

Das funktionelle Unbekanntom

Unwissenheit explizieren, um Forschung zu expandieren



Zugegeben, im Englischen klingt es besser: „Functional Unknomics“ [16]. Gemeint sind damit diejenigen der etwa 20 000 menschlichen Proteine, über die man nichts weiß, und um die es in einer spannenden, in *PLoS Biology* am 8. August 2023 erschienenen Arbeit geht. Aus der bloßen Existenz dieser Eiweißkörper kann man auf jeden Fall eines schließen: Für irgendetwas müssen sie gut sein, denn sonst wären sie nicht da. Die Evolution verschwendet weder Material noch Energie zum Bau von komplexen Molekülen, die der Organismus nicht braucht. Hinzu kommt, dass manche dieser Proteine evolutionär erstens recht alt und zweitens wichtig sein müssen, weil sie in vielen anderen Arten vorkommen.

Seit der ersten, noch recht „grobkörnigen“ Entschlüsselung¹ des menschlichen Genoms vor mehr als 20 Jahren [10] kennt man zwar viele Proteine, aber die Rate

derer, deren Funktionen neu aufgeklärt werden, nimmt eher ab, wie der fast mysteriöse Titel einer bereits vor 5 Jahren im Fachblatt *Proteomics* erschienenen Arbeit andeutet: *Dunkelheit im menschlichen Gen- und Proteinfunktionsraum: weithin bescheidene oder fehlende Erhellung durch die biowissenschaftliche Literatur und der Trend zu weniger Proteinfunktionsentdeckungen seit 2000* [18]. Die Autoren dieser Arbeit legten eine quantitative Auswertung hierzu vor, die ergab, dass 95 % aller Publikationen in den Lebenswissenschaften sich auf weniger als 5000 gut bekannte Proteine beziehen. 10 Jahre nach der Entschlüsselung des menschlichen Genoms (von der man annahm, dass sie zu einer starken Ausweitung der Forschung zu neuen Genen und Proteinen führen würde) bezogen sich 75 % der Publikationen auf die 10 % der Proteine, die bereits zuvor bekannt waren. Dies beklagten kanadische und US-amerikanische Wissenschaftler in einem Kommentar im Fachblatt *Nature*, der mit *Too many roads not taken* überschrieben war [4].

Das Ganze ist keineswegs ein rein akademisches oder grundlagenwissenschaftliches Problem: Viele Medikamente wirken, indem sie an Proteine andocken und dadurch deren Funktion modifizieren. Unentdeckte bzw. unbekannte Proteine, die noch etwa ein Drittel des menschlichen Proteoms ausmachen, könnten den Wirkungsmechanismus vieler Medikamente besser verständlich machen und damit zur Entwicklung neuer Medikamente führen [14]. Diese Datenlage veranlasste die US-amerikanische nationale Gesundheitsbehörde (*National Institutes of Health*, NIH) vor etwa 5 Jahren zur vollen Implementierung des seit 2014 bestehenden Pilotprogramms *Illuminating the Druggable Genome*

im Jahr 2018 [16].² Andere Autoren sprechen davon, dass nur etwa 2000 der etwa 20 000 menschlichen Proteine im Fokus der Forschung stehen [20].

Die Autoren der *Plos Biology* Arbeit zu „Functional Unknomics“ führen die Bevorzugung bereits bekannter Proteine, bzw. die teilweise Blindheit der biologischen Forschung gegenüber unbekanntem Proteinen auf mehrere Faktoren zurück: „Es liegt auf der Hand, dass Finanzierungs- und Begutachtungssysteme eher die Erforschung von Proteinen unterstützen, deren funktionelle oder klinische Bedeutung bereits nachgewiesen ist“ [16]. Kurz: Wenn man tut, was alle tun, hat man ein geringeres Risiko, völlig falsch zu liegen und damit Zeit und Geld zu verschwenden. So einleuchtend das klingt, in der Wissenschaft ist es langfristig fatal, wenn alle so denken (was leider einem langjährigen Trend entspricht, wie man im Fachblatt *Nature* vor gar nicht allzu langer Zeit lesen konnte [15]).

