13. AIRR data under the EU Trade Secrets Directive: aligning scientific practices with commercial realities[1]

Jacob S. Sherkow and Timo Minssen

1. INTRODUCTION

Next generation, high throughput DNA sequencing ('HTS') has led to an explosion of genetic data.[2] Today's HTS projects often yield hundreds of millions of reads of DNA, constituting terabytes of data—a large enough amount that storage, access, and computing power become significant constraints.[3] Because such data often contains medically sensitive information, safeguarding it has become an important object for privacy regulations.[4] But keeping vast troves of genetic data secret may also be important for commercial applications—to maintain intellectual property rights in their use.[5] Whether

[1] This work was generously supported by a New York Law School Summer Research Grant. Timo Minssen's research for this chapter was supported by a Novo Nordisk Foundation grant for a scientifically independent Collaborative Research Programme in Biomedical Innovation Law (grant agreement number NNF17SA0027784).

[2] Matthew B Scholz, Chien-Chi Lo and Patrick SG Chain, "Next Generation Sequencing and Bioinformatic Bottlenecks: The Current State of Metagenomic Data Analysis" (2012) 23 Curr Opinion Biotech 9.

[3] Sebastian Wandelt and others, "Data Management Challenges in Next Generation Sequencing" (2012) 12 Datenbank-Spektrum 161; Paul Muir and others, "The Real Cost of Sequencing: Scaling Computation to Keep Pace with Data Generation" (2016) 17 Genome Biol 53.

[4] Isabelle Budin-Ljøsne and others, "ELSI Challenges and Strategies of National Biobank Infrastructures" (2012) 21 Norsk Epidemiologi 155; Charles Auffray and others, "Making Sense of Big Data in Health Research: Towards an EU Action Plan" (2016) 8 Genome Med 71; Bartha Maria Knoppers and Adrian Mark Thorogood, "Ethics and Big Data in Health" (2017) 4 Curr Opinion Sys Biol 53.

[5] Robert Mullan Cook-Deegan and Stephen J McCormack, "Patents, Secrecy, and DNA" (2001) 293 Science 217 ("The intellectual property regime for DNA sequences is trade secrecy when data are most valuable, followed by government-enforced monopoly rights for the duration of a patent's term"). See also Robert Cook-Deegan

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

such data constitute trade secrets under the EU Trade Secrets Directive,[6] and under what terms, is an issue of critical importance to European researchers and businesses.

This chapter explores the application of the EU Trade Secrets Directive for a new iteration of HTS technology: Adaptive immune receptor repertoire sequencing, or "AIRR-seq." AIRR-seq uses HTS technologies to sequence substantial portions of cells constituting the human immune system, sometimes referred to as the "immunome."[7] This can be more data intensive than using HTS for cells outside the immunome, because human immune cells display substantial genetic variability.[8] Indeed, this variability is a function of how the immune system adapts to new infections: By producing a large number of different immune receptors, the immune system can better select for cells expressing immune receptors that target foreign molecules, thereby adapting to novel infections.[9] The genetic diversity of these adaptive immune receptors is immense: As many as 100 billion different immune receptors can be coded into the genes of just one type of immune receptor.[10] Protecting such large and complex datasets as trade secrets under the new Directive presents some significant challenges.

Nonetheless, trade secrecy is likely to be important for AIRR-seq data because it can be commercially valuable. Ascertaining the genetic sequences of specific immune receptors marshaled to combat new infections is important

---

and others, "The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?" (2013) 21 Euro J Human Genetics 585 (providing a robust ethical critique of the practice).

6 Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure [2016] OJ L157/1 (EU Trade Secrets Directive).

7 Felix Breden and others, "Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data" (2017) 8 Frontiers Immunol 1; Anne S De Groot, "Immunomics: Discovering New Targets for Vaccines and Therapeutics" (2006) 11 Drug Discovery Today 203.

8 Breden and others (n 7) 2.

9 Max D Cooper and Matthew N Alder, "The Evolution of Adaptive Immune Systems" (2006) 124 Cell 815.

10 Jacob Glanville and others, "Precise Determination of the Diversity of a Combinatorial Antibody Library Gives Insight into the Human Immunoglobulin Repertoire" (2009) 106 Proc Nat Aca Sci USA 20216.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

to the development of new biologic drugs, genetic diagnostics, and vaccines.[11] Other forms of intellectual property protection are less than optimal.[12]

Secrecy, however, presents problems for AIRR-seq data. Much of the AIRR-seq work currently conducted in Europe and elsewhere is mediated through academic laboratories that often share data through open protocols.[13] Open sharing allows others researchers to validate AIRR-seq data against standard reagents and protocols.[14] It also allows researchers to annotate the data in useful ways, such as identifying paired "chains" of immune receptors and noting which immune receptors seem specific to which foreign molecules, or "antigens."[15] Open sharing of AIRR-seq data is also the preferred—though not exclusive—disclosure norm of the AIRR Community (an international community of AIRR researchers), and a default provision of MiAIRR, the data standard for AIRR-seq data.[16] Beyond these scientifically important drivers for open sharing of AIRR-seq data, there are utilitarian and normative ones, too. Sharing AIRR-seq data helps researchers publish their results in scientific journals, demonstrate the superiority of their sequencing methods, and raise their status among their peers.[17] The open sharing of AIRR-seq data also better aligns with many researchers' expectations of Mertonian science and their feelings of responsibility to the grant funding public.[18] For many researchers, the open sharing of AIRR-seq data is both a practice and a norm.

Yet open sharing destroys any trade secrecy in the underlying data. The EU Trade Secrets Directive defines trade secrets as any information that "has commercial value because it is secret" and "has been subject to reasonable steps … to keep it secret."[19] Information disclosed to others "within the circles that normally deal with the kind of information in question" is not protectable as a trade secret.[20] While Member States have some leeway to craft safeguards to harsh implementations of the Directive, it does not appear that they have

---

[11] William H Robinson, "Sequencing the Functional Antibody Repertoire—Diagnostic and Therapeutic Discovery" (2015) 11 Nat Rev Rheumatology 171, 178–9.

[12] De Groot (n 7) 203; Jacob S Sherkow and Christopher Scott, "Myriad Stands Alone" (2014) 32 Nat Biotech 620.

[13] Breden and others (n 7) 2; Florian Rubelt and others, "Adaptive Immune Receptor Repertoire Community Recommendations for Sharing Immune-Repertoire Sequencing Data" (2017) 18 Nat Immunology 1274.

[14] Rubelt and others (n 13) 1275.

[15] Ibid.

[16] Ibid.

[17] Breden and others (n 7) 2; Rubelt and others (n 13) 1275.

[18] Robert K Merton, *The Sociology of Science* (U Chi Press 1973); Breden and others (n 7) 3–4.

[19] EU Trade Secrets Directive, art 2.

[20] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

authority to carve out exceptions to the Directive's definitions of secrecy.[21] Academic AIRR-seq research and the commercial development of products arising from that research therefore exist in tension: AIRR-seq data shared in the normal course of research may consequently stymie the development of therapies or vaccines based on that data.

At the same time, the nature of AIRR-seq data itself suggests several paths forward for maintaining the protectability of some aspects of that data under the Directive, even if otherwise disclosed. First, one of the most significant challenges in interpreting AIRR-seq data is the pairing of immune receptor chains, the individual components that constitute a complete immune receptor.[22] This information—which chains form a unique immune receptor—may not be readily apparent from raw AIRR-seq data itself and, if kept secret, should be separately protectable under the Directive.[23] Second, the antigen specificity of immune receptors—what the receptors bind to—is not readily apparent from AIRR-seq data itself.[24] This information is critically important for commercial developers seeking to create products and diagnostics from AIRR-seq data;[25] it too should be separately protectable under the Directive.[26] And third, the disclosure of broad AIRR-seq data should not obviate trade secret protection for modified, or recombinant, forms of immune receptor sequences.[27] Those sequences, often identified through laborious experimentation, should be separately protectable independent of how and how much related AIRR-seq data is disclosed.[28] This is all to say that information *derived* from complex, publicly disclosed information—what we term "follow-on information"—should nonetheless be protected under the Trade Secrets Directive. Allowing such information to be protectable, even where underlying raw sequences are otherwise publicly disclosed, would also have the salutary effect of further encouraging the disclosure of basic scientific information, rather than leaving researchers guessing as to the downstream effects of data sharing.

Beyond this practical assessment of protecting follow-on data, AIRR-seq projects present the occasion to think about how the Trade Secrets Directive aligns with commercial development, science, and disclosure more generally.

---

[21] Ibid, art 13.

