda and Mr Suzuki led to the foundation of Ajinomoto, one of Japan’s first venture companies deriving from cooperation between industry and academia. This example illustrates two important aspects: that science was driven to application and that Japan was amongst the pioneers of biotechnology-based technology transfer in the early 20th century. Many such firms were developed (and remain very active in all fields) such as Takeda Pharmaceutical Company Limited, Kyowa Hakko Kirin Co., Ltd., Astellas Pharma Inc. and Daiichi Sankyo Co., Ltd. to name a few.

During the 1980s, Japan initiated and developed international collaborative programmes to improve its life sciences efforts in pre-competitive research. During the period in office of Prime Minister Yasuhiro Nakasone (1982-1987), the Human Frontier Science Program was started (1987), the Japan Key Technology Center was established (1985), the Protein Engineering Research Institute was established (1986), amongst others. As a result, the life sciences in Japan became more internationalised with an increase of foreign scientists at different levels, from post-docs to senior scientists. However, this strategy did not have the desired increased impact on promoting

2 http://www.oncotherapy.co.jp/eng/corporate/enkaku.html
3 http://www.oncotherapy.co.jp/eng/corporate/business.html
biotechnology-based knowledge transfer. Japan still lagged behind the USA and Europe in this aspect. During this period of transformation, there were few efforts by researchers to obtain patents. Sometimes inventions developed at universities were transferred free of charge from faculty members to companies with which the faculty members have close relations. These practices are a far cry from transferring technologies developed at universities to the most appropriate corporations. There were no systematic technology transfer policies in place from academia to industry, nor biotech start-ups developed from academia.

4.4.3 Japanese innovation policies
Following the collapse of the Japanese economy in the early 1990s, in order to create new industries for economic growth, the decision was taken to appropriate funds to support research and development at universities, starting in 1995. The Science and Technology Basic Law was enacted in November 1995, and the Cabinet approved the first phase of the Science and Technology Basic Plan in July 1996. Universities were increasingly expected to serve as the source of technological innovation in order to revitalise the economy, and society increasingly needed to see returns on research expenditure at universities, which grew amid the economic slump. Under these circumstances, measures related to industry-university cooperation were promoted in the late 1990s. The concept of technology-licensing organisations (TLOs) emerged in 1998 as the symbol for providing patents on inventions developed at universities. The law on promotion of transfer of technology-related research results from universities and other institutions to private corporations (the law for technology transfer from universities) was enacted in 1998 and stipulates conditions for TLOs to receive government designation. Activities related to industry-university cooperation expanded following the passage of this law.

In 2000, the rules and regulations of the National Personnel Authority were revised to increase willingness to set up start-ups led by academics. This enabled faculty members at national universities to also serve as corporate executives with the aim of commercialising research results. In 2002, under Prime Minister Junichiro Koizumi, the Intellectual Property Policy Outline was established, and the Basic Law on Intellectual Property was enacted. On 8 July 2003, the government’s Intellectual Property Policy Headquarters announced the Strategic Program for the Creation, Protection and Exploitation of Intellectual Property. On 15 July, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) announced 34 institutions that qualified for the University Intellectual Property Headquarters Development Programme. These changes showed that universities had entered a new era in terms of intellectual property rights.

Along with this trend of emphasising the importance of intellectual property, a ‘Japanese cluster policy,’ aimed at promoting innovation in a specific region by linking technology seeds in university or public research institutions with corporations, was initiated by two ministries during the 2000s. The Ministry of Economy, Trade and Industry (METI4) started the Industrial Cluster Initiative in 2001 and MEXT5 started the Intellectual Cluster Initiative in 2002. An example of this

---

4 Following the Central Government Reform in January 2001, the Ministry of International Trade and Industry (MITI) was renamed METI.
5 Owing to the Central Government Reform in January 2001, the Ministry of Education, Science and Culture was merged with the Science and Technology Agency to establish MEXT.
was the Kansai region, where one could observe a rapid accumulation of laboratories of large bio-related companies, bio start-ups, public research institutions and universities.