1 Was oft als „Entschlüsselung des menschlichen Genoms“ bezeichnet wird, war ein großes internationales biologisches Forschungsprojekt – das Humangenomprojekt (englisch *Human Genome Project*, HGP), das von 1990–2003 bestand [6]. Im Jahr 2001 brachte es die ersten Ergebnisse, die die Autoren als „unvollständig und vorläufig“ [10] bezeichneten, da man damals erst 83 % des menschlichen Genoms entschlüsselt hatte. Das Projekt wurde im Herbst 1990 mit dem Ziel gegründet, die Abfolge der Basenpaare der menschlichen DNA auf ihren einzelnen Chromosomen zu erforschen. Später bekam es privat finanzierte Konkurrenz durch die Firma *Celera* und deren Chef Craig Venter. Das vorläufige Genom wurde gemeinsam im Jahr 2000 vorgestellt und getrennt in *Nature* am 15.2.2001 (*Human Genome Project*) und in *Science* am 16.2.2001 [23] publiziert. Seit April 2022 gilt das menschliche Genom mit 200 Millionen zusätzlich gefundenen und damit insgesamt 3,055 Milliarden Basenpaaren als nahezu vollständig (mit Ausnahme des Y-Chromosoms) entschlüsselt [13]. Seit dem 23. August 2023 ist auch die komplette Sequenz des menschlichen Y-Chromosoms bekannt [26]. Im Mai 2023 erschien das erste menschliche Pangenom, das auch Variationen zwischen den Genomen einzelner Individuen berücksichtigt [11].

2 In ihrem kurzen Statement zu diesem Projekt fassen die Autoren dessen Ansatz wie folgt prägnant zusammen: „Von den etwa 20 000 proteinkodierenden Genen im menschlichen Genom wurden nur wenige intensiv untersucht. Darüber hinaus gehören schätzungsweise 3000 dieser Gene zum durch Medikamente beeinflussbaren Genom (drugable genome), d. h. der Untergruppe der Gene, die Proteine kodieren und Moleküle binden können, die als Arzneimittel wirken. Weniger als 700 dieser Proteine werden von durch die FDA zugelassenen Medikamenten als Ziel beeinflusst. Dieser Mangel an Aufmerksamkeit für den größten Teil des Proteoms besteht in der akademischen Forschungsgemeinschaft fort, obwohl es Hinweise darauf gibt, dass wenig untersuchte Proteine oft sowohl interessant als auch wichtig für biologische Prozesse sind. Auch neigt die pharmazeutische Industrie dazu, sich auf relativ gut bekannte Proteine als Ziele für die Arzneimittelentwicklung zu konzentrieren, um ihr Risiko zu minimieren. Es besteht daher ein erheblicher Bedarf, sowohl das Grundlagenwissen zu erweitern als auch den Vorrat an potenziellen Angriffspunkten neuer Arzneimittel durch die funktionelle Charakterisierung von wenig untersuchten und schlecht charakterisierten Mitgliedern von arzneimittelwirksamen Genfamilien zu vergrößern“ [17].

Das Problem verschwindet nicht von allein, sondern scheint zuzunehmen. Einige Autoren halten es für einen Fall von *Streetlight-Effekt*: Man sucht nicht dort, wo man etwas verloren hat oder vermutet, sondern dort, wo am meisten Licht ist.³ Bewirkt wird dies durch die routinemäßige Verwendung von Big Data bei der genetischen Forschung: Man verwendet Datenbanken wie *Gene Ontology (GO)*, *Reactome*, *Comparative Toxicogenomics Database (CTD)*, *DrugBank*, *Protein Data Bank (PDB)*, *PubPolar* oder *NCBI GeneRIF*, um die Funktionen von Gensequenzen zu identifizieren. Diese Datenbanken sammeln jedoch genau das, was schon publiziert ist, d. h. jede auf ihnen basierende Suche erfolgt dort, wo das Licht der Erkenntnis bereits leuchtet, und schließt damit systematisch genau das aus, was noch nicht gefunden wurde. Umgekehrt werden dadurch die bereits untersuchten Gene noch besser untersucht, und es entsteht eine Art negativer „hermeneutischer Zirkel“ – mit dem Effekt, dass man nur findet, was alle schon gefunden haben [8].