[22] Brandon J DeKosky and others, "High-Throughput Sequencing of the Paired Human Immunoglobulin Heavy and Light Chain Repertoire" (2013) 31 Nat Biotech 166.

[23] Text at n 127–47.

[24] Robert A Holt, "Interpreting the T-cell Receptor Repertoire" (2017) 35 Nat Biotech 829.

[25] Robinson (n 11) 178–9.

[26] Text at n 145–55.

[27] Text at n 154–64.

[28] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

First, AIRR-seq presents a case study of the mutual exclusivity of patents and trade secrets for complex technologies that may be informative for future policymakers. It also provides some conceptual guidance about which aspects of clinical trial data—some of the most valuable information in biopharmaceutical development—should be disclosed and which should be afforded protection. Lastly, trade secrets for follow-on AIRR-seq may provide ways of reorienting trade secrecy around genetic diagnostics. AIRR-seq data shows that it is at least theoretically possible to simultaneously promote information sharing while reserving the most value for aspects of that information to encourage private development of new technologies.

Against this backdrop, Section 2 of this chapter explains the particulars of AIRR-seq data. Section 3 reviews and describes the most important provisions of the EU Trade Secrets Directive for AIRR-seq data. Section 4 then discusses some difficulties arising from the sharing of AIRR-seq data under the Trade Secrets Directive and proposes several solutions. Section 5 uses AIRR-seq data sharing under the Trade Secrets Directive to better understand other problems related to information disclosure versus sharing.

2. AIRR DATA

AIRR sequencing—short for Adaptive Immune Receptor Repertoire sequencing—is one of the more fascinating and powerful applications of HTS to date.[29] AIRR-seq data comes from the sequencing of a large number of cells of the immune system, large enough that a sequencing sample constitutes an approximate repertoire of that cell type for that particular organism.[30] This is important because, in contrast to other cell types, many immune system cells are genetically heterogeneous; they each contain different sequences of immune receptor genes.[31] Sequencing the repertoire of these cells allows researchers to follow the immune system as it adapts to an infection or some other immunological insult.[32] This allows researchers to understand how the

[29] Breden and others (n 7) 1; Rubelt and others (n 13) 1275; Robinson (n 11) 171.

[30] J Douglas Freeman and others, "Profiling the T-cell Receptor Beta-Chain Repertoire by Massively Parallel Sequencing" (2009) 19 Genome Res 1817; Harlan S Robins and others, "Comprehensive Assessment of T-cell Receptor β-chain diversity in αβ T Cells" (2009) 114 Blood 4099; Joshua A Weinstein and others, "High-Throughput Sequencing of the Zebrafish Antibody Repertoire" (2009) 324 Science 807; Breden and others (n 7) 1.

[31] Breden and others (n 7) 2.

[32] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

immune system works, generally, and develop immune receptors with commercial applications, such as vaccines and biologic drugs.[33]

2.1 The Adaptive Immune System

Humans and many other vertebrates possess two forms of biologic immunity: Innate, immune responses that we are born with, and adaptive, immune responses created to combat infections and foreign material.[34] Broadly speaking, these immune responses work through complex networks of cellular signaling.[35] Some cells, upon recognizing foreign material in the lymph, will use a series of molecular signals to cue other cells to destroy them.[36]

Two cell types are particularly critical to this signaling function: B cells, which mature in bone marrow, and T cells, which mature in the thymus gland.[37] Each produces specialized proteins that reside on the cell surface and bind to foreign material or "antigens":[38] B cells produce antibodies and T cells produce T cell receptors; collectively they are "immune receptors."[39] At their tips, antibodies and T cell receptors have remarkably diverse and intricate structures, known as complementary determining regions ('CDRs'), responsible for interacting with and binding to foreign antigens.[40] Astonishingly, each new B cell and T cell produces a new immune receptor with different CDRs.[41] This creates a cosmic diversity of immune receptors—approximately 100 billion different human antibodies are possible, for example.[42]

Astonishingly, the system has the capacity to develop immune receptors specific to virtually any foreign antigen.[43] But because the human genome comprises only 19,000 genes, such diversity must come from something other than a 1:1 relationship between gene and immune receptor.[44] The immune system achieves this by segmenting the genes required to produce immune

---

33 Robinson (n 11) 178–9.
34 Weinstein and others (n 30) 807.
35 Cooper and Alder (n 9) 815. See also William E Paul, *Fundamental Immunology* (7th edn, Wolters Kluwer 2013).
36 Ibid.
37 Ibid.
38 Ibid.
39 Ibid 815–16.
40 Paul (n 35) 134.
41 Robins et al (n 30) 4099.
42 Glanville et al (n 10) 20216. See also Robins et al (n 30) 4099 (estimating ten quadrillion possible combinations).
43 Cooper and Alder (n 9) 815.
44 Susumu Tonegawa, "Somatic Generation of Antibody Diversity" (1983) 302 Nature 575.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

receptors—namely, into variable or "V," diversity or "D," joining or "J," and constant or "C" segments—and then expanding each segment into variable cassettes—V1, V2, and V3, for example.[45] During the cellular maturation process, B and T cells randomly select a single cassette for each immune receptor gene segment, permanently editing the cellular genome in the process.[46] This resulting immune receptor gene then undergoes further changes through a process known as somatic hypermutation, which is then often further refined if the resulting immune receptor pairs with a foreign antigen.[47]

The hundreds of millions B and T cells circulating in the lymph at any given time consequently have different genetic sequences.[48] Examining these sequences as a whole has the potential to reveal much about to whom they belong, which immunities they have developed, which infections they are currently combating, the geography of where they have been and grew up, their ancestry, and so on.[49] But understanding these sequences across a large scale of subjects also informs immunologists about how the immune system works generally, and about how to best engineer the immune system, both to respond to target antigens—as with vaccines—and to create therapeutically effective antibodies and T cell receptors.[50]

2.2 AIRR Sequencing

An AIRR sequencing project will begin by obtaining a large number of immune cells, most often through the blood.[51] Particular cells of interest, such as certain types of B or T cells, can then be isolated using a variety of techniques.[52] Once immune cells of interest have been isolated, they can then

---

[45] Ibid 576.

[46] Ibid.

[47] Deborah L French, Reuven Laskov and Matthew D Scharff, "The Role of Somatic Hypermutation in the Generation of Antibody Diversity" (1989) 244 Science 1152.

[48] Glanville and others (n 10) 20216.

[49] Harlan Robins, "Immunosequencing: Applications of Immune Repertoire Deep Sequencing" (2013) 25 Curr Op Immunology 646; Poornima Parameswaran and others, "Convergent Antibody Signatures in Human Dengue" (2013) 13 Cell and Host Microbe 691; George Georgiou and others, "The Promise and Challenge of High-Throughput Sequencing of the Antibody Repertoire" (2014) 32 Nat Biotech 158; Jacob Glanville and others, "Identifying Specificity Groups in the T cell Receptor Repertoire" (2017) 547 Nature 94.

[50] Georgiou and others (n 49) 160.

[51] Elisa Rosati and others, "Overview of Methodologies for T-cell Receptor Repertoire Analysis" (2017) 17 BMC Biotech 1, 4.

[52] Ibid 8. Obtaining enough cells to ensure that a large portion of the immune repertoire will be sequenced, not just a fraction of it, remains challenging. Joseph Kaplinsky

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

be prepared for genetic sequencing. Prior to the advent of HTS, this was a laborious and time intensive process that greatly limited the number of cells that could be sequenced—and as a consequence, the fraction of the immune repertoire that could be identified.[53] HTS has since allowed researchers to sequence "tens to hundreds of millions of receptor sequences" in a single "run," including multiple sequences—or "reads"—from a single gene.[54] This redundancy best ensures both accurate and complete sequences from the derived immune cells.[55]

Today, HTS sequencing of immune receptor genes is typically accomplished in several ways, with cell barcoding and droplet sequencing constituting two of the more significant techniques. For cell barcoding, researchers separate each cell into individual wells containing unique, short fragments of DNA specific to that well—a DNA "barcode"—that can later be used to identify which cell any given sequence originated from.[56] This barcode is then affixed to copies of cloned DNA ("cDNA") from the cell and the entire genetic fragment—barcode plus cDNA—is then sequenced.[57] For droplet sequencing, also known as "Drop-seq," each immune cell is suspended in an emulsion droplet—essentially, a droplet of oil—containing a microscopic bead that binds to cDNA.[58] Once inside the droplet, the cell is lysed, and the bead binds to DNA.[59] During sequencing, each bead–DNA conjugate is sequenced, collating the sequences pertaining to each bead as coming from a unique cell.[60] These technologies have numerous advantages over their predecessors: Sequencing can be run in parallel, with multiple DNA segments sequenced

---

and Ramy Arnaout, "Robust Estimates of Overall Immune-Repertoire Diversity from High-Throughput Measurements on Samples" (2016) 7 Nat Communications 11881, 8. Nonetheless, procedures such as leukapheresis—the removal of large numbers of subject's white blood cells from whole blood—have been shown to be both safe and powerful enough to encompass the immune receptor repertoire. Peggy A English and others, "The Safety and Utility of Leukapheresis of Normal Donors for Obtaining Products for Immunologic Research" (1990) 135 J Immunological Meth 285.