According to the 2008 report of the Japan Bio-Industry Association, there were 577 bio-venture companies in Japan as of 2007, and the landscape was heterogeneous, incorporating medical research, research support, consulting, environment, agriculture, and production of bio-molecules.

4.5 Method
To summarise the methodology developed in Chapter 3, the thematic and knowledge-based aspects of the patent application and publication data are linked to each other through similarities between the bibliographic and in-text NPLRs of the patent applications and the publication corpus of the inventor. The necessary data and processes will be further explained in the following sections.

4.5.1 Data collection
For our publication and patent data, we use the European Patent Office (EPO) patent database PatSTAT (September 2011 version) and Thomson Reuters’ (ISI) Web of Science (WoS) publication database.

The sources and types of data come from:
1. Patents - we extracted all patent applications with OncoTherapy listed as an applicant from the EPO PatSTAT database (2000-2008) of all inventors;

2. Publications - we downloaded all publications with OncoTherapy listed as an institution from WoS (all entries up to 2011); and all publications with Nakamura listed as one of the authors.

These base data were parsed using SAINT (2009) and managed in a relational database. We collected further data from the patents - specifically (where found):
3. In-text non-patent literature references (IT-NPLRs) - citations to publications within the body of the patent, but not always in the front-page reference list. These IT-NPLRs were automatically extracted from the full-text versions of the patent documents by custom software.

4. Bibliographic NPLRs (B-NPLRs) - these are citations included primarily by the examiner and added as front-page references.

We grouped the collected patent documents by INPADOC family and aggregated the data

---

6 We chose patents up to 2008 as there is considered to be a delay in the completeness of patent data in PatSTAT. 2008 was chosen as the last year as we could be more certain that it included all possible patent data.

7 English language only

8 SAINT (Science-system Assessment Integrated Network Toolkit - a Rathenau Instituut open-source software suite designed to parse, clean and organise bibliometric data to be later used in relational database software such as MS Access and MySQL.

9 INPADOC extended families are grouped by Paris Convention priorities, domestic continuations and technical relations. The INPADOC family serves to aggregate patents protecting the same or related inventions, represented by different applications over time or different patenting offices.
associated with each application to the parent INPADOC family. This was done to overcome the disparities and lack of data associated with patent documents from some patent offices. As such, we choose to view the collective patent documents and INPADOC families as representing a specific technology (Martinez, 2010). We extracted the patent application metadata up to December 2008 from the EPO’s PatSTAT database.

For each of the patent documents extracted from PatSTAT, we extracted the non-patent literature references (NPLRs) using custom software. The software identified and downloaded the full-text versions from the EPO web portal, and parsed and extracted both the in-text NPLRs (IT-NPLRs) and bibliographic NPLRs (B-NPLRs).

As far as possible, we located and downloaded the ISI WoS publication equivalents of all the NPLRs and added them to the existing publication data set. Some NPLRs could not be linked to WoS publications as they had insufficient identifying data, such as the author name only, or author name and year only, or journal name and year only. However, these were minimal as most NPLRs contained enough data to accurately link them to publications found in WoS.

The parsed publication corpora were grouped into a single relational database, recording the origins of each document within the combined set.

4.5.2 Similarity calculations and clustering

Publications
The similarities between publications (both NPLR and Nakamura’s) were calculated based on their shared cited reference and title word combinations using a method developed by Van den Besselaar and Heimeriks (2006). We constructed a network using the publications as nodes and the edges representing the degree of similarity as calculated above. The research streams of publications within the network were assigned using a community detection algorithm developed by Blondel et al. (2008). Once the initial research stream assignment was completed, the general streams were isolated and the community detection algorithm was run again to produce smaller concept clusters (Gurney et al., 2012).

INPADOC families
The INPADOC families were clustered using the International Patent Classifications (IPC) codes added by the examiner to the patent application at the time of application. The IPC classification codes are internally orientated search codes for examiners to assign patent applications to different classes. The use of IPC codes as tokens in similarity calculations to determine knowledge-relatedness is an extensively developed method (Breschi et al., 2003; Jaffe, 1986). We examined in detail the patent titles and claims associated with each patent application within each INPADOC family to determine the clinical application, general methodology and target disease. Then we recorded the resulting clusters of INPADOC families.