Wie aber findet man das Unbekannte? – Hier ist zunächst einmal das bekannte Unbekannte vom unbekanntem Unbekanntem zu unterscheiden (► **Abb. 1**). Diese Unterscheidung geht auf den libanesischen ehemaligen Finanzmathematiker und gegenwärtigen Wissenschaftler und Publizisten Nassim Nicholas Taleb zurück, der sich mit den Methoden der Berechnung und Interpretation von Zufallsereignissen und dem Umgang mit besonders seltenen unvorhergesehenen, aber bedeutsamen (wirkmächtigen) Ereignissen beschäftigt hat. Er hat für solche Ereignisse den Namen „Schwarzer Schwan“ populär gemacht – nicht zuletzt in seinem Bestseller mit dem gleichen Titel. Die in ► **Abb. 1** dargestellte Einteilung zu Sachverhalten und unserem Wissen über sie ist selbst kein objektiver Sachverhalt, denn

	Wissen (Bekanntes, knowns)	Unwissen (Unbekanntes, unknowns)
bekanntes	bekanntes Wissen	bekanntes Unwissen
unbekanntes	unbekanntes Wissen	unbekanntes Unwissen

► **Abb. 1** Kenntnismatrix, die auf die Sozialpsychologen Joseph Luft und Harry Ingham [12] zurück geht und – nach einem Mix aus ihren Vornamen – auch *Johari-Fenster* genannt wird. Will man Handlungen begründen, kann man dies mit bekannten Fakten und Ursache-Wirkungs-Zusammenhängen tun, sich also auf *bekanntes Wissen* beziehen. Zum *unbekannten Wissen* zählen statistische Zusammenhänge, die man nicht weiter begründen kann. Aber es gibt sie: Die Risiken, an einem Verkehrsunfall oder Flugzeugabsturz zu sterben, gehören zu unserem Wissen; wann so etwas eintritt, weiß man jedoch nicht. Das *bekanntes Unwissen* hingegen beschreibt Fälle wie beispielsweise Gegenstände der aktuellen Forschung (wenn man das wüsste, würde man nicht von Forschung sprechen) oder die Dunkelziffer bei Kriminalität oder Rauschmittelmissbrauch: Wir wissen, dass es eine Dunkelziffer gibt, aber wir kennen sie nicht (sonst läge sie nicht im Dunkeln). Und dann gibt es noch das *unbekannte Unwissen*, z. B. einen möglichen Kometeneinschlag oder andere extrem seltene Zufälle betreffend.

sie hängt von unserem Wissen ab. Dieses kann sich erstens ändern und zweitens ist es nicht bei jedem gleich: Was heute unbekannt ist (eine neue Art in der Tiefsee), kann morgen wahrscheinlich (bekanntes Unwissen) sein (Foto einer Unterwasserkamera) und sich übermorgen (weitere Fotos und genetische Analysen) bereits in bekanntes Wissen verwandeln. Wissenschaft schafft Wissen.⁴

