ROUGH DRAFT – PLEASE DO NOT CITE OR CIRCULATE

INCENTIVES, INTELLECTUAL PROPERTY, AND BLACK BOX PERSONALIZED MEDICINE

W. Nicholson Price II[*]

ABSTRACT

Personalized medicine is reshaping the biomedical landscape. Where Big Data meets Big Health, it has been hailed as the next leap forward in health care, and is both a subject of health law and an object of innovation policy. Humans are inherently variable, and closely matching treatment to patients has the potential to save and extend lives by suggesting better treatments, to avoid unnecessary treatment, and to streamline the process of drug discovery and clinical trials—all important innovation goals. But the version of personalized medicine being implemented today is just an entrée into the realm of what huge amounts of data can tell us about our health and how to improve it. Current versions of personalized medicine rely on the simple relationships that we can explicitly identify and validate in clinical trials. But biology is complicated. This paper introduces into legal scholarship the concept of black box personalized medicine, which seeks to use more directly that biological complexity by finding and using more complex, implicit biological relationships within the troves of health data we are increasingly amassing. This new form of personalized medicine offers potentially immense benefits, but requires high investment in developing new data, models, and applications—all of which are hard to protect once they become public. The current set of intellectual property incentives, particularly after the Supreme Court's recent decisions in *Prometheus* and *Myriad*, fails to provide the necessary incentives for that investment, and instead pushes firms toward simple diagnostics paired with devices or trade secrecy and proprietary data. This paper addresses the concepts underlying black box personalized medicine, explains why the current intellectual property landscape provides inadequate and misdirected incentives, and briefly suggests policy options to better align incentives.

[*] Assistant Professor, University of New Hampshire School of Law. J.D., Columbia Law School, 2011. Ph.D. (Biological Sciences), Columbia Graduate School of Arts and Sciences, 2010. A portion of this work was completed while an Academic Fellow at Petrie-Flom Center for Health Law Policy, Biotechnology and Bioethics at Harvard Law School. I wish to thank Ana Bračič, Glenn Cohen, Matt Lawrence, Kevin Outterson, Geertrui Van Overwalle, Ben Roin, and Jeff Skopek for their helpful comments and feedback. This work benefited from feedback at the Health Law Professors' Conference and the Munich Conference on Innovation and Competition. All errors are my own.

TABLE OF CONTENTS

INTRODUCTION

Personalized medicine, where Big Data meets Big Health, has been hailed as the next leap forward in health care.[1] It is already developing and

[1] For descriptions of personalized medicine in the medical literature, s*ee generally* Isaac S. Chan & Geoffrey S. Ginsburg, *Personalized Medicine: Progress and Promise*, 12 ANNU. REV. GENOMICS HUM. GENET. 217 (2011); Edward Abrahams & Mike Silver, *The Case for Personalized Medicine*, 3 J. DIABETES SCI. TECHNOL. 680 (2009); Wylie Burke &

spreading rapidly: doctors are using increasing amounts of personal data, especially genetic diagnostic tests, to tailor treatments to individual patients.[2] Humans and diseases are inherently variable in many dimensions, genomic and otherwise; as a result, 38% of patients with depression, 40% with asthma, and 75% with cancer fail to respond to treatment, belying the efficacy of a one-size-fits-all model of medicine.[3] When medical science can determine what predicts *which* fraction of patients will respond to a treatment, that treatment can be matched to the right patients. Personalized medicine can thus save and extend lives by suggesting better treatments, and can help avoid the cost and risk of unnecessary medical interventions.[4] In addition to aiding patient care, personalized medicine can speed and streamline the process of drug discovery and clinical trials by identifying which patients a developing drug is most likely to help.[5]

But the version of personalized medicine being implemented today—what I dub explicit personalized medicine (EPM)—is just an entrée into the realm of what huge amounts of data can tell us about our health and how to improve it. Current versions of personalized medicine (and of health care in general) frequently rely on what we can explicitly understand: relatively simple relationships that can be identified and validated in clinical trials that group large numbers for statistical power. But biology is complicated; many important relationships aren't one-to-one, two-to-one, or several-to-one correspondences, but are instead networks between dozens of interacting variables, including those which are readily observable (age, weight, or sex) and those which are less so (genomic markers or metabolite levels).[6]

Bruce M. Psaty, *Personalized Medicine in the Era of Genomics*, 298 JAMA J. AM. MED. ASSOC. 1682 (2007); G. S Ginsburg & J. J McCarthy, *Personalized Medicine: Revolutionizing Drug Discovery and Patient Care*, 19 TRENDS BIOTECHNOL. 491 (2001); Margaret A Hamburg & Francis S Collins, *The Path to Personalized Medicine*, 363 N. ENGL. J. MED. 301 (2010).

[2] *See* Chan & Ginsburg, *supra* note 2.

[3] Spear BB, et al. Trends Mol Med 2001;7:201-4; Brian B. Spear et al., *Clinical Application of Pharmacogenetics*, 7 TRENDS MOL. MED. 201 (2001)

[4] *Id.*

[5] Lawrence J. Lesko et al., *Pharmacogenetics and Pharmacogenomics in Drug Development and Regulatory Decision Making: Report of the First FDA-PWG-PhRMA-DruSafe Workshop*, 43 J. CLIN. PHARMACOL. 342, 348–55 (2003).

[6] For instance, one recent technique used genetic sequence data from 5,000 genes to classify two different types of lung tumor with very high accuracy; the two types of tumor respond best to different therapies. Hojin Moon et al., *Ensemble Methods for Classification of Patients for Personalized Medicine with High-Dimensional Data*, 41 ARTIF. INTELL. MED. 197, 198, 203–04 (2007). The same team's efforts to predict distant metastasis of breast cancer tumors were less successful. *Id.* at 204–05.

This paper introduces into legal scholarship the concept of black box personalized medicine (BBPM), which seeks to use more directly that biological complexity.[7] Black box personalized medicine, pursued by geneticists, personalized medicine advocates, and other health care innovators, already does and will increasingly use the combination of large-scale high-quality datasets with sophisticated predictive algorithms to identify and use implicit, complex connections between multiple patient characteristics.[8] By capturing those implicit, complex connections—and by at least sometimes allowing their use with algorithmic validation rather than relying on clinical trials[9]—BBPM lays open far more possibilities for shaping treatment and drug development. Assuming *arguendo* that BBPM desirable, the path there presents challenges.[10]

Costs and hurdles exist at each phase of BBPM's development.[11] First, information must be gathered and vetted, which requires financial resources and navigating legal requirements including privacy and informed consent. Second, developing reliable and sensitive algorithms needs dedicated effort by sophisticated programmers. The experience of other predictive algorithms demonstrates this; for example, the movie-rental service Netflix created a three-year, multi-million dollar prize effort to improve its simple movie-prediction algorithm, in which thousands of teams managed to improve the algorithm's performance by only 11%.[12] Third, since complex implicit predictions are much less amenable to the forms of validation on

---

[7] This paper focuses on introducing the concept of black box personalized medicine and addressing the innovation policy questions surrounding its development. As a major development in health care and biomedical science, BBPM has other legal implications on, for example, the application of informed consent and patient autonomy, medical tort law, corporate practice of medicine doctrines, and regulatory oversight, but these must remain the subject of future work.

[8] Amarasingham and colleagues describe one form of BBPM, "predictive analytics," involving the use of real-time large datasets and predictive algorithms to help inform treatment decisions such as who should be sent first to intensive care units. Ruben Amarasingham et al., *Implementing Electronic Health Care Predictive Analytics: Considerations And Challenges*, 33 HEALTH AFF. (MILLWOOD) 1148 (2014). Other forms of BBPM, described below, relate to the choice of which drugs to give to patients, or complex interacting constellations of disease risk factors. *See* Section I.A.3, *infra*.

[9] Bypassing clinical trials in at least some instances is not as dramatic as it sounds. Current practices in off-label drug use frequently involve treatment based on correlations, connections, and hypotheses without the backstop of well-controlled clinical trials.

[10] Although this Article takes a sympathetic view of black box personalized medicine, its purpose is not to make the case for BBPM. Rather, assuming that BBPM is a positive form of innovation to be encouraged, it seeks to analyze the incentives available, identify their shortcomings, and suggest ways to improve those incentives.

[11] *See* Section II, *infra*.

[12] *See* http://www.netflixprize.com/; notes 78–86, *infra*, and accompanying text.

which we traditionally rely—scientific understanding, clinical trials, and postmarket surveillance—other forms of validation must be developed by the innovating firm, regulators, and/or third parties.

Overcoming these hurdles will require significant incentives, and pure market incentives are likely to be woefully insufficient. BBPM follows the classic pattern justifying intellectual property, in which firms underinvest in non-excludable information goods because they cannot capture the full social value of those goods.[13] BBPM relies principally on pure information goods: collected data, patterns discovered within that data, and validation of those patterns. Intellectual property allows firms to exclude others from the information good and therefore appropriate a higher portion—though not all—of the surplus, increasing innovation closer to optimal levels.

The current intellectual property regime not only provides inadequate incentives for BBPM, the incentives it provides push the field in counterproductive directions. The primary intellectual property incentives for technological innovation are provided by patents. Although patents are imperfect at driving algorithm development, as discussed below,[14] they still create significant incentives. Until quite recently, method patents were broadly available for diagnostic algorithms, as long as they satisfied the Federal Circuit's requirements that the invention must involve a machine or a transformation of matter—which could be satisfied by as little as performing a blood test.[15] But in 2012 the Supreme Court held in *Mayo v. Prometheus Labs* that a patent covering a standard diagnostic method—administering a drug, measuring the level of a metabolite, and knowing that

---

[13] *See, e.g.*, K. J. Arrow, *Economic Welfare and the Allocation of Resources for Invention" in RR Nelson (ed.), The Rate and Direction of Inventive Activity. Princeton, Princeton University Press*, 619 (1962) ("To sum up, we expect a free enterprise economy to underinvest in invention and research (as compared with an ideal) because it is risky, because the product can be appropriated only to a limited extent, and because of increasing returns in use."); *see also* Benjamin Roin, *Patent Effectiveness*, at 18–20 (2014).

[14] *See* Section II.A, *infra*.

[15] When the Federal Circuit first addressed the *Prometheus* case, it held that testing blood for the presence of metabolites was a "transformation" sufficient to make the invention patentable. Prometheus Laboratories v. Mayo Collaborative Ser., 581 F 3d 1336, 1347 (Fed. Cir. 2009). After the Supreme Court's decision in *Bilski v. Kappos*, the Federal Circuit's "machine or transformation test" went from a dispositive test to an "important and useful clue" as to whether the invention covers patentable subject matter. 130 S.Ct. 3218, 3226–27 (2010). However, the importance of this clue to the Federal Circuit was such that it remained practically dispositive. *See* Prometheus Laboratories v. Mayo Collaborative, 628 F 3d 1347, 1355 (Fed. Cir. 2010) (holding in *Prometheus* on remand after *Bilski* that "as applied to the present claims, the 'useful and important clue, an investigative tool,' leads to a clear and compelling conclusion, viz., that the present claims pass muster under § 101.").

certain metabolite levels suggest the need to increase or decrease the drug's dosage—was unpatentable as essentially claiming—and thus preempting—a law of nature.[16] Close on the heels of *Prometheus*, the Court decided *Association for Molecular Pathology v. Myriad* in 2013, holding isolated genomic DNA unpatentable as a natural phenomenon;[17] such DNA patents, while not essential to diagnostic testing methods, provided secondary protection to those involving genetic testing.[18]

After *Prometheus* and *Myriad*, incentives for developing personalized medicine—and especially the complex algorithms at the heart of BBPM—are much lower than they were before for two reasons.[19] First, a core set of patents on pure algorithms are likely unavailable. Second, a zone of patents around that core are now of uncertain validity or accessibility.[20]

Perhaps more importantly, the incentives which remain available now pull personalized medicine in the wrong direction. Because patents are on stronger ground when they cover inventions that closely link devices or treatments to the new correlation or algorithm, firms are likely to prioritize development of those combination products rather than pursuing broader analyses of large datasets and complex correlations within them. This pulls firms toward maintaining the current model of simple, explicit relationships rather than developing and exploiting the far larger realm of complex and often implicit relationships.

In addition, firms may increasingly turn away from the patent system and rely instead on trade secrecy law and practices to protect proprietary data and algorithms. Secrecy is problematic for medicine in general, but

---

[16] Mayo Collaborative v. Prometheus Labs., 566 US 10 (2012).

[17] Ass'n for Molecular Pathology v. Myriad, 133 Ct 2107 (2013).

[18] If a firm cannot fully protect diagnostic methods or algorithms that involve a piece of genetic information, patents on the isolated gene of interest can still prevent others from determining the gene variant and therefore practicing the method. Myriad Genetics used this strategy to protect its breast cancer diagnostic tests. *Ass'n for Molecular Pathology v. Myriad*, 133 Ct 2107. This strategy is imperfect; indeed, whole-genome sequencing likely circumvents isolated gene patents, W. Price & I. I. Nicholson, *Unblocked Future: Why Gene Patents Won't Hinder Whole Genome Sequencing and Personalized Medicine*, 33 CARDOZO REV 1601 (2011), but blocked many market entrants in Myriad's case.

[19] This is not to say that no incentives exist—first mover advantages, trade secrecy, and whatever patents are available provide some incentives—nor that BBPM is not being developed at all—a few firms are active in the space—but rather that available incentives are smaller than optimal and that BBPM is being developed less and more slowly than would be preferable.

[20] This point mirrors a frequently made inverse argument, that the existence of patents of uncertain validity and scope may block or limit genetic research or clinical testing services. *See* Robert Cook-Deegan, *Gene Patents: The Shadow of Uncertainty*, 331 SCIENCE 873 (2011); Isabelle Huys et al., *Legal Uncertainty in the Area of Genetic Diagnostic Testing*, 27 NAT. BIOTECHNOL. 903 (2009).

especially for black box personalized medicine. Because BBPM already involves complex and frequently implicit relationships, as much transparency as possible is needed for validation and oversight. In addition, cumulative innovation based on shared data and algorithms is crucial to advancing the field but is restrained by pervasive secrecy.

So how can we smooth the path for black box personalized medicine? The first reaction to inadequate innovation incentives is to throw more intellectual property protection at the issue to drive it forward, but I argue that this approach is insufficiently nuanced here. Developing BBPM involves solving multiple interconnected problems: generating and consolidating the necessary data, developing algorithms and models, and validating those models for medical use. Each of those processes require individual consideration through the lens of innovation policy; while algorithms follow familiar innovation patterns, databases—especially large, broad databases aimed to drive future innovation—are more similar to infrastructure than to inventions, and validation of algorithms requires yet another set of incentives.

This Article proceeds in three Parts. Part I describes personalized medicine and the difference between the current version of explicit personalized medicine and black box personalized medicine. It also lays out the hurdles to the development of BBPM. Part II addresses the incentives available for personalized medicine and the diagnostic tests and algorithms on which it relies. Part III discusses potential solutions and policy interventions. A few brief thoughts conclude.