Linking patent families and publications
The NPLRs were co-located within the general research streams based on the level of similarity of shared title word and cited reference combinations. The shared knowledge features, such as
concepts, knowledge bases and, to a certain extent, skill sets involved can be elucidated through the degree of similarity between the publications. By linking the INPADOC families to the general publication communities in which their NPLRs are co-located, we can infer that there is at least a degree of shared knowledge features between the publication community and the citing INPADOC families. For more specific knowledge features, the second layer of concept clusters provided a finer-grained view into the communities.

The source composition of publications varies within each concept cluster. In our case study, in which Nakamura is the primary producer of the publications, each concept can potentially contain a mixture where varying proportions of source publications imply differing levels of imparted or similar knowledge features of the publications and the INPADOC families. These concept clusters include NPLRs where a) Nakamura is the author; b) Nakamura is not the author and/or Non-NPLRs authored by Nakamura but not cited by the patent document.

**Visualisation technique**

To visualise the research streams over time, we employ a method introduced by Horlings & Gurney (2012) where ‘cognitive communities’ or ‘research trails’ are isolated and mapped over time. This method of visualisation provides a clear view of what we call in this paper different ‘research streams’ and their mutual relations. This manner of visualisation also allows closer examination of an individual’s contributions during various phases of their career trajectory, such as during their PhD, post-doc and professorial phases.

Using this method of analysis and visualisation, we are able to simultaneously place Nakamura’s scientific output (publications) and technological output (patents) within the same frame. The linking method of clustering the NPLRs and researcher/inventor publication corpus allows us to comment directly on the similarity between the scientific work undertaken by Nakamura and the technological output of which he is a primary inventor.

**4.6 Results**

**4.6.1 Patents and patent families**

In total we collected 242 patent application documents via PatSTAT with Nakamura listed as inventor and OncoTherapy as assignee. The patent documents came from 90 INPADOC families, and were composed of 115 priority patents. The priority patent applications were primarily filed in the USA (101 applications) and the rest in Japan (14 applications). The earliest patent filing date was March 2000, and the latest was November 2008. The maximum, minimum, average and median numbers of patent applications per INPADOC family were 23, 2, 5.3 and 4 respectively.

**4.6.2 Clustering of INPADOC families by IPC**

Three primary INPADOC clusters were found, using main group IPC data. The growth in the number of patent applications per INPADOC cluster is shown in Figure 1(a) and the count per year of unique families in each INPADOC cluster is shown in Figure 1(b). In Figure 1(a), the number of patent

---

10 Priority patents have been defined in this study as patent applications that have no earlier priority date.
applications in clusters 1 and 3 increased, whilst cluster 2 showed little increase. From 2003, this pattern reversed and the number of patent applications in cluster 2 eventually overtook clusters 1 and 3. This trend once again reversed from 2004 when cluster 1 became the dominant cluster until 2006 when cluster 3 overtook it. From Figure 1(b), in 2002 and 2004, the number of unique INPADOC families increased at a slower rate, suggesting a period of specialisation within OncoTherapy. Taken together, from 2004, the increased number of unique families and the increased number of patent applications suggest a period of diversification for clusters 1 and 3, whilst research in cluster 2 decreased overall. From 2006, there was less patenting overall, but a similar number of unique INPADOC families, suggesting that there was still diversity overall in the fields of technology being addressed, but no visible expansion in diversity.

**Figure 1(a)** INPADOC cluster patent application count.

---

**Figure 1(b)**

![INPADOC cluster patent application count](image-url)
Figure 2 shows a 2-mode network of the aggregated INPADOC clusters and main group level IPC technological areas. Clusters 1, 2 and 3 contain technologies from eleven shared main group level IPC code areas, whilst clusters 1 and 3, and clusters 2 and 3, share seven and zero IPC code designations respectively. The 2-mode network in Figure 2 demonstrates the specific areas shared by each INPADOC cluster but also serves to highlight which clusters have specialised technological areas that are only applicable to each cluster. INPADOC cluster 2 does not address any unique areas, whereas both clusters 1 and 3 do.
As Figure 2 shows: at the main group IPC levels, the subject areas addressed by the INPADOC clusters relate primarily to the use of micro-organisms, enzymes, peptides and growth factors, recombinant DNA technologies and medicinal preparations using the peptides and RNA.