Die Autoren der neuen Studie untersuchten Gene mit unbekannter Funktion zunächst im Hinblick auf den Grad der Unbekanntheit der von ihnen kodierten Proteine. Sie untersuchten also *nicht* die Funktion selbst, sondern deren (Un-)Bekanntheit! Hierzu verwendeten sie Daten der weit verbreiteten Datenbank *GO*, in der verzeichnet ist, was Wissenschaftler zu verschiedensten Genen und Proteinen zusammengetragen haben. Weil alle eukaryontischen⁵ Lebewesen sehr viel „biologische Maschinerie“ gemeinsam haben, ist es sinnvoll, deren Genetik (Genom) und Proteine (Proteom) auch über die verschiedensten Arten hinweg vergleichend zu untersuchen. Hierzu wurde 2000 die genannte Datenbank *GO* ins Leben gerufen, wie im Fachblatt *Nature Genetics* berichtet wurde [1]. Zur besseren Erforschung der biologischen Rolle von Proteinen in einem bestimmten Organismus kann mit ihrer Hilfe oft auf andere Organismen und das Wissen über sie zurückgegriffen werden. Das Ziel der Datenbank *Go* und des dafür gegründeten Konsortiums war es, „ein dynamisches, kontrolliertes Vokabular zu erstellen, das auf alle Eukaryonten angewendet werden kann, auch wenn sich das Wissen über die Rolle von Genen und Proteinen in Zellen ständig erweitert und verändert“, charakterisieren die Autoren ihre Bemühungen [1]. Um dieses Ziel zu erreichen, wurden 3 unabhängige Ontologien

- zum biologischen Prozess,
- zur molekularen Funktion und
- zur zellulären Komponente entworfen,

5 Eukaryonten sind Lebewesen mit Zellkern und anderen Kompartimenten (Mitochondrien, Golgi-Apparat, Endoplasmatisches Retikulum, etc. – man spricht auch von „Organellen“). Zu diesen Lebewesen gehören Einzeller, Algen, Pflanzen, Pilze, Tiere und damit auch der Mensch. Der Gegenbegriff ist *Prokaryont*, wobei es sich um Lebewesen ohne Zellkern und weitere interne Kompartimente handelt. Hierzu zählen Bakterien. In evolutionärer Hinsicht entstanden vor etwa 3,5 Milliarden Jahren die einfacheren Prokaryonten (Bakterien und Blaualgen). Über die nächsten 2 Milliarden Jahre entwickelten sich dann Organismen mit Zellkern, die vor 1,5 Milliarden Jahren erstmals nachgewiesen wurden. Vor etwa einer Milliarde Jahren entstanden dann die ersten vielzelligen Organismen (Pflanzen und Tiere), wobei noch unklar ist, wann genau die ersten Tiere auftraten; vor etwa 500 Millionen Jahren waren sie jedenfalls da.

3 In den Sozialwissenschaften des anglo-amerikanischen Sprachraums sprach man auch vom „principle of the drunkard’s search“ (dem Prinzip von der Suche des Betrunkenen). Der Ausdruck „Streetlight-Effekt“ wurde explizit vom US-amerikanischen Autor David H. Freedman in einem Artikel geprägt der den Titel „Why Scientific Studies Are So Often Wrong: The Streetlight Effect“ trug und im Magazin *Discover* im Jahr 2010 erschien [5].

4 Bekannt wurde das unbekanntes Unbekanntes allerdings erst durch eine Aussage des damaligen US-Verteidigungsministers Donald Rumsfeld während einer Pressekonferenz im Februar 2002, als es um die Evidenz für Massenvernichtungswaffen ging, deren Herstellung dem Irak unterstellt wurden, was den später gegen diese Nation geführten Angriffskrieg rechtfertigen sollte. Man kann davon ausgehen, dass er mindestens die Zuhörer verwirren wollte.

die über das World-Wide-Web allen Wissenschaftlern frei zugänglich sind [7].

Man konnte nun dieses Wissen nutzen, um Gene zu finden, zu denen fast nichts oder gar nichts zum Wissen um ihre Funktion existiert, und um auf diese Weise eine *Unbekanntheits-Datenbank* menschlicher Proteine zu erstellen, die mittlerweile publiziert und für alle zugänglich ist (Unknowndatabase; <http://unknowndatabase.org>). In ihr kann man nach Proteinen suchen, die zwar einerseits kaum oder gar nicht bekannt sind, jedoch andererseits evolutionär sehr konserviert sind, also in vielen heute existierenden Lebewesen vorkommen. Die *Suchkriterien*, um unbekanntes Interessantes zu finden, bestanden also in bisheriger Unbekanntheit und hohem Grad der Konservierung.

