Klinisch-bioinformatische Analyse bei Intelligenzminderung

Aufgrund der hohen Heterogenität von Intelligenzminderungen (ID) handelt es sich bei dieser Erkrankungsgruppe um das Paradebeispiel für den Einsatz genomweiter, molekulargenetischer Untersuchungsmethoden. Zahlreiche Studien belegen, dass sich mit der sorgfältigen Analyse von Genomen eine Diagnosequote von bis zu 60 % erreichen lässt [1, 2]. Bei der hohen Anzahl der zu beurteilenden Sequenzvarianten kommt der computergestützten Dateninterpretation jedoch eine besondere Bedeutung zu und viele Humangenetiker haben bereits damit begonnen, komplexe bioinformatische Methoden und Priorisierungsverfahren, die auf maschinellem Lernen beruhen, in ihren Arbeitsalltag zu integrieren. Ein fairer Vergleich dieser „Pipelines“ ist jedoch schwierig, da sie meist auf das jeweilige Nasslabor angepasst sind. Zudem findet meist ein iterativer Verbesserungsprozess der Analysesoftware statt, der eng an die im Labor prozessierten Proben gekoppelt ist. Ringversuche zur Qualitätssicherung wären wünschenswert und können nun auch im Rahmen des Versorgungsprojektes Translate NAMSE (<https://translate-namse.charite.de/>) durchgeführt werden, welches unter anderem die Evaluation der Exom-Diagnostik in der Regelversorgung zum Ziel hat. Da die Ergebnisse hierzu noch ausstehen, wird in diesem Beitrag kein umfangreicher Überblick über die existierenden Werkzeuge gegeben. Es werden hingegen nur Eindrücke vermittelt, die über die letzten Jahre gesammelt wurden, und Überlegungen dazu angestellt, wie unser Fach die Schnittstelle zwischen

Regelversorgung und Forschung in einem lernenden Gesundheitssystem so definieren könnte, dass mit den für die Krankenversorgung bereitgestellten Mitteln das mögliche Maximum erzielt und gleichzeitig effektiv und kollaborativ Wissenschaft betrieben werden kann.

Datengenerierung und Analysestrategie

Eine effiziente Stufendiagnostik würde idealerweise mit dem Test beginnen, der im Verhältnis zur erwartenden Diagnosequote am günstigsten ist. Bei ID beginnt man in der Regel mit der Zytogenetik, in der strukturelle Chromosomenstörungen abgeklärt werden. Ist diese Analyse unauffällig, wird mit der Sequenzierung von einzelnen Genen oder kleinen Gen-Panels fortgefahren. Als letzte Stufe im einheitlichen Bewertungsmaßstab (EBM) der Kassenärztlichen Bundesvereinigung (KBV) ist die Mutationssuche in >25 kb vorgesehen, die üblicherweise als Exom-Analyse (WES) vorgenommen wird, sich aber natürlich auch als Genom-Analyse (WGS) realisieren ließe. Der jeweilige Geldwert, der in der Gebührenordnung pro Analyse erstattet wird, spiegelt nicht die realen Kostenstrukturen wider, die zudem von Labor zu Labor stark variieren. Bei der WES ist für die Kosten entscheidend, auf welcher Sequenzier-Plattform gearbeitet wird. Die Rohdatengenerierung auf einem Next-Seq ist beispielsweise mehr als doppelt so teuer wie auf einem NovaSeq, der wiederum ein hohes Probenaufkommen voraussetzt und mit deutlich höheren Investivkosten verbunden ist. Die meisten

niedergelassenen Humangenetiker und viele universitäre Standorte können daher nicht konkurrenzfähig Trio-Exome sequenzieren, was sich bereits jetzt zu Ungunsten der Patienten auswirkt. Wenn bei der Datenerzeugung stärker in partnerschaftlichen Verbänden zusammengearbeitet würde, so hätte dies vermutlich auch eine Produktivitätssteigerung in der klinischen Bioinformatik und einen zügigeren Umstieg auf die WGS zur Folge, die die Lücke bei der Mutationssuche im nicht kodierenden Bereich schließen würde. Die WGS bietet zudem den Datensatz, der alle Analysen aus der Zytogenetik erlauben würde und darüberhinaus die Detektion von Strukturvarianten im Bereich 20 bp–2 kb ermöglichen würde.