The overall description is in line with OncoTherapy’s official stated research theme: the development of technologies related to gene expression analysis, identification of target genes, cancer peptide vaccines, antibodies for treatment and diagnosis, small molecule drugs and RNA medicines.

4.6.3 Publications and NPLRs
The methodological approach taken in this study considers, in tandem, the publications of Nakamura and the NPLRs found in the patents of OncoTherapy. The wider body of Nakamura’s publications is discussed first, and followed by the NPLRs found both in the bibliographic (B-NPLR) section of the patent applications and the in-text references (IT-NPLRs).
**Nakamura publications**
Nakamura has published extensively with 931 publications over 33 years. His first publication was in 1977 and he published less than five publications per year until 1987. Between 1988 and 1994, he published between 5 and 10 publications a year and in his current phase, his publication count jumped to about 50 a year. His earliest phase of publishing coincided with his MD and PhD, and research fellowships. His middle phase was composed of an assistant Professorship at the University of Utah and becoming Head of Department at the Cancer Institute in Tokyo. His current phase coincides with his Professorship at the University of Tokyo and his directorships of both the RIKEN Center for Genomic Medicine and of the Human Genome Center at the University of Tokyo.

**NPLRs**
In total we were able to isolate 2037 NPLRs from the 242 patent applications. Of these 2037, there were 842 unique NPLRs. We were able to successfully match 525 of these NPLRs with the WoS. Of the matched NPLRs, there were 147 unique NPLRs found only in the bibliography, 313 unique NPLRs found only in the text and an overlap between the two of 65 NPLRs. Of the 525 matched NPLRs, 259 NPLRs are cited more than once, 73 are cited 5 or more times and 20 are cited 30 or more times. The most cited publication is cited by 41 different patent applications. The most cited publications come from the time period of 1996-2004 with less than 10% of NPLR citations going to publications older than 1996.

**Publication clustering**
We clustered the publications as described in the methods section. The resulting clusters represent research streams that differ in size and duration. The largest research stream (by count of Nakamura publications) is stream 2. Streams 1 and 2 contain both high numbers of Nakamura's publications and NPLRs. 55 (10.5%) of Nakamura's publications are cited as NPLRs - 19 in-text NPLRs, 18 bibliographic NPLRs, and 18 cited both in-text and bibliographic.

Streams that are almost exclusively NPLR-based include streams 4, 8, 9, 10 and 12. Nakamura's publications are the only publications in streams 15, 16 and 18. There are a few streams composed of only two publications, and these are 3, 6, 17 and 19. Stream 3's two NPLR publications are extremely highly cited (approximately 40 000 citations each). Table 1 provides a summary of the streams including the relative presence of NPLRs in each stream.
Figure 3 shows a similarity network of the publications and NPLRs to demonstrate the longitudinal aspects of the streams and the degree of similarity between them. Figure 3 and Table 1 reflect the same data but Figure 3 provides an overview of how the different streams are linked. The lines between the nodes represent citation relations. Figure 3 identifies Nakamura’s publications, the NPLR, and papers that belong to both these categories. As table 1 shows, only two of Nakamura’s research streams (7 and 13) are cited frequently by the patent applications, and two others are cited very incidentally (2 and 5). The other NPLR is located within streams that also contain (many) Nakamura publications.

Stream 18 is a large set of publications (183 in total) exclusively by Nakamura, which show little similarity to any of his previous or concurrent works. Finally, stream 7 is interesting in that the NPLRs precede the stream, with no similar publications by Nakamura. However, 5 years after the start of the stream, more and more publications by Nakamura appear and are increasingly cited by the patent applications.