Wie aber definiert man „unbekannt“? Um Proteinen einen Bekanntheitsgrad zuzuweisen, genügt die subjektive Bewertung definitiv nicht, denn „bekannt“ kann alles meinen, von „hat wahrscheinlich die Wirkung X“ bis hin zu „hat den molekularen Wirkungsmechanismus Y“. Dazwischen liegt, wie man gerne sagt, ein weites Feld! Man benötigte daher klare Kriterien für „bekannt“. Dieses Problem lösten die Autoren dadurch, dass sie die Annotationen von Proteinfunktionen in der Datenbank GO zählten – und nichts weiter.

Aber ist das hierdurch gefundene Unbekannte wirklich interessant? Um hier weiterzukommen, verwandten die Autoren einen Trick: Sie suchten 260 Proteine aus, die bei der Fruchtfliege (*Drosophila melanogaster*) und beim Menschen vorkommen, deren Funktion (bei Mensch und Drosophila) aber unbekannt war. Sie verwendeten die Methode der *RNA-Interferenz*, die auch als *posttranskriptionelles Gene-Silencing* bezeichnet wird. Hierbei handelt es sich um einen in den Zellen von Eukaryonten vorkommenden natürlichen Mechanismus, durch den ganz gezielt einzelne Gene abgeschaltet werden können. Dieses Abschalten nennt man „Silencing“, was man mit „Stilllegen“ übersetzen könnte. Als Folge der Stilllegung wird die mRNA in Bruchstücke gespalten und die zu übertragende Information damit zerstört. Die Proteine können somit nicht mehr produziert werden, d. h. deren Translation wird verhindert.

TRANSKRIPTION UND TRANSLATION; MRNA UND IRNA

Bekanntermaßen müssen Gene zunächst „eingeschaltet“ werden, um ihre Funktionen zu entfalten. Hierzu wird die Erbinformation (DNA) zunächst in RNA „abgeschrieben“. Dieses Abschreiben nennt man *Transkription*. Weil diese damit die Informationen („Nachrichten“, engl. „message“) enthält, um Proteine herzustellen, wird sie auch als messenger-RNA (kurz: *mRNA*) bezeichnet. Im Normalfall werden die mRNA-Informationen dann von den Ribosomen verwendet, um Aminosäuren aneinanderzuhängen und um auf diese Weise Proteine „zusammenzuhäkeln“ wie eine Schnur. Dieses Zusammenhäkeln nennt man *Translation*. Beim posttranskriptionellen *Gene-Silencing* werden also Gene, die bereits abgelesen sind, um Proteine zu erzeugen, wieder stillgelegt, indem sie zielgerichtet abgeschaltet werden. Dies geschieht durch kurze RNA-Moleküle, die sich an die mRNA anlagern und diese dadurch an ihrer Funktion hindern. Man sagt auch, dass sie mit der Funktion *interferieren*, weswegen man diesen Typ RNA auch als *iRNA* bezeichnet. Aber wozu das alles – und warum so kompliziert? Mit *iRNA* lässt sich die Funktion von Genen noch beeinflussen, wenn sie schon abgelesen sind, was in der Natur vielfach vorkommt (siehe Haupttext).

In den Biowissenschaften hat sich das Stilllegen einzelner Gene durch RNA-Interferenz („Gen-knock-down“) seit Jahren als eine neue Möglichkeit zur Untersuchung der Funktion von Genen und deren Genprodukten etabliert. Dieser Mechanismus wurde ursprünglich beim Facenwurm *C. elegans* von den US-Wissenschaftlern Andrew Z. Fire und Craig C. Mello entdeckt [27], wofür sie im Jahr 2006 den Nobelpreis für Physiologie und Medizin erhielten.⁶ Durch zielgerichtetes Abschalten einzelner

6 Seit etwa 5 Jahren sind iRNA-basierte Medikamente zugelassen und weitere Medikamente dieser Art sind in der Entwicklung.