I. A NEW CONCEPTION OF PERSONALIZED MEDICINE

Personalized medicine represents a tremendous step forward for modern medicine. Doctors are already using increasing amounts of personal data, especially genetic diagnostic tests, to tailor treatments to the individual. These variations in treatment reflect the variation inherent among humans, and the link between those variations is carefully examined, tested, and clinically validated. Personalized medicine has the potential to save and extend lives, to avoid unnecessary treatment, and to speed and streamline the process of drug discovery. But far more links are available to be used than the current version of personalized medicine addresses. This section describes the next phase of personalized medicine, which has received significant attention among genomic researchers[21] and health technology

[21] *See, e.g.*, Amarasingham et al., *supra* note 9; Xiaoqian Jiang et al., *Calibrating Predictive Model Estimates to Support Personalized Medicine*, 19 J. AM. MED. INFORM. ASSOC. 263 (2012); Jesse Davis et al., *Machine Learning for Personalized Medicine: Will*

companies,[22] but has gone largely unnoticed by legal scholars. Because this concept is largely unexplored in the legal literature, it will be described in some detail.

Section A describes the concept of personalized medicine in general, discusses the defining characteristics of the current model, and lays out the next step in personalized medicine. Section B discusses the benefits of BBPM, primarily for patient care and for pharmaceutical development. Finally, Section C discusses the hurdles that must be overcome to develop BBPM.

A. Revolution in personalized medicine

Before turning to what's coming next, it's important to know the state of the art. This section describes the current version of personalized medicine—itself still developing and having a major impact on health care—and then addresses the changes coming in the shift to black box personalized medicine.

1. What is personalized medicine?

While doctor-patient relationships have historically focused on the patient, and in that sense have long been personal, new advances in medical science under the name of personalized medicine have been heralded as revolutionary.[23] Although there are many slightly varying definitions of

---

This Drug Give Me a Heart Attack in THE PROCEEDINGS OF INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML) (2008), *available at* http://www.ualberta.ca/~szepesva/ICML2008Health/Davis.pdf; Moon et al., *supra* note 7.

[22] Companies working in this field include Knome, www.knome.com; Foundation Medicine, www.foundationmedicine.com; and 23andMe, www.23andme.com; and Illumina, www.illumina.com.

[23] *See generally* James P. Evans et. al., *Preparing for a Consumer-Driven Genomic Age*, 363 New Eng. J. Med. 1099 (2010) (discussing personalized health care in the direct-to-consumer genetic testing context); Eric D. Green & Mark S. Guyer, *Charting a Course for Genomic Medicine from Base Pairs to Bedside*, 470 Nature 204 (February 2011) (discussing a 2011 vision for moving towards an era of genomic medicine); Margaret A. Hamburg & Francis S. Collins, *The Path to Personalized Medicine*, 363 New Eng. J. Med. 301 (2010) (discussing the hurdles in moving from concept to clinical use); The Case for Personalized Medicine, Personalized Medicine Coalition, *available at* http://cllcanada.ca/2010/pdfs/TheCaseforPersonalizedMedicine_5_5_09.pdf (discussing the benefits of personalized medicine and the necessary steps for widespread implementation).

[23] A. Jamie Cuticchia, *Existing Ethical Principles and Their Application to Personalized Medicine*, 2 Open Ethics J. 29 (2008), *available at* www.bentham.org/open/toj/openaccess2.htm.

personalized medicine, the heart of it is this: all patients are different, and treatment can and should be tailored to the patient to the extent possible. This Article adopts this broad definition of personalized medicine, though other terms exist with contested and more specific meanings.[24]

Personalized medicine contrasts with much of contemporary evidence-based medicine. Evidence-based medicine as a paradigm has led to tremendous advances, identifying which drugs and treatments work, which do not, and which are better than others.[25] However, evidence-based medicine relies principally on clinical trials designed to be broadly applicable across populations, so drugs are similarly approved broadly, not for small sub-populations.[26] This approach develops strong scientific

---

[24] The President's Council of Advisors on Science and Technology defines personalized medicine most closely to the broad version used here: "tailoring of medical treatment to the individual characteristics of each patient." PRIORITIES FOR PERSONALIZED MEDICINE : REPORT OF THE PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY. 1 *available at* http://cdm16064.contentdm.oclc.org/cdm/ref/collection/p266901coll4/id/1735, www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf. (hereinafter, PRIORITIES FOR PERSONALIZED MEDICINE). The FDA defines personalized medicine more narrowly as selecting "the best medical outcomes by choosing treatments that work well with a person's genomic profile or with certain characteristics in the person's blood proteins or cell surface proteins," Michelle Meadows, *Genomics and Personalized Medicine.*, 39 FDA CONSUM. 12 (2005); and the NIH more narrowly still as "an emerging practice of medicine that uses an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease." Genetics Home Reference Glossary, http://ghr.nlm.nih.gov/glossary=personalizedmedicine. For a criticism of equating personalized medicine with genomic medicine, *see* Leigh Ann Simmons et al., *Personalized Medicine Is More than Genomic Medicine: Confusion over Terminology Impedes Progress towards Personalized Healthcare*, 9 PERS. MED. 85 (2011).

[25] David L. Sackett et al., *Evidence Based Medicine: What It Is and What It Isn't*, 312 BMJ 71 (1996).

[26] This characterization is indisputably overbroad. Drugs typically require separate clinical trials to be specifically approved for pediatric use, though they are frequently used in children without those trials. Roberts R et al., *Pediatric Drug Labeling: Improving the Safety and Efficacy of Pediatric Therapies*, 290 JAMA 905 (2003). Acknowledging the limitations of one-size-fits all clinical trials, clinical trial guidelines have shifted to address those concerns. NIH Policy and Guidelines on The Inclusion of Women and Minorities as Subjects in Clinical Research (Oct. 2001), *available at* http://grants1.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm; ("[W]omen and members of minority groups and their subpopulations must be included in all NIH-funded clinical research, unless a clear and compelling rationale and justification establishes . . . that inclusion is inappropriate with respect to the health of the subjects or the purpose of the research"); *see also* NIH Revitalization Act of 1993, PL 103-43, §492B(a)(1). These efforts have met with some success; while in a study of 72 drugs approved in 2011 for 31 separate indications, more than a third of the indications included clinical study populations that were over 90% self-identified white patients, not all trials

evidence of average treatment efficacy, but misses much of the variation among patients.

Personalized medicine, in modern usage, aims to remedy this problem by demonstrating scientific links between biological patient characteristics, diagnoses, and treatment options. It aims to allow physicians and patients to better choose treatment options in light of this. The analysis provides the ability to "classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a particular treatment."[27]

Personalized medicine offers substantial benefits. It can lower the costs and improve the efficiency of the healthcare system,[28] allowing doctors to provide better diagnoses and more effective treatments.[29] In addition, the pharmaceutical and biotechnology industries can focus drug development efforts on subpopulations who have the same critical genetic variants.[30] A treatment paradigm that acknowledges the variants' role in treatment and disease (although the molecular pathways need not be fully understood) should lead to better health outcomes, rather than treating all patients with the same disease in the same way.[31]

2. Explicit personalized medicine

---

are so racially homogeneous. U.S. FDA, COLLECTION, ANALYSIS, AND AVAILABILITY OF DEMOGRAPHIC SUBGROUP DATA FOR FDA-APPROVED MEDICAL PRODUCTS 20 (August 2013), *available at* http://www.fda.gov/downloads/regulatoryinformation/legislation/federalfooddrugandcosmeticactfdcact/significantamendmentstothefdcact/fdasia/ucm365544.pdf. Indeed, drugs have been approved for use in specific subpopulations; the combination drug BiDil, for instance, was approved by the FDA in 2005 to treat congestive heart failure in African-American patients, FDA Approves BiDil Heart Failure Drug for Black Patients, FDA News, June 23, 2005, http://www.fda.gov/bbs/topics/NEWS/2005/NEW01190.html, though that decision has generated its own controversy. *See, e.g.*, Howard Brody & Linda M. Hunt, *BiDil: Assessing a Race-Based Pharmaceutical*, 4 ANN. FAM. MED. 556 (2006); Susan M. Wolf, *Debating the Use of Racial and Ethnic Categories in Research*, 34 J. LAW. MED. ETHICS 483 (2006); Sara R. Jordan, *Race, Medicine, and Social Justice: Pharmacogenetics, Diversity, and the Case of BiDil*, 25 REV. POLICY RES. 53 (2008). Nonetheless, the vast majority of clinical trials have historically been conducted on undifferentiated patient bases, and most drugs are approved for broad use. Mahvash Hussain-Gambles et al., *Why Ethnic Minority Groups Are under-Represented in Clinical Trials: A Review of the Literature*, 12 HEALTH SOC. CARE COMMUNITY 382 (2004).

[27] PRIORITIES FOR PERSONALIZED MEDICINE 1.

[28] *Id.* at 7–8.

[29] *Id.* at 8.

[30] Ginsburg & McCarthy, *supra* note 2.

[31] *Id.*

The form of personalized medicine described briefly above is the standard model. However, to differentiate it from what is on the horizon—and what is described below—I call it "explicit personalized medicine." Explicit personalized medicine relies on relatively simple and explicit links between patients' information and treatment options. Knowledge about these links is carefully generated through scientific and clinical research, resulting in well-characterized and hopefully well-understood links.

Explicit personalized medicine is nothing to sniff at. It uses previously unknown links between individual biomarkers and medical responses to determine treatment plans and sometimes diagnoses.[32] Frequently, these biomarkers are genomic variations, and genetic diagnostic tests are correspondingly the most explored version of EPM. However, other sets of biomarkers—different "'omics"—are also used in EPM, including measurements of RNA transcription levels (transcriptomics), the presence and level of various proteins (proteomics), levels of non-protein small metabolic molecules (metabolomics), and the presence of DNA modifications that affect gene expression levels (epigenomics).[33] Each type of biomarker can help direct treatment of patients or improve the drug development process.

Explicit personalized medicine is already used to calibrate treatment options. One prominent example is the anticoagulant drug warfarin, which can lead to heavy bleeding if used at an improper dosage. Some patients metabolize the drug faster, and some slower; giving too much to a slow-metabolizer results in an overdose.[34] Earlier dosing regimens relied on trial and error combined with some easily measurable patient characteristics such as age, weight, and gender. Recently, however, researchers discovered that two proteins are particularly relevant in warfarin metabolism: the Cytochrome P450 enzyme CYP2C9 which metabolizes warfarin and the Vitamin K epoxide reductase gene VKORC1.[35] Those proteins come in different versions, which work less or more efficiently, and genetic tests can determine which version a particular patient has. Now, warfarin dosing can be determined after genetic testing, with far more accurate results than the prior regime.[36] Similar tests can improve drug response and reduce side

[32] Chan & Ginsburg, *supra* note 2

[33] *Id.* at 222–24

[34] Schwarz, U.I., "Clinical relevance of genetic polymorphisms in the human CYP2C9 gene," Eur. J. Clin. Invest. 33. Suppl 2: 23–30 (2013).

[35] Chan & Ginsburg, *supra* note 2, at 227

[36] *Id.* The information used to provide warfarin dosing information is collected at www.warfarindosing.org. According to the website, the calculator uses "clinical factors and (when available) genotypes of two genes: cytochrome P450 2C9 (CYP2C9) and vitamin K epoxide reductase (VKORC1)." Recommendations are based on data from a

effects in schizophrenia patients.[37]

Diagnostic testing can identify not only patient characteristics but also the nature of the disease itself. In oncology, genetic tests can determine the specific variant of a cancer and consequently, how best to attack it.[38] The monoclonal antibody Herceptin (trastuzumab), exemplifies this approach: patients who overexpress the HER2/neu receptor in breast cancer patients can usefully be treated with Herceptin,[39] while in other patients, the side effects of the drug outweigh any benefits.[40]

EPM also promises benefits to conducting clinical trials leading to drug approval. If only certain genetically identified participants in a Phase I or Phase II clinical trial respond to an investigational drug, Phase III trials can focus on individuals with that genotype.[41] This approach can potentially lower the expense of the trial and generate a more focused indication and label much earlier in the process.[42]

Throughout these uses—dosing, diagnosis, and drug development—runs the common thread of EPM: it uses biological relationships that are typically simple, explicit, and carefully validated in laboratory clinical trial settings.[43] These characteristics are also reflected in evidence-based

---

cohort of over 1,000 patients, and the information used for the calculator can explain 53% of the variation in response to warfarin doses. *Id.*

[37] Cichon, S., Nöthen, M.M., Rietschel, M., Propping, P., "Pharmacogenetics of schizophrenia," Am. J. Med. Genet. 97 (1): 98–106 (2000) (describing relevant variants of genes encoding Cytochrome P450 enzymes CYP2D6, CYP2C19, and CYP2C9).

[38] Mansour, J.C., Schwarz, R.E., "Molecular mechanisms for individualized cancer care," J. Am. Coll. Surg. 207 (2): 250–8 (2008).

[39] Carney, W., "HER2/neu Status is an Important Biomarker in Guiding Personalized HER2/neu Therapy," Connection 9: 25–27 (2006).

[40] Telli, M.L., Hunt, S. A., Carlson, R. W., Guardino, A. E., "Trastuzumab-Related Cardiotoxicity: Calling Into Question the Concept of Reversibility," Journal of Clinical Oncology 25 (23): 3525–3533 (2007).

[41] Ginsburg & McCarthy, *supra* note 2

[42] *Id.*. The FDA has issued guidance on the use of pharmacogenomic data in the context of clinical trial development, and trials increasingly include such data. U.S. FDA, *Guidance for Industry: Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling* (2013), *available at* http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM337169.pdf.

[43] For warfarin dosing, for example, several clinical trials were used to evaluate the efficacy of using genotype to guide warfarin dosing over standard protocols. *See*, *e.g.*, Anderson JL, Horne BD, Stevens SM et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. Circulation 2007;116:2563-70; Caraco Y, Blotnick S, Muszkat M. CYP2C9 genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. Clin Pharmacol Ther 2008:460-70; and Lenzini PA, Grice GR, Milligan PE, et al. Laboratory and clinical outcomes of pharmacogenetic vs. clinical protocols for

medicine in general—clinically validated, explicit links leading to better-supported treatment choices.

These traits are frequently virtues but also have negative implications. In particular, requiring that links be explicit and clinically validated limits the complexity and identity of the links can be used. More complex links and links based on hard-to-observe characteristics are unavailable to EPM, even if they have significant biological implications. And the more we learn, the more we understand that biology and pathology rely on incredibly complex networks and pathways.

One potential response is to wait. Eventually, even very complex relationships may be fully explained and validated.[44] But that day is very far in the future. Much nearer is the leverage of combining large datasets and sophisticated algorithms to make predictions and improve treatment, without explaining or even identifying the underlying complex relationships: black box personalized medicine.

