Patent mining and landscaping of emerging recombinant factor VIII through network analysis

Cristiano Gonçalves Pereira[1], Virgínia Picanco-Castro[2], Dimas Tadeu Covas[2] & Geciane Silveira Porto[1,3]

A landscape analysis of the recombinant factor VIII patent sector using network analysis to map the cooperation network among assignees shows emerging technologies that focus on improved methods for rFVIII production and bioactivity performance.

The standard treatment for patients with hemophilia A is replacement therapy with intravenous infusion of a plasma-derived concentrate or recombinant factor VIII (rFVIII)[1]. Although there are several products currently available and included in the treatment protocol, many challenges still remain.

In the past five years some efforts to extend the half-life of rFVIII have been made, and these have included the use of PEGylation. The GlycoPEG-FVIII (Turoctocog alfa) has a truncated B-domain of 21 amino acids; it was the first product used in clinical trials[2]. Two years later a PEGylated full-length rFVIII, PEG-rFVIII (BAX 855) was designed to increase half-life and reduce the frequency of infusions[3]. Other products were developed by fusing rFVIII with the Fc region of human antibodies (rFVIII-FC)[4] or changing the FVIII structure to create a stable single-chain rFVIII (rFVIII-single-chain)[5] (**Fig. 1**). These products demonstrated a slight increase of FVIII half-life in circulation. More recently, non-coagulant factors have been developed to achieve hemostasis in patients with hemophilia, such as ACE910, a recombinant humanized bispecific antibody that binds to activated factor IX and factor X and mimics the cofactor function of FVIII managed to show a half-life of four to five weeks[6].

Therefore, improvements in the pharmacokinetic profile and efficacy of FVIII products, as well as in their safety and convenience of administration, continue to be the focus of bioengineering. Technological advances have led in many instances to new patented molecules that have presented surprising results in clinical trials, relieving the heavy burden for the patient in the treatment of hemophilia[7].

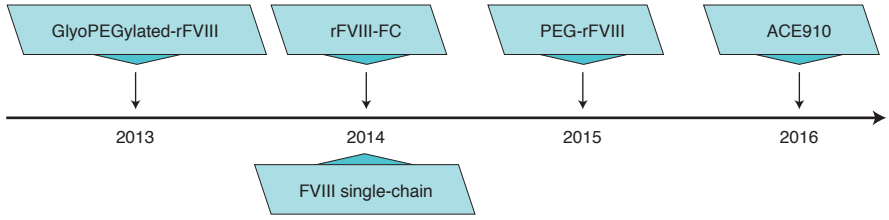

Figure 1 New FVIII molecules targeting increased half-life by the first report year. The molecules GlycoPEG-FVIII[2]; rFVIII-FC[4]; rFVIII single-chain[5]; PEG-rFVIII[3]; ACE910 (ref. 6) aimed to extend the half-life of rFVIII.

The development of better and more innovative technologies, not only in recombinant proteins, is robustly leveraged by the establishment of cooperation among institutions. An interesting approach in studying cooperation is taking advantage of social network analysis (SNA), whereby the flow of information from cooperative structures can be interpreted by SNA metrics and graphs, which have been extensively applied in innovation studies[8–10]. Since key indicators used by researchers to evaluate research and development and innovation results can be found in patent data[11,12], information from the assignees and citation links are useful in designing networks that can bring insights to the efforts promoting innovation and the knowledge flows of the technology development process. The bibliometric analysis based on patent citation (co-citation) is an important tool to forecast emerging technologies[13,14], and the study of the relevant paths in citation networks can depict technological trajectories of great relevance in the network context[15–17].

Here, we mine patents related to rFVIII used in hemophilia treatment to design a patent landscape of the sector indicating technological trends, and use the SNA approach to map a cooperation network among patent assignees. The same approach has also enabled us to forecast and build the trajectories of emerging technologies based on the main path of patent citations. The information resulting from this study may be of value to both R&D managers and researchers in the field by identifying the most relevant technologies and giving an overview of the rFVIII inventions in the last 20 years.

Results and discussion

The group of rFVIII patents studied was selected from the Derwent Innovation platform using a search query specific to the subject, and comprised 3,424 patent families dealing exclusively with medicinal preparations and peptides of rFVIII (**Supplementary Methods**). The technological fields represented by IPC (International Patent Classification), covered by the rFVIII patents selected in this study, validated our data set. The most frequent fields

[1]*School of Economics, Business Administration and Accounting, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.* [2]*Center for Cell-based Therapy CTC, Regional Blood Center of Ribeirão Preto, University of São Paulo. Ribeirão Preto, São Paulo, Brazil.* [3]*Institute of Advanced Studies of the University of São Paulo IEA/USP, Ribeirão Preto, São Paulo, Brazil.*
e-mail: geciane@usp.br