Gene lässt sich auch ganz einfach untersuchen, ob das, was sie kodieren, zum Leben gebraucht wird. Zurück zu den 260 *Drosophila*-Proteinen mit unbekannter Funktion, die auch beim Menschen vorhanden sind. Wie sich zeigte, war das Abschalten von 62 dieser Proteine (24% der 260) mit dem Leben nicht vereinbar. Das ist bemerkenswert, wenn man bedenkt, dass über *Drosophila* seit Jahrzehnten vielfach geforscht wird. Zudem gehen die Autoren aus verschiedenen Gründen davon aus, dass die Letalität mit ihrer Methode eher unter- als überschätzt wird.⁷

„Natürlich ist mehr am Leben dran als nur am Leben zu sein,“ fahren die Autoren in ihrer Arbeit fort⁸ [16] und untersuchten die anderen 198 Gene, deren Ausfall zwar nicht zum Tode führt, aber vielleicht zugleich auch nicht folgenlos ist. Sie führten dazu 7 recht komplexe Untersuchungen zu unterschiedlichen wichtigen biologischen Funktionen durch, die *nicht* zu den üblichen Standardverfahren zum genetischen Screening gehörten und zum Teil neu entwickelt werden mussten. Bei diesen Funktionen handelte es sich um weibliche und männliche Fruchtbarkeit, gewebliches Wachstum, Protein-Qualitätskontrolle, Resilienz gegenüber oxidativem Stress und Stress durch reduzierte Nahrung sowie motorische Fähigkeiten. „In der Praxis bedeutete dies, phänotypische Screens zu entwerfen, die komplexer waren als üblich. Wir hofften, auf diese Weise einen größeren Anteil von Genen zu identifizieren, die in den üblichen *Drosophila*-Screens nicht gefunden wurden“, beschreiben die Autoren ihr Vorgehen. Und tatsächlich wurden sie bei weiteren 59 Genen fündig:

2 Gene reduzierten die weibliche Fruchtbarkeit, 7 Gene führten zu nahezu vollständiger männlicher Sterilität und 5 weitere zu reduzierter männlicher Fruchtbarkeit. Weitere Gene betrafen das Wachstum und die Entwicklung sowie die „Qualitätskontrolle“ und die Reaktion auf Stress. Störungen dieser Prozesse fallen nur dann auf, wenn man wirklich genauer hinschaut, weswegen die

7 „it is likely that using RNAi to knockdown gene function underestimates lethality“ [16].

8 „Of course, there is more to life than being alive.“

Autoren genau hier suchten: „[Wir hatten die] Vermutung, dass eine unverhältnismäßig hohe Anzahl der unbekannt Gene an der Qualitätskontrolle und an Stressreaktionen beteiligt sein könnte, die bei vielen traditionellen experimentellen Ansätzen wahrscheinlich übersehen wurden“ [16]. Manche der für Qualitätskontrolle (produzierter Proteine) zuständigen Gene führen in der Säugetierversion zu neurodegenerativen Erkrankungen, was das Vorgehen der Autoren sehr deutlich validiert. 8 Gene standen im Zusammenhang mit Stress durch Nahrungsreduktion und 11 Gene drehten sich um oxidativen Stress. 6 weitere Gene führten beim Ausschalten zu motorischen Störungen.

In ihrer Diskussion gehen die Autoren nochmals auf die Besonderheit ihres methodischen Vorgehens ein: Sie verwendeten Werkzeuge der Bioinformatik, um zunächst einmal herauszufinden, wo andere Wissenschaftler noch nicht gesucht hatten. Sie befassten sich also zunächst mit dem Suchraum, und nicht mit dem, was man in ihm sucht. Hierzu ist anzumerken, dass genau hier auch Künstliche Intelligenz in sehr unterschiedlichen wissenschaftlichen Kontexten wichtige Beiträge geliefert hat [19]: Durch ihre geschickte Anwendung lässt sich der Suchraum verkleinern, was im Falle von Grenzkontrollen in Griechenland während der Corona-Pandemie zu einer Verdopplung der Aufklärungsrate von Corona-Tests [2] und im Falle der Suche nach neuen Antibiotika gegen resistente Keime zu einer Verkleinerung des Suchraums auf ein Millionstel seiner ursprünglichen Größe geführt und damit die Suche überhaupt erst ermöglicht und zum Erfolg geführt hat [21].