3. Black box personalized medicine

Black box personalized medicine is the next stage of personalized medicine.[45] It differs from EPM in three principal ways. First, the information used to develop the relationships and predictions used in treatment recommendations comes from a much larger, broader set of information. Second, while there are still some simple, explicit links, a large and rich dataset and machine learning techniques enables many predictions based on complex and often implicit connections between patient information and expected treatment results. For example, as opposed to the relatively simple links described above, a BBPM prediction might be that patients who have a set of linked variations in a dozen different genes, smoke, and have middling-high blood pressure might predictably respond better to one medication than another—even if those factors could not be explained or even explicitly identified.[46] Third, the

warfarin initiation in orthopedic patients. J Thromb Haemost. 2008;6:1655-62..

[44] One challenge even in a waiting approach arises from the current nature of clinical trials; with increasingly complex hypotheses, the number of patients needed grows rapidly, and suitable patients simultaneously grow rarer as the definition of "suitable" becomes more constrained.

[45] *See, e.g.*, Amarasingham et al., *supra* note 9; Jiang et al., *supra* note 22; I. Glenn Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, 33 HEALTH AFF. (MILLWOOD) 1139 (2014).

[46] In some sense, BBPM seems to be a throwback to traditional reliance on the experience and intuition of doctors: "I've tried this on patients like you before and it's worked, so that's what I'll recommend for you." Inasmuch as both this model and BBPM

complex and implicit nature of these relationships means that they will be developed and validated without—or at least far faster than—the development of explicit, well-characterized models through scientific and clinical research.

a. Big data

Far more health data are collected today than ever before, and that collection increases rapidly. Data are collected for several reasons, including improving patient care, documenting care to ward off malpractice threats, increasing the efficiency of care, and keeping records to support insurance and payment claims.

The tremendous growth of recorded data has been facilitated by the gradual transition to electronic health records (EHRs).[47] EHRs not only have capacity to record more data, they are more readily accessible and can be combined into larger databases more easily than scattered paper patient records.

The types and volume of data collected and included in EHRs are also ballooning. Genetic testing for single-nucleotide polymorphisms, which measures some genetic variation, is now inexpensive and frequent, and

---

rely on implicit links, the analogy is apt. However, BBPM relies on far broader sets of information in making connections, and will involve quantitative validation of those models in a fashion atypical of physician experience/intuition-based treatment. For these reasons, BBPM is not a true step away from evidence-based medicine, but rather takes a somewhat orthogonal tack.

[47] The terms EMR (electronic medical record) and EPR (electronic patient record) are also used, frequently interchangeably. The differences between them, such as they are, are largely not important for this Article. The growth in EHRs is attributable to several factors, a system of penalties and incentives as part of the Health Information for Economic and Clinical Health (HITECH) Act, enacted as Title XIII of the American Recovery and Reinvestment Act of 2009, Pub. L. 111–5, and other potential cost savings; Dwight C. Evans et al., *Effect of the Implementation of an Enterprise-Wide Electronic Health Record on Productivity in the Veterans Health Administration*, 1 HEALTH ECON. POLICY LAW 163 (2006); Richard Hillestad et al., *Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs*, 24 HEALTH AFF. (MILLWOOD) 1103 (2005). Improved patient care—another motivation for adopting EHRs, Jeffrey A Linder et al., *Electronic Health Record Use and the Quality of Ambulatory Care in the United States*, 167 ARCH. INTERN. MED. 1400 (2007)—has received mixed reviews, with some finding no substantial improvement, Max J Romano & Randall S Stafford, *Electronic Health Records and Clinical Decision Support Systems: Impact on National Ambulatory Care Quality*, 171 ARCH. INTERN. MED. 897 (2011); Ashly D Black et al., *The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview*, 8 PLOS MED. e1000387 (2011); and others observing some improvement, Randall D. Cebul et al., *Electronic Health Records and Quality of Diabetes Care*, 365 N. ENGL. J. MED. 825, 828–30 (2011).

whole-genome sequencing continues to drop in price and to approach widespread clinical use.[48] Other "omics" technologies like the testing of large panels of metabolites, gene expression levels, and protein levels, are similarly becoming more accessible.[49] Each new broad analytical tool creates large amounts of data which can be captured in EHRs and linked to patient health outcomes.

All of these data can be used to understand and improve the practice of medicine (after overcoming substantial hurdles, discussed below[50]). And indeed, providers and health care firms are already using the data to improve efficiency and patient outcomes.[51] But beyond the relatively simple links available to regression modeling and other forms of explicit analysis,[52] many complex relationships are impossible to observe or use without a different set of algorithmic tools.

b. Black-box algorithms

To discover new complex relationships, BBPM relies on computer systems which improve their performance over time by trying a certain solution, evaluating the outcome, and then modifying that solution accordingly to improve the outcome.[53] For a familiar example to illustrate the novel features of this approach, consider the familiar examples of the music service Pandora and the video service Netflix, both of which make

---

[48] Stories of imminent whole-genome sequencing for under $1,000 have existed for years without fruition, *see* John A. Robertson, *The $1000 Genome: Ethical and Legal Issues in Whole Genome Sequencing of Individuals*, 3 AM. J. BIOETH. 35 (2003); Simon T Bennett et al., *Toward the $1000 Human Genome*, 6 PHARMACOGENOMICS 373 (2005); Erika Check Hayden, *Is the $1,000 Genome for Real?*, NATURE (2014), http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530, but the costs have been dropping at a rapid rate, Kris Wetterstrand, *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), available at* www.genome.gov/sequencingcosts. However, quality concerns exist; as of 2014, using whole-genome sequencing as a common clinical tool must cope with high false negative rates. Frederick E. Dewey et al., *Clinical Interpretation and Implications of Whole-Genome Sequencing*, 311 JAMA 1035 (2014).

[49] Chan & Ginsburg, *supra* note 2, at 222–24.

[50] *See* Section I.C, *infra*.

[51] *See* n. 46, *supra*.

[52] This is not to denigrate explicit modeling, or to understate the tremendous effort needed to develop those models, the knowledge benefit that comes from developing them, or their potential benefits for patients and the system. I intend rather to point to a different form of analysis, which opens many additional possibilities.

[53] Davis et al., *supra* note 22. This approach is frequently referred to as "machine learning." For a general overview of the field, *see* PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA (2012).

recommendations to their users.

Pandora relies on a technique called content-based filtering that is simpler and resembles the current practice of personalized medicine. This technique uses discrete characteristics about an object and knowledge about the user's relationship to those characteristics to make recommendations. In the music service Pandora, for instance, experts characterize songs based on a set of explicit criteria, like major vs. minor key or the presence of vocals. When a customer selects a song, Pandora identifies the traits of that song and suggests other songs which share those traits. In the medical context, content-based filtering maps closely onto the explicit science-based paradigm of modern medicine. If a patient presents with fever and cough and tests positive for strep throat, a doctor would likely prescribe an antibiotic such as amoxicillin to treat a likely strep infection. Content-based filtering requires a relatively small set of information—in this case, a positive test for strep might be enough—but can only make recommendations based on already known explicit links to that information

Netflix, on the other hand, uses a technique called collaborative filtering, which is more complex and more closely resembles BBPM. Collaborative filtering uses information groups of similar users to construct an underling predictive model and makes recommendations based on that model. Netflix uses this approach, predicting which movies a user might like based on the customer's ratings of watched movies and comparing that set of data to similar data from other customers. This allows predictions without any explicit knowledge; for instance, it might be true that the vast majority of people who liked Notting Hill, Casino Royale, and the television show Dr. Who turn out to like the cult foodie film Tampopo. Someone who likes the first three would be offered Tampopo as a recommendation, despite the lack of any clear or identified link. In the medical context, data might reveal, for example, that male patients diagnosed with schizophrenia who have several specific genetic markers and are between the ages of 22 and 27 might respond significantly better to cognitive behavioral therapy when combined with low doses of caffeine. Why? The model couldn't tell us—though it could suggest that research into the mechanism might eventually be of interest—but it could suggest treatment contours in a way previously unavailable.[54]

[54] It is worth noting that informal versions of comparison-based recommendations are currently in use, though they are not typically well regarded under the modern medical paradigm. Sites like patientslikeme.com, where patients describe symptoms and successful treatments, frequently without any specific scientific basis for the treatment choice or its success, essentially show collaborative filtering in action, though without quantitative or algorithmic analysis, and generally with much less data. Also note that purely

To summarize, correlations can be roughly classified in the two-by-two grid below: simple or complex and implicit or explicit.[55] BBPM can help us find correlations in all four boxes, but its novelty increases with darker shading; the realm of complex, implicit linkages is almost completely the purview of BBPM. To the extent that important biological relationships dwell in the realm of the complex and un-understood—a situation we should expect—BBPM can be expected to bear significant fruit. However, even without relying on full-blown BBPM, when EPM looks more like BBPM, involving big data, more sophisticated algorithms, and informational discoveries difficult to lock to particular physical products, the challenges described below also apply to EPM.

Table 1: Types of biomedical correlations

	Simple	Complex
Explicit	Alleles of the HER2 gene predict whether Herceptin will effectively treat breast cancer.[56]	Alleles of the CYP2C9 and VKORC1 genes, combined with patient weight, age, height, and gender predict patient response to the blood thinner Warfarin.[57]
Implicit	Use of BiDil to treat congestive heart failure in self-identified black patients.[58]	A set of patients with specific values for a dozen genes and a dozen physical factors, whose relationships are unknown, respond strongly to monoclonal

retrospective data analyses come with a significant set of issues, including the possibilities of overspecification, latent variables, endogeneity problems, and other complexities. Anup Malani and colleagues, among others, have described these problems in the context of FDA approval for drugs based on post-hoc subgroup analysis. ANUP MALANI ET AL., REFORMING SUBGROUP ANALYSIS (Social Science Research Network, SSRN Scholarly Paper ID 1119970, April 13, 2008), *available at* http://papers.ssrn.com/abstract=1119970. Reliable BBPM would need to compensate for these issues, as does any primarily data-mining approach; Malani and colleagues suggest, as does this article, that independent third-party validation may help counteract some problems of post-hoc analysis. *Id* at 13–16.

[55] Of course, each axis is actually a continuum.

[56] *See* notes 39–40, *supra*, and accompanying text.

[57] *See* note 36, *supra*, and accompanying text.

[58] *See* Brody & Hunt, *supra* note 27.

		antibody therapies for lung cancer.[59]

B. The benefits of black box personalized medicine

BBPM has the potential to bring tremendous benefits to the practice of medicine and to the health care system more generally. As with EPM, these benefits fall into two linked main categories: improving patient care and increasing possibilities of drug discovery or drug repurposing.

1. Patient care

Black box personalized medicine is key to realizing the next-generation health benefits of genomics, electronic health records, and "big data" in the health care sector. Currently, we are amassing patient data, but are only able to use a relatively small fraction of that data because of the sheer complexity of biological systems. BBPM promises to make at least some of that complexity available for medical purposes, as described above.[60] And if new connections are available for medical use, that promises to allow different treatment options, and to fine-tune provider responses and treatments which are already in use. BBPM can also potentially not only new treatment recommendations, but also prophylactic recommendations based on individual patient data. More specific predictions are hard to make, since BBPM by definition focuses on complex and implicit links. A look at aspects of current practice which are closes to BBPM offers some suggestions, however. In particular, the major benefits that have arisen from increased understand of warfarin dosing, using individualized predictors to avoid a trial-and-error approach, are suggestive of how BBPM could improve drug treatments. BBPM might be especially helpful for other drugs with narrow therapeutic indices like warfarin.[61] As such, BBPM can significantly improve patient care. In addition, BBPM promises to improve care indirectly through its effects on drug development and use.

2. Drug discovery and development

[59] Moon et al., *supra* note 7.

[60] *See* Section I.A.3, *supra*.

[61] A drug's therapeutic index measures the range within which the drug is effective; higher doses are likely to have toxic effects, and lower doses to be ineffective. In instances where the therapeutic index is relatively narrow and is also impacted by patient characteristics—like warfarin—BBPM is particularly likely to offer useful guidance.

Black box personalized medicine can help resolve a major challenge facing the drug industry, the question of new uses for already-approved drugs. It is extremely costly to develop new drugs. A significant portion of that cost goes to ensuring basic safety and administrability, and a larger fraction goes to demonstrating efficacy through clinical trials. Therefore, there is a substantial advantage to developing new uses for old drugs;[62] most drugs have multiple uses.[63] Finding a new use avoids repeated costs in demonstrating safety, but still requires costly clinical trials to demonstrate efficacy. However, recouping those costs is hard because patents and regulatory exclusivity tend not to protect new uses effectively.[64]

BBPM offers less expensive routes to discover and validate new uses.[65] The wealth of data available in electronic health records of patients suffering from different ailments and responding to drugs they take for other purposes may be usefully mined by big-data algorithms and can suggest new uses. BBPM would broaden the already-widespread concept of off-label use beyond those uses based today on practitioner experience or limited clinical trials.[66] In the BBPM paradigm, useful links need only be

---

[62] At least in part because many drugs are relatively crudely targeted, they typically have multiple effects on the human body. The simplest example is the existence of side effects. Better-validated multiple effects enter clinical practice as off-label use; for instance, many of the drugs used in chemotherapy have not been regulator-approved for that use, and some such uses have never even been the subject of clinical trials. *See* Rena M. Conti et al., *Prevalence of Off-Label Use and Spending in 2010 Among Patent-Protected Chemotherapies in a Population-Based Cohort of Medical Oncologists*, J. CLIN. ONCOL. JCO.2012.42.7252 (2013) (finding 30% rate of off-label use among 10 common patent-protected intravenous chemotherapies).

[63] *See* BENJAMIN N. ROIN, SOLVING THE PROBLEM OF NEW USES 19–20 (Social Science Research Network, SSRN Scholarly Paper ID 2337821, October 1, 2013), *available at* http://papers.ssrn.com/abstract=2337821.

[64] Rebecca S. Eisenberg, *Problem of New Uses, The*, 5 YALE J HEALTH POL ETHICS 717 (2005) at 724–25 (finding patents ineffective at incentivizing new use clinical trials) and 728–35 (finding regulatory efforts similarly ineffective).

[65] A complex subsidiary issue, on which this Article does not take a stand, is whether such new uses for old drugs should remain off-label or be added to the drug label. Personalized medicine and the increased granularity of post-market surveillance raise such questions generally, about how much data is necessary for label modifications and how practice should best be guided by information which is below that threshold.

[66] The best-validated multiple effects of drugs come when firms decide the limited protection available for new uses is worth the cost and effort of undertaking full clinical trials and acquiring regulatory approval for those uses. Whatever profits available on marketing older drugs for new uses may provide an end-stage incentive which drives the creation of earlier innovations needed to get there; such incentives could potentially drive the development of BBPM. However, those incentives are attenuated by the difficulty in enforcing patents on new uses. *See* Eisenberg, *Problem of New Uses, The*, *supra* note 65, at 725–28.

correct, not explicit or extensively validated in clinical trials; this could facilitate the off-label use of drugs which are approved as safe but not approved—and which may never be approved—for the algorithmically-suggested purpose.