The data found in Table 1 and Figure 3 are used for descriptors 1 (the reputational and applica-
Figure 3  Longitudinal research stream clustering of Nakamura and NPLR publications

Note: edges between publications signify degree of similarity based on shared combinations of title words and cited references. Nodes have been coloured to indicate source.
bility aspects) and 2 (the overall research trajectory of Nakamura overlapping with the required knowledge of the technologies). Nakamura’s research is cited prominently in two research streams and co-occurs in multiple streams. His research may be considered vitally necessary in streams 7 and 13 (from Table 1), and his overall research trajectory is well embedded in the technologies (from Figure 3). There are some aspects of the technologies that do not draw extensively from Nakamura’s expertise, such as those found in streams 4, 8 and 9 (from both Table 1 and Figure 3). The research streams in which Nakamura’s work is cited by the patent applications are also widely cited by other publications (Table 1).

4.6.4 Patents and publications
We constructed a 2-mode network map using the research streams cited by the INPADOC family IPC clusters. This is presented in Figure 4. The publication communities in which the cited NPLRs are co-located are represented by circles, and the square nodes represent the INPADOC clusters. INPADOC clusters 1, 2 and 3 have NPLRs co-located within seven research streams; whilst INPADOC clusters 1 and 3 have four shared streams. INPADOC cluster 2 does not link to any unique publication communities.

Figure 4  Annotated 2-mode map of INPADOC family clusters and research streams

Note: streams not cited by patent applications are shown in the top-left corner. For INPADOC clusters, size of node indicates count of INPADOC families. For research streams, size of node indicates count of publications in stream. Edge weight is the proportional count of NPLR citations from INPADOC clusters to streams. Annotations are summarised from the most used title words and journal names/categories for each research stream.
These results address specific descriptors, namely (1) the reputational aspect (as defined by number of citations) and applicability aspect (as defined by the proportion of the researcher/inventor’s work cited by the patent applications versus their total corpus) and (2) Nakamura’s trajectory being co-located in the fields of research necessary for the technologies. Table 1 gives the average citations for each of Nakamura’s publications. In four streams, his publications are cited over 45 times, and in seven of the other streams his publications are cited at least 25 times on average. Just over 10% of his publications are cited by the patent applications, and are found in both the IT-NPLRs and B-NPLRs. Even if Nakamura’s publications are not cited by the patent applications, they are intimately co-located within the same topic and background as the NPLRs. Aspects of Nakamura’s research are utilised in all three INPADOC clusters, and there are some sub-fields of his research that are utilised by only one, or two, INPADOC clusters. Other fields/sub-fields such as research streams 8, 10 and 12, are used for the technologies (as cited by INPADOC clusters 1 and 3) but are not part of, or similar to, Nakamura’s research. This indicates that creating these technologies also requires knowledge that is outside of the expertise of Nakamura.

Co-inventors and partner institutes
Figure 5 shows the distribution of Nakamura’s co-inventors in the publication corpus. Many of Nakamura’s publications are co-authored by his co-inventors. Some of the NPLR-cited publications by his co-inventors do not include Nakamura as an author. This could indicate that the knowledge utilised by the patent applications stems not only from Nakamura, but also from his co-inventors. However, the relative scarcity of NPLRs that do not include Nakamura as author but with one of his co-inventors authoring instead would suggest that the knowledge does come from within Nakamura’s own research group.

Within the 77 INPADOC families that list OncoTherapy as assignee and Nakamura as inventor, Nakamura has 10 recurring co-inventors, and 4 of these co-inventors also patent without Nakamura. OncoTherapy has 6 researchers that patent without Nakamura, but the vast majority of INPADOC families primarily stem from patent applications with Nakamura listed as inventor.

OncoTherapy only collaborates on patents with two organisations: the University of Tokyo in 26 different INPADOC families, and Sentan Kagaku Gijutsu Incubation Center in 1 INPADOC family. The University of Tokyo is present in just under a third of OncoTherapy’s INPADOC families, which - considering Nakamura is based at the university - does not seem particularly high.