Die Bedeutung des verbesserten Suchens für das Finden ist für die Autoren der wesentliche Punkt ihrer Arbeit, wie sie gegen Ende der Diskussion nochmals betonen: „Am wichtigsten ist vielleicht, dass unsere Datenbank eine leistungsfähige, vielseitige und effiziente Plattform für die Identifizierung und Auswahl wichtiger Gene mit unbekannter Funktion bietet, wodurch die Schließung der Lücke im biologischen Wissen, die das Unknown darstellt, beschleunigt wird. In praktischer Hinsicht stellt diese Datenbank eine Ressource für Forscher dar, welche die mit unerforschten Bereichen der

Biologie verbundenen Möglichkeiten nutzen wollen. Solche Vorhaben sind natürlich mit einem gewissen Risiko verbunden, da die Ergebnisse ungewiss sind, und es gibt Hinweise darauf, dass Nachwuchswissenschaftler mit geringerer Wahrscheinlichkeit etablierte Wissenschaftler (principle investigator) werden, wenn sie an Genen arbeiten, die wenig beachtet wurden“ [16]. Sie halten daher „die Anerkennung und Bewertung von Unwissenheit für einen wichtigen Faktor bei Entscheidungen über die relative Priorität, die der Beantwortung der verbleibenden grundlegenden Fragen in der Biologie gegenüber der Umsetzung und Nutzung des vorhandenen Wissens eingeräumt wird“ [16]. Die Tatsache, dass während des Entstehens ihrer Arbeit eine ganze Reihe der unbekannt Gene genauer charakterisiert werden konnten, und die von den Autoren generierten Erkenntnisse – wie sie zeigen – sich vielfach nahtlos in dieses neue Wissen integrieren ließen, spricht eindeutig für das von ihnen gewählte Vorgehen. Kennt man erst einmal das Unbekannt, kann man es untersuchen – und damit definitionsgemäß verkleinern.

Solange man dagegen in der Genforschung nur auf die Ausweitung bekannter Methoden mit immer mehr Daten setzt, die zudem weitgehend automatisiert (beispielsweise mittels Künstlicher Intelligenz) produziert und in das vorhandene Wissen integriert werden, handelt man sich den Autoren zufolge die Gefahr ein, dass sich falsche, widersprüchliche oder oberflächliche Ergebnisse einschleichen, die in absehbarer Zeit kaum Aussicht auf Korrektur haben. Die systemimmanente Bevorzugung bekannter Gene behindert daher eher den Forschungsfortschritt. „Zusammenfassend stellen wir fest, dass die genaue Bewertung der Unkenntnis über die Genfunktion eine wertvolle Ressource für die Ausrichtung biologischer Studien darstellt und sogar wichtig für die Festlegung von Strategien zur effizienten Finanzierung der Wissenschaft sein kann“ [16]. *Big Science* kostet eben *Big Money*. Je bürokratisierter die Forschungsförderung, desto eher werden Großprojekte mit geringem Risiko (aber auch geringem wirklichen Erkenntniszuwachs) gefördert. Aus meiner Sicht ist es sehr erfrischend, wie hier ein *David* durch sehr klare Gedanken zu seiner Unwissenheit gegen den *Goliath* der etablierten *Big Science*

ce sehr erfolgreich antritt. Wissenschaft lebt vom Risiko, sich mit Unbekanntem zu beschäftigen. Die Strategie, erstmal aus unbekanntem Unbekanntem bekanntes Unbekanntes zu machen und dieses dann genauer zu untersuchen, ist durchaus mit Risiko verbunden. Aber wie heißt es so schön: No risk, no gain.

Interessenkonflikt

Es besteht kein Interessenkonflikt.