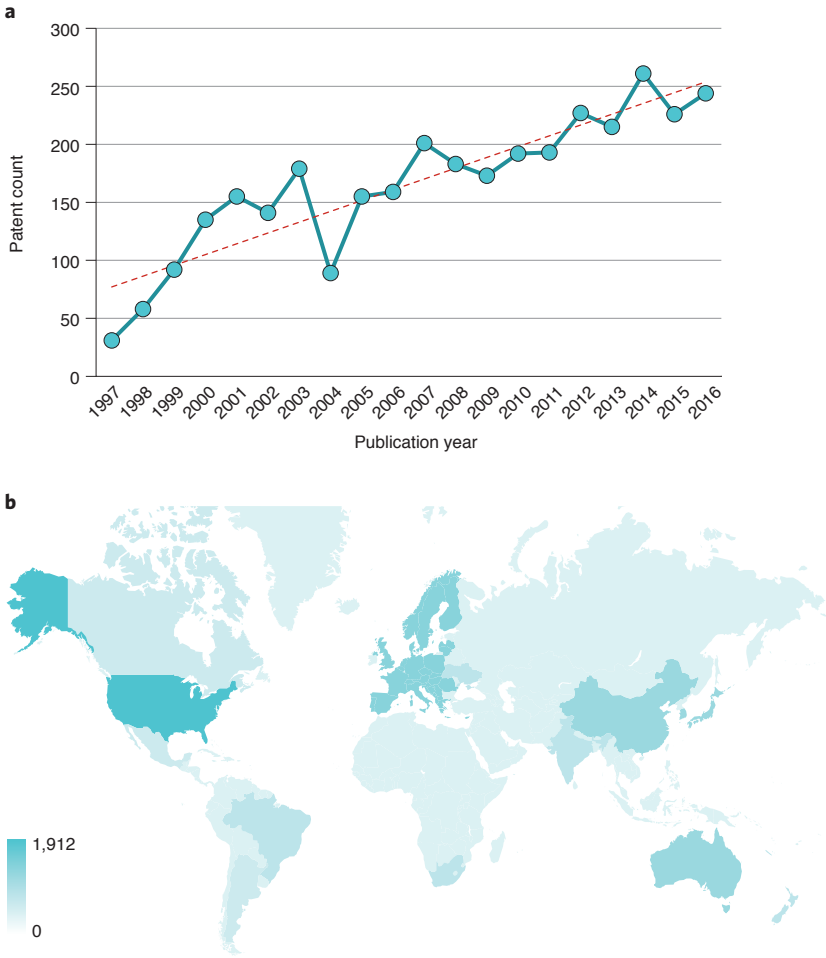

Figure 2 Publication trend and geographical coverage of rFVIII patents. (**a**) The number of patents by publication year showing an increased trend in rFVIII inventions since 1997. (**b**) Geographic coverage of patents in each jurisdiction's patent office according to priority country code of the patent records. The color intensity denotes the number of patent families. Deposits as PCT are not shown.

concerned factor VIII peptides and its medicinal preparations, drugs for disorder of blood, and recombinant DNA technology (**Supplementary Fig. 1**). Only 14% of rFVIII patents from the data set have been granted (n = 490), and 85% of them are still under examination (n = 2,905). A similarly low percentage of granted patents was also observed in a recent patent landscape study of induced pluripotent stem cell inventions[18] and patents dealing with CRISPR–Cas gene editing technology[19]. Among many factors that contribute to the lag in granting approval, a very important one is the number of claims per patent, as a larger number of claims is directly linked to a greater delay in examination, since each claim must be checked and validated by the examiner[20,21]. In our analysis, we found 18% of patents had more than 50 independent and dependent claims, which has likely impaired the speed of granting patents (**Supplementary Fig. 2**).

The patent publication trend has a rising tendency over a period of 20 years, starting from 31 published patents in 1997 to 241 in 2016 (**Fig. 2a**). More than half of the patents were first deposited in the United States (n = 1,912), the country with the highest number of patients worldwide (**Fig. 2b**). The state members of the European Patent Office come in second place, with 302 patents representing 9% of the total. Japan is the third priority country, that is, the country where the patent is first filed before being extended to other countries, by number of deposits (n = 175). India, the country with the second highest number of hemophilia patients[22], showed 15 patents as a priority country. There was a considerable number of PCT (Patent Cooperation Treaty) deposits with priority international application (n = 45).

Regarding the top patent inventors, Peter Turecek holds the leading position with 74 patents followed by Juergen Siekmann with 43 patents (**Fig. 3a**). Both are researchers at Baxalta. Among the main patent assignees, Baxter appears in first place with 287 patents (8.4%), which includes 87 patents as a single assignee and 200 in cooperation with other assignees (**Fig. 3b**). Baxter develops products for the treatment of hemophilia, kidney diseases, immune diseases and other chronic and acute medical conditions, maintaining intensive joint R&D efforts with Baxalta, which is responsible for 70% of their patented rFVIII in cooperation. Baxalta (now part of Shire), which was spun off from Baxter and inherited its parent company's hemophilia treatments, is the third major assignee with a total of 155 rFVIII patents, together with Baxter. The second largest assignee is the Danish pharma Novo Nordisk, with 222 patents (6.5%), showing a rate of cooperation of 31%; the company recently launched Novoeight, an FDA-approved rFVIII with more stability and less immunogenicity than other products. The German company Bayer sits in fourth with 149 inventions (4.35%), 41% of them being in cooperation. More than 37% of its patents in cooperation were with Maxygen, a biopharma focused on developing improved versions of protein drugs which was acquired by Bayer in 2008. Biogen comes in at fifth with 74 patents (2.16%), including 23 patents in cooperation (31%) (**Fig. 3b**).