Interessanterweise könnte schon jetzt in der Übergangsphase zu WGS die parallele Datengenerierung von WES und „array-based comparative genomic hybridization“ (Array-CGH) wirtschaftlich sinnvoll sein, wenn die potenzielle Einsparung an Arbeitszeitkosten für die Befunderstellung über denen der zusätzlichen Rohdatengenerierung liegt. Dies kann an einem Gedankenbeispiel veranschaulicht werden: Wenn Ihnen eine Liste hochqualitativer Varianten aller Größenordnungen (circa 30.000 SNVs („single nucleotide variants“) und Indels von 1–20 bp aus der WES und 10–20 Insertionen, Duplikationen und Deletionen >2 kb aus der Array-CGH) von einem Trio (Patient plus nicht betroffene Eltern) vorläge und Sie die Diagnosequote in einer vorgegebenen Analysezeit maximieren müssten, wie würden Sie vorgehen? Ein Abgleich mit Datenbanken, die pathogene Mutationen enthalten, lässt sich

gut automatisieren und kann insbesondere bei den rezessiven IDs schnell einen Treffer landen. Bei den weiteren seltenen Varianten lohnt sich zunächst ein Blick auf bekannte pathogene Mutationen in Datenbanken und de novo Mutationen (DNMs) aus WES und Array-CGH; diese stellen die häufigste Ursache bei ID dar und sind üblicherweise auf eine geringe Anzahl im Indexpatienten beschränkt. Die Interpretation ist zudem vergleichsweise einfach, da bereits eines der ACMG-Kriterien starker Evidenz für Pathogenität erfüllt ist [3]. Labore mit ausgefeilten Prozessabläufen berichten davon, dass sich auf diese Weise circa die Hälfte der Fälle, in denen letztlich eine molekulargenetische Diagnose benannt werden kann, in unter einer halben Stunde lösen lassen [4]. Bei den restlichen Fällen ist insbesondere eine effiziente und umfassende Kommunikation über den Phänotyp zwischen dem einsendenden Kliniker und dem Labor entscheidend. Nur in wenigen Labors erfolgt eine elektronische Auftragsvergabe, bei der auch Phänotypinformation zum Beispiel aus dem Klinikinformationssystem (KIS) übertragen wird, sodass viel Zeit beim Datenimport und der Ergänzung unvollständiger Angaben verloren geht. Eine gute KIS/Labor-Informationssystem(LIMS)-Integration spart nicht nur Arbeitszeit in der Diagnostik, sondern ist auch die Grundlage für Phenom-weite Assoziationsstudien, die bei Varianten unklarer Signifikanz und insbesondere ungelösten Fällen von großem wissenschaftlichen Interesse sein werden.

Datenbankabgleich und Neumutationsanalyse

Die zum Teil frei zugänglichen Datenbanken wie ClinVar, LOVD und HGMD umfassen derzeit schon über 75.000 pathogene Mutationen, die monogenen Krankheitsbildern zugeordnet sind [3, 4]. Anhand der phänotypischen Auffälligkeiten des Patienten, die die Indikationsstellung begründen, kann die Abfrage zudem auf Krankheitsgene beschränkt werden, für die vergleichbare Symptome bereits beschrieben wurden. Umfangreiche Feature-Gen-Zuordnungen stellt die Human Phenotype Ontology [5]

zur Verfügung (<https://hpo.jax.org/app/download/annotation>). Wenn in dieser Datenbank-Abfrage auch alle im Labor als pathogen befundenen Mutationen berücksichtigt werden, lassen sich in den folgenden Filterschritten zudem alle generierten Datensätze verwenden und die Genotyp-Frequenz-Filter stringenter einstellen. Man braucht also keine Sorge haben, dass eine pathogene bereits befundene Mutation fälschlicherweise herausgefiltert wird, nur weil sie in der hauseigenen Datensammlung gehäuft aufgetreten ist.