Korrespondenzadresse

Prof. Dr. Dr. Manfred Spitzer
Universität Ulm
Abteilung für Psychiatrie
Leimgrubenweg 12–14
89075 Ulm, Deutschland

Literatur

- [1] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29
- [2] Bastani H, Drakopoulos K, Gupta V, et al. Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature* 2021; 599: 108–113
- [3] Cardon T, Fournier I, Salzet M. Shedding Light on the Ghost Proteome. *Trends Biochem* 2021; 46: 239–250
- [4] Edwards A, Isserlin R, Bader G, et al. Too many roads not taken. *Nature* 2011; 470: 163–165
- [5] Freedman DH. Why Scientific Studies Are So Often Wrong: The Streetlight Effect. *Discover Magazine* 10.12.2020 www.discovermagazine.com/the-sciences/why-scientific-studies-are-so-often-wrong-the-streetlight-effect; abgerufen am 9.8.2023
- [6] Gannett L. The Human Genome Project. *Stanford Encyclopedia of Philosophy* 2008 <https://plato.stanford.edu/entries/human-genome/#Bib>; abgerufen am 9.8.2023
- [7] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res* 2021; 49: D325–D334. doi.org/10.1093/nar/gkaa1113
- [8] Haynes WA, Tomczak A, Khatri P. Gene annotation bias impedes biomedical research. *Sci Rep* 2018; 8: 1362. doi.org/10.1038/s41598-018-19333-x

- [9] Kustatscher G, Collins T, Gingras AC, et al. Understudied proteins: opportunities and challenges for functional proteomics. *Nat Methods* 2018; 19: 774–779
- [10] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2021; 409: 860–921
- [11] Liao WW, Asri M, Ebler J et al. A draft human pangenome reference. *Nature* 2023; 617: 312–324
- [12] Luft J, Ingham H. The Johari Window, a Graphic Model of Interpersonal Awareness. *Proceedings of the Western Training Laboratory in Group Development* 1955; 246
- [13] Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science* 2022; 376: 44–53
- [14] Oprea T, Bologa C, Brunak S et al. Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov* 2018; 17: 317–332
- [15] Park M, Leahey E, Funk RJ. Papers and patents are becoming less disruptive over time. *Nature* 2023; 613: 138–144
- [16] Rocha JJ, Jayaram SA, Stevens TJ, et al. Functional unknomics: Systematic screening of conserved genes of unknown function. *PLoS Biol* 2023; 21: e3002222. doi.org/10.1371/journal.pbio.3002222
- [17] Rodgers G, Austin C, Anderson J, et al. Glimmers in illuminating the druggable genome. *Nat Rev Drug Discov* 2028; 17: 301–302
- [18] Sinha S, Eisenhaber B, Jensen LJ, et al. Darkness in the human gene and protein function space: widely modest or absent illumination by the life science literature and the trend for fewer protein function discoveries since 2000. *Proteomics* 2018; 18: e1800093. doi.org/10.1002/pmic.201800093
- [19] Spitzer M. *Künstliche Intelligenz*. München: Droemer, 2023
- [20] Stoeger T, Gerlach M, Morimoto RI, et al. Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol* 2018; 16: e2006643. doi.org/10.1371/journal.pbio.2006643
- [21] Stokes JM, Yang K, Swanson K, et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* 2020; 180: 688–702
- [22] Taleb NN. *Der Schwarze Schwan*. München: Patheon, 2018
- [23] Venter JC, Adams MD, Myers EW et al. The Sequence of the Human Genome. *Science* 2001; 291: 1304–1351
- [24] Victor D. Climate change: Embed the social sciences in climate policy. *Nature* 2015; 520: 27–29
- [25] Fire A, Xu S, Montgomery MK, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998; 391(6669): 806–11
- [26] Rhie A, Nurk S, Cechova M, et al. The complete sequence of a human Y chromosome. *Nature* 2023. doi: 10.1038/s41586-023-06457-y

Bibliografie

Nervenheilkunde 2023; 42: 727–731
 DOI 10.1055/a-21106-0196
 ISSN 0722-1541
 © 2023. Thieme. All rights reserved.
 Georg Thieme Verlag KG, Rüdigerstraße 14,
 70469 Stuttgart, Germany