Among the top 15 assignees, only one—Fresenius Kabi, a German healthcare company—does not establish cooperation measured by co-ownership of patents. The top five assignees showed rates of cooperation above 30%, which means that an intensive collaborative effort to generate new technologies is important. Five of these companies—Baxalta, Neose Technologies, Aventis Behring, Centeon Pharma and Novozymes—develop their patent portfolio mostly by establishing partnerships. Considering this point, we sought to analyze the cooperation networks that resulted in the development of rFVIII technologies that led to patent applications. To this end, we generated the cooperation network from a bipartite graph, which included nodes of both patent and assignee categories (**Supplementary Fig. 3**). Baxter, Novo Nordisk, Bayer and Baxalta stood out in the network due to the high volume of patents.

The full cooperation network, derived from the bipartite graph, has 192 nodes (assignees) and 300 edges (collaboration among assignees), consisting of 163 connected components (**Fig. 4a**). From this network, we filtered its giant component (GC) to select the most interconnected group of assignees (**Fig. 4b**). The cooperation intensity between two assignees is represented by the edge thickness. We find

that Baxter has worked on many joint efforts with Baxalta, as has Novo Nordisk with Neose Technologies (they have a thicker edge). The larger the nodes the greater is the diversity of technological partners. Baxter, Aventis Behring, the US Department of Health & Human Services, INSERM and the American National Red Cross are those that have established more partnerships.

In the cooperation network, we identified 14 communities, which are each shown in a different color (**Fig. 4b**). Using the assignee's information, it was possible to predict some interesting features. The dark green community was formed mostly of European companies, mainly in France, including INSERM (Institut National de la Santé et de la Recherche Médicale) cooperating with many different institutions. The light green community has Aventis Behring, which cooperates with organizations that are currently its subsidiaries. The pink community is constituted mainly of US government institutions and universities. The blue community is characterized by Baxter and its partners, showing the strong link with Baxalta.

Supplementary Table 1 shows the top ten assignees in order of number of distinct partners (degree), and includes information regarding the intensity of the connections, represented by the weighted degree. For instance, Novo Nordisk has fewer connections than Aventis; however, the intensity of its connections is stronger than all the assignees below Baxter. These data demonstrate that, for the matter of cooperation, not only the number of connections itself is relevant, but the strength of the links must also be considered. The measurement of eigenvector centrality (EC), suitable for finding prominent or key authors in relationship networks[23], showed that although Baxter is a player with the highest degree and weighted degree score, its EC score is not so prominent (0.59), and so neither is its influence in the network. However, Aventis Behring has the maximum EC score, followed by INSERM (0.71) and CNRS (Centre National de la Recherche Scientifique) (0.62), meaning these players have the biggest influence in the network despite having fewer connections than Baxter, as they are connected with other influential actors.

To be able to retrieve valuable information about the content of the patents, we used large-scale text and data processing analysis based on text clustering and the ThemeScape thematic text-mining tool, to discover meaningfully implicit subjects throughout all the patent documents. The text clustering list represented by the word cloud (**Fig. 5a**), showed the most frequent