Sollte sich der Fall nicht mithilfe pathogener Mutationsdatenbanken lösen lassen, so steht als Nächstes die Suche nach DNMs an. Bei gut eingestellter Pipeline liegt der Erwartungswert einer Einzelbasensubstitution oder eines kleinen Indels im Exom nahe an der „wahren Basisrate“ von 1 DNM pro Individuum der Allgemeinbevölkerung (Mutationsrate von $1,2 \cdot 10^{-8}$ /bp mal Größe der Zielregion mal 2 für diploid, ca. 60–100 Mb, je nach „enrichment kit“). Der wichtigste Kniff für hochqualitative „DNM calls“, insbesondere bei Indels, ist neben einem guten „realignment“ ein Abgleich mit „variant calls“ von Proben, die mit der gleichen Pipeline prozessiert wurden. Dies ist erstaunlicherweise selbst bei nur einigen Hundert verfügbaren Kontrollen bereits effektiver als ein Filter mit „Genome Aggregation Database“ (gnomAD) [6].

Bei diesen großen, aus über 100.000 Individuen aggregierten Daten ist es aufgrund von Sequenzier-Artefakten und Mosaiken außerdem zu empfehlen, eine DNM erst zu verwerfen, wenn sie häufiger als fünf Mal beobachtet wurde. So wird das Risiko reduziert, dass eine Variante, die als Mosaik nicht krankheitsverursachend ist und daher in gesunden gnomAD-Kontrollen auftreten könnte, herausgefiltert wird.

In einer Kohorte von Patienten mit ID spiegelt die Differenz der beobachteten DNM-Rate zur Basisrate übrigens auch annähernd die Diagnosequote wider, die mittels De-Novo-Analyse erzielt werden kann. In vielen ID-Kohorten beträgt diese Rate ca. 1,3–1,5, d. h. es müssen in der Regel pro Fall weniger als 2 De-novo-Mutationen begutachtet werden, um damit

bis zu 50 % der Fälle zu diagnostizieren. Wenn wie bereits erwähnt in diese Analyse ebenfalls noch die De-novo-Mikrodeletionen und -Duplikationen aus der Array-CGH eingeschlossen werden, die äußerst selten in der gesunden Bevölkerung anzutreffen sind, ist die Diagnosequote bei kaum verändertem Arbeitsaufwand entsprechend höher.

Priorisierungsansätze bei vielen Varianten

Bioinformatisch deutlich anspruchsvoller ist die Analyse von Varianten, wenn kein Trio zur Analyse zur Verfügung steht oder von einem rezessiven Erbgang ausgegangen wird.

Dennoch können für Varianten in Genen, für die der Vererbungsmodus des Krankheitsphänotyps sowie dessen Prävalenz bekannt ist, Allel-Frequenz- oder besser noch Genotyp-Frequenz-Filter passgenau eingestellt werden. Es sollte dabei beachtet werden, dass die Schwellenwerte hierbei immer auf die Subpopulationen anzuwenden sind. Eine Variante, die zum Beispiel in der finnischen Bevölkerung in 10 aus 1000 homozygot anzutreffen ist, verursacht auch in keiner anderen Subpopulation ID, selbst wenn sie in der gesamten gnomAD-Kohorte nur einen Anteil von unter 0,0001 hat. Auf diese Weise verbleiben von zum Beispiel initial ca. 2000 in SysID [7] gelisteten ID-Genen üblicherweise deutlich weniger als 100 Kandidaten übrig [2].

In die Priorisierung dieser Gene sollte die Information von populationsgenetischer, molekularer und phänotypischer Ebene einfließen. Methoden des maschinellen Lernens sind gut geeignet, Scores unterschiedlicher Vorhersagewerkzeuge, wie zum Beispiel MutationTaster, „combined annotation dependent depletion“ (CADD) oder Phenomizer zu verarbeiten [8–10]. Zu einer ganz wichtigen Ressource für die Interpretation von DNMs in noch nicht beschriebenen ID-Genen sind auch die „pLI scores“ geworden. Mit diesen Werten, die aus den Abweichungen von beobachteten zu erwarteten „nonsense“-Mutationen im gnomAD-Datensatz berechnet wurden, lässt sich die Wahrscheinlichkeit einer

Haploinsuffizienz abschätzen [11]. Wenn der Phänotyp rezessiv in Erscheinung tritt, ist die zu erwartende Abreicherung an „loss of function“-Mutationen deutlich geringer, sodass der Ansatz von Samocha et al. [11] bei diesen Genen gute Vorhersagen erst bei Kohortengrößen weit über eine Million liefern würde.