53 Robinson (n 11) 171.

54 Rubelt and others (n 13) 1274.

55 Robins (n 48) 649.

56 Christopher Vollmers and others, "Genetic Measurement of Memory B-cell Recall Using Antibody Repertoire Sequencing" (2013) 110 Proc Nat Aca Sci USA 13463.

57 Ibid.

58 Simon Friedensohn, Tarik A Khan, and Sai T Reddy, "Advanced Methodologies in High-Throughput Sequencing of Immune Repertoires" (2017) 35 Trends Biotech 203, 210.

59 Ibid.

60 Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

simultaneously;[61] even in large sequencing runs, they have the ability to identify which DNA sequences came from which cells;[62] and for immune receptors, specifically, these technologies can be easily modified to ensure that the entire CDR of an immune receptor—not just a single chain—is sequenced.[63]

There are limitations to these technologies, to be sure. The same process that gives rise to immune receptors' great variability also confounds easy sequencing: AIRR researchers often use reference gene data—IMGT is a popular database—to validate their findings.[64] But gaps in reference databases, or thin data, make finding new immune receptor variants that much more difficult.[65] In other cases, it is simply too difficult to ensure that the entirety of an immune receptor's CDR is sequenced; researchers may only be able to obtain a single chain, for example.[66] In addition, despite recent advances in barcoding and droplet sequencing, there remain latent errors in the sequencing process itself.[67]

Nonetheless, AIRR data is a remarkably precise instrument for determining how the immune system responds and evolves to infections.[68] When coupled with clinical data—information about patients' infections and how they have responded—AIRR researchers can use their data to develop powerful tools for the treatment of disease.[69] A recent study, for example, examined the immune repertoire of patients suffering from a particular type of cancer, cutaneous T cell lymphoma.[70] Often, it is unclear whether a patient truly has lymphoma or if it is instead a benign skin disease.[71] Immune receptor repertoire sequencing, however, was able to correctly predict true lymphomas in 46/46—100 percent—of patients,[72] divined from a sea of hundreds of millions of strings of

---

[61] Freeman and others (n 30) 1817.

[62] Friedensohn, Khan, and Reddy (n 58) 210.

[63] Dekosky and others (n 22) 166.

[64] Shuo Li and others, "IMGT/HighV QUEST Paradigm for T cell Receptor IMGT Clonotype Diversity and Next Generation Repertoire Immunoprofiling" (2013) 4 Nat Communications 2333.

[65] Corey T Watson and Felix Breden, "The Immunoglobulin Heavy Chain Locus: Genetic Variation, Missing Data, and Implications for Human Disease" (2012) 13 Genes Immunity 363.

[66] Georgiou and others (n 49) 158.

[67] Nicolas Fischer, "Sequencing Antibody Repertoires: The Next Generation" (2011) 3 mAbs 17.

[68] Xueling Wu and others, "Focused Evolution of HIV-1 Neutralizing Antibodies Revealed by Structures and Deep Sequencing" (2011) 333 Science 1593.

[69] Ilan R Kirsch and others, "TCR Sequencing Facilitates Diagnosis and Identifies Mature T cells as the Cell of Origin in CTCL" (2015) 7 Sci Translational Med 308ra158.

[70] Ibid.

[71] Ibid.

[72] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

DNA sequences.[73] AIRR sequencing is a triumph of HTS and "big data" technologies, and heralds the possibility of both understanding and manipulating the immune system in granular detail.

2.3 AIRR Data

Data underlying AIRR-seq projects come in a variety of forms. There are the raw, genomic sequences themselves, spanning the entirety of the cassette that gives rise to each chain of the cell's antibody or T cell receptor.[74] Similarly, there are the complete, raw sequences of each cell's immune receptor's mRNA, the intermediate precursor from DNA to producing the immune receptor itself.[75] For any given AIRR experiment from a single subject, these can come from hundreds of thousands or millions of cells, producing the raw sequences of millions of chains of immune receptors.[76]

This raw sequence data can be valuable in and of itself. Bioinformatics software allows AIRR researchers not involved in the principal sequencing efforts to infer which *V*, *D*, and *J* genes contributed to producing each cell's immune receptor, and even to infer new genes not yet known in databases, such as IMGT.[77] Others can make use of the data to "pair" immune receptor chains together that were not originally paired when sequenced.[78] This is crucial, for example, in the development of antibody-based therapies that typically require the paired sequences of both an antibody's heavy and its light chains.[79]

Beyond the raw sequence itself, information concerning sequences' functional properties, such as what each component of the immune receptor is likely to bind to, can also be valuable. While the field is still nascent, some researchers are attempting to derive information from raw sequences to predict immune receptors' likely targets.[80] Databases such as Epitome contain

---

[73] Ibid.

[74] Rubelt and others (n 13) 1275.

[75] Ibid 1276.

[76] Ibid 1274.

[77] Scott D Boyd and others, "Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements" (2010) 184 J Immunol 6986; Watson and Breden (n 65) 363; Dekosky et al (n 22) 166.

[78] Dekosky and others (n 22) 166.

[79] James S Blachy and others, "Immunoglobulin Transcript Sequence and Somatic Hypermutation Computation from Unselected RNA-seq Reads in Chronic Lymphocytic Leukemia" (2015) 112 Proc Nat'l Aca Sci USA 4322, 4325 ("[R]econstruction of the entire *V-D-J* sequence and thus definition of the complete CDR3 is important to study BCR specificity and stereotypy").

[80] Avner Schlessinger and others, "Epitome: Database of Structure-Inferred Antigenic Epitopes" (2006) 34 Nucleic Acids Res S1, D777; Randi Vita and others, "The Immune Epitope Database 2.0" (2010) 38 Nucleic Acids Res S1, D854.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

information on computationally inferred binding sites from immune receptor sequence data.[81] Given that validation studies confirming the binding specificity of an immune receptor are difficult and take time, computational techniques such as these are likely to both produce more data related to AIRR-seq projects and to make the original raw sequences increasingly valuable.

Broadly speaking, AIRR-seq data is thus of two types: (1) "raw" sequence information, typically in the file format FASTQ that enumerates the string of nucleotide sequences from a single sequencing project; and (2) what we refer to as "follow-on" information about that sequence data: Information concerning the pairing of immune receptor chains; inferred antigen specificity; data on the "populations" or "clones" of certain B or T cells; or other functional or descriptive information about the cells and immune receptors that have been sequenced. As we detail in this chapter, the protectability of raw versus follow-on AIRR data as trade secrets raises different sets of issues under the EU Trade Secrets Directive and for academic science.

3. THE EU TRADE SECRETS DIRECTIVE

Similar to intellectual property rights—such as patents, copyrights, and trademarks—trade secrets protect economically valuable, secret information.[82] Until recently, trade secrecy law in the EU was a matter entirely for each individual Member State, with an additional patchwork of legal protections such as unfair competition, contract, and criminal laws.[83] There had been no harmonized EU definition of a "trade secret"; nor did the European Union have a harmonized protection system for trade secrets.[84]

This changed recently with the EU Directive on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure.[85] The Directive, following a proposal from the European Commission, was finally adopted by the European Parliament and the European Council on June 8, 2016, aimed at standardizing the national laws in EU Member States against the unlawful acquisition, disclosure, and use of trade secrets.[86] The Directive also establishes a harmonized definition of "trade secret" and makes uniform the responsibilities of trade secrets holders

---

[81] Schlessinger and others (n 80) D777; Vita and others (n 80) D854.

[82] Michael Risch, "Why Do We Have Trade Secrets?" (2007) 11 Marquette Intell Prop L Rev 1; Mark A Lemley, "The Surprising Virtues of Treating Trade Secrets as IP Rights" (2008) 61 Stan L Rev 311.

[83] EU Trade Secrets Directive, preamble ¶ 6.

[84] Ibid ¶ 10.

[85] Ibid.