BBPM has other potential benefits for drug discovery and development. To the extent that BBPM identifies correlations between complex sets of variables which are largely implicit, BBPM suggests potential new research pathways to make those implicit connections explicit. Finally, BBPM could also aid discovery of new drugs, or clinical validation of secondary uses for old drugs, by targeting clinical trials. EPM and pharmacogenomic testing generally can already be used to streamline clinical trials, though this practice is not without its critics. BBPM could further expand these possibilities by suggesting participant populations that meet a more complex set of criteria for as-yet-unknown reasons. In addition, BBPM could more radically reduce the cost of clinical trials by avoiding more of them, especially in the context of off-label uses for already-approved drugs as described above.

II. HURDLES TO DEVELOPMENT

While BBPM offers large benefits, getting there will not be easy. Developing complex predictive models requires the ongoing generation and consolidation of very large datasets about individuals and their health. This requires significant costs in the collection of data—both from modern sources, such as electronic health records, and, more expensively, from paper-filed records located in widely dispersed doctors' offices. Genetic sequence data collection will be required to complement and inform collected health records. Once the datasets are gathered, the development of accurate predictive and analytical algorithms is expensive. Finally, validating those algorithms—whether through independent testing, repeats on separate datasets, or clinical validation—will require additional funds and effort.

A. Data gathering

BBPM will require large sets of high-quality health data to find the complex correlations at its heart. Notably, some firms have already amassed significant health information databases, but these are traditionally aimed either directly at immediate care or at insurance reimbursement, and

lack many types of data necessary for BBPM development.[67] Generating large datasets suitable for BBPM presents a significant challenge for two principal reasons. First, the gathering, cleaning, and assembling of data from many different sources is an expensive endeavor. Second, legal hurdles exist for assembling data, most notably in the form of HIPAA's privacy protections.

1. Expense

Assembling large datasets of high quality health information is expensive. While the costs of actually possessing and storing even tremendous amounts of data are relatively low, the data must be gathered in the first place, checked for quality and "cleaned," and then put into compatible formats for a unified database. As a first point, the health data necessary to fill databases for BBPM will come from two different sources: electronic health records and paper health records. Each raises different challenges.

Collecting electronic information should theoretically be much less expensive, because it does not require encoding new data. This does not mean that collecting electronic information is cheap or free; many health records are kept in mutually incompatible data formats, and the information included is highly heterogeneous.

The practical challenges with assembling paper health records are very high—likely higher—and of a different nature. Paper records are scattered throughout healthcare facilities, from doctors' offices to hospitals.[68] Once located—and once relevant permissions are obtained, as discussed below—information must be encoded to electronic format, either by hand or by optical character recognition.[69] One response is to begin with electronic health information and proceed later to paper records, though this may lead to the absence of older, longer-term data and possible population selection effects.

Practical challenges with data gathering are compounded by the

[67] *See, e.g.*, Madelyn Kearns, *Returning Patients to Data Aggregation*, MED. PRACT. INSID. (2013) ("The healthcare industry is still very much in the pubescent stages of data management and storage — experimenting with its new data capture proficiencies and what the general breadth of digital medical information means for care delivery"). One of the largest and best-known health data aggregators is IMS Health, www.imshealth.com. The data from IMS Health is deidentified, strictly proprietary, and extremely expensive.

[68] Roy Schoenberg & Charles Safran, *Internet Based Repository of Medical Records That Retains Patient Confidentiality*, 321 BMJ 1199, 1199 (2000).

[69] The notoriously poor handwriting of doctors on medical records may make this task even more expensive.

complex and ideally comprehensive nature of the data being gathered. More data leads to greater capacity to tease apart complex implicit relationships. Thus, an ideal database might include, for instance, not only typical physical measurements (blood pressure, heart rate, height, weight, symptoms, &c.), but also medications being taken (both over-the-counter and prescription, including frequency and duration) and genetic information. As metabolite screens, RNA expression profiles, and other biomarker sets become more readily available, those will become increasingly useful data. For the last forms of data—those forms which rely on new technologies to measure—collecting data obviously requires that those technologies be adopted by practitioners.[70]

2. HIPAA and other legal hurdles

Gathering data requires not only surmounting practical but also legal hurdles, the most obvious of which is the Health Insurance Portability and Accountability Act (HIPAA).[71] In relevant part, HIPAA limits the disclosure of patients' Protected Health Information (PHI) by "covered entities" and their "business associates." Covered entities include providers, health insurance plans, and health care clearinghouses.[72] Business associates include anyone who assists or performs any HIPAA-regulated activity on behalf of a covered entity, or provides services to the entity that involve individually identifiable health information.[73] The

[70] This presents something of a chicken-and-egg problem: practitioners are unlikely to implement broad biomarker screens without significant uses, while BBPM is unlikely to develop those uses without those data linked to health information. This problem may be amenable to a bootstrapping solution, where limited use in some circumstances, possibly compensated by the innovator firm, are used to develop at least some applications which then lead to further use and better applications.

There exists at least a potential but practically very unlikely exception for genetic testing or sequencing, since all states require blood testing of newborns, and store the blood spots taken; those blood spots could be used as samples for genetic sequencing. However, an immense change in opinion would be needed to permit either the public or private sequencing of newborn blood spots taken under state mandate. For an in-depth treatment of this issue, *see* Sandra J. Carnahan, *Biobanking Newborn Bloodspots for Genetic Research without Consent*, 14 J HEALTH CARE POL 299 (2011).

[71] Health Insurance Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

[72] 45 C.F.R. § 160.102. Only providers who transmit health information in electronic form in connection with certain transactions are "covered entities." *Id.* Health care clearinghouses process information between different formats. 45 CFR § § 160.103, 164.500(b).

[73] 45 C.F.R. § 160.103.

protected information includes medical records and billing records.[74] Generally, PHI may only be disclosed by a covered entity with the patient's permission or for certain narrowly defined permitted purposes.[75] Finding ways to work around HIPAA barriers,[76] as well as negotiating informed consent and other issues, will require impose additional costs on data aggregation.[77]

B. Algorithm generation and validation

The second set of challenges lies in the actual generation of predictive algorithms for BBPM; that is, the task of parsing the data, identifying correlations, and making sure those correlations suggest real and useful health measures. While predictive algorithms have become increasingly sophisticated, they still require extensive development and specialization to adapt them to particular contexts and specific concerns.

To take a recent example of the complexity in developing predictive algorithms, consider the Netflix Prize.[78] Netflix's movie recommendation algorithm is a core part of Netflix's business. Netflix developed its own predictive algorithm,[79] sought to harness outside expertise, and in 2006 offered a $1,000,000 prize for a team which could beat the performance of

---

[74] 45 C.F.R. § 164.501.

[75] HIPAA Privacy Rule, 45 CFR Part 160, Part 164 Subparts A & E.

[76] HIPAA places substantial legal limits on data aggregation, but in practice seems more of a source of additional costs than an actual barrier. First, "covered entities" fails to include many relevant entities with health information, including providers who do not transmit electronic information (admittedly likely a small set) and self-administered employer health plans with fewer than 50 participants. Thus, substantial amounts of data are not covered by HIPAA. In addition, PHI can be disclosed with patient consent, which may be obtained by payment or simple request. An initial reaction might be that patients are particularly jealous of their medical privacy; the readiness with which personal information is shared on social media suggests that at least going forward, this may be a less universally applicable assumption. Finally, of practical concern if not a typical policy consideration, HIPAA is only lightly enforced, since it typically requires administrative action for enforcement.

[77] Clearly, these issues require much more exploration, which is outside the scope of this work. For an initial overview, *see* Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, *supra* note 46

[78] *See* http://www.netflixprize.com.

[79] Despite Netflix's high incentives, its initial algorithm, Cinematch, used "straightforward statistical linear models with a lot of data conditioning." http://www.netflixprize.com/faq. This model provides only a 10% better prediction of user scores than a trivial algorithm which predicted as the user score the average score the movie had received from other users. *Id.*

its in-house algorithm by 10%.[80] Over 20,000 teams registered.[81] Three years and tens of thousands of submissions later, Netflix awarded the prize to a team consisting of several researchers from different institutions—after three years of work, they had improved Netflix's performance by 10.06%.[82] The winning algorithm involved dozens of separate collaborative filtering algorithms[83] and the training of 100 parallel predictors with results blended through eleven different computational methods.[84] Despite the technology available, and the simplicity of the dataset, the team still ran into substantial computational limitations.[85]

The experience of Netflix demonstrates that while the development of algorithms might seem an easy step, it remains in fact extraordinarily complex and requires close and careful involvement from sophisticated programmers. And the Netflix dataset was almost laughably simple compared to that necessary for personalized medicine; while it involved data from a substantial 480,189 users on 17,770 movies (for a total of 100,480,507 ratings), the ratings were simple 1–5 integer scores.[86] Health information databases would involve between dozens and thousands of potentially relevant variables, as described above.[87] The stakes and error costs are higher, but the potential rewards are much higher as well: while the entertainment industry makes billions of dollars, U.S. health care expenditures are over 17% of GDP, for a 2012 total of approximately $2.8 trillion.[88] The challenge—as discussed throughout this Article—is how to develop incentives appropriate to the potential rewards.

---

[80] Linyuan Lü et al., *Recommender Systems*, 519 PHYS. REP. 1, 3–5 (2012)

[81] *Id.*

[82] http://www.netflixprize.com/community/viewtopic.php?id=1537. Notably, the first- and second-place teams submitted their results within 24 minutes of each other. *Id.*

[83] *See* Andreas Töscher et al., *The Bigchaos Solution to the Netflix Grand Prize*, NETFLIX PRIZE DOC., 6–16 (2009), Yehuda Koren, *The Bellkor Solution to the Netflix Grand Prize*, NETFLIX PRIZE DOC., 9 (2009).

[84] Töscher et al., *supra* note 84, at 16–20

[85] *Id.* at 3, 9, 15, 17.

[86] In fact, even with a far simpler dataset, Netflix never ended up implementing the winning solution, finding that "the additional accuracy gains . . . did not seem to justify the engineering effort needed to bring them into a production environment." Xavier Amatriain, *The Netflix Tech Blog: Netflix Recommendations: Beyond the 5 Stars (Part 1)*. Netflix did end up implementing some simpler solutions developed earlier in the competition. *Id.*

[87] *See* note 70, *supra*, and accompanying text.

[88] U.S. Centers for Medicare and Medicaid Services, National Health Expenditure Highlights 2012, http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf, 1.

C. Validation

The third challenge in developing BBPM is the question of validation; that is, making sure that the algorithmic models developed by firms are actually accurate and useful. For typical new treatment methods—whether drugs or otherwise, validation comes in several forms. First, the treatment is generally scientifically understood.[89] Second, clinical trials are used to demonstrate the validity of a treatment method. Third and finally, the validity of the treatment is confirmed by actors other than the sponsoring company, whether by other clinical trials (conducted by health agencies, for instance), or by post-marketing surveillance mechanisms and the experience of clinicians using the drug.

The complex and implicit models at the heart of BBPM face challenges at each of these stages. First, the nature of the algorithms means that they are highly unlikely to be well understood on a scientific level. Typical clinical trials are also likely to be challenging for two reasons: the complex relationships of BBPM are unlikely to be susceptible to classic gold-standard clinical trial methodology,[90] and some of the principal benefits of BBPM—high speed and relatively low cost of specialized treatment recommendations—relies on avoiding a large and lengthy clinical trial process.[91]

This reality increases the need for external validation of BBPM algorithms at a computational level, based on the same data. There are two principal concerns, roughly equivalent to the well-trodden concepts of analytical and clinical validity in diagnostic testing. First a model may not

---

[89] There are notable exceptions. For instance, lithium, used to treat mood disorders, has an unknown mechanism of action. Gin S. Malhi et al., *Potential Mechanisms of Action of Lithium in Bipolar Disorder. Current Understanding*, 27 CNS DRUGS 135 (2013). Similarly, although aspirin has been commonly available since the beginning of the twentieth century, its mechanism of action was only discovered in 1971. John Robert Vane, *Inhibition of Prostaglandin Synthesis as a Mechanism of Action for Aspirin-like Drugs*, 231 NATURE 232 (1971). However, the norm is that for a new drug—and especially for a new treatment regime—at least a basic understanding of the relevant science is needed for adoption.

[90] Because personalized medicine relies on very specific patient profiles, it is hard to aggregate similar patients. The expectation of different results among different patients runs counter to the average treatment effects observed in randomized clinical trials.

[91] Firms could, and likely should, run broad clinical trials on BBPM algorithms overall; that is, does a group of patients treated according to algorithm X have significantly better clinical outcomes, in general, than patients treated according to the standard of care? But this broad form of clinical trial shows some overall validation for the full complex algorithm set, not for any particular treatment option.

predict what it says it does.[92] Second, and more specific to BBPM, a model may predict what it aims to, but for reasons based on idiosyncrasies of the dataset or overspecification rather than true biological phenomena. With enough data[93] and enough searching, patterns will emerge that look real but do not reflect true underlying relationships.[94]

Model validation could be performed by the initial innovator or the FDA, but each has problems. The initial innovator faces strong financial incentives not to disprove its own algorithm once marketed, and retains whatever biases or errors may have created problems in the first place.[95] Regulatory oversight could serve some validation role, but the FDA currently lacks the expertise to independently replicate a company's algorithmic results; at most, it could provide procedural oversight—that is, that the data collection, consolidation, and analysis methods are appropriate.

Validation by private third parties would better confirm performance.[96] Agreement between different firms, using different methodologies, on recommended treatment options would go a long way to demonstrate that

---

[92] Into this category fall the most basic form of errors: errors in the coding of the program ("bugs"), corrupt or flawed data, and other such challenges. These problems' mundane nature does not diminish their importance. Like in other places in the health care system, simple and mundane errors can have tremendous and costly consequences.

[93] Spurious patterns can appear in both small and large samples; the latter is the concern here.

[94] Consider the identification of constellations; given enough stars, some creativity, and a lot of squinting, one can find lions and crabs amid the randomness. Validation serves to ensure that lions are ignored, but that the gravitational patterns showing the presence of an invisible black hole are still observed and understood.

[95] While the initial innovator has reputational incentives to ensure a high-quality product, as well as duties under tort law, prior experiences with drug company behavior in the past shows that reputational and tort incentives cannot assure uniform validation and disclosure of problems. In the most high-profile example, Merck failed for years to disclose information about risks of its blockbuster drug Vioxx, resulting in nearly 30,000 tort claims amid an estimated 88,000 to 120,000 excess cases of serious heart disease in the U.S. Harlan M. Krumholz et al., *What Have We Learnt from Vioxx?*, 334 BMJ 120 (2007); David J Graham et al., *Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with Cyclo-Oxygenase 2 Selective and Non-Selective Non-Steroidal Anti-Inflammatory Drugs: Nested Case-Control Study*, 365 THE LANCET 475, 480 (2005). Thus, while validation by the initial innovator is important, and should be demonstrated—especially in the regulatory context—it is unlikely to suffice.

[96] *See* Malani et al., *supra* note 55; Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, *supra* note 46. To some extent, noncommercial third parties—primarily foundations and academic researchers—can validate BBPM models in the same way that they currently check some drug trials. But that current role is quite limited; the resources necessary to conduct independent clinical trials are significant, and even performing reanalysis of clinical trial data requires time, money, and expertise.

those implicit correlations are medically valid and not merely artifacts of the dataset or the specific choices of the algorithm developers.[97] Because such validation will also be expensive, policy regarding personalized medicine should also consider incentives for third-party validation.[98] Finally, postmarket surveillance and provider experience can also help bolster the case for an algorithm's overall efficacy, but these rely on the wide deployment of the algorithm in the first place.