In der PEDIA (Prioritization of Exome Data by Image Analysis)-Studie (www.pedia-study.org) fließt erstmals neben den bereits genannten Scores auch ein Ähnlichkeitswert aus der computergestützten Gesichtsanalyse ein. Bislang wird hierfür das künstliche neuronale Netzwerk DeepGestalt eingesetzt, welches auch in Face2Gene Verwendung findet [12]. Eine weitere Initiative, die einen ähnlichen Ansatz des maschinellen Lernens verwendet, ist Minerva&Me (<https://www.minervaandme.com/>).

In der PEDIA-Studie wurden über 600 molekulargenetisch gelöste Fälle zusammengetragen, um einen frei zugänglichen Trainingsdatensatz zu bilden. Insgesamt sind darin 105 unterschiedliche monogene Erkrankungen vertreten, davon mehr als die Hälfte mit ID, die sich alle mittels Exom-Analyse diagnostizieren ließen. Wir haben diese Daten zum Training eines Klassifikators verwendet, der im Testmodus über 98 % der krankheitsverursachenden Mutationen an erster Stelle listet [13].

Der Beitrag der einzelnen Eingabewerte zum PEDIA-Score veranschaulicht deren Bedeutung in der Diagnosefindung. Auch wenn faziale Auffälligkeiten in der PEDIA-Studie bislang ein Einschlusskriterium darstellten, ist das optimale Gewicht der Gesichtsanalyse am PEDIA-Score von knapp einem Drittel doch beachtlich und unterstreicht den Wert der Dysmorphologie. Interessanterweise ließen sich hierbei keine Unterschiede in der Performanz zwischen rezessiven und dominanten Formen der ID beobachten. Diese nächste Generation der Phänotypisierungswerkzeuge (next-generation phenotyping, NGP) ermöglicht es nun erstmals auch Ähnlichkeiten zu quantifizieren und die zunehmend hinderliche Unterscheidung zwischen syndromaler und nicht-syndromaler ID damit zu beenden. In der DDD-Studie war der Anteil, der bis dahin ungelöst

medgen 2019 · 31:20–23 <https://doi.org/10.1007/s11825-019-0233-7>
© Der/die Autor(en) 2019

P. Krawitz

Klinisch-bioinformatische Analyse bei Intelligenzminderung

Zusammenfassung

Hintergrund und Methoden. Durch den Einsatz von Hochdurchsatz-Sequenzierungstechnologie kann bei der Mehrheit von Patienten mit Intelligenzminderung (ID) eine molekulare Ursache gefunden werden. Für die Integration der Ganzgenomsequenzierung in die Regelversorgung ist der Einsatz effektiver Filter- und Priorisierungsverfahren unerlässlich, um die Datenmengen effizient sichten zu können. Entscheidend ist dabei die Kommunikation zwischen Klinik und Labor, die die Kombination von phänotypischer und molekularer Information erst ermöglicht. In der PEDIA (Prioritization of Exome Data by Image Analysis)-Studie wurden erstmals Verfahren des maschinellen Lernens in der Bildanalyse eingesetzt, um syndromale Ähnlichkeiten zu quantifizieren und für die weitere bioinformatische Auswertung nutzbar zu machen.

Ergebnisse und Schlussfolgerungen. Bei vielen monogenen Erkrankungen kommen faziale Auffälligkeiten vor, die sich für die

computergestützte Bildanalyse eignen. Aktuell kann ein „gestalt score“ bereits für ca. 300 unterschiedliche Phänotypen, viele davon mit ID, anhand eines Porträt-Fotos berechnet werden. Bei diesen Erkrankungen erhöht sich die Genauigkeit der Priorisierung (top 1 accuracy rate) um ca. 20%, wenn solche Ähnlichkeitswerte zusätzlich berücksichtigt werden.