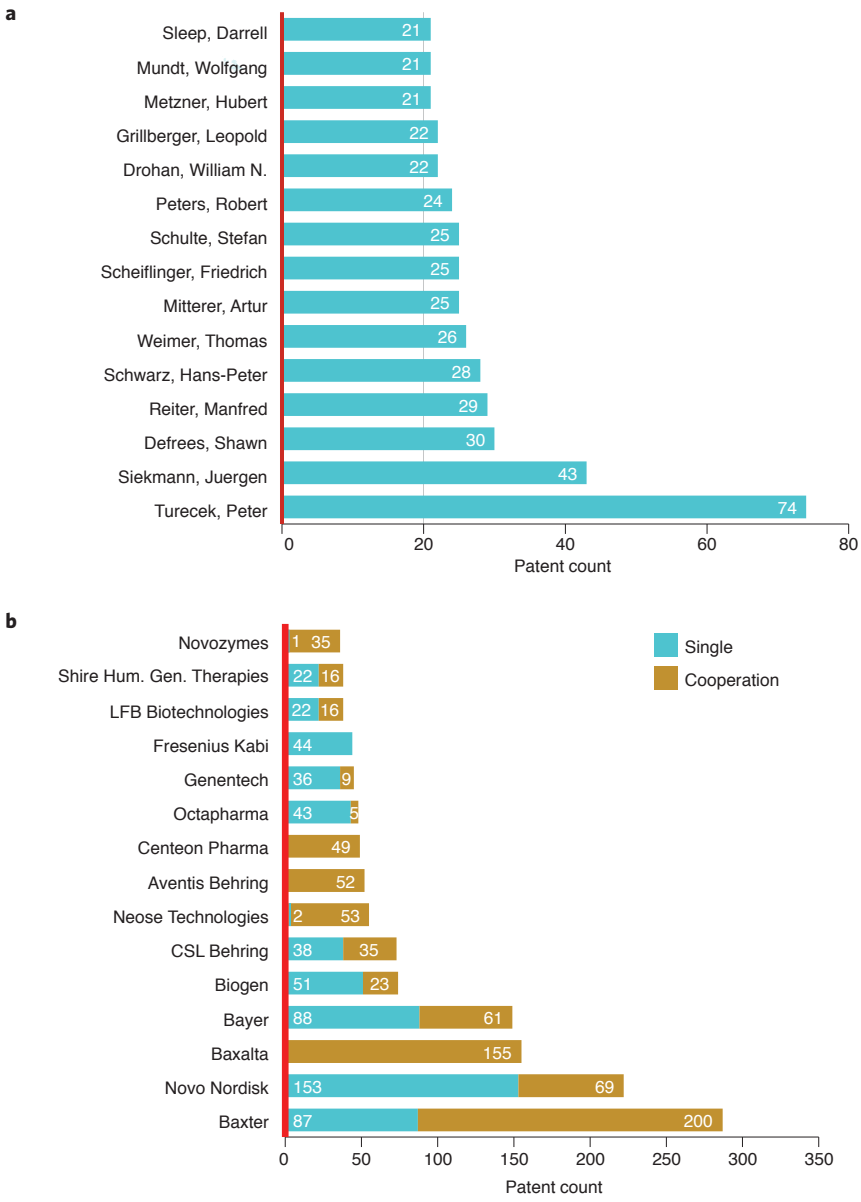

Figure 3 The top 15 inventors and assignees of patents related to rFVIII. (**a**) Top 15 inventors in terms of numbers of patent family applications. The first two main inventors are researchers who work for Baxter–Baxalta (now Shire). (**b**) The top 15 assignees in terms of number of patent families split by the proprietary technology as single assignee and co-assignee patents or in cooperation show that most of the companies established partnerships that resulted in patent application. Patents from Baxalta are entirely in cooperation with Baxter.

set of words comprising "vector, gene, promoter" (n = 336), including the terms "cassette, expression cassette, lentiviral" in the subgroup (**Supplementary Table 2**), which means that the majority of inventions deal with recombinant proteins, which use DNA vectors containing specific promoters for stable gene expression. The second set of words with a high frequency was "polypeptide, variant, amino" (n = 280), with subcategories "variant, FIX, FVIIA" and "albumin, fusion, fusion protein" having a higher frequency than others. The complete text clustering list is shown in **Supplementary Table 2**.

The patent landscape shows fields with high patent (dots) activity labeled by the peaks "Conjugate Polymer" (Amino Acid Modification Relative), "Flow Fluid" (Tissue Culture Fluid), "Gene Expression" (AAV Vector), "Inflammation Rheumatoid" and "Cancer Fragment" (Rheumatoid) (**Fig. 5b**). An emerging area was identified in the map formed mostly by recent patents published

between 2012 and 2016 (n = 110), which are related to "Covalent Complex" peak (**Fig. 5b**, yellow inset). The "Polypeptide Host" mountain within this region represents a new field that arose more recently. Most of the patents in this field are from Biogen (**Supplementary Fig. 4**) and deal with chimeric, polypeptide and conjugated molecules of rFVIII intended to improve protein half-life. In this regard, Biogen has recently launched a new rFVIII and Fc fusion, commercially named Eloctate, which is a conjugated molecule of modified FVIII and the Fc domain of human IgG, and has a half-life two times longer than that of rFVIII itself[24]. Moreover, Biogen researchers in cooperation with Amunix developed another conjugated molecule of FVIII, which includes the Fc domain with an addition of a polypeptide named XTEN, known to prolong the half-life of fused proteins[25], and a thrombin cleavage site. The new conjugated molecule achieved a fourfold extension of half-life compared to conventional FVIII in hemophilia A mice[26].