In general, the implicit nature of BBPM's predictions increases the need for other types of validation than the standard mechanisms of scientific understanding and clinical trials. Ensuring robust validation systems will be key to ensuring that BBPM develops fully.

Overall, BBPM promises to be a significant step beyond—or at least in addition to—current versions of personalized medicine. The benefits of BBPM require substantial effort in collecting, analyzing, and validating the large datasets of high-quality health information. The next section analyzes the way the legal system structures those incentives, and points out major flaws in the current regime.

III. FAILURES OF THE CURRENT INTELLECTUAL PROPERTY REGIME

The basic justification for intellectual property is a well-told story.[99] Society derives tremendous benefits from innovation, but absent intellectual property, ideas are frequently expensive to produce but hard to protect. In addition to the initial discovery, the process of taking an idea through the development into a commercial product can be costly, and similarly is frequently subject to free-riding. Since firms cannot capture much of the value of their investments in innovation, they invest at a socially suboptimal level. Intellectual property allows firms to capture some of that surplus, increasing the incentives for invention by allowing firms to exclude others from the invention. BBPM follows this pattern closely: databases, algorithms, and the knowledge that algorithms are reliable are all information goods which are difficult to keep exclusive once known. Accordingly, intellectual property—or a substitute incentive set—is likely necessary for its socially optimal development.

---

[97] Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, *supra* note 46

[98] *See* Section III.E, *infra*.

[99] *See, e.g.*, Mark A. Lemley, *Ex Ante versus Ex Post Justifications for Intellectual Property*, UNIV. CHIC. LAW REV. 129, 129 (2004); Kevin Outterson, *Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets*, 5 YALE J. HEALTH POLICY LAW ETHICS 193, 195–98 (2005); Benjamin N. Roin, *Unpatentable Drugs and the Standards of Patentability*, 87 TEX REV 503, 507–09 (2008).

The patent system fills this role by guaranteeing significant protection in exchange for disclosure of the technology, thus increasing the type of protection available and enabling more cumulative innovation.[100] Patents provide an alternative to either not developing an appropriable innovation or keeping it secret. Secrecy prevents appropriation and is bolstered by the mostly state-law doctrine of trade secrecy,[101] but works poorly for innovations which can be reverse-engineered or which are unavoidably public.[102] In addition, trade secrecy tends to prevent cumulative innovation, where different innovators build off the inventions and innovations of a first firm.

Patents are particularly important in the biomedical fields, playing crucial roles for the development of new drugs and biologics.[103] Patents have also been the subject of significant dispute in those fields.[104] In

[100] Patents also hamper cumulative innovation, if the patent on the original invention blocks the second innovator from developing her innovation; the extent to which this occurs is something of an open empirical question. *See generally* Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSPECT. 29 (1991); *see also* Arti K. Rai, *Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust*, 16 BERK TECH LJ 813, 838–44 (2001) (describing how patent doctrine can facilitate cumulative innovation). However, effective trade secrecy can very effectively prevent cumulative innovation because the initial innovation never becomes known.

[101] For a general overview of trade secrecy law, *see* Robert G. Bone, *A New Look at Trade Secret Law: Doctrine in Search of Justification*, CALIF. LAW REV. 241, 247–51 (1998).

[102] See RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 43 (1995) (listing reverse engineering as a proper means of acquiring a trade secret). Trade secrecy in the context of BBPM will be discussed below in Section III.C.2.

[103] *See, e.g.*, Eisenberg, *Problem of New Uses, The*, *supra* note 65; Roin, *Unpatentable Drugs and the Standards of Patentability*, *supra* note 100. Many have also criticized this view and the dominance of pharmaceutical patents. *See, e.g.*, Ellen 't Hoen, *TRIPS, Pharmaceutical Patents, and Access to Essential Medicines: A Long Way from Seattle to Doha*, 3 CHIC. J. INT. LAW 27 (2002); Tim Hubbard & James Love, *Medicines without Barriers*, 178 NEW SCI. 2995 (2003).

[104] Various policy arguments around patents have included the use of patents to extend drug monopolies for longer terms than contemplated in the patent term, *see, e.g.*, C. Scott Hemphill & Bhaven N. Sampat, *Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals*, 31 J. HEALTH ECON. 327 (2012); alleged antitrust violations when brand-name drug companies and generic companies agree to delay generic market entry in patent litigation, *see, e.g.*, Ronald W. Davis, *Reverse Payment Patent Settlements: A View into the Abyss, and a Modest Proposal*, 21 ANTITRUST 26 (2006); Daniel A. Crane, *Per Se Illegality for Reverse Payment Patent Settlements*, 61 ALA. LAW REV. 575 (2009); Lisa Allen, *Reviewing the Legality of Pharmaceutical Reverse Payment Settlements: The FTC Doesn't Get It Right*, 8 GEORGET. J. LAW PUBLIC POLICY 245 (2010), and the Supreme Court's take on the issue in Federal Trade Commission v. Actavis, Inc., 133 Ct 1310 (Supreme Court 2013); and international patent protection hindering the access of patients

general, however, patents appeared to provide at least some incentives for the development of the algorithms central to BBPM. However, recent changes to patent law have severely limited those incentives, leaving an incentive landscape driving personalized medicine away from BBPM.

A. Intellectual property before Mayo v. Prometheus

Intellectual property law heavily influences the data consolidation and analysis processes, since the data and models—which require major investments—are classic non-excludable goods which can be cheaply and easily copied once developed. Thus, they are traditionally fruitful targets for innovation policy. However, even as firms have begun to slowly develop models and underlying datasets, the relevant intellectual property landscape has become especially challenging.

Before 2012, firms could and did use patents to protect models for prediction and treatment in the United States. Patents could not validly claim a natural law itself—for instance, the correlation between the effect of a drug and the blood level of that drug's metabolites. However, patents could claim essentially the same subject matter by drafting a broad claim for treating a patient with the drug, while measuring the metabolite levels and knowing the relevant correlation. Thus, essentially all uses of the algorithm could be protected.

Prior to the Supreme Court's decision in *Prometheus*, the Federal Circuit had held generally that a broad diagnostic method was patentable so long as it was either linked to a machine or resulted in a transformation of matter (the "machine or transformation test.")[105] This test was disapproved in *Bilski* but remains an "important clue" to patentability and generally supported the patentability of diagnostic tests until *Prometheus*. Thus,

---

in developing nations to lifesaving drugs, *see, e.g.*, Amir Attaran & Lee Gillespie-White, *Do Patents for Antiretroviral Drugs Constrain Access to AIDS Treatment in Africa?*, 286 JAMA 1886 (2001); Bryan C. Mercurio, *TRIPS, Patents, and Access to Life-Saving Drugs in the Developing World*, 8 MARQ INTELL PROP REV 211 (2004); Sigrid Sterckx, *Patents and Access to Drugs in Developing Countries: An Ethical Analysis*, 4 DEV. WORLD BIOETH. 58 (2004). Patents for surgical techniques also raised a furor when they were introduced, but have since been statutorily limited. *See* Robert M. Portman, *Legislative Restriction on Medical and Surgical Procedure Patents Removes Impediment to Medical Progress*, 4 U BALT INTELL PROP LJ 91 (1996), 35 U.S.C. § 287(c); *but see* http://271patent.blogspot.com/2007/07/medical-technique-patents-in-spotlight.html (medical technique patents are becoming more prevalent and being used in more lawsuits).

[105] *See Metabolite Laboratories, Inc. v. Laboratory Corp. of America Holdings*, 370 F.3d 1354 (Fed. Cir. 2004); *Prometheus*, 628 F.3d 1347 (Fed. Cir. 2010); *Association for Molecular Pathology v. U.S. Patent & Trademark Office*, 653 F.3d 1329 (Fed. Cir. 2011); and *Classen Immunotherapies, Inc. v. Biogen IDEC*, 659 F.3d 1057 (Fed. Cir. 2011).

patent incentives were typically available for diagnostic algorithms.[106]

All this is not to say that patents provided ideal incentives for the algorithm development at the heart of BBPM. Complex and especially implicit algorithms are more difficult to describe sufficiently than other inventions, making it harder satisfy § 112's written description requirement. In addition, complex algorithm patents are hard to enforce, especially when those algorithms are embedded in medical practice, due to difficulties detecting when the patented algorithm is being used. However, the patent incentives for diagnostic algorithms, while imperfect, were at least still available prior to 2012.

B. Mayo v. Prometheus

Patent protection for algorithms and diagnostic methods changed dramatically with the Supreme Court's decision in *Mayo v. Prometheus Labs*,[107] which held essentially diagnostic methods claims unpatentable.

Prometheus involved two patents of the type described above related to the use of thiopurine drugs to treat autoimmune diseases.[108] Claim 1 of patent 6,355,623 (the '623 patent), which both the Federal Circuit and the Supreme Court took as exemplary,[109] teaches "a method of optimizing therapeutic efficacy for treatment . . . comprising (a) administering a drug . . . and (b) determining the level of [the metabolite] . . . wherein [metabolite levels below a certain threshold] indicates a need to increase the amount of said drug . . . and [metabolite levels above a different threshold] indicate a need to decrease the amount of said drug." Prometheus, the exclusive licensee of the patents, sells diagnostic kits which embody the patented process.[110] Mayo Clinic Rochester and Mayo Collaborative Services (collectively, "Mayo") bought those kits and used them until 2004, when Mayo decided to start making, using, and selling its own kit, with slightly different metabolite level limits.[111] Prometheus brought an infringement action in district court, which held the patents infringed but invalid as claiming a natural law.[112] The Federal Circuit reversed on the grounds that the patents' "administering" and "determining" steps satisfied the "machine

---

[106] Of course, the patentable subject matter inquiry is only part of the patentability inquiry. The claimed invention must also be useful, novel, nonobvious, and enabled. 35 U.S.C. §§ 101, 102, 103, and 112, respectively.

[107] 132 S. Ct. 1289 (2012).

[108] U.S. Pat. Nos. 6,355,623 (filed Apr. 8, 1999) and 6,680,302 (filed Jan. 27, 2001).

[109] *Prometheus*, slip op. at 5.

[110] *Id.* at 6.

[111] *Id.*

[112] *Id.* at 6–7.

or transformation" test and held that the patents did not encompass laws of nature.[113]

On certiorari, the Supreme Court held the invention was not patentable subject matter under § 101 of the Patent Act.[114] Rather than a "genuine application" of an unpatentable natural law, the court said that the patent merely provides the natural law and "tells doctors to engage in well-understood, routine, conventional activity"—namely, measuring metabolite levels and then using Prometheus's new information to inform treatment decisions. Accordingly, the patent was invalid. Broadly, *Prometheus* means firms cannot identify a useful natural law (a correlation, set of correlations, or other predictive/diagnostic relationship) and then patent a general method for its use in diagnostics or treatment.

Under the decision's strikingly broad general analysis, many if not most biological diagnostic tests can be characterized as only involving steps that measure levels of biological molecules and relating that measurement to an underlying natural connection to provide information about the patient's biological characteristics, including genes and their expression levels.[115] Under *Prometheus*, "routine, obvious" pre- or post-solution activity cannot make a claim patentable if it is primarily directed to a law of nature; thus, combining diagnostic methods with standard practice procedures will typically not aid patentability.[116]

Notably, though the correlation in *Prometheus* was quite simple, nothing in the opinion limits it to simple relationships, and the Court explicitly eschewed choosing among different laws of nature.[117] Thus, the complexity of relationships in BBPM are unlikely to make them patentable under § 101.[118]

---

[113] 628 F.3d 1347, 1355 (2010). The full procedural history is somewhat more complex; the Federal Circuit held in 2009 that the patents claimed patentable subject matter under its then-dispositive "machine or transformation test;" the Supreme Court granted certiorari, vacated, and remanded with instruction to reconsider in light of its holding in *In re Bilski*, 561 U.S. __ (2011) that that test was not dispositive, but merely an important and useful clue to the patentable subject matter inquiry; on remand, the Federal Circuit found the satisfaction of the test sufficient as a clue to patentability and again held the patents to claim patentable subject matter. *Prometheus*, slip op. at 7–8.

[114] 35 U.S.C. § 101.

[115] *See, e.g.*, Schwarz, U.I., "Clinical relevance of genetic polymorphisms in the human CYP2C9 gene," Eur. J. Clin. Invest. 33. Suppl 2: 23–30 (2013).

[116] *Prometheus*, 132 S.Ct. at 1298.

[117] *Id.* at 1303 ("[O]ur cases have not distinguished among different laws of nature according to whether or not the principles they embody are sufficiently narrow. . . . [T]he cases have endorsed a bright-line prohibition against patenting laws of nature, mathematical formulas and the like.") (internal citations omitted).

[118] The complexity of BBPM relationships, and their implicit nature, may in fact make

The Patent and Trademark Office (PTO) has issued guidance to its examiners espousing a broad interpretation of *Prometheus*,[119] and both district courts and the Federal Circuit have invalidated diagnostic test patents based on *Prometheus*.[120] Since most if not all diagnostic tests center on identifying new laws of nature and inserting them into the normal flow

---

them harder to patent under the written description requirement of 35 U.S.C. § 112.

[119] Memorandum of Andrew H. Hirshfeld, Deputy Commissioner for Patent Examination Policy, USPTO, "2014 Procedure for Subject Matter Eligibility Analysis of Claims Reciting or Involving Laws of Nature/Natural Principles, Natural Phenomena, and/or Natural Products" (Mar. 2014), *available at* http://www.uspto.gov/patents/law/exam/myriad-mayo_guidance.pdf (hereinafter "2014 Guidance"). The 2014 Guidance emphasizes that all claims reciting or involving "laws of nature/natural principles, natural phenomena, and/or natural products" must be analyzed using a three part method, analyzing whether: (i) the claimed invention is directed to one of the four statutory patent-eligible subject matter categories of process, machine, manufacture, or composition of matter; (ii) the claim recites or involves one or more judicial exceptions, for example, abstract ideas, laws of nature/natural principles, natural phenomena, and natural products; and (iii) the claim as a whole recites something *significantly different* than the judicial exceptions.

[120] In *PerkinElmer, Inc. v. Intema Ltd.*, the Federal Circuit held invalid a claim over a test that established the risk of fetal Down's syndrome. 496 Fed. Appx. 65 (2012). The claimed methods compared marker measurements with each other to predict the risk of Down's syndrome. *Id.* at 69. The claim was analogous to that in *Prometheus*, and thus invalid because it merely claims a co-occurrence between biological molecules and a natural statistical relationship. *Id.* at 69.

Similarly, in *SmartGene, Inc. v. Advanced Biological Laboratories*, the Federal Circuit considered and held unpatentable—admittedly, in a nonprecedential opinion—a system paradigmatic of BBPM-type diagnostics, though relying on expert rules rather than implicit relationships. Slip op. (Fed. Cir., Jan 14, 2014), *available at* http://www.cafc.uscourts.gov/images/stories/opinions-orders/13-1186.Opinion.1-22-2014.1.PDF.