Deep-Learning-Verfahren werden zukünftig auch bei der Beurteilung weiterer medizinischer Bilddaten eine wichtige Rolle spielen. Um diese Technologie in der Entscheidungsfindung in der Klinik nutzen zu können, muss die IT-Infrastruktur entsprechend angepasst werden.

Schlüsselwörter

Maschinelles Lernen · Ganzgenomsequenzierung · Priorisierungsverfahren · Computergestützte Bildanalyse · Next-generation phenotyping

Clinical bioinformatics analysis in intellectual disability

Abstract

Background and methods. Using high-throughput sequencing technology, a molecular cause can now be found in the majority of patients with intellectual disability. For the integration of whole genome sequencing into standard care, however, the use of effective filtering and prioritization methods is indispensable to be able to efficiently view quantities of data. Communication between clinic and laboratory, which enables the combination of phenotypic and molecular information, is crucial. In the Prioritization of Exome Data by Image Analysis (PEDIA) study, deep learning was used for the first time in image analysis to quantify syndromic similarities and make them usable for further bioinformatic evaluation.

Results and conclusions. In many monogenic diseases, facial abnormalities occur that are

suitable for computer-assisted image analysis. Currently, a gestalt score can be calculated for approximately 300 different phenotypes, many of them with ID, using a portrait photo. In these diseases, the accuracy of the prioritization (top 1 accuracy rate) increases by approximately 20% if such similarity values are also taken into account.

Deep learning will also play an important role in the evaluation of other medical image data in the future. To be able to use this technology in clinical decision support, it is necessary to adapt the IT infrastructure accordingly.

Keywords

Deep learning · Whole genome sequencing · Variant prioritization · Computer-assisted image analysis · Next-generation phenotyping

Patienten mit Mutationen in „syndromalen“ Genen überraschend hoch und es wäre interessant zu untersuchen, ob der PEDIA-Ansatz auch auf dieser Kohorte mit milden fazialen Auffälligkeiten in ähnlichem Maße die Performanz steigern könnte [14].

Die zentrale Aufgabe der klinischen Bioinformatik ist es, den Entscheidungsprozess in der Diagnostik, der meist ein Zusammenspiel von Genetikern, Klinikern anderer Fächer und Biologen darstellt, zu unterstützen. Diese unterschiedlichen Expertenmeinungen

stützen sich meist auf unterschiedliche Evidenzen und lassen sich daher statistisch betrachtet mittels Bayes verknüpfen [15]. Mit computergestützter Bildanalyse lässt sich nun erstmals auch das ACMG-Kriterium PP4 „using phenotype to support variant claims“ anhand von Gesichtsaufnahmen quantifizieren. In Zusammenschau aller Evidenzen wird der Grad der Glaubwürdigkeit, dass eine Mutation krankheitsverursachend ist, aus allen zur Verfügung stehenden Scores berechnet. Die Güte eines solchen Ansatzes lässt sich auch anhand großer Patientenkollektive ermitteln und würde dadurch die Konsensbildung erleichtern, an welchem Punkt aus einem Diagnostikfall ein Forschungsfall werden darf. Wenn kostendeckend gearbeitet werden soll, müsste bei den aktuellen Vergütungsmöglichkeiten ein Diagnostikbefund für Array-CGH plus WES in unter 3 Stunden erstellt werden. Die beschriebenen bioinformatischen Analyseansätze können maßgeblich dazu beitragen, dass bei der Indikationsstellung Intelligenzminderung eine hohe Diagnoserate erreicht wird. Wenn mit einer wohldefinierten SOP („standard operating procedure“) bereits 98 % aller lösbaren, also über 50 % aller vorstelligen Fälle befundet werden können, ist es akzeptabel einem „clinician scientist“ den Rest für eine zeitaufwendigere Nachbereitung zu überlassen. Im Rahmen von Translate NAMSE gibt es hierfür ein klares Stufenkonzept von Fallkonferenzen und Reanalysen, in dem zusätzlich alternative bioinformatische Workflows zur Anwendung kommen. Am Ende steht die Abfrage von Datenbanken wie Matchmaker und VarWatch zur Beurteilung von Varianten unklarer Signifikanz und zur Identifikation neuer Gen-Phänotyp-Beziehungen im Forschungskontext [16].