[86] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

to guard their information.[87] The Directive is not self-executing, however; Member States were required to bring into force the laws and administrative provisions necessary to comply with the Directive by June 9, 2018.[88] Remarkably, only the United Kingdom, Denmark, and Slovakia transposed the Directive by this deadline.[89] Since then, however, 23 Member States have transposed the Directive, in one form or another, into national law.[90]

Perhaps the principal achievement of the EU Trade Secrets Directive is its establishment of a harmonized definition of "trade secrecy," as well as common measures aimed at preventing the wrongful disclosure of confidential commercial information.[91] Regarding the definition of "trade secrecy," the Trade Secrets Directive largely borrows from Article 39 of the Agreement on Trade-Related Aspects of Intellectual Property Rights ("TRIPS"), first ratified in 1994.[92] Like Article 39 of TRIPS, Article 2 of the Trade Secrets Directive defines a trade secret as any information that is "secret," has "commercial value because it is secret," and "has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret."[93] "Secret," under the Trade Secrets Directive, is further defined as "not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question."[94]

This definition is extraordinarily broad and leaves wide room for interpretation. For example, the extent of the notions of "generally known" and "readily accessible" is not entirely clear. Nor is it clear what it means to have taken "reasonable steps" to keep information secret. Past practices in individual Member States in the EU—with different requirements—may not necessarily be a good guide for future practice. Much will have to be determined in due course through judgments by the Court of Justice of the EU ("CJEU"). Nonetheless, according to the recitals to the Directive, the definition is meant to cover *any* information where the secret holder has both a legitimate interest

---

87 Ibid, preamble ¶ 10.

88 Ibid, art 19(1).

89 Will Smith and Robert Williams, "Implementation of the Trade Secrets Directive – The Deadline Has Passed!" (June 11, 2018) www.twobirds.com/en/news/articles/2018/global/implementation-of-the-trade-secrets-directive-the-deadline-has-passed accessed June 10, 2018 (archived at https://perma.cc/B62Q-ZVJB).

90 See National transposition measures communicated by the Member States https://eur-lex.europa.eu/legal-content/EN/NIM/?uri=CELEX:32016L0943 accessed Oct 16, 2019 (archived at https://perma.cc/W5RR-JAJK).

91 EU Trade Secrets Directive, art 2; Ibid art 4.

92 Ibid, art 2.1.

93 Ibid.

94 Ibid, art 2.1(a).

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

in keeping the information confidential and a legitimate expectation that such confidentiality will be preserved.[95]

While the preservation of trade secrets tends to involve contracts, for example nondisclosure agreements among collaborators, enforcing them requires litigation.[96] Under the Trade Secrets Directive, the cause of action for "infringement" is a claim to misappropriation, "such as theft, unauthorized copying, economic espionage, [the] breach of confidentiality requirements," or any other method contrary to honest commercial practices.[97] The Directive also includes the potential for third party liability for those who obtain trade secrets from unscrupulous parties. The Directive proscribes use or disclosure of a trade secret by a person who knew, or should have known, that the trade secret was obtained from someone using it or disclosing it unlawfully.[98] In addition, unlike other jurisdictions, including the United States, the Directive establishes liability for producing, offering, or placing on the market "infringing goods," if the defendant knew or should have known that the trade secret was used unlawfully in their manufacture.[99] In other words, it enables an action against persons further down the supply chain, even if they have no knowledge of the confidential information itself.

The Directive also provides defenses to these claims. One can defend against claims of misappropriation if the information was *lawfully* obtained. As defined by the Trade Secrets Directive, this includes independent discovery, reverse engineering of publicly available products, discovery pursuant to a license, observation of the item in public use or on public display, or derivation from published literature.[100] This also includes boundary situations, where the secret was nonetheless obtained from the secret holder otherwise "in conformity with honest commercial practices."[101]

Since many of the concepts in the Trade Secrets Directive are not concretely defined, the CJEU will, in all likelihood, need to interpret them. The Directive further protects "whistleblowers" who disclose trade secrets to reveal misconduct, wrongdoing, or illegal activity for the purpose of protecting the general public interest.[102] This protection is extended irrespective of whether the trade secret disclosure is in confidence.[103] In these senses, the Trade Secrets

---

95 Ibid, preamble ¶ 1.
96 Ibid, preamble ¶ 24.
97 Ibid, preamble ¶ 4.
98 Ibid, art 4.
99 Ibid, art 4.5.
100 Ibid, art 3.
101 Ibid, art 3(1)(d).
102 Ibid, art 5.
103 Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

Directive makes trade secrets somewhat unlike traditional forms of intellectual property,[104] and the Directive itself makes clear in recitals 1 to 3 and 39 that trade secret protection should be differentiated from traditional intellectual property rights.[105] What exactly this entails is not entirely clear, and it remains to be seen how Member States' legislatures and courts, including possibly the CJEU, will interpret the Directive for new technologies.

4. SHARING AIRR-SEQ DATA AND TRADE SECRETS: PROBLEMS AND SOLUTIONS

AIRR-seq data is at its most robust when it is shared among researchers and annotated.[106] Researchers can use others' data to pair together immune receptor chains from the same B or T cell, annotate sequences for antigen specificity, and use the data in the development of recombinant immune receptors.[107] Sharing also allows outside researchers to validate the results of prior sequencing efforts, an important task in ensuring the reproducibility of results.[108] But the sharing of AIRR-seq data practiced by academic researchers complicates efforts by commercial developers to protect aspects of the underlying data as trade secrets.[109] This Part describes the norms and practices that researchers employ in sharing AIRR-seq data and several aspects of that data that may nonetheless be protectable under the Trade Secrets Directive as follow-on data.

4.1 Sharing Norms and Practices of AIRR-seq Data

AIRR researchers typically deposit their AIRR-seq data in a variety of public archives managed by government research agencies.[110] In Europe, the European Bioinformatics Institute manages the European Nucleotide Archive

---

[104] Ibid, preamble ¶ 2.

[105] Ibid, preamble ¶ 39.

[106] Rubelt and others (n 13) 1275.

[107] Sai T Reddy and others, "Monoclonal Antibodies Isolated Without Screening by Analyzing the Variable-Gene Repertoire of Plasma Cells" (2010) 28 Nat Biotech 965; Dekosky and others (n 22) 166; Anastasios Spiliotopoulos and others, "Sensitive Recovery of Recombinant Antibody Clones After Their *in silico* Identification Within NGS Datasets" (2015) 420 J Immunological Methods 50.

[108] Github, Recommendations from the AIRR Common Repository Working Group (2017) https://github.com/airr-community/common-repo-wg/blob/master/recommendations.md accessed June 1, 2018 (archived at http://perma.cc/6F9L-WWT2).

[109] EU Trade Secrets Directive, art 2.1.

[110] Georgiou and others (n 49) 163; Breden and others (n 7) 1; Rubelt and others (n 13) 1275.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

(ENA), a repository for "the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation."[111] GenBank, a service of the U.S. National Institutes of Health, compiles information from the ENA and similar U.S. and Japanese efforts.[112] In addition, because HTS data tend to be more complex than traditional methods, the Sequence Read Archive serves as a worldwide bank of HTS data produced in the FASTQ format, specifically.[113] All of these databases are publicly available; they can be accessed by anyone with virtually no restrictions.[114]

Like public deposit, AIRR researchers often share data among themselves without restrictions. Nondisclosure agreements—so important for trade secrecy—are virtually unheard of among academic researchers. AIRR researchers have many reasons to share their data openly among themselves. The higher-ranked scientific journals, including *Nature* and *Science*, often require the deposit of sequence data as a condition of publication;[115] journal editors' requirements are perhaps one of the strongest influences on scientific practices.[116] Open sharing and public deposit of AIRR-seq data also sends strong signals to researchers' peers that the sequencing methods and protocols met community standards and are otherwise of good quality.[117] Sharing also comports with many researchers' senses of scientific communalism, the Mertonian norm of common ownership of research products and data.[118] Many researchers further see the sharing and deposit of AIRR-seq data as fulfilling their obligations to the public at large, stemming from their funding by public agencies.[119] In short, AIRR researchers openly share and deposit their data because doing so aligns with academic norms and scientific practices.

For this reason, members of the AIRR Community have developed the Minimum Information AIRR standard, or MiAIRR, "a community-based

---

[111] European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), "European Nucleotide Archive" www.ebi.ac.uk/ena accessed June 1, 2018.

[112] National Center for Biotechnology Information, "GenBank Overview" www.ncbi.nlm.nih.gov/genbank/ accessed 1 June 2018.

[113] National Center for Biotechnology Information, "SRA" www.ncbi.nlm.nih.gov/sra accessed June 1, 2018.

[114] Sources at nn 111–113.

[115] Mark Blaxter et al, 'Reminder to Deposit DNA Sequences' (2016) 352 Science 780; Steven L Salzberg, "Databases: Reminder to Deposit DNA Sequences" (2016) 533 Nature 179.