In the patent at issue, the representative claim 1 recited the steps of:

> (a) providing patient information to a computing device comprising three different knowledge bases: therapeutic treatment regimens, expert rules and advisory information useful for the treatment of a particular disease or medical condition;
>
> (b) generating a ranked listing of the treatment regimens; and
>
> (c) generating advisory information.[120]

The Federal Circuit held the patent to cover abstract ideas, relying both on prior Federal Circuit precedent, *id.* at 7–9, and *Prometheus*, *id.* at 9–10. Recitation of steps that are well known in the art and that manipulate a naturally occurring relation cannot claim patent-eligible subject matter.[120] *Slip op.* at *2–3. For further discussion, *see* Timo Minssen & David Nilsson, *The US Supreme Court in Mayo v Prometheus–taking the Fire from or to Biotechnology and Personalized Medicine?*, QUEEN MARY J. INTELLECT. PROP. 376, 383 (2012)

of clinical practice, *Prometheus* strikes directly at the patentability of diagnostics and the personalized medicine of which they are an integral part.[121]

This change was strengthened in 2013 by the Court's decision in *Association for Molecular Pathology v. Myriad Genetics*, which held isolated DNA sequences unpatentable.[122] Previously, isolated DNA patents had provided a potential complement to patents on genetic diagnostic methods, as they protected another step in the process and were easier to enforce. *Myriad* was also a case about diagnostic testing—in particular, genetic testing for a predisposition to breast cancer. The Supreme Court held that the patent claims claiming isolated genomic DNA, the hardest form of DNA to invent around, were invalid under § 101.[123] The patents also claimed genetic diagnostic methods, which the Federal Circuit had held below to be unpatentable subject matter under *Prometheus*.[124] Finally—though less significantly, in 2014, *Alice Corp. v. CLS Bank Int'l* reaffirmed *Prometheus* by clarifying that abstract inventions, such as an algorithm, were not made patentable merely by implementation on a computer.[125] Prior to *Alice*, some scholars had suggested that *Prometheus* might be tightly cabined, but *Alice* reaffirmed its principles and broad reach.[126] *Prometheus*, buttressed by *Myriad* and *Alice*, now drives firms toward incremental improvements and away from BBPM.

C. The impact of Prometheus *on personalized medicine*

Prometheus—with support from *Myriad* and *Alice*—has major real-world effects on the industry and BBPM. In addition to general negative reactions—decreased venture capital investments in the diagnostic industry and pessimistic outlooks[127]—firms have already begun to shift product

---

[121] *See, e.g.*, *Id.* at 384 ("[C]laims that are broadly directed to what may be considered to be a typical method exploited in personalized medicine will probably be held to be unpatentable under the Prometheus principles." This is not to argue that the patent system before created ideal incentives for BBPM, a point discussed further below. However, under prior law at least some patent protection was available.

[122] 133 S. Ct. 2107 (2013).

[123] *Id.* at 2117–18. Myriad's claims covering complementary DNA ("cDNA"), which are DNA that complement protein-coding messenger RNA, were valid under § 101; cDNA is patentable subject matter. *Id.* at 2119.

[124] Ass'n for Molecular Pathology v. USPTO, 689 F 3d 1303, 1335 (Fed. Cir. 2012).

[125] __ U.S. __ (2014) (No. 13–298, June 19, 2014), slip op at 10–16.

[126] *Id.* Even if *Prometheus* were tightly limited, it would still cast doubt on the patentability of the sort of the diagnostic tests it directly addressed.

[127] *See, e.g.*, Heidi Ledford, *Software Patents Await Legal Fate*, 507 Nature 410, 410 (2014).

focuses to those which can still be successfully protected by patents. In addition, trade secrecy, with its problems for oversight and cumulative innovation, becomes more attractive by comparison.

1. Paired diagnostics

Because *Prometheus* makes unpatentable natural laws and routine applications of those laws, firms may seek to link newly discovered natural laws strongly to machines or to specific treatments. Indeed, firms have done just that, by increasing their emphasis on combination products that pair a diagnostic with a device and/or drug.[128] Since such pairings involve substantially more than just stating a natural law, they are likely still patent-eligible under *Prometheus.*[129] Combination products tend to focus on simple, explicit links, and are tested and brought through an FDA approval process which focuses on validating those links in clinical trials.[130] Thus, while this change in focus by firms may be entirely rational, it means that the contours of intellectual property rights are pushing to keep the industry focused on explicit personalized medicine, rather than devoting energy to the broader algorithms, models, and datasets necessary to bring about the benefits of black box personalized medicine.

The major problem with moving to a combination product model is that is keeps personalized medicine firmly locked into the current regime of incremental steps. This is not to disparage the potential benefit of combination devices or EPM in general. However, to the extent that firms attempt to maintain patentability by focusing on simple, explicit links associated with devices, they leave untapped the larger datasets and more complex algorithms needed for BBPM, and which themselves can help drive EPM forward.

A focus on paired diagnostics is one potential response to an inability to patent pure diagnostic tests, by shifting from patenting the relationship itself to a combination of the relationship and some other physical machine or process. An alternate pathway is to eschew the patent system entirely, and to rely on trade secrecy.

---

[128] Aaron Kesselheim & Jason Karlawish, *Biomarkers Unbound – The Supreme Court's Ruling on Diagnostic-Test Patents*, 366 New Eng. J. Med. 2338, 2340 (2012).

[129] PTO Guidance, *supra* n. 119.

[130] U.S. FDA, Draft Guidance for Industry and Food and Drug Administration Staff - In Vitro Companion Diagnostic Devices (July 14, 2011), *available at* http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262292.htm.

2. Trade secrecy and proprietary data

Trade secrecy is the principal private law alternative to the patent system to protect technological innovation.[131] Keeping information secret prevents competitors from acquiring or using the innovation. Trade secrecy can protect information which is unpatentable, and lasts as long as the information is secret.[132] Knowledge which is reasonably kept secret and which derives independent economic value from its secrecy is protected from misappropriation by state and federal trade secret law.[133] However, secret information can legally be reverse-engineered.[134]

Sets of data have never been protectable with patents or copyrights, but can be kept secret.[135] Algorithms based on those data, however, were eligible for such protection until *Prometheus*, as described above. Without other intellectual property protection, secrecy is the strongest method available to keep competitors from accessing the data and algorithms at the heart of personalized medicine. Many companies have chosen to rely on proprietary data and secrecy to maintain a competitive advantage, and this practice is likely to increase.

The clearest example is Myriad Genetics itself.[136] After its loss in the Supreme Court, the company has sought to keep other information about genetic variation secret.[137] Myriad's gene testing process reveals combinations of alleles present in patients; the company then offers free testing to family members, and analyzes family variation to determine

---

[131] *See* Bone, *A New Look at Trade Secret Law*, *supra* note 102, at 243

[132] *Id.* at 248

[133] Forty-seven states have enacted some form of the Uniform Trade Secrets Act; the exceptions are New York, North Carolina, and Massachusetts. http://www.uniformlaws.org/LegislativeFactSheet.aspx?title=Trade%20Secrets%20Act. Under federal law, the Economic Espionage Act of 1996 makes the theft or misappropriation of a trade secret a federal crime. 18 U.S. C. § 1832.

[134] See RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 43 (1995) (listing reverse engineering as a proper means of acquiring a trade secret).

[135] In Europe, a *sui generis* system of database protection has existed since 1996. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *available at* http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31996L0009&from=EN.

[136] Barbara J. Evans, *Economic Regulation of Next-Generation Sequencing*, in SPECIAL ISSUE: POLICY ISSUES IN NEXT-GENERATION SEQUENCING (Amy L. McGuire, David J. Kaufman & Margaret A. Curnutte, eds.), 42 J. L., MED. & ETHICS (forthcoming 2014).

[137] Monya Baker, *Policy Paper: Myriad Turns Cancer Genetic Data into Trade Secrets*, NATURE NEWS BLOG, (Oct. 31, 2012), http://blogs.nature.com/news/2012/10/policy-paper-myriad-turns-cancer-genetic-data-into-trade-secrets.html.

significantly linked genetic patterns.[138] Since Myriad has a substantially greater set of data on BRCA1/2 variants, only 3% of its samples have variants of unknown significance;[139] for competitors, roughly 20% to 30% of samples have variants of unknown significance.[140] Test samples sent to Myriad are therefore much less likely to be returned to the physician as "uninterpretable" than samples sent to their competitors,[141] providing a robust competitive advantage. While Myriad's data advantage could be overcome as other firms slowly assemble their own databases, the fact that Myriad currently possesses a much larger database—amassed from its period of patent protection—is self-reinforcing.[142] Myriad can provide more results, and is therefore likely to continue receiving more test samples; the resulting larger database would still be kept as a trade secret.[143] Myriad's business plan includes retaining and expanding this secrecy-based advantage of mutation data and algorithms.[144]

Other genetic testing firms have similarly relied on proprietary data. Quest Diagnostics has licensed and introduced a proprietary biomarker for the diagnosis of rheumatoid arthritis, "IdentaRA with 14-3-3η."[145] GeneDx offers its "TessArae" service to test for Noonan Syndrome; it uses proprietary software algorithms to interpret its test results.[146] And Good Start Genetics combines a range of proprietary technologies, processes, and

---

[138] *Id.*

[139] In a genetic test like Myriad's, the physical process first determines which alleles of a gene the patient has. That identification must then be interpreted to convey useful medical information; are the alleles associated with a higher or a lower risk of cancer, or with no change? Douglas F. Easton et al., *A Systematic Genetic Assessment of 1,433 Sequence Variants of Unknown Clinical Significance in the BRCA1 and BRCA2 Breast Cancer–Predisposition Genes*, 81 AM. J. HUM. GENET. 873, 873 (2007)When the interpreting entity lacks sufficient information about a particular allele to provide a useful interpretation, it is termed a "variant of unknown significance," and that part of the test is inconclusive.

[140] Baker, *supra* n. 137.

[141] *Id.*

[142] *See* Evans, *supra* n. 136.

[143] *Id.*

[144] *Id.;* Myriad Genetics, Inc., *United States Securities and Exchange Commission Form 10-K*, (Fiscal Year 2013), *available at* http://files.shareholder.com/downloads/MYGN/3108552224x0xS1193125-13-334245/899923/filing.pdf.

[145] Quest Diagnostics, Inc., *United States Securities and Exchange Commission Form 10-K*, (Dec. 31, 2013), *available at* http://ir.questdiagnostics.com/phoenix.zhtml?c=82068&p=irol-irhome.

[146] GeneDx, *TessArae Targeted Resequencing at GeneDx Provides More Comprehensive Analysis, Lower Cost, and Faster Results*, http://www.tessarae.com/downloads/Literature/TessArae_GeneDx_CaseStudy.pdf.

algorithms, into its proprietary—and partially secret—"platform" for detecting Tay-Sachs, which includes a broader set of mutations than standard platforms.[147]

Among other concerns,[148] keeping data and algorithms secret in this area may significantly hamper the overall development of personalized medicine in general and black box personalized medicine in particular. Trade secrecy slows cumulative innovation and promotes duplicative investment.[149] Predictive ability increases with dataset size and variety; the "big data" nature of BBPM is what enables the discovery of complex correlations. When the data shrinks and fragments, fewer relationships are available, and those which are found are less robust.

In sum, though *Prometheus* and its kin may or may not be justified on substantive patent law grounds—a debate into which this Article does not wade—the cases seriously decrease the patent incentives available in the United States for personalized medicine in general, and for black box personalized medicine in particular.[150]

---

[147] Good Start Genetics, *About Us*, http://www.goodstartgenetics.com/about/; Stephanie Hallam, *Validation for Clinical Use of, and Initial Clinical Experience With, A Novel Approach to Population-Based Carrier Screening Using High-Throughput, Next-Generation DNA Sequencing*, 16 JOURNAL OF MOLECULAR DIAGNOSTICS 180 (2014). Approximately one quarter of mutations identified by Good Start Genetics are not included in more traditional mutation panels. *Id.*

[148] Keeping data proprietary raises several potential concerns. On the ethical side, the Chairwoman of the European Society of Human Genetics' Professional and Public Policy Committee described herself as "very concerned that such important data is being withheld from those who most need it." Martina Cornel, *Statement*, (Oct. 31, 2012), https://www.eshg.org/13.0.html. She suggested that "[p]olicymakers take an urgent look at the regulatory and reimbursement issues involved in genomic testing in order for all the data that is essential to understanding…the clinical significance of [mutations] to be made public, to the benefit of patients and healthcare providers alike." *Id.* Others have noted that keeping data proprietary removes them from the potential of peer review and makes us less certain of their accuracy. Monya Baker, *Policy Paper: Myriad Turns Cancer Genetic Data into Trade Secrets*, Nature News Blog, (Oct. 31, 2012), http://blogs.nature.com/news/2012/10/policy-paper-myriad-turns-cancer-genetic-data-into-trade-secrets.html. Other concerns arise with respect to transparency, oversight, and the blocking of future research directions.

[149] Bone, *A New Look at Trade Secret Law*, *supra* note 102; Robert G. Bone, *The (Still) Shaky Foundations of Trade Secret Law*, TEX. REV (2014). For a defense of treating trade secrecy as intellectual property, *see* Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN REV 311 (2008).

[150] This Article considers only domestic protection and incentives; international analyses are beyond its scope. The situation in Europe appears to differ substantially; methods like that in *Prometheus* would likely be patentable subject matter under Article 52 of the European Patent Convention (EPC), because they involve *in vitro* diagnostic tests performed on human subjects. Timo Minssen and David Nilsson, *The US Supreme Court*

IV. IMPROVING INCENTIVES

The absence of appropriate incentives significantly impedes the development of BBPM; more directly, the incentives available actively drive the development of personalized medicine in unhelpful directions. Accordingly, the final section of this Article briefly proposes potential improvements to the existing incentive structure.

Because the development process involves distinct forms of innovation, incentives are most usefully considered separately for each form. Generating large and well-curated datasets likely requires the greatest investment. Patents are unavailable to datasets, and trade secrecy is relatively ill-suited to consolidation and cumulative innovation. Instead, the amassing of high-quality datasets might better be conceived as an infrastructure for further innovation, which suggests a role for more direct government involvement. An additional possibility is the implementation of a tailored *sui generis* dataset protection regime, such as exists under EU law.

The second phase of BBPM development is the generation of algorithms. As described above, algorithms were previously patent-eligible, so one potential incentive for algorithm would come from reinstating patent protection for them. This solution, however, comes with its own set of complications. Regulatory exclusivity might be preferable, though that would require a regulatory preapproval regime that currently does not exist. Prizes are another potential solution; although they are subject to many of the same general innovation considerations as patents, they are typically more flexible to implement.