To address the research question through descriptor (4) - the role of collaborators in the development of the technologies - we conclude that whilst there is input from Nakamura’s co-inventors in the NPLR, the number of NPLRs authored without Nakamura is very low. However, we do find instances of research conducted by collaborators (stream 9 from Figure 5) necessary for all three INPADOC clusters. In this instance, whilst Nakamura does not possess the necessary expertise, his collaborators fill the necessary gap.
Figure 5  Co-inventor location in research streams

Note: Only edges between patent applications and cited publications are shown (both IT-NPLR and B-NLPR.
4.6.5 Concept clusters
From the 19 research streams, we extracted 66 concept clusters that contain NPLRs. We linked these to the citing INPADOC families and the designated INPADOC clusters. Presented in Figures 6(a)-(c) are citations to concept clusters from the INPADOC clusters. The three figures show different concept clusters being cited by the patent applications within the INPADOC clusters. Figure 6(a) shows the concept clusters that contain no publications authored by Nakamura. Figure 6(b) shows concept clusters containing publications cited by the patent applications but not authored by Nakamura, and publications authored by Nakamura but not cited by the patent applications. Finally, Figure 6(c) shows concept clusters that contain publications cited by the patent applications and are authored by Nakamura. This was done to clarify the specific research ideas and associated knowledge and skill sets that are necessary for the technologies, rather than using the broad research streams, and the presence of Nakamura's publications within these streams. The presence (or lack thereof) of Nakamura’s publications within a concept cluster indicates at a specific level Nakamura’s contributions to the technologies, in terms of direct citations, and through similar knowledge and skill sets. The timing of Nakamura’s publications’ entry into the concept clusters is also important. If Nakamura’s publications were present from an early stage we can assume that the necessary knowledge for the technologies was always present within Nakamura from the beginning. Following this line of reasoning, if publications by Nakamura appear later on in the concept cluster, we assume that Nakamura realised the importance of conducting his own research in the topics that he deemed to be necessary for the technologies. We can also assume that through performing the research, he gained a greater understanding of, and thus ability to conduct, further necessary research.

Figure 6(a) shows that the patent applications located in all the INPADOC clusters rely heavily, and from an early stage, on stream 9, especially concept cluster 0 (research related to the cytotoxic effect of lymphocytes and leucocytes in human cells). Similarly, the research in stream 9 (CC9/1), cited by INPADOC clusters 1 and 3 addresses the same general topics of lymphocytes, melanomas, peptides and antigens, but focuses more specifically on the characteristics of the human leucocyte antigens (HLA) such as identification, populations and susceptibility. INPADOC cluster 1 is the only cluster to cite research from stream 7 (CC7/2, increasing rates of bile duct cancer) and from stream 1 (CC1/2, mRNA binding proteins expression and cancer proteins).
What is important to take away from Figure 6(a) is that the patent applications are citing research outside of Nakamura’s immediate expertise. The cited publications may be providing background or introductory information (as in stream 7, CC7/2), and these publications do not link to any of Nakamura’s publications through their title words and cited references.

Figure 6(b) shows concept clusters containing publications cited by the patent applications but not authored by Nakamura, and publications authored by Nakamura but not cited by the patent applications. This combination of sources indicates that there is some immediate similarity between research performed by Nakamura and the cited publications. Compared to the relative sparseness of Figure 6(a), the figure contains many more concept clusters. In many cases, the research is cited from an early stage (grey edges) but there is a fair degree of research cited later in the technologies’ development phases (dashed edges).

Concept clusters from stream 9 (CC9/2, CC9/3 and CC9/4) are cited early by all three clusters, but Nakamura only starts to publish much later in these topics (indicated by shading of the concept cluster nodes). We can confirm this by examining stream 9 in Figure 3, which shows a large corpus of NPLRs, with Nakamura only appearing as author 10-15 years after the date of the first NPLR in that stream. Fig 6(b) shows that he started publishing late in most of the relevant cited concept clusters of stream 9. All three INPADOC clusters cite research in stream 1 (CC1/4 and CC1/5), but again Nakamura’s publications related to those topics are only published later.