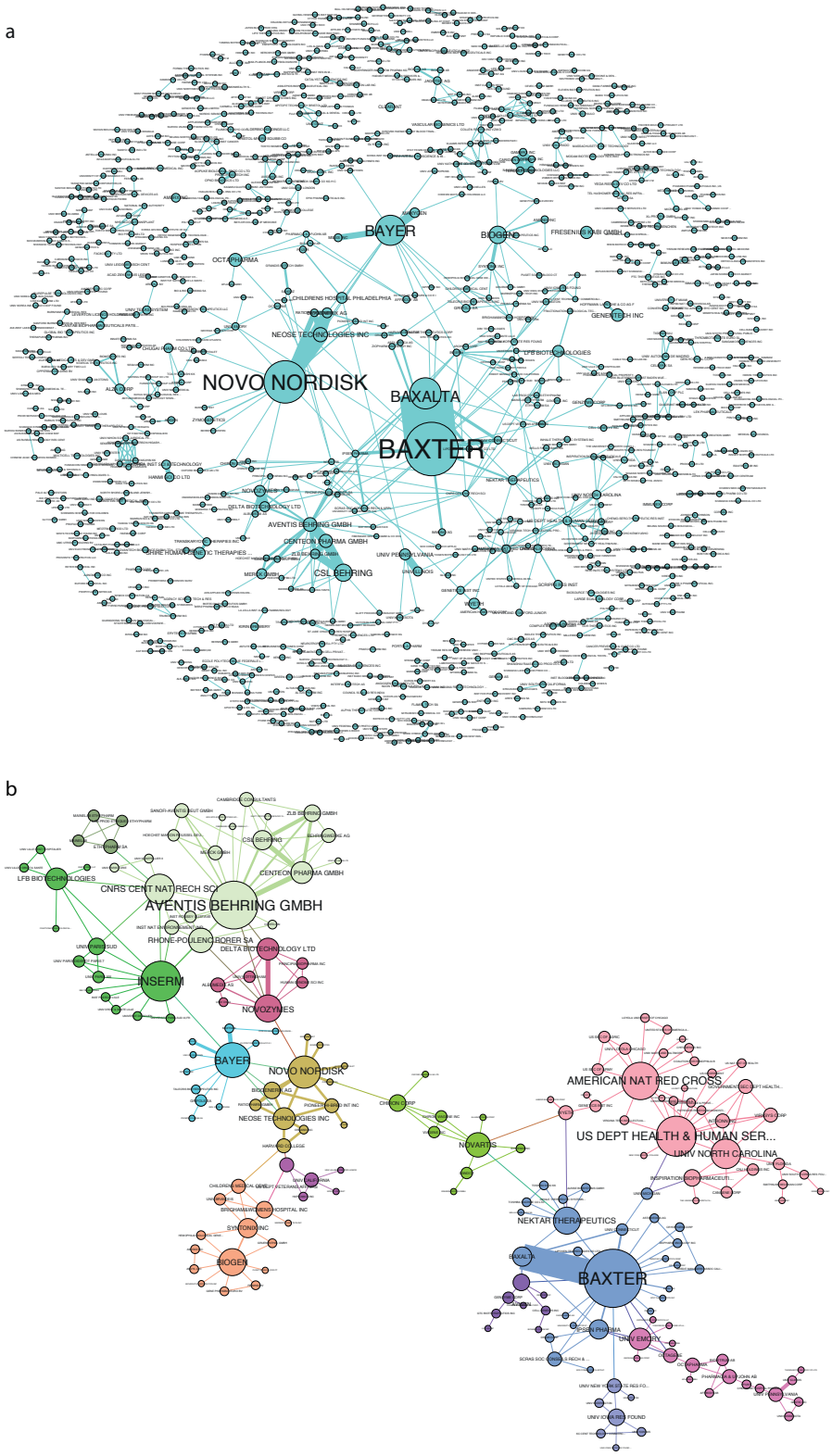

Figure 4 Cooperation network among assignees of rFVIII patents. (**a**) A full cooperation network among assignees, with 163 components. (**b**) The isolated giant component of the cooperation network showing different communities identified by the metrics of modularity.

Proceeding with the investigation for emerging technologies, we built the principal technological route derived from the patent citation network through the mapping of the main citation path by the algorithm Search Path Link Count (SPLC)[15,17,27,28] to uncover technological trend clusters in rFVIII patents. The global network of patent citations consists of 13,316 nodes and 15,045 edges (**Fig. 6a**). The node size was set according to its in-degree value (cited patent), that is, the greater the in-degree, the larger the node size, and the more citations a given patent received. The largest community, colored in green, represents 28.8% of the nodes, the second, in orange, comprises 15.2% of nodes and the third, in purple, 9%. The green community includes patents with many bigger nodes, which highlights its importance to the network since it gathered the most cited patents. The main technological route calculated by SPLC shows a directed, linear and temporal path (latest technology to oldest) with higher relevance to the network (**Fig. 6b**) and encompasses 19 nodes and 20 edges, derived from the largest community. Patents included in the top of this path can likely be considered in this study as having a high likelihood of becoming emerging technologies in the near future.