Third, incentives are needed for validation. An ever-present concern in complex implicit models—especially when developed via black-box

---

in Mayo v Prometheus – *Taking the Fire* from *or* to *Biotechnology and Personalized Medicine?*, 2 Queen Mary J. Intell. Prop. 376, 385–86 (2012). *See* Paul Cole, *Prometheus v Mayo – A European View*, Patently-O , http://patentlyo.com/patent/2012/04/guest-post-prometheus-v-mayo-a-european-view.html (last visited April 30, 2014). In fact, patents on the same methods as in *Prometheus* were granted by the EPO. *Id*. Those patents were not the subject of opposition proceedings in the EPO, which therefore did not rule on their patentability (nor, to the author's knowledge, has the EPO ruled on the patentability of precisely analogous claims). However, Article 52(2) of the EPC states that "discoveries, scientific theories and mathematical methods . . . [and] schemes, rules and methods for performing mental acts" are not patentable inventions. Thus, it is unclear whether purely algorithmic innovations would be patentable. Even if such algorithms are patentable in Europe, their inability to receive patents in the U.S. increases incentives for firms to keep them secret, or pursue other innovations.

methods—is ensuring that they are valid and generally applicable, rather than just statistical artifacts arising from over-specification in large datasets. The paper will thus present potential structures for regulatory "bounties:" rewards provided to competitor firms for either validating or substantially falsifying the BBPM algorithms of the innovator firms.

A. Incentives for datasets

As described above, significant hurdles exist in the collection of large, high quality datasets available for the development of BBPM algorithms.[151] Patents are unavailable, and trade secrecy presents problems mirroring those above: it lends itself to fragmenting rather than consolidating information, restricts cumulative innovation, and creates advantages for incumbents—like Myriad Genetics—which may continue indefinite specific monopolies.[152] To increase incentives, therefore, policymakers could turn to direct government intervention or a public-private partnership focused on data as infrastructure, or could alternately use a different intellectual property regime to drive private development, likely modeled on the European *sui generis* system of database protection.

In the context of genetic testing, the secrecy which protects the databases of incumbent firms has been analogized to an infrastructure problem, wherein specific sets of correlations—namely, the significance of individual genetic variations—have several features of essential facilities.[153] Datasets for BBPM development may similarly take the role of common infrastructure for further innovation.

Under this view, direct or indirect government intervention could usefully aid the generation of datasets. On a direct level, collecting data shows a *prima facie* advantage for government. In the U.S., over 95 million patients participate in Medicare and Medicaid, where the government provision of insurance allows access to patient medical records. The Department of Defense and Veteran's Administration provide direct health care, and consequently collects data, for over 11 million military personnel, veterans, and their families.[154] In other nations, the government concentration of data is even stronger; the U.K.'s National Health Service

---

[151] *See* Section II.A, *infra*.

[152] For a detailed description of this problem, *see* Evans, *supra* n. 136.

[153] *Id.*

[154] Tricare, which offers health for military personnel, has 9.2 million eligible beneficiaries. http://www.tricare.mil/stakeholders/statistics.cfm. The Veteran's administration had 8.9 million enrollees in 2013. http://www.va.gov/HEALTHPOLICYPLANNING/enroll02/Fnl925Doc.pdf.

(NHS) provides free health care to over 63 million UK residents, and consequently accumulates tremendous amounts of data.[155]

Government possession of data brings its own challenges. For instance, it is emphatically *not* the case that the data gathered by the Department of Defense and the Veteran Administration's data are neatly available in high-quality interoperable formats. In fact, the two agencies have spent billions trying and failing to upgrade their electronic records systems, which remain incompatible.[156] And the U.K.'s NHS, while it has a great deal of data, is prevented by strict privacy rules from using much of that.[157] However, government entities are taking steps in the direction of data collection, even if not in sharing: the Veteran's Administration is well into its effort to collect genetic and phenotypic information on a million veterans for research purposes.[158] However, there are currently indications that this information—or other information like it—will be made available for further innovation by private entities.

This last caveat could be changed: government could enable BBPM (and other personalized medicine) by simplifying the data collection step, generating the dataset-infrastructure, and then allowing private parties to compete in the analysis and validation steps.[159] The data could be leveraged in one of at least two ways. First, the data could be used exclusively for some time; firms could bid for access to segments of the data, coupled with a commitment to make any resulting algorithms public after some period of exclusivity. Second, the data could be made freely available, but with the caveat that firms using the data disclose their algorithms. This would enable a broader set of concurrent developments, while still allowing firms

---

[155] http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx.

[156] *Billions Wasted on Fruitless Bid to Create Paperless Vet Health Records*, NBC NEWS (August 26, 2013), http://www.nbcnews.com/news/other/billions-wasted-fruitless-bid-create-paperless-vet-health-records-f8C11001233 (last visited May 9, 2014).

[157] Others have noted the richness of the NHS's data and the challenge of its privacy rules. *See* Wayne Parslow, *How Big Data Could Be Used to Predict a Patient's Future*, THE GUARDIAN (2014) ("Although currently shielded by privacy rules, the personal data that can risk score every NHS patient already exists. And it is already far more centralised and normalised than in countries such as the US, giving the UK the opportunity to become the world leader.")

[158] *See* U.S. Dept. of Veteran's Affairs, Office of Research and Development, *Million Veteran Program*, http://www.research.va.gov/MVP/.

[159] This approach has potential political economy problems with implementation; the specter of government collecting health records and turning them over to private parties for their exclusive benefit would likely meet substantial resistance. This dynamic certainly exists in other frameworks—notably in the patenting of government-funded innovation under the Bayh-Dole Act—but may be even more politically sensitive in the context of health information.

to capture benefits of their (reduced) innovation investments.[160]

A similar path could be pursued with somewhat less direct government intervention. If political economy concerns or other issues counseled against direct government intervention, a public-private partnership could fill a similar role. deCODE Genomics famously exemplifies such a partnership: the Icelandic biopharmaceutical firm successfully lobbied the Icelandic Parliament to create a population-wide Health Sector Database including genomic, genealogical, and health information.[161] Court challenges shifted the database from mandatory to voluntary,[162] and the effort was highly controversial,[163] but over 120,000 individuals still volunteered, and the company has published extensively on the explicit genomic links it has found.[164] Similarly, the Human Genome Project provides another clear precedent: there, a collaboration between government and private researchers sequenced the human genome with the intention of providing it freely to future researchers and innovators as a common infrastructure resource.[165]

Other fully private options for adding incentives for databases could rely on large-scale prizes[166] or could follow Europe's example by creating *sui generis* intellectual property for databases.[167]

B. Incentives for algorithms

The heart of BBPM is the development of biomedically useful

---

[160] This plan would also potentially avoid some of the political economy problems mentioned above, *supra* n. 159. However, it would exacerbate privacy and reidentification concerns. *See* Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, *supra* note 46.

[161] V. Árnason, *Coding and Consent: Moral Challenges of the Database Project in Iceland*, 18 BIOETHICS 27 (2004).

[162] Renate Gertz, *An Analysis of the Icelandic Supreme Court Judgement on the Health Sector Database Act*, SCR.-ED 241 (2004)

[163] *See, e.g.*, Árnason, *supra* note 162; J. R Gulcher & K. Stefansson, *The Icelandic Healthcare Database and Informed Consent*, 342 N. ENGL. J. MED. 1827 (2000).

[164] *See, e.g.*, Unnur Styrkarsdottir et al., *Nonsense Mutation in the LGR4 Gene Is Associated with Several Human Diseases and Other Traits*, 497 NATURE 517 (2013); Thorlakur Jonsson et al., *A Mutation in APP Protects against Alzheimer's Disease and Age-Related Cognitive Decline*, 488 NATURE 96 (2012).

[165] Francis S. Collins et al., *The Human Genome Project: Lessons from Large-Scale Biology*, 300 SCIENCE 286 (2003).

[166] *See* Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND REV 115, 225–31 (2003) (Proposing that prizes might succeed in the closely related area of genomics research).

[167] *See* n. 135, *supra*.

algorithms by plumbing the masses of health data.[168] However, as described above, this process is neither easy nor inexpensive.[169] And current intellectual protection is both inadequate and skewed away from BBPM. Accordingly, better incentives are needed to drive algorithmic development. Potential incentives could come in at least three forms: patent protection, regulatory exclusivity, or prizes.

1. Patents

Patents are an obvious source of incentives for algorithms, as they were generally available for algorithms until the Supreme Court's decision in *Prometheus*. Congress could override that statutory interpretation decision by amending the statute to, for instance, allow the patenting of biomedical laws of nature. While this approach is initially attractive, challenges arise in both enactment and enforcement.

First, overruling *Prometheus* may face problems of overbreadth. In particular, BBPM is similar to straightforward computer software patents and algorithms, which are criticized by academics, frequently disliked by the software industry itself, and a target of reform efforts.[170] Broad-brush patent changes to revive algorithmic patents may therefore face considerable resistance and may also have negative impacts on other industries.[171] Finally, of course, the Supreme Court may have correctly judged the innovation incentives regarding laws of nature and determined

[168] *See* Section I.A.3.b, *infra*.

[169] *See* Section II.B, *infra*.

[170] *See, e.g.*, Jay Dratler Jr, *Does Lord Darcy Yet Live-The Case against Software and Business-Method Patents*, 43 ST. CLARA REV 823 (2002); Colleen V. Chien, *Reforming Software Patents* (2012); Pamela Samuelson, *Benson Revisited: The Case against Patent Protection for Algorithms and Other Computer Program-Related Inventions*, 39 EMORY LJ 1025 (1990); Robert E. Thomas, *Debugging Software Patents: Increasing Innovation and Reducing Uncertainty in the Judicial Reform of Software Patent Law*, 25 ST. CLARA COMPUT. HIGH TECH LJ 191 (2008); *but see contra* Martin Campbell-Kelly & Patrick Valduriez, *Technical Critique of Fifty Software Patents, A*, 9 MARQ INTELL PROP REV 249 (2005) (finding most frequently-cited software patents worth protection).

[171] Merely restoring patent law incentives for BBPM might also result in problems for BBPM itself. Although this Article begins with the position that *Prometheus* and *Myriad* problematically reduced the patent incentives available for BBPM, those initial incentives had flaws as well. In particular, since BBPM resembles software in many ways, we might expect to see some of the same problems that infect patents in the software industry, including frequent issuance of patents on obsolete technologies, mismatch between patent claims and actually invented subject matter, broad and vague claim language, and substantial transaction costs. *See, e.g.*, Mark A. Lemley, *Software Patents and the Return of Functional Claiming*, 2013 WISC REV 905 (2013); Chien, *supra* note 171; Thomas, *supra* note 171.

that patents on relationship-based algorithms may be harder to invent-around and therefore overall innovation-blocking.[172]

Second, patents granted on BBPM algorithms face significant difficulties in enforcement. Especially for more complex algorithms, knowing whether infringement is occurring and proving that it has occurred are both likely to be difficult.[173] Thus, though restoring the patent system to its status before *Prometheus* has some initial appeal for driving the development of algorithms, other possibilities may better align incentives.

2. Regulatory Exclusivity

Regulatory exclusivity could provide incentives better tailored to algorithms. Instead of relying on the patent system to provide the incentive of excludability, in regulatory exclusivity a regulator excludes competitors from selling a product by withholding market pre-approval.[174] Thus, regulatory exclusivity requires the existence of a premarket approval regime. In multiple contexts where such preapproval requirements exist, regulatory exclusivity is used as an innovation incentive;[175] in others, it has been proposed.[176] The majority of extant applications of regulatory exclusivity are administered by the FDA, primarily around the marketing of small-molecule drugs and biologics.[177]

---

[172] The scholarly debate on these issues is extensive and need not be recapped here. *See*, *e.g.*, Mark A. Lemley et al., *Life After Bilski*, 63 STANF. REV 1315. For discussion of the relevant knowledge/embodiment distinction in patent law, *see* Kevin Emerson Collins, *The Knowledge/Embodiment Dichotomy*, 47 UC DAVIS REV (2014).

[173] For an analogous situation, *see*, *e.g.*, W. Nicholson Price II, *Making Do in Making Drugs: Innovation Policy and Pharmaceutical Manufacturing*, 55 BOSTON COLL. LAW REV. 491, 526–27 (2014) (describing the difficulties secrecy creates in enforcing manufacturing process patents).

[174] *See*, *e.g.*, R. S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH TELECOMM TECH REV 345 (2006); Yaniv Heled, *Regulatory Competitive Shelters—An Emerging Class of Administrative Properties Yaniv Heled*. Heled proposes the more general term "regulatory competitive shelter" to describe this phenomenon, but I will continue to use "regulatory exclusivity," largely for simplicity's sake. To the extent that regulatory competitive shelters could take on more shades than pure exclusivity—for instance, purely higher costs to market entry, in the nature of regulator-enforced mandatory license fees—the broader phenomenon might more appropriately describe alternate solutions.

[175] *See* Heled, *supra* note 175 (describing regulatory exclusivity regimes for drug, biologic, and pesticide development).

[176] *See* Price II, *supra* note 174 (proposing regulatory exclusivity for drug manufacturing innovations to promote such innovation).

[177] Heled, *supra* note 175 (listing 14 such regimes, of which 13 are administered by the FDA and one by the Environmental Protection Agency).

Assuming the existence of a preapproval regime, regulatory exclusivity would function the same way for BBPM predictive models as for other innovations within preapproval regimes. In the context of drugs, the FDA will not approve a generic drug within five years of the approval of a drug based on a new chemical entity; for biologics, the period is twelve years.[178] Similarly, if FDA approval were required for BBPM models to be commercially marketed and used, FDA could withhold that approval for a fixed period of time as a reward to the innovator company. The main advantages for regulatory exclusivity come in flexibility, ease of enforcement, and strong disclosure.

Regulatory exclusivity is more flexible than the patent system for two principal reasons. First, it is administered by an expert agency with experience in the specific technology and—ideally—an innovation mandate as well as a regulatory health/safety mandate.[179] Even without substantial changes to the statutory contours of exclusivity, the agency can apply it flexibly. Second, statutory changes are made easier because regulatory exclusivity is not bound by the same treaty requirements as patent law; it can be flexible across products and across industrial sectors in a way that patent law cannot.[180]

The second advantage is that regulatory exclusivity is substantially easier to enforce than patents, with consequently more uniform enforcement. The default of a market preapproval regime is the inability to enter the market; thus, if regulatory exclusivity exists for a particular product, competitors can be prevented from entering that market simply by denying approval for the competitors' products for the appropriate period of time.[181] This contrasts with the difficulty and expense of enforcing

---

[178] 21 U.S.C. § 355(j)(5)(F)(ii) (granting five years of market exclusivity for new chemical entities); 42 U.S.C.A. § 262(k)(7) (West 2011 & Supp. 2013) (granting four years of market exclusivity and an additional eight years of data exclusivity to biologics).

[179] Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 175

[180] The treaty on Trade Related Aspects of Intellectual Property (TRIPS), to which the United States is a party, requires that patent systems be relatively uniform across different countries. *Id.* at 365. For the purposes of this Article, the most important requirement of TRIPS is that patent terms cannot be technology-specific. *Id.* (suggesting that regulatory exclusivity may help tailor innovation policy without violating TRIPS). Note, however, that countries may choose to disallow the patentability of medical techniques (as Europe has largely done, *see* EPC Art. 52(2), and as the U.S. has effectively done, *see* 35 U.S.C. § 287(c)). BBPM and algorithmic medicine in general could potentially be excluded from patentability—but probably not given an intermediate or differently-structured set of incentives from other technological areas.