Stream 1, CC1/1, is cited exclusively by INPADOC cluster 1 in the middle phase of its development (dashed edges). Notably, Nakamura - whilst having published extensively in that concept cluster - is not cited at all.
Figure 6(b) shows the concept clusters in which publications by Nakamura are directly cited by the patent applications. These are considered to contain the most specifically necessary aspects of Nakamura's research for the technologies encompassed by INPADOC clusters 1-3. In most cases, the INPADOC clusters cite them from an early stage (as indicated by the grey lines) but in several other cases Nakamura did not start publishing in the topics from the beginning. This is the case with stream 13 (CC13/4), stream 7 (CC 7/1, CC7/4), and stream 2 (CC2/0). This is a strong indicator that Nakamura at some stage recognised that he had to perform his own research in these topics. As the concept clusters were cited before Nakamura began publishing in them, we attribute this observed pattern to Nakamura's ability to acquire and assimilate the required knowledge - and recognise the need to do so.

The entry period of Nakamura's publications and NPLRs to the overall publishing landscape (as found in Figures 6(b) and (c)) is important. If Nakamura is present from the beginning of the concept cluster date range (indicated by white nodes), we assume Nakamura possessed the knowledge base and skill sets as the technology was being prepared. If Nakamura entered later, we assume he did not possess the knowledge and skill sets at the time, but acquired them later.
The two final descriptors used to address the research question can now already be seen: (5) the degree of knowledge features utilised by the technologies that come from Nakamura in relation to other sources; and (6) the incorporated and applied skill sets acquired during the development of the technologies. We shall discuss a few typical examples: research streams 1, 7, 9 and 13.

In stream 1, Nakamura is not cited by the patent applications at all. However, he publishes extensively in the sub-field represented by the stream, and he was present at the start or early stages of all but one of the concept clusters cited by the INPADOC clusters. CC1/2 is the only part of stream 1 where Nakamura does not have a presence. This is reflected in Figure 6(a). Two (CC1/4 and CC1/5) are cited by all three clusters; one (CC1/3) is cited by clusters 1 & 3; and one (CC1/1) is cited exclusively by INPADOC cluster 1. From this information, the degree of exogenously-generated knowledge is high in stream 1, with no direct contributions by Nakamura. However, the shared knowledge base and shared minimum skill set is also high because only one of the five cited did not contain any Nakamura publications.

Stream 7 is initially hardly cited by the INPADOC clusters. Only CC7/3 and CC7/4 are cited from an early stage by INPADOC cluster 2, and CC7/0 by cluster 3 (see Figure 6(c)). Confirming this, from Figure 3, up to 2004 the first cited NPLRs were all non-Nakamura NPLRs. From 2004 onwards, almost all of the NPLRs cited in stream 7 were authored by Nakamura, and he went on to publish prolifically on that subject. Of the NPLRs that Nakamura authored, all three INPADOC clusters cite stream 7 concept clusters varyingly. Nakamura does play a direct role in the technologies of the clusters as evidenced by the proportionally large number of Nakamura-NPLRs. Nakamura’s
knowledge base and skill sets, whilst being developed at a later stage have now become integral to the technologies.

The role of Nakamura in the subject areas of stream 9 is limited. He has no NPLRs in stream 9. He has a limited number of publications in all of the concept clusters cited and in two of these (CC9/2 and CC9/4) his publications appear only in later stages. Additionally, the technologies in the clusters cite the stream extensively, as seen in Figure 6(a). However, his co-inventors are cited in the NPLRs. As such, the necessary scientific aspects derived from stream 9 are outside his expertise but from within his scientific network. Stream 9 NPLRs are cited extensively by all three INPADOC clusters, suggesting that the research published in community 9 may be more than base or legacy research, i.e. research that is required to be cited by necessity in a historical development sense (an example of this is fundamental research conducted many decades prior).

As seen in Figure 3, Table 1 and Figure 6(c), stream 13 is dominated by publications authored by Nakamura. Almost a third of his publications are cited by the patent applications. From Figures 6(a)-(c), three (CC13/4, CC13/1 and CC13/0) are cited significantly by all three clusters. In one, however, (CC13/4) Nakamura is not the first to publish, with some NPLRs existing before he published in that stream.