[116] Stephen M Maurer, *Self-Governance in Science* (Cambridge University Press 2017) 1719.

[117] Rubelt and others (n 13) 1275.

[118] Merton (n 18); Breden and others (n 7) 3–4.

[119] Rubelt and others (n 13) 1277.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

standard for the reporting of experimental results" concerning AIRR-seq data.[120] While the standard focuses on technical aspects of reporting—requiring the production of certain datasets in specific formats—it is also designed to encourage AIRR researchers to maximally share their data.[121] The standard is designed in a way that enables incorporation into GenBank, the public repository of genetic sequence information hosted by the U.S. NIH.[122] MiAIRR also allows AIRR-seq data to be transferred across platforms, enabling groups of researchers to access and use AIRR-seq data across labs and experiments.[123] And lastly, the adopters envision that the establishment of MiAIRR will encourage scientific journals and funding agencies to adopt it, implicitly requiring the sharing of experimental AIRR-seq data.[124] By creating a data standard, the AIRR Community has forged tools to advance the norms of open and free data sharing.

4.2 Protecting Aspects of AIRR-seq Data under the Directive

Under the Trade Secrets Directive, public sharing of AIRR-seq data, such as that through the MiAIRR standard without further restrictions, destroys any protectability in the underlying information itself.[125] Trade secrets are only information that is, in fact, secret.[126] But the commercial value of AIRR-seq data lies not in the raw data itself but in information derived from the data: what we term "follow-on information." Elucidating this follow-on information is itself a difficult task and not readily apparent even from MiAIRR-compliant AIRR-seq data.[127] This section discusses three, commercially important, types of follow-on information—receptor chain pairing, antigen specificity data, and recombinant immune receptor sequences—and how they should be treated under the Trade Secrets Directive even where the underlying AIRR-seq data is disclosed.

4.2.1 Receptor chain pairing

Because immune receptors are genetically coded as individual chains, it is often difficult to obtain the entire sequence of a given immune receptor—for

---

120 Ibid 1275.
121 Ibid 1277.
122 Ibid.
123 Breden and others (n 7) 3–4.
124 Rubelt and others (n 13) 1275.
125 EU Trade Secrets Directive, art 2.
126 Ibid.
127 Rubelt and others (n 13) 1275.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

example, both the heavy and light chains of an antibody.[128] Researchers have begun to solve this problem, however, by employing single cell sequencing.[129] By sequencing individual cells through droplet sequencing, scientists can ensure that any genetic immune receptor sequences consist of all of the chains constituting a single cell's immune receptor.[130] Separate sequences are then later annotated as having originated from a single immune cell—the implication being that they collectively constitute a set of chains from a single immune receptor.[131]

Pairing immune receptor chains is crucially important for understanding the specificity of a given immune receptor; frequently, all of an immune receptor's chains are required to model an immune receptor's CDR.[132] Partial chains do little to help researchers understand to what an individual immune receptor binds.[133] Further, knowledge of chain pairing has the potential to help researchers understand how the adaptive immune receptor repertoire evolves in response to infections—significant information for the development of vaccines.[134]

In this way, immune receptor chain pairing follow-on data has the potential to meet prongs (1)(a) and (1)(b) of the Directive's definition trade secrets: Chain pairing data would not be "generally known among or readily accessible" to other AIRR-seq researchers without this annotation;[135] and the information, at least potentially, "has commercial value because it is secret."[136] Obtaining immune receptor chain pairing information from unsorted data would be tremendously difficult,[137] and freely releasing immune receptor chain pairing data to the larger research community would likely dissuade commercial developers from working with any of the public sequences.[138] Indeed, because patent protection for natural antibodies is becoming increasingly diffi-

---

[128] Dekosky and others (n 22) 166.

[129] Ibid; Friedensohn, Khan, and Reddy (n 58) 204–5.

[130] Dekosky and others (n 22) 166; Friedensohn, Khan, and Reddy (n 58) 204–5.

[131] Dekosky and others (n 22) 166; Friedensohn, Khan, and Reddy (n 58) 204–5.

[132] Dekosky and others (n 22) 167; Friedensohn, Khan, and Reddy (n 58) 206–7.

[133] Anne Eugster and others, "Measuring T Cell Receptor and T Cell Gene Expression Diversity in Antigen-Responsive Human CD4+ T Cells" (2013) 400 J Immunological Methods 13–14.

[134] Fischer (n 67) 19.

[135] EU Trade Secrets Directive, art 1.

[136] Ibid.

[137] Dekosky and others (n 22) 166.

[138] Panel Discussion, IP Issues in Data Sharing, Adaptive Immune Receptor Repertoire (AIRR) Community Meeting (4 Dec 2017) (Rockville, Maryland, USA) (IP Issues in Data Sharing).

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

cult, both in Europe and the US, trade secrecy for immune receptor sequences is often thought of as the only intellectual protection available.[139]

To the degree that AIRR researchers wish to protect the commercial viability of chain pairing, then, they would only need to take "reasonable steps under the circumstances … to keep it secret."[140] Practically, this would mean that MiAIRR-compliant disclosures of unsorted sequences would still be permissible, while the particular annotations concerning chain-pairing could be kept sequestered. This would both preserve the economic viability of follow-on chain pairing data and meet the Trade Secrets Directive's requirement that the information sought to be protected be kept secret. Researchers wishing to commercialize specific antibodies or TCRs could then share chain-pairing annotations under the cover of nondisclosure agreements. In some ways, the Trade Secrets Directive here provides clarity as to both what can be protected and what can continue to be reasonably shared.

To be clear, however, there is one possible technological complication: Inferred paired sequence chains from otherwise unsorted data.[141] Researchers have begun to develop some techniques to computationally infer chain pairings from unsorted data.[142] While the technology is, as of this writing, still rudimentary, it holds the promise to make valuable a large quantity of otherwise "loose" AIRR-seq data. But it would also constitute "reverse engineering," thus destroying the protectability of any paired-chain follow-on data that researchers would have otherwise liked to have kept secret.[143] Whether such technology will come to pass such that chain-pairing information, even from unsorted data, is otherwise "readily accessible" remains to be seen—and is likely part of a larger story of reverse engineering technology outstripping engineers' abilities to sequester trade secrets.[144] But for now researchers can be confident that disclosing unpaired AIRR-seq, MiAIRR data will still allow

---

[139] *Amgen Inc v Sanofi* (2017) 872 F.3d 1367 (Fed. Cir. USA); W Nicholson Price II and Arti K Rai, "Manufacturing Barriers to Biologics Competition and Innovation" (2015) 101 Iowa L Rev 1023; Theresa Gresl, Ulrich Storz, and Colin Sandercock, "An Update on Obtaining and Enforcing Therapeutic Antibody Patent Claims" (2016) 34 Nat Biotech 1242.

[140] EU Trade Secrets Directive, art 1.

[141] Bryan Howie and others, "High-Throughput Pairing of T Cell Receptor α and β Sequences" (2015) 7 Sci Translational Med 301ra131.

[142] Ibid.

[143] EU Trade Secrets Directive, art 3(1).

[144] Marie E Ceste and John C Doyle, "Reverse Engineering of Biological Complexity" (2002) 295 Science 1664; Pamela Samuelson and Suzanne Scotchmer, "The Law and Economics of Reverse Engineering" (2002) 111 Yale LJ 1575; Colin Bradley and Bernadette Currie, "Advances in the Field of Reverse Engineering" (2005) 2 Computer-Aided Design & Applications 697.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

them to protect as trade secrets any valuable chain-pairing information that comes with it.

4.2.2 Antigen specificity data

AIRR-seq data itself, even with chain-pairing, does not necessarily inform researchers which antigens bind to which immune receptors.[145] To obtain such information, researchers often need to engage in time-consuming validation studies: Expressing any immune receptors of interest in cultures of cells, isolating the immune receptors themselves, and then measuring which ones bind to an antigen of interest.[146] In some instances—if, for example, the sampled cells have been collected from a patient with a known infection—antigen specificity can be inferred: Sets of immune receptors produced at significantly higher volumes are likely to be specific to an antigen related to the infection.[147] But in most cases, the antigen specificity of any given AIRR sequence, standing alone, is a mystery.[148]

Some recent work by AIRR researchers, such as Lindsay Cowell at the University of Texas Southwestern and others, has begun to attempt to computationally infer antigen specificity from gene sequence data.[149] By assigning a biochemical "score" to given stretches of DNA sequences, researchers can zero in on the physical structure of the resulting immune receptor and predict to which antigen the immune receptor is likely to bind.[150] Accurate analogies are hard to come by—but the technique is, perhaps, akin to using the sound made by a key fitting into a lock to predict the shape of both the lock and the key. The technique can be medically powerful, too. In some circumstances, it can be used to accurately diagnose patients exhibiting symptoms of two closely related diseases—such as two variations of multiple sclerosis.[151] At the same time, the power of the technology is—to date—limited. Computationally inferred antigen specificity data still must be validated by experiment, and the system still needs to be trained on "known" samples—samples derived from patients with known illnesses, for example.[152]

---

[145] Rubelt and others (n 13) 1275.