Korrespondenzadresse

Prof. Dr. med. Dipl. Phys. Peter Krawitz
Institut für Genomische Statistik und
Bioinformatik, Universitätsklinikum Bonn,
Rheinische-Friedrich-Wilhelms-Universität
Bonn
Sigmund-Freud-Str. 25, 53127 Bonn,
Deutschland
pkrawitz@uni-bonn.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. P. Krawitz gibt an, dass kein Interessenkonflikt besteht.

Alle beschriebenen Untersuchungen am Menschen wurden mit Zustimmung der zuständigen Ethik-Kommission, im Einklang mit nationalem Recht sowie gemäß der Deklaration von Helsinki von 1975 (in der aktuellen, überarbeiteten Fassung) durchgeführt. Von allen beteiligten Patienten liegt eine Einverständniserklärung vor.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Mattick JS, Dinger M, Schonrock N, Cowley M (2018) Whole genome sequencing provides better diagnostic yield and future value than whole exome sequencing. *Med J Aust* 209(5):197–199 (<http://www.ncbi.nlm.nih.gov/pubmed/29621958>. Accessed September 9, 2018)
- Wright CF, McRae JF, Clayton S et al (2018) Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med*. <https://doi.org/10.1038/gim.2017.246>
- Landrum MJ, Lee JM, Riley GR et al (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42(D1):D980–D985. <https://doi.org/10.1093/nar/gkt1113>
- Henrie A, Hemphill SE, Ruiz-Schultz N et al (2018) ClinVar miner: demonstrating utility of a web-based tool for viewing and filtering ClinVar data. *Hum Mutat* 39(8):1051–1060. <https://doi.org/10.1002/humu.23555>
- Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S (2008) The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* 83(5):610–615. <https://doi.org/10.1016/j.ajhg.2008.09.017>
- Lek M, Karczewski KJ, Minikel EV et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536(7616):285–291. <https://doi.org/10.1038/nature19057>
- Kochinke K, Zweier C, Nijhof B et al (2016) Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *Am J Hum Genet* 98(1):149–164. <https://doi.org/10.1016/j.ajhg.2015.11.024>
- Schwarz JM, Rödelberger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8):575–576. <https://doi.org/10.1038/nmeth0810-575>
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310–315. <https://doi.org/10.1038/ng.2892>
- Köhler S, Schulz MH, Krawitz P et al (2009) Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 85(4):457–464. <https://doi.org/10.1016/j.ajhg.2009.09.003>
- Samocha KE, Robinson EB, Sanders SJ et al (2014) A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 46(9):944–950. <https://doi.org/10.1038/ng.3050>
- Gurovich Y, Hanani Y, Bar O et al (2018) DeepGenet—identifying rare genetic syndromes using deep learning. January. <http://arxiv.org/abs/1801.07637>. Zugegriffen: 8. Febr. 2018
- Hsieh T-C, Mensah MA, Pantel JT, consortium P, Krawitz P (2018) PEDIA: Prioritization of Exome Data by Image Analysis. *bioRxiv* November 2018:473306. <https://doi.org/10.1101/473306>
- Deciphering Developmental Disorders S (2017) Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542(7642):433–438. <https://doi.org/10.1038/nature21062>
- Tavtigian SV, Greenblatt MS, Harrison SM et al (2018) Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. <https://doi.org/10.1038/gim.2017.210>
- Sobreira NLM, Arachchi H, Buske OJ et al (2017) Matchmaker exchange. In: *Current protocols in human genetics*, Bd. 95. John Wiley & Sons, Hoboken, S 9.31.1–9.31.15. <https://doi.org/10.1002/cphg.50>