The analysis of the technological route reveals some interesting information about trends in rFVIII inventions (**Supplementary Table 3**). The route starts with two recent PCT applications, one published in 2016 and the other in 2015, both from Biogen. The most recent (WO2016070152A1) describes the use of hypotaurine, gaba, beta-alanine and choline as supplements for cell culture medium to support the large-scale expression of the polypeptide of interest, such as rFVIII. The other starting node of this trajectory (WO2015120056A1) is related to improved methods in the separation of recombinant

polypeptides with post-translational modifications from complex mixtures. Both patents cited one from Amunix, published in 2011 (US20110046061A1), related to compositions comprising coagulation factor VII linked to XTEN (which prolongs half-life in fusion proteins) for treatment of coagulation-factor-related disorders. Despite the invention not being related directly to rFVIII, the claims indicate that an effective amount of this product is sufficient to bypass the need for exogenously administered FVIII to yield a comparable therapeutic effect. Thus, the patent hereby proposes an improvement in hemostatic efficacy and stability of rFVII molecules in order to optimize the dosing to bypass the need for rFVIII and to treat patients that have developed antibodies to FVIII[29].

The subsequent patents include inventions related to improving the *in vivo* stability of rFVIII as well as the development of less immunogenic products. Four patents in the main path are the most cited: (i) US20100189682A1 (three citations), which is related to increasing *in vivo* and/or *in vitro* stability of biologically active protein by encoding into it unstructured recombinant polymers, from Amunix; (ii) US20050042721A1 (three citations), which deals with enhanced production of biologically active polypeptides *in vitro* and *in vivo* by using peptide cleavage site at the vector constructs, from Biosante and Cell Genesys; (iii) WO1997003195A1 (six citations), which describe a FVIII:C analog with improved properties from Chiron Pharma; and (iv) US6093392A (three citations), which relates to a method of treating hemophilia comprising a recombinant adeno-associated viral vector associated with Factor IX and expression elements, from the Philadelphia Children's Hospital. We also identified the edges with highest edge betweenness (EB) in the global network, which denotes a crucial connection that links the largest number of communities in the network. The first (highlighted in the blue box), second and third edges (highlighted in the red box) with the highest EB in the global citation network were also present in the technological route, supporting the significance of the method to forecast relevant technologies (**Fig. 6b**).

The technological trends shown in this study consisted of strategies and solutions to improve the production of rFVIII to improve its short half-life and to avoid the development of alloantibodies by the patient. Recently, several clinical trials have shown the importance of the extended half-life of clotting factors and the benefits for adults and children, and many products have already been tested for their safety and efficacy in patients[30].

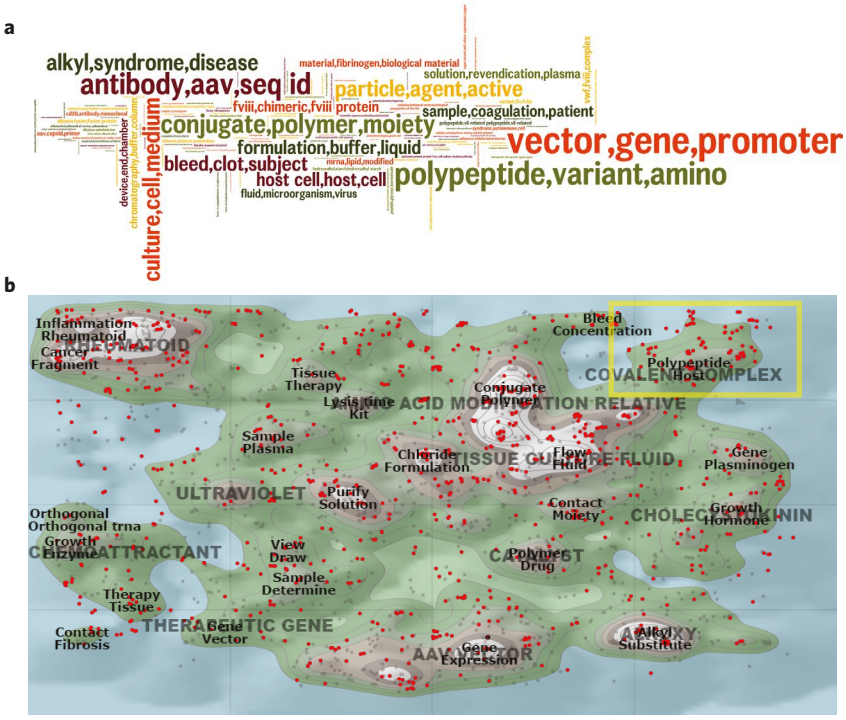

Figure 5 Analysis of patent documents content by text clustering and ThemeScape. (**a**) A word cloud from the text clustering list of most frequent terms in title, abstract and claims of patent families related to rFVIII patents. (**b**) A landscape by ThemeScape using information included in title abstract and claims of the rFVIII patents. Recent patents published between 2012 and 2016 are indicated as red dots, and the yellow rectangle shows an emergent area relating to covalent complex and polypeptide host, mostly from Biogen.