[146] Jennifer Bordeaux and others, "Antibody Validation" (2010) 48 Biotechniques 197.

[147] Schlessinger and others (n 80) D777; Vita and others (n 80) D854.

[148] Schlessinger and others (n 80) D777; Vita and others (n 80) D854.

[149] Jared Ostmeyer and others, "Statistical Classifiers for Diagnosing Disease from Immune Repertoires: A Case Study Using Multiple Sclerosis" (2017) 18 BMC Bioinformatics 401.

[150] Ibid.

[151] Ibid.

[152] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

Consequently, this information, like chain-pairing data, is clearly "follow-on data" arising from the interpretation of larger AIRR-seq datasets. And it, too, is likely to fit within the Trade Secrets Directive's definition of a "trade secret." Antigen specificity—without more information about the patient source of the data or use of computational inference techniques—would not be "generally known" among AIRR researchers. It would also have significant commercial value to developers interested in developing predictive diagnostics using immune receptor sequence data.[153] And the data could easily be subjected to reasonable steps to keep it secret. Nor does it seem that sequestering antigen specificity data violates any of the norms of AIRR-seq data sharing or the MiAIRR standard. Researchers could still freely share the underlying AIRR-seq data itself with others without disclosing its antigen specificity. And there is nothing in the MiAIRR standard that appears to compel the disclosure of antigen specificity even if known. Conceiving of antigen specificity data as protectable follow-on data, even if derived from publicly available MiAIRR datasets, seems to comport with both the purposes of the Trade Secrets Directive as well as scientific practices.

4.3 Recombinant Immune Receptors

Perhaps the most commercially significant follow-on information derived from AIRR-seq data is the development of "recombinant" immune receptors—immune receptors, or fragments of immune receptors, that are then further modified, or "recombined," to have an enhanced or superlative effect relative to the underlying "natural" sequence.[154] Recombinant immune receptors are critical in the development of vaccines and other therapies. They allow developers to modulate the binding affinity between a given immune receptor and its antigen; to develop immune receptors that bind more specifically to particular antigens than their native versions; to bind to certain areas of an antigen relative to others; and to recombine—or "fuse"—immune receptors with other proteins to give them functions they did not otherwise possess.[155] There are other potential applications as well.[156]

Researchers and commercial developers use AIRR-seq data to create recombinant immune receptors in a variety of ways. Targeted mutagenesis is one particular technique: Researchers iteratively "mutate," or change, certain

---

[153] Ilan Kirsch, Marissa Vignali, and Harlan Robins, "T-Cell Receptor Profiling in Cancer" (2015) 9 Molecular Oncology 2063.

[154] Robinson (n 11) 172.

[155] Daniel M Czajkowsky and others, "Fc-Fusion Proteins: New Developments and Future Perspectives" (2012) 4 EMBO Molecular Med 1015.

[156] Robinson (n 11) 175–6.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

sequences along the immune receptors' chains to produce a certain effect, such as a stronger binding affinity to a particular antigen.[157] Directed evolution is yet another technique, where researchers intentionally create small, random mutations in sequences of interest, and continually screen the products of their work until they achieve a particular result.[158]

Under any method, however, researchers are left with a sequence that is often subtly different from the original sequence—and one that would have neither been predictable nor readily apparent from the natural sequence itself.[159] It is often this sequence itself—the recombinant sequence—that is of particular commercial value.[160] To date, several commercial companies have been established on the foundation of recombinant immune receptor sequences derived from AIRR-seq data.[161]

Protecting the commercial viability of these sequences remains important for commercial developers.[162] To that end, it seems to be clear that the disclosure of underlying AIRR-seq data would not upend trade secrecy for any recombinant sequences derived from it. The AIRR-seq data from which any recombinant sequence is derived is not the same as the sequence information produced by commercial developers. Thus, disclosure of the underlying AIRR-seq data would not, as a matter of law, make public any recombinant sequences sought to be protected. Further, the recombinant sequences would neither be known nor readily accessible simply based on the disclosure of the underlying sequence itself.[163] In the case of targeted mutagenesis, while researchers may very well have an intuition of which sequences would be ideally modified, it would not be clear what those modifications were or whether they would actually work to produce the effect sought. And as for directed evolution, because the process is, essentially, random, predicting sequence variations ahead of time is highly unlikely.

In these cases, researchers seeking to protect, as trade secrets, recombinant immune receptor sequences should not fear the disclosure of the underlying AIRR-seq data from which they derive. This leaves researchers free to disclose largescale AIRR datasets, even under public sharing protocols such as

---

[157] Tomoyuki Igawa and others, "Engineering the Variable Region of Therapeutic IgG Antibodies" (2011) 3 mAbs 243.

[158] Yi Li and others, "Directed Evolution of Human T-cell Receptors with Picomolar Affinities by Phage Display" (2005) 23 Nat Biotech 349.

[159] Ibid.

[160] IP Issues in Data Sharing (n 138).

[161] Corbeau Biotech, LLC (Nashville, Tennessee, USA); Oxford BioTherapeutics (Abingdon, UK); Symphogen (Ballerup, Denmark).

[162] IP Issues in Data Sharing (n 138).

[163] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

MiAIRR, without losing otherwise valuable protection to follow-on information. Recombinant sequences, in this way, seem to best comport with the ideals of the Trade Secrets Directive, to both promote the "dissemination of knowledge and information" and "protect access to, and exploit, knowledge that is valuable … and not widely known."[164]

5. THE TRADE SECRETS DIRECTIVE AND GENETIC DATA SHARING

While the issues surrounding trade secrecy and follow-on AIRR-seq data are themselves worthy of exploration, they also shed light on several other quandaries related to trade secrets and the disclosure of scientific information. They inform us about the dichotomous nature of trade secrets versus patent protection. Because the election of secrecy and patent protection are mutually exclusive, the current absence of robust patent protection for AIRR-seq provides a natural experiment, of sorts, regarding the value of trade secrecy for scientific information. The issues concerning trade secrecy and follow-on AIRR-seq data also highlight some of the difficulties regarding the disclosure and protection of other largescale datasets in Europe, such as clinical trials. Lastly, the issues above are informative about protecting diagnostic technologies—underlying technologies that create the follow-on information sought to be protected in the first instance. Whether follow-on AIRR-seq will be protected by trade secrets under the Trade Secrets Directive, and the extent to which this may be the case, will likely be informative for these other areas as well.

5.1 Trade Secrecy vs Patent Protection

The importance of trade secrets in follow-on AIRR-seq data arises, in part, due to the lack of available patent protection. Patent protection for genetic sequences is, in many instances, suboptimal.[165] Given advances in HTS technology, patent protection is unlikely to compel the elucidation of genetic sequence technology.[166] It is unclear whether HTS technologies, like AIRR-seq, would even infringe genetic sequence patents.[167] And gene sequence patents

[164] EU Trade Secrets Directive, preamble 3; Ibid 1.

[165] Robert Cook-Deegan and Christopher Heaney, "Patents in Genomics and Human Genetics" (2010) 11 Ann Rev Genomics Human Genetics 383.

[166] Jacob S Sherkow and Ryan Abbott, "Fortune and Hindsight: Gene Patents' Muted Effect on Medical Practice" (2018) 126 Br Med Bull 37.

[167] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

are, for a variety of reasons, difficult to enforce.[168] But AIRR-seq data seems to demonstrate that an absence of patent protection may channel researchers interested in protecting valuable sequences toward trade secrecy.[169] Whether protection as trade secrets for AIRR-seq related data is ever enforced (as opposed to merely kept as de facto secrets) remains unclear.[170] But AIRR-seq data seems to provide a case study of the disclosure tradeoffs between patents and trade secrecy.