[181] On the other hand, the definition of a "product" might be particularly flexible in the context of BBPM; fluid boundaries would raise some of the same enforcement challenges that exist in patent law.

patents.[182]

The third and final advantage to regulatory exclusivity comes only if exclusivity is coupled with a disclosure requirement. In the context of drug development, regulatory exclusivity demands the production of knowledge (that a drug is safe and effective, as measured by clinical trials), and requires at least some disclosure of that knowledge.[183] Although clinical trial data are not fully disclosed now,[184] the basic results of trials—that a particular drug is safe and effective for a particular indication—becomes public and can be relied upon by generic companies to secure approval.[185] In general, since the regulator who approves the product is the same entity which enforces regulatory exclusivity, innovators have an incentive to be forthcoming and candid in their disclosures, rather than facing the incentive to obscure useful technical information in patents to minimize disclosure to competitors.

The principal challenge with implementing regulatory exclusivity is that it relies on a market-spanning regulatory pre-approval regime, which does not currently exist for data-driven diagnostic tests.[186] A full analysis of FDA's diagnostic test regime and what is most appropriate for BBPM must await future work. In brief, however, while FDA does currently regulate some diagnostic tests, many exist outside its current scope, and there is certainly not a comprehensive regime in place.[187] Were such a regime implemented, regulatory exclusivity would be an attractive possibility.

---

[182] *See* American Intellectual Property Law Association, 2011 Report of the Economic Survey (July 2011) (For patent infringement claims under $1 million, median legal costs are $650,000; for claims from $1 million to $25 million, costs are $2.5 million; for claims over $25 million, median costs are $5 million.

[183] Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 175, at 366–72.

[184] A rapidly growing movement focuses on the disclosure of clinical trial data. *See* Mary Beth Hamel et al., *Preparing for Responsible Sharing of Clinical Trial Data*, 369 N. ENGL. J. MED. 1651 (2013); Kamran Abbasi, *Compulsory Registration of Clinical Trials: Will Be a Requirement before Submission to the BMJ from July 2005*, 329 BMJ 637 (2004); Richard Lehman & Elizabeth Loder, *Missing Clinical Trial Data*, 344 BMJ d8158 (2012). However, these disclosures are not tied to FDA approval. Arguments have long been made that information submitted for regulatory approval should be disclosed, *see* Thomas O. McGarity & Sidney A. Shapiro, *The Trade Secret Status of Health and Safety Testing Information: Reforming Agency Disclosure Policies*, 93 HARV. LAW REV. 837 (1980), but those arguments have not succeeded.

[185] 21 U.S.C. § 305(b)(2) (allowing generic drug applicants to rely on the finding of safety and efficacy of the pioneer drug).

[186] The FDA does regulate diagnostic testing kits, but does not regulate testing services provided by individual laboratories. Steven Gutman, *The Role of Food and Drug Administration Regulation of In Vitro Diagnostic Devices—Applications to Genetics Testing*, 45 CLIN. CHEM. 746 (1999).

[187] *Id.*

Other problems with regulatory exclusivity are inherent in the name and the concept: it, like the patent system, focuses on exclusivity. To the extent that BBPM models rely on underlying natural laws, excluding others from using those laws faces the same problems that the Supreme Court named as problematic for innovation in *Prometheus*. Additionally—and problematically—applying regulatory exclusivity relies on defining the contours of a specific models. When models are multifaceted, complex, and implicit, defining the contours of a model and knowing whether another model overlaps those contours may be an insurmountable hurdle.

3. Prizes

A third possibility to enhance innovation in algorithms is reliance on prizes and/or grants as a reward for innovation.[188] Grants and prizes each typically rely on the award of money—typically a fixed sum—to solve a defined problem. Under a grant regime, firms compete for monetary incentives which are then to be used to develop an innovation. Under a prize regime, a monetary prize is offered to whichever firm can develop a solution to a defined problem. Typically, the prize amount is fixed, though it need not be.[189] Such devices can avoid the requirement of exclusivity,

[188] An extensive literature examines prizes and grants as alternatives to patents. *See, among many others,* Joseph Stiglitz, *Give Prizes Not Patents*, 16 NEW SCI. (2006); Abramowicz, *supra* note 167; Marlynn Wei, *Should Prizes Replace Patents-A Critique of the Medical Innovation Prize Act of 2005*, 13 BUJ SCI TECH L 25 (2007); Michael Kremer, *Patent Buyouts: A Mechanism for Encouraging Innovation*, 113 Q.J. ECON. 1137 (1998); Steven Shavell & Tanguy Van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525 (2001) (arguing for the superiority of an optional prize system); *see contra* F. Scott Kieff, *Property Rights and Proprietary Rules for Commercializing Inventions*, 85 MINN. L. REV. 697 (2001) (arguing against prizes). For an overview of grants and prizes which places them in a taxonomy with patents and tax incentives—and argues that all four can set economic incentives that should be at base indistinguishable to rational firms—*see* Daniel J. Hemel & Lisa Larrimore Ouellette, *Beyond the Patents–Prizes Debate*, 92 TEX REV, 310–13 (2013). For an argument that patents and prizes, at least as applied, are largely indistinguishable, *see* Benjamin N. Roin, *Intellectual Property versus Prizes: Reframing the Debate*, 81 U CHI REV __ (2014). Roin also offers an extensive bibliography. *Id.* at 3–5. This literature has typically not included regulatory exclusivity among the menu of options, perhaps because its exclusivity model parallels that of patents; to the extent that regulatory exclusivity has benefits over patents for certain fields of technological innovation, it may obviate certain criticisms that lead at least some scholars to prefer prizes.

[189] For instance, instead of a fixed sum of money, a prize could be defined as a fraction of identifiable government savings attributable to the innovation. *Cf.* Earl L. Grinols & James W. Henderson, *Replace Pharmaceutical Patents Now*, 25 PHARMACOECONOMICS 355, 356 (2007) (proposing drug prizes tied to sales); Nancy Gallini & Suzanne Scotchmer,

either in situations where it is unavailable (as when the innovation is unpatentable) or where free distribution is mandated as part of the incentive regime (as, for instance, where entering a prize competition or winning a grant requires relinquishing patent rights and committing to disclosure).

Prize and grant systems each require knowing the approximate contours of a defined problem with a defined solution and knowing the rough value of a solution to the problem.[190] Since personalized medicine in general and BBPM in particular are broad endeavors with significant implicit knowledge, clearly defining problems and solutions appears particularly difficult. Prizes could be defined very generally—for instance, any algorithm which decreases costs while maintaining or increasing health measures. Such very broad (and very valuable) algorithms might be most useful, but might also be hardest to overcome the incentives of private parties with competitive incentives to keep the algorithms secret. Goals could also be defined more narrowly: for instance, any algorithm which decreases the frequency adverse reactions to taking a drug with a narrow therapeutic index by 10%.

But the challenge remains of determining the optimal incentive size. The advantage of patents and other exclusivity regimes is that—at least ideally—the size of the reward should track the social value of the innovation.[191] Firms can use market information to project that value and invest accordingly. For prizes, whoever sets the prize, typically the government, usually must determine the eventual social value in advance; governments are typically not well suited to this task.[192] Potentially, this problem could also be solved by basing the reward not on a specific dollar amount but rather on a fraction of savings to government health programs like Medicare or Medicaid; this would scale with social value without requiring pre-estimation of the eventual size of the reward. However, many medically valuable uses are not particularly economical; keeping a patient

---

Intellectual Property: When Is It the Best Incentive System?, in 2 INNOVATION POLICY AND THE ECONOMY 54 (Adam Jaffe et al. eds. 2002) (prizes should be tied to social value).

[190] Roin, *Intellectual Property versus Prizes*, *supra* note 189, at 26–29

[191] Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA REV 1575, 1580 (2003).

[192] Grants face the same type of problem, though the grant-making organization must accurately estimate the cost of the innovation rather than the social value. Grants have other advantages—they are frequently used in biomedical research to incentivize innovation, and are therefore familiar; they leverage a social discount rate which is typically lower than private discount rates, and they avoid capital constraints and risk aversion. Hemel & Ouellette, *supra* note 189, at 308. However, like prizes, they do not increase with the size of the eventual social welfare gain of the innovation, and therefore face additional steps in guiding the allocation of innovative effort among projects.

alive may lead to more costs in the future. Finally, prizes face considerable political economy problems; though many medical prizes have been proposed, implementation follows far behind.

Overall, although the specifics of implementation will require considerable care, prizes appear to be an attractive alternative to more traditional exclusivity incentives for the development of BBPM models. Achieving the right level of specificity and project definition is challenging, but that challenge also arises with patents and regulatory exclusivity regimes; moreover, prizes can be precisely tailored and can be structured to require disclosure so as to enable continued cumulative innovation.

C. Incentives for validation

Finally, incentives are needed for model validation.[193] Unlike both traditional medical development and explicit personalized medicine, BBPM cannot readily be validated in standard, straightforward ways. However, BBPM still needs validation to ensure reliability. Instead of scientific understanding, clinical trials, and postmarket surveillance, the validation of complex, implicit BBPM models require validation through other computational mechanisms. Developing methods for that validation, and ensuring they are consistently applied, is an important piece of the innovation policy picture.

Innovation policy should ensure that appropriate incentives exist to drive validation. A bounty could be implemented for external validation (with standards likely set by the FDA). Bounties could be set as a small fraction of revenues of the model overall—set as part of the initial regulatory exclusivity bargain, if one exists. The size of the reward would then roughly scale with the overall value of the model.[194] Rewards for confirmatory validation would ideally decrease asymptotically, so that initial validation would be much more valuable than further confirmation, but that any confirmation over a particular validity threshold received at least *some* reward. This could be set to ensure that the overall fraction of originator revenue which could be siphoned to incentivize validation would

[193] *See* Section II.C, *infra*.

[194] One challenge is that focusing on monetary goals, whether revenue-based or savings-based, might focus incentives on models which deal primarily with costs rather than health improvements. If the principal goal of BBPM is cost-reduction, this focus would be unproblematic. However, if—as seems likely—improving health outcomes is either a primary objective of BBPM or at least an important ancillary objective, then an alternate path to valuing validation would be needed. An alternate possibility would be to offer rewards based on a combination of monetary savings and QALYs or DALYs.

remain constant.[195] On the contrary side, rewards for finding problems should also exist, and should likely not decrease with repetition.[196]

As an additional factor, concerns about validation are exacerbated when data and models are kept secret and proprietary. Implicit models are difficult to validate for the reasons described above, more difficult without access to the modeling code, and extremely difficult without access to the data on which the model was based. Thus, ensuring disclosure, as discussed above, is important to enable not only development and cumulative innovation, but also validation of existing models.

Overall, the appropriate balance of innovation incentives for the development of BBPM requires significant and detailed further work. However, an optimal final landscape might include some push to assemble useful information, either via a public or public-private enterprise, tailored prizes to help drive algorithm development, and bounties for the purposes of third-party validation. In the latter two categories, the prosaic solution of increased grant funding for academic model development may also provide a significant boost in an area where the incentives need are significant but not excessively large.

CONCLUSION

Overall, black box personalized medicine offers immense promise for changing the way medicine is practiced and the way medical technologies are created and deployed. However, the growth of BBPM requires an active and effective innovation policy. The current intellectual landscape in the United States creates problematic incentives which encourage firms to keep data secret and to focus on simple drug-device linkages, rather than developing the necessary capabilities to develop BBPM. The paper has suggested a few ways to change that innovation policy on the path to the major economic and health benefits of the next step in personalized medicine.

More generally, this paper stands along previous work to suggest that our broad-brush innovation system has problematic implications on the ground as it is applied to different questions of innovation in different

[195] For instance, for a validation cap of 2%, the first validator to pass a certain threshold could receive 1%, and each subsequent validator could receive half the amount of the previous validator; the sum of these fractions converges to 2%.

[196] The incentives available for challenges to models might be expected to decrease naturally; if a model is called into question, its value presumably decreases and any fixed fraction of that value would also decrease.

industries.[197] The pharmaceutical and biomedical industries are typically characterized as an area where patents work fairly well; other industries, like software, are characterized as areas where patents work much less well to drive innovation.[198] This Article argues for greater nuance and granularity even within industries: drug manufacturing responds differently to patent and regulatory incentives than drug discovery and development,[199] development of new uses responds differently than developing initial uses,[200] and, as I have argued here, simple diagnostic tests respond differently to patent incentives than complex diagnostic algorithms.

Though there are substantial theoretical arguments both for and against technology-specific patent law,[201] the reality on the ground appears to be that overall innovation policy—a complex amalgam of patents, grants, prizes, tax incentives, regulatory exclusivity, and regulatory barriers to innovation—already varies significantly be industry.[202] If innovation policy already has technology-specific variance, both intentional and accidental, then reshaping policy to address both cross-industry and within-industry variation becomes more palatable, though no less challenging.

---

[197] *See*, *e.g.*, Michael W. Carroll, *One for All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM UL REV 845 (2005); Michael W. Carroll, *One Size Does Not Fit All: A Framework for Tailoring Intellectual Property Rights*, 70 OHIO ST LJ 1361 (2009); Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific*, 17 BERKELEY TECH LJ 1155 (2002); Price II, *supra* note 174; William Fisher III, The Disaggregation of Intellectual Property, HARV. L. BULLETIN, (Summer 2004) (arguing for more industry-specific patent laws); Peter S. Menell, A Method for *Reforming the Patent System*, 13 MICH. TELECOMM. TECH. L. REV. 487 (2007) (same); ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS, AND WHAT TO DO ABOUT IT (2004); (defending industry-neutrality of patent laws); Clarisa Long, *Our Uniform Patent System*, 55 FED. LAWYER 44 (2008) (same);

[198] *See*, *e.g.*, Edwin Mansfield, Patents and Innovation: An Empirical Study, 32 MGMT. SCI. 173 (1986) (reporting different rates of patent importance in different industries); WESLEY M. COHEN ET AL., PROTECTING THEIR INTELLECTUAL ASSETS: APPROPRIABILITY CONDITIONS AND WHY U.S. MANUFACTURING FIRMS PATENT (OR NOT) (National Bureau of Economic Research, Working Paper 7552, February 2000), *available at* http://www.nber.org/papers/w7552 (same).

[199] Price II, *supra* note 174.

[200] Roin, *Solving the Problem of New Uses*, *supra* note 64.

[201] *See supra* note 197.

[202] *See*, *e.g.*, Burk & Lemley, *Is Patent Law Technology-Specific*, *supra* note 198; Burk & Lemley, *Policy Levers in Patent Law*, *supra* note 192 (courts apply enablement and written description requirements differently for software and biotechnology industries); ERIC BUDISH ET AL., DO FIXED PATENT TERMS DISTORT INNOVATION? EVIDENCE FROM CANCER CLINICAL TRIALS (National Bureau of Economic Research, 2013), *available at* http://www.nber.org/papers/w19430 (fixed patent terms distort the incentives for different types of drug development).