Conclusions

Our study shows an increase in patent publication during the last 20 years related to rFVIII, which means that the treatment for hemophilia A based on replacement therapy has been continuously improved. There are some dominant companies in this market. Baxter is a giant player in coagulation factor development and appears as the biggest assignee with intensive cooperation with Baxalta.

Strategic alliances are one of the key drivers to success. The sector is marked by frequent mergers and acquisitions. One of the top 15 assignees of rFVIII patents identified in our study is Shire, which is is the target of a pending acquisition by Takeda Pharmaceuticals. This study highlights the importance of establishing partnerships in order to develop new and improved technologies. The cooperation network of rFVIII patent assignees was dense and interconnected, with links among many different players, such as public organizations, companies and universities. Some regional clusters evidenced the importance of geographic proximity in establishing cooperation. Organizations take advantage of these cooperative agreements to contribute mutually to R&D projects to develop more innovative and competitive technologies.

The emerging technologies identified in this study showed patents that were concerned with the increase in the half-life of rFVIII and the avoidance of the development of auto-antibodies against rFVIII, as well as some improvements in recombinant protein production, expression and purification from mammalian cells. The development of conjugate forms of rFVIII seems to be very promising for future therapies as well.

Patent data mining using the SPLC method resulted in a technological trend path that includes patents with the greatest future exploratory potential, but we must not discard the uncertainty of technological change. Tracking the inventions in the main paths over the next years remains the best way to validate them as promising technologies.

Analyses of emerging technologies have implications that reduce uncertainty and contribute to guiding critical choices when prioritizing R&D projects and making strategic decisions on technology licensing, while also encouraging cooperation. This study provides valuable information for R&D managers who can better understand how technological developments are advancing and changing in their priority fields, and thus prepare themselves to better face competitions. Additionally, the