This tradeoff arises from the dichotomous nature of trade secrets and patents. In the classical account, patents encourage scientific disclosure by giving inventors *ex post* protectable rights to the inventions disclosed.[171] The act of disclosure itself, therefore, does not destroy the economic value of the shared information.[172] Because trade secrets are antithetical to patents, however, inventors must choose beforehand whether to protect their information through either patents or trade secrets—they cannot choose both.[173]

Several legal shifts, in both Europe and the United States, have curtailed the availability of patents for genetic sequences and other data-driven inventions. First, a series of three US Supreme Court decisions delivered from 2012 to 2014 created enormous legal uncertainty as to whether genetic data, as well as data and algorithm-based technologies, are considered patentable subject matter.[174] These have been implemented at the USPTO with the adoption of several important "guidances" on examining claims related to genetic material, imposing increasingly strict limitations on the patentability of natural products and methods using laws of nature.[175] While patents for genetic sequences are still granted in Europe, broad claims covering "naked" genetic sequences are no longer available in light of scientific advances in sequencing; such claims now often fail to meet traditional patentability requirements, such as novelty and inventive step.[176] These developments, in combination with stricter guide-

---

168 Ibid.

169 IP Issues in Data Sharing (n 138).

170 Ibid.

171 Jeanne C Fromer, "Patent Disclosure" (2009) 94 Iowa L Rev 539, 548.

172 Ibid.

173 Alan Devlin, "The Misunderstood Function of Disclosure in Patent Law" (2010) 23 Harv JL & Tech 401, 417–18.

174 *Bilski v Kappos* [2010] 561 U.S. 593 (S. Ct.) (U.S.); *Mayo Collaborative Servs v Prometheus Labs* [2012] 566 U.S. 66 (S. Ct.) (U.S.); *Alice Corp v CLS Bank Int'l* [2014] 134 S Ct 2347 (U.S.).

175 USPTO, Nature-Based Products (2014) www.uspto.gov/sites/default/files/documents/mdc_examples_nature-based_products.pdf accessed June 3, 2018 (archived at https://perma.cc/LKB4-Q7HG).

176 Timo Minssen, "Patenting Human Genes in Europe and How It Compares to the US and Australia" in Duncan Matthews and Herbert Zech (eds), *Research Handbook on Intellectual Property and the Life Sciences* (Edward Elgar 2017) 26–39.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

lines on the application of patent criteria and the absence of sufficient parallel protection in the United States, often make reliance on traditional patent strategies unattractive.[177] This is one of the primary reasons why AIRR researchers rarely, if ever, seek patent protection for valuable immune receptor sequences and may rather be inclined to focus on trade secret protection.

Trade secrecy law has also become increasingly harmonized—making it an increasingly attractive form of worldwide IP protection.[178] This is, of course, the purpose of the Trade Secrets Directive: "to ensure … a sufficient and consistent level of civil redress" in the EU. But the United States has also recently passed its version of the Trade Secrets Directive, the Defend Trade Secrets Act—a federal law largely viewed as supplanting (although not explicitly so) a prior regime of individual trade secrecy laws in all 50 US states.[179] The rise of AIRR-seq data, beginning in 2009 and continuing into the present, parallels these legal developments: As the AIRR research community has grown, so has the international harmonization of trade secrecy law. Coupled with the absence of patent protection, international developers of AIRR-seq follow-on data can take increasing solace in trade secrecy as both an adequate and uniform form of IP protection.

Lastly, *de jure* protection of trade secrets is substantially easy to obtain—like copyright, trade secrecy protection comes into existence at the moment of creation, without any particularized examination process.[180] By extension, this means that trade secrets that would otherwise fall foul of various patentability requirements—such as proving an "inventive step" or "industrial application"—can still be protected without an application or other administrative formalities. This absence of a requirement for a trade secret to demonstrate any statutory requirements before qualifying for protection makes trade secrets cheap, easy, and flexible. This, too, is an attractive option for AIRR-seq follow-on data, created rapidly, sometimes en masse, and without the immediate knowledge of the full value of follow-on information. Above all, AIRR-seq data teaches that, in the absence of patent protection, trade secrecy may be an attractive option for economically valuable follow-on data.

---

[177] Sherkow and Abbott (n 166) 2.

[178] Anand B Patel and others, "The Global Harmonization of Trade Secret Law: The Convergence of Protections for Trade Secret Information in the United States and European Union" (2016) 83 Def Counsel J 472.

[179] EU Trade Secrets Directive, preamble ¶10.

[180] Neil Wilkof, "The Cost of Trade Secrets: Don't Overlook the Psychological Price Being Paid" (2017) 12 J Intell Prop L & Practice 715, 715.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

5.2 Clinical Trial Data Disclosure and Trade Secrets

Data from clinical trials—largescale, human studies conducted by therapeutic developers to obtain regulatory approval—remain some of the most valuable data in the world.[181] Almost universally, commercial developers have protected their clinical trial data as secrets.[182] Somewhat like the tension surrounding disclosing raw AIRR-seq data, the worry behind releasing clinical trial data to the public is that the disclosure would obviate protection for any follow-on information derived from it.[183] This even includes cases concerning "negative" data—results of an examined therapeutic development having no beneficial clinical effect or no statistically significant effect at all.[184] Like AIRR-seq data, the perceived value to follow-on information produces friction to sharing raw information.

Several jurisdictions have responded to these difficulties in disclosure by proposing numerous regulations mandating, or strongly encouraging, the sharing of clinical trial information.[185] In the European Union this is governed by Regulation No 536/2014, On Clinical Trials on Medicinal Products for Human Use (the Clinical Trial Regulation).[186] The Clinical Trial Regulation requires companies to submit clinical trial data to a publicly accessible EU database.[187] At the same time, the Regulation provides that a portion of the information can be withdrawn on the grounds of "protecting commercially confidential information."[188] While the Regulation has had its critics, it has largely been lauded as a compromise between disclosure and secrecy of commercially valuable information.[189]

Several recent rulings by the CJEU interpreting the Clinical Trial Regulation align well with distinctions between basic scientific information, such as raw AIRR-seq data, and follow-on information. In *Pari Pharma v EMA*, for

---

[181] Rebecca S Eisenberg, "The Role of the FDA in Innovation Policy" (2006) 13 Mich Telecomm & Tech L Rev 345.

[182] Ibid.

[183] Ibid.

[184] Ibid, 347 n 5.

[185] Michelle M Mello and others, "Preparing for Responsible Sharing of Clinical Trial Data" (2013) 369 New Engl J Med 1651.

[186] Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on Clinical Trials on Medicinal Products for Human Use, and Repealing Directive 2001/20/EC [2014] OJ R158/1 (Clinical Trial Regulation).

[187] Ibid.

[188] Ibid.

[189] Sabine Atzor, Surendra Gokhale, and Michael Doherty, "Will the EU Clinical Trials Regulation Support the Innovative Industry in Bringing New Medicines Faster to Patients?" (2013) 27 Pharm Med 75.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

example, the CJEU upheld the EMA's decisions to release documents from clinical trials under its older "Transparency Regulation."[190] While the Court rejected the companies' objections to the data's release, it also provided a roadmap for future efforts of nondisclosure of clinical trial data context. Specifically, the CJEU allowed the nondisclosure of "new scientific conclusions" or "an inventive strategy" derived from clinical data.[191] In *PTC Therapeutics International v EMA*, the CJEU similarly required the disclosure of a "clinical study report" ("CSR")—a summary of clinical trial data—on the grounds that the CSR was not follow-on information, but coextensive with the underlying data itself, the "latter being accessible to the public and containing data emanating directly from the report at issue."[192]

Recently, however, this interpretation of the Clinical Trial Regulation has been challenged. In the course of an appeal of *PTC Therapeutics International v EMA*, the Advocate General recommended that the court set aside the CJEU's judgment and refer the matter back to the General Court.[193] The Advocate General's recommendation stemmed from the fact that the information disclosed in a CSR—even if coextensive with that given to the EMA—"would … be of considerable advantage to any potential competitor … [providing] insight into the working methods, methodologies, etc. … perhaps even to the point of providing a 'road map' for future [marketing authorization] applications—not least in a commercial environment which is exceptionally competitive."[194] Whether the CJEU will adopt the Advocate General's analysis—despite its earlier precedent—remains unclear as of this writing.

The tension highlighted by the Advocate General's opinion mirrors some advantages and disadvantages of the Regulation's approach. As one example, the case of AIRR-seq data and recombinant sequences gives some credence

---

[190] Regulation (EC) No 1049/2001 [2001] OJ L145/43; see also Case T-235/15, *Pari Pharma v EMA* [2018] CJEU 65 (in relation to the disclosure of similarity and superiority reports on an orphan medicine, prepared by the Committee for Medicinal Products for Human use (CHMP)); Case T-718/15, *PTC Therapeutics Intl v EMA* [2018] CJEU 66 (on the disclosure of a clinical study report); Case T-729/15, *MSD Animal Health Innovation and Intervet Intl v EMA* [2018] CJEU 67 (regarding five toxicology study reports for a veterinary medicine).