4.7 Summary and conclusion

Professor Nakamura’s past research forms the backbone of the technologies of OncoTherapy. Given that Nakamura founded the firm based on his research at the University of Tokyo, at a superficial level this is to be expected. However, the depth of his knowledge that is utilised by the firm is extensive, and that extensiveness was only found through the methodology deployed in this paper. There are aspects of the technologies that lie outside Nakamura’s (and his co-inventors’) expertise, and Nakamura has adopted and adapted these necessary aspects into his own research and the output of the firm. In some small sense, his co-inventors have bridged the knowledge gap between Nakamura’s expertise and what is required for the technologies, but Nakamura’s uptake of these research areas has filled his expertise gap.

The research question can be addressed by means of the series descriptors developed in this paper: (1) Nakamura conclusively adds to the reputational and applicability aspects of the scientific base work of the technologies; (2) His overall trajectory is closely intertwined with the fields of research necessary for the technologies, and in many cases the trajectories of both his university research and firm application have been in lock-step, with (3) fields outside his expertise initially contributing to the technologies. (4) There is a low level of input from his collaborators, with very little overlap between his co-authors and co-inventors, with (5) only some shared knowledge features shared between them. The most important aspect of Nakamura’s links between his academic research and industrial applications is that (6) he incorporated skills acquired during research, in both his academic and industrial trajectories, and applied them to new research endeavours.

From the point of view of the absorptive capacity dimensions of Zahra and George (2002) and their respective sources of knowledge, be they generated exogenously or endogenously, we observed:
a. **Acquisition** where Nakamura’s co-inventors were considered part of the acquisition dimension of absorptive capacity as they provided their own expertise and skill sets, which added to the knowledge infrastructure in place. Nakamura, however, in his role as Professor at the University of Tokyo, is the most active member of the firm in terms of acquisition.

b. **Assimilation** processes, where ‘learning by doing’ seemed to be prevalent. By conducting research in the topic areas required for the technologies, whether through a non-concerted approach (such as in exploratory research) or a cumulative directed approach (such as in a strategic, application-driven approach to research), the newly developed skills and insights impacted the development of the technologies at OncoTherapy. This suggests a high degree of assimilation by Nakamura and his co-inventors, including tentative evidence of knowledge creation feedback between Nakamura’s work as an inventor and his research as an academic scientist.

c. **Transformation** and **exploitation**, where patent applications cited research outside Nakamura’s expertise, but later became deeply embedded in his research trajectory. In other words, the knowledge upon which the technologies are dependent was previously externally sourced, but has been incorporated, transformed and exploited by Nakamura.

On a methodological level, our approach benefits from its ability to encompass both the macro knowledge environment and the micro knowledge capture processes. Our approach can isolate and highlight specific aspects of utilised knowledge in relation to the knowledge features already locally in place. We are able to co-locate the knowledge features of individuals who contribute to the publications and patent applications, not through the direct citations of NPLRs, but through the co-location of NPLRs in their wider knowledge environment.

A possible disadvantage of our method is the complexity of the process. Due to this complexity we chose to aggregate the technologies into clusters of INPADOC families. This limits our attention to detail within the technologies but allows a thorough examination of the contributions of an individual (in this case, Professor Nakamura). The alternative strategy would be to aggregate on the publication side and examine in detail the characteristics of the technologies being produced. This is partially done in figure 2, where we identify the relevant patent families in more detail. To do both at the same time would require more space than is possible in this publication.

Our method makes it possible to position technologies and the knowledge contributions of those involved in the development of those technologies. With the addition of funding information in the metadata extracted from WoS, it would be possible to trace the results of such funding to their exploitation. The method can also be used to examine firms that have more than one bridging scientist, and operate between multiple universities or firms. The scaling up of this method allows research groups, departments or entire research institutes or infrastructures to map their contributions in the early stages of the development of a technology right through to its exploitation or implementation. This would be useful for funding agencies and universities for reporting on their research achievements, as in many cases the end-point of fundamental and applied research may be so far removed from the origin as to be unrecognisable.
4.8 References


Martinez, C. (2010). *Insight into different types of patent families*: OECD.