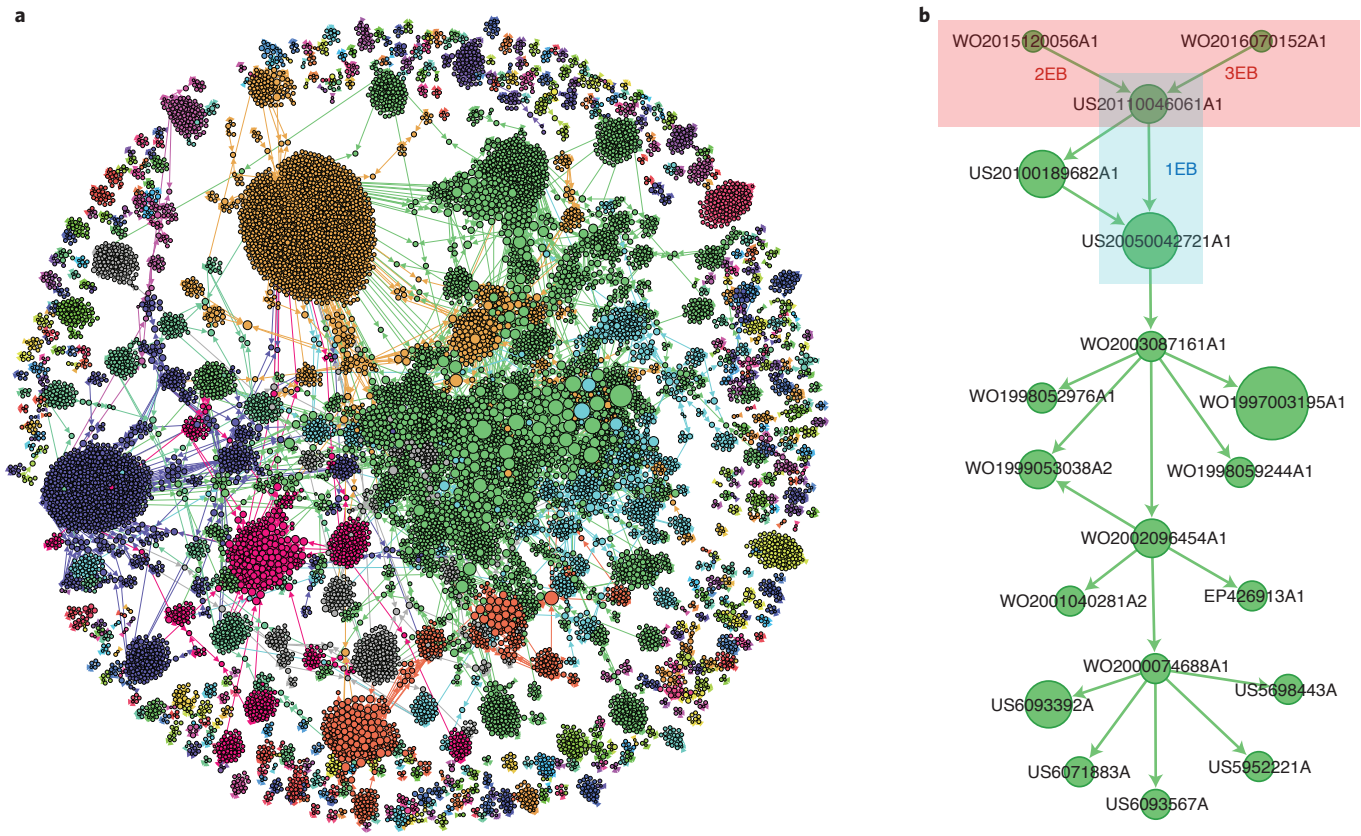

Figure 6 Patent citation network of rFVIII inventions and its main path. (**a**) Global citation network including all patents and its citation links. Bigger nodes represent highly cited patents. (**b**) The technological route derived from the global network based on the main path of patent citation. The edges with the three highest EB were highlighted in blue (highest as 1EB) and red transparent box (second and third highest 2EB and 3EB).

study contributes to the improvement in the quality of decision-making in research strategic planning by the government, academia and industry, providing the technological trends based on the knowledge flows that attempt to anticipate future innovation pathways and understand the potential direction and characteristics of technological change, particularly in rFVIII inventions.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This study is supported by research funds from São Paulo Research Foundation – FAPESP grant number 2012/22686-9; 2014/22500-8; 2015/13816-4; 2013/08135-2 and 2016/02433-0.

AUTHOR CONTRIBUTIONS

C.G.P., V.P.-C. and G.S.P designed the experiments, analyzed the data and wrote the manuscript; G.S.P. and D.T.C. discussed the data and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

1. Srivastava, A. *et al. Haemophilia* **19**, e1–e47 (2013).
2. Tiede, A. *et al. J. Thromb. Haemost.* **11**, 670–678 (2013).
3. Konkle, B.A. *et al. Blood* **126**, 1078–1085 (2015).
4. Mahlangu, J. *et al. Blood* **123**, 317–325 (2014).
5. Zollner, S.B. *et al. Thromb. Res.* **132**, 280–287 (2013).
6. Uchida, N. *et al. Blood* **127**, 1633–1641 (2016).
7. Pabinger-Fasching, I. & Négrier, C. *Thromb. Res.* **141** Suppl 3, S1 (2016).
8. Cantner, U. & Graf, H. *Res. Policy* **35**, 463–480 (2006).
9. Owen, R., Macnaghten, P. & Stilgoe, J. *Sci. Public Policy* **39**, 751–760 (2012).
10. Van Der Valk, T. & Gijsbers, G. *Innov. Manag. Policy Pract.* **12**, 5–17 (2010).
11. Trajtenberg, M. *Rand J. Econ.* **21**, 172–187 (1990).
12. Jaffe, A.B. & Trajtenberg, M. *Patents, Citations & Innovations: a Window on the Knowledge Economy* (MIT, 2002).
13. Narin, F. *Scientometrics* **30**, 147–155 (1994).
14. Karki, M.M.S. *World Pat. Inf.* **19**, 269–272 (1997).
15. Verspagen, B. *Adv. Complex Syst.* **10**, 93–115 (2007).
16. Fontana, R., Nuvolari, A. & Verspagen, B. *Econ. Innov. New Technol.* **18**, 311–336 (2009).
17. Porto, G.S., Kannebley, S., Jr. & Baroni, J.P.M.T. in *Economia de Baixo Carbono: Impactos de Novos Marcos Regulatórios e Tecnológicos sobre a Economia Brasileira*, (eds. Toneto, Jr., R. & Pinho, M.) **452** (FUNPEC-Editora, 2017).
18. Roberts, M. *et al. Nat. Biotechnol.* **32**, 742–748 (2014).
19. Egelie, K.J., Graff, G.D., Strand, S.P. & Johansen, B. *Nat. Biotechnol.* **34**, 1025–1031 (2016).
20. Régibeau, P. & Rockett, K. CEPR Discussion Paper No. 6178. https://EconPapers.repec.org/RePEc:cpr:ceprdp:6178 (2007).
21. Popp, D., Juhl, T. & Johnson, D.K.N. *Top. Econ. Anal. Policy* **4**, 1–48 (2004).
22. World Federation of Hemophilia. *World Federation of Hemophilia Report on the Annual Global Survey 2014* (World Federation of Hemophilia, 2015).
23. Bihari, A. & Pandia, M.K. in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)* 510–514 (IEEE, 2015).
24. Dumont, J.A. *et al. Blood* **119**, 3024–3030 (2012).
25. Schellenberger, V. *et al. Nat. Biotechnol.* **27**, 1186–1190 (2009).
26. Drager, D. *et al. Blood* **126**, 3492 (2015).
27. Érdi, P. *et al. Scientometrics* **95**, 225–242 (2013).
28. Hummon, N.P. & Dereian, P. *Soc. Networks* **11**, 39–63 (1989).
29. Astermark, J. *et al. Blood* **109**, 546–551 (2007).
30. Young, G. & Mahlangu, J.N. *Haemophilia* **22** Suppl 5, 25–30 (2016).