[191] *Pari Pharma* (n 190) at para 154. At the same time, the Court imposed a high standard of proof on companies seeking to prevent disclosure by requiring that they "describe in specific terms the professional and commercial importance of the information … and the utility of that information for other undertakings which are liable to examine and use it subsequently," for example "to show specifically and actually how, once the documents have been disclosed, competitors would be able to enter the market." Ibid.

[192] *PTC Therapeutics* (n 190) at para 89.

[193] Case C-175/18 P, *PTC Therapeutics Intl v EMA* [2019] Op Adv Gen 709.

[194] Ibid at para 76.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

to developers' arguments that negative data is, too, commercially valuable. Negative clinical trial data, as much as positive clinical trial data, can lead to 'readily available' commercially valuable information about the direction of therapeutic development: improvements for compositions, new medical uses, and successful dosage regimens, for example.[195] This can be contrasted, for example, with the concerns surrounding the release of native versus recombinant sequences of immune receptors. Because recombinant sequences would not be readily available from their native cousins, it makes little sense to protect the latter as trade secrets; but because drug development information can, in some instances be readily available from negative trial data, it *does* make sense to protect the latter.

The Clinical Trial Regulation also provides a cautionary tale about Member State flexibility. The Trade Secrets Directive does, despite its unifying effect, allow Member States to provide for their provisions concerning enforcement that include the "public interest." Given that different Member States are likely to have different views about this public interest provision, it may cow commercial developers into overextending their claims to secrecy—as was the case in *PTC Therapeutics*. If the effect of the Trade Secrets Directive is to provide a safe space for the disclosure of raw sequence information, the effect of the Clinical Trial Regulation's amorphous "public interest" exemption may be the opposite.

5.3 Trade Secrecy, Disclosure, and Diagnostics

Follow-on AIRR-seq data, under the Trade Secrets Directive, may also be informative to best practices for the protection of genetic diagnostics—the tools used to create and interpret genetic data in the first instance. While patents for diagnostics have waned, trade secrecy for diagnostics is on the rise.[196] This is problematic for several reasons: trade secrecy interferes with practitioners' understanding of diagnostics' outputs; it hides whether diagnostic results are reproducible; it risks public health.[197] At the same time, information concerning how diagnostics arrive at the outputs they do and how, in other instances, they interpret those outputs often constitutes the principal commercial value of the diagnostic.

The treatment of AIRR-seq follow-on data suggests that follow-on information can be kept secret while disclosing the raw AIRR-seq data. This

---

[195] Eisenberg (n 181) 347 n 5.

[196] Robert Cook-Deegan and others, "The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?" (2012) 21 Eur J Human Genetics 585.

[197] Ibid.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

parallels the disclosure of diagnostics *inputs* while keeping their *outputs* secret. This makes sense given that diagnostics must rely on interpretive software.[198] Diagnostic platforms typically focus on largescale screening of genes, proteins, and metabolites, and derive value from their ability to distinguish between multitudes of diseases when paired with bespoke analytical software.[199] As such, it is this analytical software that could represent valuable trade secrets.[200] For example, a diagnostic device, although having a unique combination of patent-protected biomarkers, may only provide a meaningful diagnosis when paired with relevant software.[201] This is akin to bioinformatics tools used to infer AIRR sequences, create chain pairs, or attempt to predict the binding affinity of immune receptors. If the follow-on AIRR-seq data is any guide, then best practices suggest protecting outputs rather than inputs.

6. OPEN QUESTIONS

Much like the CJEU's future interpretation of the Trade Secrets Directive, the Directive's applicability to AIRR-seq data leaves open a number of questions for further investigation. First, as the technology progresses, it is unclear how tenable the distinction between "raw" and "follow-on" data will remain. As this gap shrinks, it may very well be the case that disclosure of raw data to relevant "circles" of scientists may ultimately prevent the availability of trade secrecy protection on even follow-on data under the Trade Secrets Directive.[202] The commercial consequences of such an interpretation are unknown, as is whether this would encourage researchers to abandon disclosure of even basic AIRR-seq data.

Second, it is not entirely clear whether the protectability of raw data could also be challenged under the "commercial value" prong, especially given the Advocate General's expansive interpretation of "commercial value" in another context under the Clinical Trial Regulation.[203] As we have described above, the true commercial value of the raw data is normally first revealed during a follow-on process, that is, when skilled researchers routinely apply specialized algorithms to the raw AIRR-seq data. Here too, it could also be argued that the potential value, that is, the mere possibility of extracting such infor-

---

[198] Rochelle C Dreyfuss and James P Evans, "From *Bilski* Back to *Benson*: Preemption, Inventing Around, and the Case of Genetic Diagnostics" (2011) 63 Stan L Rev 1349.

[199] Ibid.

[200] Ibid.

[201] Ibid.

[202] See EU Trade Secrets Directive, art 2.1(a)–(c).

[203] *PTC Therapeutics* (n 190).

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

mation from raw AIRR-seq data, would make any dataset valuable enough to fulfill the 'commercial value' prong.

Third, if this view prevails, where raw data itself would be considered "commercially valuable" under the Trade Secrets Directive, this may open the door to claims from universities regarding closing access without formal agreements in place. Under some national regimes, some universities might even be *obliged* to pursue commercial strategies—a potential that, oddly, finds support in the Trade Secrets Directive itself, in Recital 1.[204] Much will depend on the particular circumstances of each research project, but courts, lawyers and legal researchers, scientists, and universities will certainly have to consider such questions as the technology begins to become more robust and the CJEU, and national courts, begin to consider similar cases.

7. CONCLUSION

On the one hand, the EU Trade Secrets Directive attempts to reconcile two competing goals: the protection of commercially valuable information and the promotion of open innovation. On the other, certain stipulations, such as Recital 1 of the Preamble, would seem to suggest that even raw data may be valuable to noncommercial research institutions. Largescale genetic datasets, such as those created by AIRR sequencing, serve as one concrete case study of how both goals can—or cannot—be achieved. AIRR-seq data is enormously complex, costly to create, and difficult to interpret. Academic researchers, in turn, are usually bound to norms of open sharing of raw sequence data, which, in typical circumstances, would destroy any protection of the underlying data itself as being made public under the Directive. In light of this, the temptation to reorient these norms toward commercial enterprise might be strong. And in

[204] See EU Trade Secrets Directive, recital 1 of the preamble: "(1) Businesses and non-commercial research institutions invest in acquiring, developing and applying know-how and information which is the currency of the knowledge economy and provides a competitive advantage. This investment in generating and applying intellectual capital is a determining factor as regards their competitiveness and innovation-related performance in the market and therefore their returns on investment, which is the underlying motivation for business research and development. Businesses have recourse to different means to appropriate the results of their innovation-related activities when openness does not allow for the full exploitation of their investment in research and innovation. Use of intellectual property rights, such as patents, design rights or copyright, is one such means. Another means of appropriating the results of innovation is to protect access to, and exploit, knowledge that is valuable to the entity and not widely known. Such valuable know-how and business information, that is undisclosed and intended to remain confidential, is referred to as a trade secret."

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY

an era where political and democratic norms are quickly dissolving, it is not entirely clear how binding and robust the norms of open science really are.

But even if university norms and noncommercial scientific ideals of open sharing of raw sequence data prevail, the complexity of AIRR-seq data leaves an opening for the protection of commercially valuable follow-on information that arises from the interpretation of underlying data. Follow-on information of this sort includes pairing data of immune receptor chains, antigen specificity, and the development of recombinant proteins. In this way, the Trade Secrets Directive shows that—despite the conflicting goals of protectable secrecy and open innovation—it might be possible to develop open innovation platforms while allowing protection, and secrecy, for related information. Much will depend on the contextual traditions, (user-generated) standards, and habits of enforcement.

Viewed through this lens, this case study on AIRR-seq data under the Trade Secrets Directive is informative about other problems in the law concerning disclosure and secrecy: the tradeoff between patents and trade secrets, the sharing of proprietary clinical trial data, and the protection of genetic diagnostics. AIRR-seq data shows that it is at least theoretically possible to simultaneously promote information sharing and reserve the most value for aspects of that information to encourage private development of new technologies. At the same time, trade secrets and traditional intellectual property rights will continue to be only one part of the equation reconciling the aspirations and opportunities of data sharing with the harsh realities of business and research competition in biopharmaceutical development. Successful sharing strategies will need to take into account other *sui generis* rights, such as data protection, regulatory exclusivities, and anticompetition laws—many of which are discussed in Chapter 14 of this book.

Jacob S. Sherkow and Timo Minssen - 9781788973342
Downloaded from Elgar Online at 08/19/2020 11:13:53AM by emma.penton@e-elgar.co.uk
via AUTHOR COPY - NOT TO BE POSTED IN AN OPEN ONLINE REPOSITORY